

Characterization of Gut Butyrate Producers and Plasmidome in First-onset Pediatric Inflammatory Bowel Disease

Turki Abujamel

A thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the
PhD degree in Microbiology and Immunology

Department of Biochemistry, Microbiology and Immunology
Faculty of Medicine
University of Ottawa

© Turki Abujamel, Ottawa, Canada 2016

ABSTRACT

Inflammatory bowel disease (IBD) is a growing disorder with unknown etiology. However, increasing evidence strongly highlights the role of gut microbiota with possible involvement of microbial plasmidome in the inflammatory process. Although the composition of the gut microbiota has been extensively studied, important functional groups such as butyrate producers remain poorly characterized, particularly in pediatric IBD. Furthermore, evaluation of the gut plasmidome in healthy and IBD children is missing. In this study, we used molecular techniques involving quantitative PCR (qPCR) and next-generation sequencing of functional and 16S rRNA genes to analyze the level and composition of butyrate-producing microbes in mucosal washes collected from the right colon of healthy children and Crohn's disease (CD) patients during diagnostic colonoscopy. Also, we isolated and characterized the gut plasmidome from the right colon mucosal washes collected from pediatric non-IBD control, ulcerative colitis (UC), and CD subjects. Although no difference was observed in the total amount of butyrate producers that utilize the butyrate kinase (BUK) pathway for butyrate synthesis, butyrate producers that use the butyryl CoA:acetate CoA-transferase (BCoAT) pathway were decreased in CD patients with inflamed colon as compared to controls. This functional gene approach shows that pediatric CD is characterized by generalized decreased abundance of *Eubacterium rectale* and increased abundance of *Faecalibacterium prausnitzii* in patients with inflamed colon. Also, phylogenetic analysis highlighted 15 Operational Taxonomic Units (OTUs) as potential novel butyrate producers, five of which were decreased in CD patients. Using 16S rRNA sequencing approach validated the

functional gene results and showed decreased abundance of *Coprococcus* in CD patients with inflamed colon. Furthermore, non-IBD plasmidome has higher level of genes involved in butyrate synthesis and regulation of different cellular processes and stress response. On the other hand, IBD plasmidome is enriched with antibiotic resistance genes and phage elements, and pediatric CD plasmidome in particular has higher abundance of the adenosine-5'-phosphosulfate reductase gene. Altogether, our study represents the first comprehensive description of gut butyrate producers and plasmidome of pediatric subjects that emphasize a characteristic dysbiosis of butyrate producers in pediatric CD and a potential link between the gut plasmidome and IBD pathogenesis.

ACKNOWLEDGEMENT

The completion of my Ph.D. work and thesis would not have been possible without the support and encouragement of several special people. I would like to take this opportunity to show my gratefulness to those who have helped me in countless ways.

I would like to start by expressing my sincere gratitude to my supervisor Dr. Alain Stintzi for the continuous support, motivation, valuable advices, and guidance throughout my Ph.D. study. I also remain indebted for the immense knowledge that he shared with me, which made my experience productive and stimulating. I can never say thank you enough for all that he has done for me.

Besides my supervisor, I would like to thank my thesis advisory committee, Dr. David Mack, Dr. Marc-André Langlois, and Dr. Martin Pelchat, for their insightful comments, feedbacks, encouragement, and helpful suggestions.

To the inflammatory bowel disease research team at Children Hospital of Eastern Ontario (CHEO), particularly Dr. David Mack, Ruth Singleton, and Amanda Mack, thank you for all the time and effort that you put in collecting and delivering the patient samples involved in this work.

I would like to extend my sincerest thanks and appreciation to Dr. Walid Mottawea, a true friend and fantastic colleague, for helping me establishing this project. I am very grateful for his valuable discussions and cooperation throughout this research. I would like also to thank him as well as my dear friend, Sohیب Ansari for undertaking the tedious task of proofreading my thesis.

I thank my fellow labmates, Dr. Annika Flint, Dr. Christina Xueqi Wang, Dr. Genevieve Coulombe, Dr. Hai Nguyen, Dr. Ibrahim Taher, Dr. James Butcher, Dr. Kelly Grzywacz, Dr. Martin Stahl, Dr. Momen Askora, Dr. Olle de Bruin, Guillaume Romain, Jennifer Li, and Zack Li for the stimulating discussions and for all the fun we have had in the last seven years.

I would also like to sincerely thank King Abdulaziz University and the Saudi Arabian Cultural Bureau in Canada for the granted financial support to pursue my graduate studies, which made my life easier and helped me reach my educational goal.

I would like to thank my parents to whom I owe a great deal. To my late father Saleh, I wish you were here to witness this moment. You are my true inspiration. You were always there for me, supporting me in all means possible. To my mother Khadija, I would never have been able to achieve any goal in my life without your continuous prayers, support, and encouragement.

To my wife Kholood, you are the true reason behind this success. You always have been there for me during difficult times, and you made countless sacrifices to make me who I am. I can never say thank you enough for all your support, encouragement, patience and unwavering love. To my beloved daughters Farrah and Hana, thank you for always cheering me up. I am truly blessed to have you both in my life.

Finally, I would like to thank my brothers, sisters, and all my friends and relatives in Saudi Arabia and Canada for their continuous support.

DEDICATION

IN MEMORY OF MY FATHER

TO
MY MOTHER,
WIFE, DAUGHTERS,
FAMILY, AND
FRIENDS

TABLE OF CONTENT

ABSTRACT	ii
ACKNOWLEDGEMENT	iv
DEDICATION	vi
TABLE OF CONTENT	vii
LIST OF FIGURES	xii
LIST OF TABLES.....	xv
LIST OF ABBREVIATIONS	xvi
 CHAPTER 1	 1
INTRODUCTION	1
1.1. Inflammatory bowel disease	1
1.2. Clinical symptoms of IBD	1
1.3. The epidemiology of IBD in Canada	2
1.4. The etiology of IBD	3
1.4.1. The immune mechanisms involved in IBD	3
1.4.1.1. Normal mucosal immunity	3
1.4.1.2. IBD immunopathology	5
1.4.2. Genetics and IBD	6
1.4.3. The environmental risk factors for IBD	7
1.4.3.1. Tobacco smoking	7
1.4.3.2. Diet	8
1.4.3.3. Infections	10
1.4.3.4. Antibiotic consumption.....	10
1.4.3.5. Other environmental risk factors.....	12
1.5. Intestinal microbiota in healthy humans	12
1.6. Functional redundancy in intestinal microbiota	15
1.7. Complex carbohydrate fermentation in the intestine and dominant butyrate producers in the human gut	17
1.8. Studies in IBD microbiota highlight a dysbiosis in butyrate producers.....	18
1.9. The sample types used for studying intestinal microbiota	23

1.10. Other factors that contribute to functional redundancy	23
1.11. LGT and functional redundancy	24
1.12. Plasmid contribution to intestinal functions in health and disease	27
1.13. The influence of bacteriophages on microbiota diversity and functional redundancy	29
1.14. The link between bacteriophage and IBD	30
1.15. Hypothesis and objectives	31
CHAPTER 2.....	33
Diversity of Butyrate-producing Bacteria in Pediatric Crohn's Disease Revealed by Characterizing Butyryl-CoA:acetate CoA-transferase and 16S rRNA Genes	33
2.1. Introduction	33
2.2. Material and method	42
2.2.1. Ethics statement	42
2.2.2. Patient selection	42
2.2.3. Samples collection.....	43
2.2.4. Extraction of metagenomic DNA	44
2.2.5. Preparation of 16S rRNA sequencing libraries from MLI samples metagenomic DNA for deep sequencing	44
2.2.6. 16S rRNA sequencing data analysis	48
2.2.7. Preparation of BCoAT gene libraries from MLI samples metagenomic DNA for deep sequencing	49
2.2.8. BCoAT sequencing data analysis.....	53
2.2.9. Relative quantification of butyrate producers and confirmation of BCoAT gene sequencing results using qPCR.....	54
2.2.10. Statistical Analysis	57
2.3. Result.....	58
2.3.1. Patient cohort and disease characteristics	58
2.3.2. BCoAT but not BUK butyrate producers are reduced in CD patients with colonic inflammation	61
2.3.3. The diversity of BCoAT butyrate producers, revealed by 16S rRNA sequencing, is different in CD patients as compared to the control	

subjects	66
2.3.4. The diversity of BCoAT butyrate producers is different in CD patients as compared to the control subjects	71
2.3.5. Composition of butyrate-producing bacteria revealed by BCoAT sequencing.....	76
2.3.6. <i>Eubacterium rectale</i> and <i>Coprococcus</i> are depleted and <i>Faecalibacterium prausnitzii</i> thrives in the microbiota of the CD patients as compared to control subjects	78
2.3.7. Phylogenetic analysis of unclassified OTUs	80
2.3.8. Relative quantification of key butyrate producers revealed by qPCR.....	83
2.4. Discussion.....	87

CHAPTER 3

GUT PLASMIDOME IN EARLY ONSET PEDIATRIC INFLAMMATORY BOWEL DISEASE

3.1. Introduction	95
3.2. Material and methods	100
3.2.1. Patient cohort	100
3.2.2. Plasmidome isolation.....	100
3.2.3. Restriction digestion screening of plasmid metagenome	103
3.2.4. Construction of the plasmidome Illumina library.....	104
3.2.5. Data analysis	108
3.2.6. 16S rRNA sequencing data analysis	110
3.2.7. Statistical analysis	111
3.3. Results	112
3.3.1. Patient cohort and disease characteristics	112
3.3.2. Restriction digestion profile highlights group specific differences in the plasmidome composition	114
3.3.3. Sequencing of plasmidome DNA.....	116
3.3.4. The plasmidome host taxa is different than the 16S rRNA-identified phylotypes.....	118
3.3.5. Functional analysis of the plasmidome data.....	121
3.3.6. Characterizing the plasmidome function using the NCBI-NR reference	

databases	126
3.3.7. Characterizing the plasmidome function using other reference databases	137
3.4. Discussion.....	141
CHAPTER 4	150
GENERAL CONCLUSION	150
Future Directions	156
 CHAPTER 5	159
REFERENCES	159
 Collaborators Contributions	180
 Appendix	181
Appendix I: Clinical data of participating subject in the study..	181
Appendix II: Diversity of all bacteria revealed by 16S rRNA gene sequencing.....	187
Appendix III: Number of isolated plasmid clones from each sample.	188
Appendix IV: Clinical data of participating subject in the study.....	189
Appendix V: Rarefaction curves calculated for each plasmidome sample based on the NCBI-NR assigned genes.....	190
Appendix VI: NCBI-NR genes with differential abundance.	191
Appendix VII: Alpha diversity of plasmidome revealed by alignment to the HMP-GIT database.	196
Appendix VIII: HMP-GIT genes with differential abundance.....	197
Appendix IX: Plasmidome differentially abundant genes identified in IBD by the HMP-GIT database..	198
Appendix X: Beta diversity of the differentially abundant plasmidome genes identified by comparing the relative abundance of the HMP-GIT database assigned genes between groups.....	199
Appendix XI: Alpha diversity of plasmidome revealed by alignment to the ACLAME database..	200
Appendix XII: ACLAME genes with differential abundance.	201
Appendix XIII: Plasmidome differentially abundant genes identified in IBD by the ACLAME database.	202

Appendix XIV: Beta diversity of the differentially abundant plasmidome genes identified by comparing the relative abundance of the ACLAME database assigned genes between groups.....	203
Appendix XV: Recent publication as primary author.....	204

LIST OF FIGURES

Figure 2.1.1:	Summary of beneficial effects of butyrate on intestinal cells	35
Figure 2.1.2:	The simplified overall pathway for butyrate synthesis by gut butyrate producers, highlighting the two main pathways at the terminal step	40
Figure 2.2.1:	Illustration of 16S rRNA library construction strategy	46
Figure 2.2.2:	Illustration of BCoAT library construction strategy	51
Figure 2.3.1:	Relative quantification of butyrate kinase (BUK) and butyryl-CoA:CoA transferase (BCoAT) genes to total bacteria 16S rRNA from right colon MLI samples	63
Figure 2.3.2:	Relative quantification of butyrate kinase (BUK) and butyryl-CoA:CoA transferase (BCoAT) genes to total bacteria 16S rRNA from biopsies	65
Figure 2.3.3:	Diversity of butyrate-producing bacteria revealed by 16S rRNA gene sequencing	68
Figure 2.3.4:	Butyrate producers, at the genus level, identified by V6 hypervariable region of 16S rRNA sequencing	70
Figure 2.3.5:	Diversity of butyrate-producing bacteria revealed by BCoAT gene pyrosequencing	74
Figure 2.3.6:	Identified butyrate-producing bacteria at the species level using BCoAT sequencing	77

Figure 2.3.7:	BCoAT butyrate-producers with differential abundance in CD.....	79
Figure 2.3.8:	Rooted phylogenetic analysis of but database sequences with the unclassified Operational Taxonomic Units (OTUs) obtained from BCoAT 454 pyrosequencing	82
Figure 2.3.9:	Quantitative PCR analysis of key butyrate producers. <i>Eubacterium rectale</i> and <i>Faecalibacterium prausnitzii</i> were quantified using BCoAT and 16S rRNA primers	84
Figure 2.3.10:	Quantitative PCR analysis of <i>Faecalibacterium prausnitzii</i> from stool samples	86
Figure 3.2.1:	Illustration of plasmidome capture and its isolation from metagenomic DNA	101
Figure 3.2.2:	Construction of Illumina library from plasmidome DNA	105
Figure 3.3.1:	Plasmidome restriction digest profile	115
Figure 3.3.2:	Taxonomic analysis of the gut plasmidomic data	120
Figure 3.3.3:	Plasmidome functional categories in each group	123
Figure 3.3.4:	Predicted metagenome COGs functional categories from the 16S rRNA sequencing data	124
Figure 3.3.5:	Comparative analysis of the COGs functional categories of the predicted metagenome vs. the plasmidome	125
Figure 3.3.6:	Diversity of pediatric gut plasmidome	128
Figure 3.3.7:	Genes identified as differentially abundant in the plasmidome of IBD patient	130

Figure 3.3.8:	Beta diversity of the differentially abundant plasmidome genes identified by comparing the relative abundance of the NCBI-NR database assigned genes between the control and the IBD.....	132
Figure 3.3.9:	Plasmidome differentially abundant genes identified by comparing the relative abundance of the NCBI-NR database assigned genes between ulcerative colitis (UC), Crohn's disease (CD) and control samples	134
Figure 3.3.10:	Beta diversity of the differentially abundant plasmidome genes identified by comparing the relative abundance of the NCBI-NR database assigned genes between ulcerative colitis (UC), Crohn's disease (CD) and control samples	136
Figure 3.3.11:	General annotation of differentially abundant genes identified by comparing the relative abundance of the HMP-GIT database assigned genes between ulcerative colitis (UC), Crohn's disease (CD) and control samples	138
Figure 3.3.12:	General annotation of the differentially abundant genes identified by comparing the relative abundance of the ACLAME database assigned genes between ulcerative colitis (UC), Crohn's disease (CD) and control samples	140
Figure 4.1:	Proposed gut butyrate producers and plasmidome interactions in pediatric IBD	155

LIST OF TABLES

Table 1.1:	The summary of gut microbiota characterization studies on IBD	20
Table 2.2.1:	Primers used for 16S rRNA library construction	47
Table 2.2.2:	Primers used for BCoAT library construction	52
Table 2.2.3:	Primers used for qPCR	56
Table 2.3.1:	Cohort characteristics	60
Table 2.3.2:	Summary of BCoAT pyrosequencing data	73
Table 3.2.1:	Oligonucleotide sequences used for plasmidome library construction	106
Table 3.3.1:	Participating patients and disease characteristics	113
Table 3.3.2:	Number of plasmidome reads generated from each sample	117

LIST OF ABBREVIATIONS

ACLAME	A CLAssification of Mobile genetic Elements database
APS	Antigen-presenting cells
BCD	Butyryl-CoA dehydrogenase
BCoAT	Butyryl CoA:acetate CoA-transferase
BF	Burkina Faso
BHBD	β -hydroxybutyryl-CoA dehydrogenase
BLAST	Basic Local Alignment Search Tool
BMI	Body Mass Index
bp	base pair
BUK	Butyrate-kinase
CARD15	Caspase Recruitment Domain-Containing Protein 15
CCD	Colonic CD
CD	Crohn's Disease
CHEO	Children's Hospital of Eastern Ontario
CoA	Coenzyme A
COGs	Clusters of Orthologous Groups
COX2	Cyclooxygenase 2
CRO	Crotonase
Ct	Threshold Cycle
DCs	Dendritic Cells
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
DSS	Dextran Sodium Sulfate
EAEC	Enteraggregative <i>Escherichia coli</i>
FACS	Fluorescence-activated cell sorting
FLASH	Fast Length Adjustment of Short Reads
GI	Gastrointestinal
GIs	Genbank identifiers
GIT	Gastrointestinal tract
GWAS	Genome-Wide Association Study
HDAC	Histone Deacetylase
HMM	Hidden Markov Model
HMP	Human Microbiome Project
HMP-GIT	Human Microbiome Project Gastrointestinal_tract reference sequence
HPLC	High-performance liquid chromatography
IBD	Inflammatory Bowel Disease
IBD-U	IBD-unclassified
IBS	Irritable Bowel Syndrome
IC	Indeterminate colitis
ICD	Ileal CD
IgA	Immunoglobulin A
IL	Interleukin

INF	Interferon
iNOS	Nitric oxide synthase
JAK2	Janus Kinase 2
LB	Luria broth
LGT	Lateral gene transfer
MAPK	Mitogen-activated protein kinases
MDP	Muramyl dipeptide
MetaHIT	Metagenomics human intestinal tract
MGEs	Mobile genetic elements
µg	Microgram
mg	Milligram
MIDs	Multiplex Identifiers
µl	Microliter
ml	Milliliter
MLI	Mucosal-luminal interface
mM	Millimolar
NFκB	Nuclear factor kappa B
NK	Natural killer
NOD2	Nucleotide-Binding Oligomerization Domain
ORFs	Open reading frames
OriV	Origin of plasmid replication
OTU	Operational Taxonomic Unit
PCDAI	Pediatric Crohn's Disease Activity Index
PCoA	Principle Coordinate Analysis
PCR	Polymerase chain reaction
PTB	Phosphate butyryltransferase
PUCAI	Pediatric Ulcerative Colitis Activity Index
QIIME	Quantitative Insights Into Microbial Ecology
qPCR	Quantitative polymerase chain reaction
RC	Right Colon
ROS	Reactive Oxygen Species
SCFA	Short Chain Fatty Acids
SOC	Super Optimal broth with Catabolite repression
SRB	Sulfate Reducing Bacteria
STAT3	Signal Transducer and Activator of Transcription 3
TA	Toxin/anti-toxin
TCAG	Centre for Applied Genomics
TGFβ	Transforming Growth Factor Beta
Th	T helper cells
THL	Thiolase
TJs	Tight junctions
TNF	Tumor Necrosis Factor
TRACA	Transposon-aided capture
T _{reg}	Regulatory T Cells
UC	Ulcerative Colitis
VFDB	Virulence Factors Database

CHAPTER 1

INTRODUCTION

1.1. Inflammatory bowel disease

Inflammatory bowel disease (IBD) is a group of uncontrolled chronic inflammation of the gastrointestinal tract [1]. Based on the intestinal localization and features of the inflammation, about 85–90% of IBD cases are classified as either Crohn's disease (CD) or ulcerative colitis (UC). The remaining 10–15% of patients are difficult to categorize; therefore, they are referred to as IBD-unclassified (IBD-U) [2]. Dr. Burrill Bernard Crohn, an American gastroenterologist, described the first case of CD in 1932 [3]. In general, CD is characterized by patchy transmural inflammation that can strike any part of the gastrointestinal tract. The term UC, on the other hand, was introduced well before the term CD, in 1859, by Sir Samuel Wilks, a British physician and biographer [3]. Typically, UC inflammation is continuous, limited to the intestinal mucosa, and mainly restricted to the colon with involvement of the rectum in 95% of patients [4].

1.2. Clinical symptoms of IBD

The symptoms of IBD include fever and weight loss, in addition to intestinal manifestations that include abdominal pain and diarrhea [4]. In UC, the stool

occasionally includes blood and mucus [4]. Extraintestinal manifestations include arthritis, oral ulcers, skin lesions, ocular complications, kidney stones, cholestatic liver disease, and bone abnormalities [4]. Pediatric IBD is different than adult onset IBD [5] in that pediatric IBD patients have a more severe clinical presentation that is associated with rapid disease progression and detrimental effects on their growth and development [5-8]. In addition, the intestinal inflammation in pediatric IBD patients mainly involves the ileum and colon (ileocolonic) in CD and the whole colon (pancolitis) in UC [5-8]. Together, this indicates that pediatric onset represents a distinct group of IBD.

1.3. The epidemiology of IBD in Canada

The prevalence and incidence of IBD are continually rising, causing significant morbidity in various parts of the world, especially the United States, Canada, Europe, the United Kingdom, and Scandinavia [9, 10]. Here in Canada, IBD afflicts about 0.67% of the population, and it is diagnosed more in Canada than anywhere else in the world [11]. IBD alone is estimated to cost Canada more than \$1.2 billion per year in direct medical care costs and to burden the country a total of \$2.8 billion per year [11]. The prevalence of IBD in Canada is estimated to be 279 cases per 100,000 and 211 cases per 100,000 for CD and UC, respectively, with about 10,200 newly diagnosed cases every year [11]. The province of Ontario has the highest prevalence of both CD and UC among all Canadian provinces and territories. It is estimated that 53,249 and 40,809 of Ontarians suffer from CD and UC, respectively [11, 12]. Moreover, the incidence of IBD in Canadian children and adults is increasing alarmingly [11, 12]. In adult IBD, CD is more common in females

than in males, and both genders are equally affected by UC. However, pediatric IBD has different gender predominance than for adults -it has higher male/female ratio, especially for CD [5-8]. Although IBD is most commonly diagnosed in the second and third decades of life, about 20–25% of IBD cases are diagnosed during childhood [11]. Pediatric IBD is associated with a higher degree of positive family history for both CD and UC [11]. Also, hospitalization due to pediatric CD is more common than for adult onset CD [13]. These epidemiological findings further highlight the differences between pediatric and adult onset IBD.

1.4. The etiology of IBD

The exact etiology of IBD remains unclear. However, it is thought to be a result of interactions between the immune response, genetic predisposition, environmental factors, and gut microbiota dysbiosis [14]. The contribution of each of these elements to IBD is discussed in more detail below.

1.4.1. The immune mechanisms involved in IBD

1.4.1.1. Normal mucosal immunity

The intestinal tract represents the largest surface area in the human body that is in direct contact with the outer environment [15]. This enormous surface area requires the interaction of the innate and adaptive immune system to prevent pathogenic infections or unnecessary reactions to foreign antigens. The intestinal tract is lined with a closely attached, single layer of epithelial cells. These cells are held together via a network of gate-like proteins known as tight junctions (TJs) that

also regulate the trafficking of macromolecules across the intestinal lining [16]. Normally, the intestinal epithelia are covered with a mucus gel produced by the goblet cells. Colonic mucus is composed of two layers—a firm inner layer that is stratified but thin and an outer layer that is thick and loose [17]. The inner layer acts as a physical barrier that prevents direct contact between the luminal antigen and the intestinal lining. At the same time, the inner mucus membrane is rich in antimicrobial peptides, phospholipases, and lysozymes secreted by Paneth cells, as well as the immunoglobulin A (IgA) antibody, which serves as another barrier against intestinal bacteria [18-20]. However, the less dense nature of the outer mucin makes it possible for the intestinal microbiota to penetrate this mucus layer and to use its glycans as an energy source [17].

Beneath the intestinal lining resides the largest group of immune cells in the human body. These cells include antigen-presenting cells (APCs), which are mainly dendritic cells (DCs) and macrophages, B cells, T cells, and natural killer (NK) cells [21]. Foreign antigens, including commensal bacteria, can be directly sampled from the outer mucus layer through the extension of DCs between epithelial cells [22]. M cells also can transfer the lumen antigen across the intestinal lining, through transcytosis, to underneath the lymphoid follicles, including Peyer's patches and isolated lymphoid follicles [23]. Normally, a low level of antigens crosses the intestinal barrier [24-26]. Subsequently, these antigens are engulfed, processed, and displayed by APCs, which results in a low level of co-stimulatory molecules. These molecules include interleukin IL-10, transforming growth factor β (TGF- β), and retinoic acid [24-26]. A subset of activated APCs migrate to secondary lymphoid

organs, including Peyer's patches and mesenteric lymph nodes, and stimulate the induction of regulatory T (T_{reg}) cells [27, 28]. The T_{reg} cells then mainly produce interleukin 10 (IL-10) and TGF- β [29, 30], which will eventually suppress the unnecessary activation of immune cells and maintain a state of tolerance to intestinal commensal microbes.

1.4.1.2. IBD immunopathology

In IBD, both the intestinal barrier and the adaptive immune responses are affected, which results in chronic inflammation. IBD patients exhibit increased intestinal permeability compared to normal individuals and their first-degree relatives [31-34]. In UC, there is a depletion of both the goblet cells and mucin secretion [35, 36]. Also, MUC1 and MUC2 genes, which encode important components of intestinal mucin, have been highlighted as risk loci for CD [37, 38]. Antimicrobial peptides, which are the second line of defense against the luminal antigen, are also affected in IBD. Simms *et al.* [39] reported a reduction of intestinal defensins and the number of Paneth cells in CD. Furthermore, tight junction expression is reduced in IBD and is also correlated with inflammation [40]. Ultimately, the defective physical barrier in IBD results in a leaky gut. However, it is still unknown whether the defective intestinal barrier is a cause of IBD or merely sustains the inflammation.

The adaptive immune system is also involved in IBD immunopathology [41]; however, the exact mechanisms of this involvement are still not clear. Originally, CD was thought to be the result of Th1 response, characterized by increased levels of INF- γ and IL-2 [42], while UC was thought to be Th2 dependent, characterized by high levels of IL-5 and IL-13 [43-45]. In contrast to these theories, biopsies from UC

patients showed a higher level of INF- γ and a lower level of IL-13 *ex vivo* than biopsies from CD patients [46, 47]. Interestingly, both forms of IBD share the same downstream pathway that involves the production of IL-21 and IL-23 [48, 49]. This crucial finding shifted the original paradigm to one in which Th17 cells are known to be involved in IBD immunopathology. The importance of Th17 in IBD pathology is further highlighted by the identification of a number of IBD risk loci that are involved in Th17 cell differentiation and expansion, such as IL-23R, IL-12B, TNFSF15, JAK2, STAT3, and CCR6 [37, 50]. Th17 cells produce the proinflammatory cytokine IL-17 that activates the NF- κ B and MAPK pathways [51, 52], resulting in the activation of enterocytes, phagocytes, and granulocytes and the production of an array of cytokines and chemokines that exacerbate inflammation [53-56].

1.4.2. Genetics and IBD

To date, genome-wide association studies (GWASs) have identified more than 160 IBD risk loci [57]. The first and most strongly association IBD genetic risk locus identified was within the nucleotide binding oligomerization domain (*NOD2*) gene, also known as the caspase recruitment domain family member 15 (*CARD15*), on chromosome 16 [58, 59]. The leucine-rich repeats in CARD15 are known to bind muramyl dipeptide (MDP), a basic component of the peptidoglycan moiety in bacterial cell walls, and help in microbial defense via the NF- κ B signaling pathway [60]. In fact, most of the identified risk loci are directly or indirectly involved in cellular processes related to microbial defense and clearance such as barrier function, intestinal restitution, innate and adaptive immunity, reactive oxygen species generation, endoplasmic reticulum stress, and autophagy [61]. Interestingly, about

30% of the IBD risk loci are shared between UC and CD, highlighting a number of common pathways in both IBD phenotypes [57]. Also, more than half of the IBD risk loci are shared with other inflammatory and autoimmune disorders such as rheumatoid arthritis, celiac disease, asthma, multiple sclerosis, and diabetes [62]. These findings highlight the importance of the cellular pathways involved in microbial defense in IBD pathogenesis in particular and in other chronic inflammatory and autoimmune diseases in general. However, the genetic association with IBD only explains a minority of the observed variance (13.6% in CD and 7.5% in UC) [57], which suggests that other factors are of higher importance in the etiology of IBD.

1.4.3. The environmental risk factors for IBD

Several environmental factors have been associated with an increased risk of IBD, including smoking tobacco, diet, infections, antibiotic consumption, as well as other environmental factors [63]. Each of these environmental factors is discussed in more detail below.

1.4.3.1. Tobacco smoking

Tobacco smoking is one of the environmental factors with a clear association with the IBD subtypes; however, this association is markedly different for CD and UC [64-67]. In CD, tobacco smoking is not only correlated with an increased risk of the disease but also with worse clinical outcomes, increased relapse, and reduced response to treatment [64-67]. Similarly, smoking cessation is associated with improved CD outcomes [68, 69]. On the other hand, smoking tobacco has been shown to be protective against the development of UC [67] and smoking cessation

has been linked to the deterioration of UC clinical outcomes [70, 71]. The reasons behind the opposite effects of tobacco smoking on CD and UC remain obscure.

1.4.3.2. Diet

Diet is an important environmental risk factor for IBD. In general, consumption of a diet high in fat, carbohydrates, and meat, which is typical of a Western diet, increases the risk of IBD [63]. On the other hand, a diet rich in fiber, fruit, and vegetables protects against IBD [63]. In fact, the diet can directly shape the gut microbiota. Feeding mice with different genetic backgrounds a high-fat and high-sugar diet introduced consistent changes in the gut microbiota, which were also dose dependent [72]. The effect of diet on gut microbiota has also been highlighted in other studies. Comparing the fecal microbiota of European children to that of children from Burkina Faso (BF) showed that the latter exhibited a higher abundance of Bacteroidetes and short chain fatty acids (SCFAs)-producing bacteria and a lower abundance of Firmicutes, Proteobacteria, and Actinobacteria [73]. The author attributed the observed difference in microbiota to the distinct diets between the two populations. Typically, the BF diet is composed of unprocessed polysaccharide-rich foods. On the other hand, the typical European diet is rich in sugar and fat [73]. Another study compared the gut microbiota of a large cohort of children who live in urban areas in the United States to those who live in rural villages in Malawi and Venezuela using 16S rRNA sequencing and metagenomics [74]. Interestingly and despite their geographical distance, there were high similarities in the gut microbiota of Malawian and Venezuelan children with a high abundance of *Prevotella*. Furthermore, Malawian and Venezuelan infants showed enrichment in microbial

genes involved in utilization of glycans and nitrogen from breast milk [74]. On the contrary, children in the United States had a higher relative abundance of *Bacteroides* and their microbiota composition could distinguish formula-fed from breast-fed infants [74].

Furthermore, another study collected the dietary habits of 98 volunteers and divided them into two main groups: high protein and animal fat consumers and high carbohydrate consumers [75]. Subsequently, Wu *et al.* characterized the gut microbiota from stool samples collected from the participating individuals using 16S rRNA sequencing [75]. Interestingly, the gut microbiota clustered according to the individuals' dietary habits. The microbiota of high protein and animal fat-consuming individuals was characterized by a higher abundance of *Bacteroides* compared to the high carbohydrate-consuming individuals, with the microbiota of the latter group enriched with *Prevotella* [75]. More recently, David *et al.* [76] characterized the gut microbiota of nine volunteers after the consumption of either animal-based or plant-based diets for five consecutive days. Consuming the animal-based diet resulted in an increased abundance of bile-tolerant microorganisms, including *Alistipes*, *Bilophila*, and *Bacteroides*. Interestingly, increased abundance of *Bilophila* in the intestine of the *Il10^{-/-}* mouse model of colitis has been shown to cause gut inflammation that resembled IBD [77]. Based on this observation it is tempting to speculate that the animal-based diet might contribute to the pathogenesis of IBD. On the other hand, the plant-based diet increased the abundance of polysaccharides-digesting bacteria, including *Roseburia*, *Eubacterium rectale*, and *Ruminococcus bromii* [76]. Generally, polysaccharides-digesting bacteria are

beneficial bacteria to the host and their abundance is reduced in IBD [78-81]. In conclusion, these studies clearly show that diet has an influence on intestinal microbiota diversity.

1.4.3.3. Infections

Gastrointestinal (GI) infection, especially early in life, has been shown to increase the risk of developing IBD [82-85]. This increased risk of IBD is not only associated with GI infection but also with general and upper respiratory tract infections and otitis media [86-88]. However, it is not known whether the increased IBD risk that is associated with these infections is a result of the original infection or a consequence of antibiotic treatment, because antibiotic use is also associated with IBD [89-91].

1.4.3.4. Antibiotic consumption

Antibiotic consumption, particularly early in life, can have short- and long-term impacts on intestinal microbiota, including the development of antibiotic resistance [92-95]. For instance, the monitoring of the adult fecal microbiota of ciprofloxacin-treated patients using 16S rRNA sequencing showed significant alterations in the abundance of most phylotypes of intestinal microbiota [96, 97]. Although the microbiota reverted to its original composition four weeks following ciprofloxacin treatment, some bacteria taxa were never recovered [96, 97]. In addition, the antibiotic-mediated disturbance of the gut microbial ecology was more prominent in pediatric cohorts. Administration of ampicillin and gentamicin to newborns resulted in

an increased abundance of Proteobacteria and Actinobacteria, which lasted for up to four weeks after completion of the antibiotic treatment [98].

Importantly, there is evidence of an increased association between antibiotic consumption and chronic inflammatory diseases such as IBD, especially early in life [99-101]. For instance, Gevers *et al.* [81] characterized the gut microbiota of 668 children under the age of 17, including 447 newly diagnosed CD cases and 221 non-IBD controls. Among those participants, 77 were on antibiotics [81]. This study revealed an increased relative abundance of *Enterobacteriaceae*, *Pasteurellaceae*, *Veillonellaceae*, and *Fusobacteriaceae* in CD patients as compared to control subjects. On the other hand, the relative abundance of *Erysipelotrichales*, *Bacteroidales*, and *Clostridiales* was significantly decreased in the guts of the CD patients. Most interestingly, the microbiota of the CD patients exposed to antibiotics prior to the onset of the disease exhibited a higher level of microbial composition imbalance or “dysbiosis” as compared to the microbiota of the unexposed CD patients. In addition, CD patients receiving antibiotics had about a 10-fold increase in the relative abundance of *Enterobacteriaceae* and *Fusobacteriaceae* as compared to the unexposed CD patients [81]. Newly diagnosed cases of IBD are more likely to have used antibiotic 2–5 years before diagnosis [99]. These findings highlight the association between antibiotic intake and the diagnosis of IBD. Furthermore, antibiotic administration compromises the colonization barrier, which opens doors to pathogenic bacteria such as *Clostridium difficile* and *Salmonella typhimurium* [102, 103].

Antibiotic consumption also results in the long-term persistence of antibiotic resistant bacterial species [104]. Fecal samples collected from patients treated with clarithromycin against *Helicobacter pylori* infection contained enterococci with high levels of the clarithromycin resistance gene *erm(B)* [104]. On the other hand, no resistance to clarithromycin was detected in members of the control group, who had not received antibiotic treatment. Interestingly, this antibiotic resistance was shown to persist for three years post-treatment [104]. In another study by the same group, fecal *Staphylococcus epidermidis* showed resistance to clarithromycin for four years after the treatment of an *H. pylori* infection with clarithromycin [105].

1.4.3.5. Other environmental risk factors

Other IBD risk factors include psychological stress [88, 106], the use of oral contraceptives [107], and the use of non-steroidal anti-inflammatory drugs [108, 109]. Most of these IBD environmental risk factors are known to alter the intestinal microbiota or their metabolic activity [73, 102, 110-114]. On the other hand, vitamin consumption (include vitamins C and D), having been raised with pets (especially in early life), and living in rural areas have been shown to be protective against the development of IBD [115-118].

1.5. Intestinal microbiota in healthy humans

Using culture-independent molecular techniques and culture-dependent methodologies, gut microbiota has been extensively studied over the past few years. It is now known that the gut microbiota is composed of about 100 trillion microbial cells, totaling about 1,000 species with at least 160 different species in each person

[119]. This enormous microbial community expresses about 3.3 million genes, which exceed the human genes 150-fold [119]. All of these observations support the astonishing fact that the composition of the human body is a combination of 10% human and 90% microbes. Undoubtedly, gut microbiota are crucial for maintaining well-being, because they are involved in multiple processes, including the resistance against the colonization of pathogenic infections, the modulation of the mucosal immune system, the stimulation of angiogenesis, and the production of beneficial metabolites such as short-chain fatty acids (SCFAs), vitamins, and lipids [120, 121].

The gut microbiota develops rapidly within the first two years of life [122]. At birth, the infant gut is colonized mainly by the microbiota from the mother's birth canal, in addition to a few bacteria from the surrounding environment [120]. Therefore, the mode of delivery has a great impact on the diversity and composition of the infant intestinal microbiota [123, 124]. Several other factors influence the development of intestinal microbiotas during childhood, the most important of which include genetic background, nutritional habits, and the consumption of antibiotics, especially early in life [122]. At birth, the infant gut environment is aerobic and will gradually shift toward an anaerobic environment and become fully anaerobic a few days after delivery [125]. Therefore, the early gut microbiota is essentially composed of facultative anaerobic bacteria that mainly belong to the *Enterobacteriaceae* family. Later, as the microenvironment turns anaerobic, the infant gut is colonized by more obligate anaerobic bacteria that resembles the mother's skin microbiota (in cesarean section-delivered infants) or vaginal microbiota (in vaginally-delivered infants) [125]. These bacteria include *Clostridiaceae*, *Bifidobacteriaceae*,

Lactobacillaceae, *Enterococcaceae*, and *Streptococcaceae* [125]. During feeding, the infant gut is dominated by milk fermenters that mainly include *Bifidobacteria* and *Lactobacillus* [126]. The shift from milk to solid foods during the weaning process results in a decreased abundance of *Bifidobacteria* and *Lactobacillus*. This is followed by a concomitant increase in the abundance of *Clostridium*, *Ruminococcus*, and *Bacteroides*, which are SCFA-producing bacteria [127, 128].

The gut microbiota increases in complexity with age until adulthood, with approximately 90% of healthy adult gut microbiota belonging to the Bacteroidetes and the Firmicutes phyla, with a lesser abundance of Proteobacteria, Verrucomicrobia, Actinobacteria, Fusobacteria, and Cyanobacteria [120, 129]. Firmicutes in adult guts is mainly dominated by *Clostridial* clusters IV, XIVa, and XVI. On the other hand, Bacteroidetes is mainly dominated by *Prevotellaceae*, *Bacteroides vulgatus*, and *Bacteroides thetaiotamicron* [120, 129]. With age, Bacteroidetes and Firmicutes remain the dominant phyla, representing about 95% of the microbial community [130]. The elderly show an expansion of *Clostridium* cluster IV when compared to younger adults [131]. In addition, centenarians show a significant decrease in *Clostridium* cluster XIV and an increase in opportunistic pathogens that belong to the Proteobacteria phylum [130]. The intestinal microbial communities have been proposed to fall within three main clusters named enterotypes [132]. These enterotypes are dominated by either *Bacteroides*, *Prevotella*, or *Ruminococcus* species and are thought to be found in all individuals, regardless of gender, age, and weight [132]. However, this hypothesis has been

recently challenged by the demonstration of a continuous shift in enterotypes between and within individuals [133, 134].

The number of bacterial cells varies along the gastrointestinal tract (GIT). Generally, as low as 10 to 1,000 bacteria per gram of content are present in the stomach and duodenum. This number gradually increases as we move down the GIT, reaching 10^4 to 10^7 cells per gram of content in the jejunum and ileum, and the highest colonization is observed in the colon, which reaches up to 10^{12} cells per gram of colonic content [120]. The difference along the GIT is not only in bacterial load but also in the composition. By studying the bacterial composition of the small intestine and colon biopsies, Frank *et al.* [78] demonstrated that there is a substantial difference in the bacterial composition between both sites in healthy individuals. They found that the small intestine is mainly colonized by the phyla Firmicutes (especially bacteria from the class *Bacilli*) and Actinobacteria. On the other hand, the colon is heavily colonized with bacteria from the *Lachnospiraceae* family [78]. Interestingly, the composition of the gut microbiota is also different when moving from the intestinal mucosa to the lumen [129, 135, 136], in which the mucosal microbiota is enriched with *Lachnospiraceae* while the lumen harbors mainly *Bacteroidaceae* [137]. The composition of the mucosal microbiota is important because it is thought to be responsible for the modulation and maturation of the mucosal immune system [138].

1.6. Functional redundancy in intestinal microbiota

The comparison of the gut microbiota composition from different individuals showed that only a small percentage of the bacterial species is shared [139, 140],

which indicates a high degree of interpersonal variations. Furthermore, each person carries distinctive strains of gut bacteria [141, 142]. This highlights the presence of a dynamic group of microbiota that is personalized to each person based on their age, genetics, environment, and diet [143]. More interestingly, the characterization of the microbiota metagenome from 242 individuals revealed that the metabolic pathways encoded by the gut microbiota metagenomes were common between individuals [140]. Therefore, it is the function encoded by the microbiome that is important and not the taxa composition [140, 144].

Qin *et al.* [119] studied the functions of the human gut microbiota and identified a wide-range of functional categories. In total, they predicted about 3.3 million non-redundant genes from the gut metagenome of 124 individuals; and 99.69% of the phylogenetically classified genes were found to belong to bacteria and archaea [119]. A subset of these genes was found in the genome of almost all bacterial species; among this cluster were the genes involved in the main metabolic pathways and protein complexes. Another subset of genes was found in the metagenomes of all individuals; genes in this category harbored 6,313 functions, 33% of which were categorized into known functional orthologous groups, and the remaining majority (67%) had uncharacterized functions. A large number of genes that were common to the metagenomes of all individuals involved the digestion of complex carbohydrates [119]. Complex carbohydrates are not digested by the host and poorly absorbed. Therefore, they represent a significant energy source for bacteria capable of digesting complex carbohydrates which is a crucial function of the gut ecosystem [145, 146].

1.7. Complex carbohydrate fermentation in the intestine and dominant butyrate producers in the human gut

Complex carbohydrates in the human diet come mainly from edible plants, which contain both soluble and insoluble carbohydrates [147]. Complex plant carbohydrates include resistant starch, hemicelluloses, pectins, cellulose, inulin, and arabinoxylan oligosaccharides [147, 148]. Complex carbohydrates cannot be absorbed in the human gut, which lacks the necessary enzymes to digest them. Therefore, they pass through the small intestine and accumulate in the large intestine [148]. There, they become ready for fermentation by various members of the large intestine anaerobic microbiota that produce short chain fatty acids (SCFAs) as a fermentation end product [121]. SCFAs are organic acids that mainly include butyrate, propionate, and acetate [149]. The vast majority of SCFAs produced in the intestine are promptly absorbed by intestinal cells and used as energy source by the human body [150, 151]. Propionate and acetate are largely used as substrates for gluconeogenesis and lipogenesis by the liver and peripheral tissues [152, 153]. On the other hand, butyrate is oxidized by colonocytes' mitochondria and constitutes about 10% of their energy source [121, 154].

In the human gut, members of the Clostridial clusters *XIVa* and *IV* are the major producers of butyrate. Within Clostridial cluster *XIVa*, *Eubacterium* and *Roseburia* are the dominant butyrate-producing bacterial genera, and *Faecalibacterium prausnitzii* is the most abundant butyrate producer within cluster *IV* [155]. However, Clostridial clusters *XIVa* and *IV* are estimated to account for 30–60% of the total colonic microbiota and contain numerous non-butyrate producers

[156]. Moreover, minor producers of intestinal butyrate include members of the *Clostridium* clusters I, XV, and XVI [157].

1.8. Studies in IBD microbiota highlight a dysbiosis in butyrate producers

Recent studies have demonstrated an association between the gut microbiota and IBD [158, 159]. In CD patients, divergence of the fecal stream by ileostomy to bypass the inflamed areas results in better outcomes for patients, while restoration of the flow worsens clinical outcomes [158]. In animal models, the development of intestinal inflammation requires intestinal microbiota, because various spontaneous colitis susceptible mouse models do not develop colitis when kept in germ-free conditions [160-162] and the inflammation is evident only following colonization with bacteria [161]. Also, mice with adaptive immune systems deficient in the T-bet transcription factor, which regulates intestinal inflammation, resulted in the development of a colitogenic intestinal microbial community [163]. Interestingly, transfer of this colitogenic microbial community to immune intact mice induced colitis demonstrating the causal role of the microbiota in colitis development [163]. Lastly, the administration of broad-spectrum antibiotics to a dnKO mouse model that developed intestinal inflammation similar to human UC eliminated the inflammation [159]. Altogether, these findings clearly demonstrate that the intestinal microbiota influence intestinal inflammation.

Dysbiosis of intestinal microbiota in adult and pediatric IBD patients has been demonstrated in several studies [78, 80, 81, 164-175]. The main findings of these studies are summarized in Table 1.1. In general, IBD microbiota studies demonstrated a link between butyrate-producing bacteria and IBD pathogenesis.

Most studies showed that the relative abundance of major butyrate-producing bacteria (including *Eubacterium*, *Roseburia*, and *Faecalibacterium*) in IBD patients was significantly different than in healthy controls (Table 1.1). However, IBD microbiota studies have shown significant variation in their results, which may be attributed to the sample types and the microbiota characterizing approach (Table 1.1). Furthermore, almost all of the published reports to date have studied the diversity of butyrate producers based on 16S rRNA sequencing and quantification, which identifies only known butyrate producers. The use of a functional gene sequencing approach targeting butyrate biosynthetic genes might circumvent these limitations as described in Chapter 2 in more details.

Table 1.1: The summary of gut microbiota characterization studies on IBD. HC, healthy control; CD, Crohn's disease; ICD, ileal CD; CCD, colonic CD; UC, ulcerative colitis; SSCP, single-strand conformation polymorphism; FISH, fluorescence *in situ* hybridization; DGGE, denaturing gradient gel electrophoresis; NGS, next-generation sequencing.

Cohort	Number of participants			Sample type	Technique	Observed dysbiosis				Ref.
	HC	CD	UC			Firmicutes	Bacteroidetes	Actinobacteria	Proteobacteria	
Adult	31	26	31	Biopsy	SSCP 16S rRNA sequencing qPCR	↓ <i>Eubacterium</i> and ↓ <i>Lactobacillus</i> (IBD)	↓ <i>Bacteroides</i> (IBD)	-----	-----	[164]
Adult	34	35	55	Tissue	16S rRNA sequencing qPCR	↓ <i>Lachnospiraceae</i> (IBD) ↑ <i>Bacillus</i> (IBD)	↓ (IBD)	↑ <i>Bifidobacteriaceae</i> and <i>Coriobacteriaceae</i> (CD)	↑ (IBD)	[78]
Adult	7	13	0	Biopsy	16S rRNA sequencing qPCR FISH	↓ <i>Faecalibacterium</i> and ↓ <i>Subdoligranula</i> (ICD)	-----	-----	↑ <i>Enterobacteriaceae</i> and <i>E. coli</i> (ICD)	[79]
Adult	139	68	0	Stool	DGGE	↓ <i>Clostridium</i> cluster XIVa and <i>F. prausnitzii</i> (CD) ↑ <i>Ruminococcus gnavus</i> (CD)	-----	↓ <i>Bifidobacterium adolescentis</i> (CD)	-----	[176]
Adult	28	0	8	Biopsy	16S rRNA sequencing	↓ <i>Lachnospiraceae</i> and <i>Ruminococcaceae</i> (UC)	-----	↑ (UC)	↑ (UC)	[165]
Adult	6	6	0	Stool	16S rRNA Microarray 16S rRNA sequencing FISH	↓ <i>Clostridial</i> clusters IV and XIVa (CD)	-----	-----	-----	[80]

Adult	35	29	16	Stool Biopsy	16S rRNA NGS	↓ <i>Faecalibacterium</i> and <i>Roseburia</i> (ICD) ↑ <i>Faecalibacterium</i> and <i>Ruminococcus</i> (CCD)	↓ <i>Prevotella</i> (UC) ↓ <i>Hallella</i> (CCD)	↑ <i>Bifidobacteriaceae</i> and <i>Coriobacteriaceae</i> (CCD)	↑ <i>Enterobacteriaceae</i> (ICD) ↓ <i>Streptococcaceae</i> (UC)	[166]
Adult	5	6	5	Biopsy	16S rRNA sequencing	↓ (CD)	↑ (CD)	↑ (IBD)	↑ (CD)	[167]
Adult	6	6	0	Stool	16S rRNA Microarray qPCR	↓ <i>Eubacterium rectale</i> , <i>R. albus</i> , <i>R. callidus</i> , <i>R. bromii</i> , and <i>F. prausnitzii</i> (CD)	↓ <i>Bacteroides fragilis</i> , and <i>B. vulgatus</i> (CD)	-----	↑ <i>Enterococcus</i> , <i>C. difficile</i> , <i>Escherichia coli</i> , <i>Shigella flexneri</i> , and <i>Listeria</i> (CD)	[168]
Adult	5	6	6	Biopsy	16S rRNA sequencing qPCR	↓ (IBD)	↑ (IBD)	-----	↑ <i>Enterobacteriaceae</i> (CD)	[169]
Adult	27	121	75	Biopsy Stool	16S rRNA NGS Metagenomic sequencing	↓ (CD)	-----	↑ (IBD)	↑ <i>Enterobacteriaceae</i> (CD)	[170]
Adult	21	21	34	Stool	qPCR	↓ <i>F. prausnitzii</i> (IBD) ↑ <i>Lactobacillus</i> (CD)	-----	↑ <i>Bifidobacterium</i> (UC)	↑ <i>E. coli</i> (CD)	[171]
Adult	32	16	16	Mucosal lavage	16S rRNA NGS	↓ (IBD)	-----	↑ (IBD)	↑ (IBD)	[172]
Pediatric	12	13	12	Biopsy	16S rRNA NGS qPCR	↑ <i>Clostridial</i> cluster IV and <i>Faecalibacterium</i> (CD)	↓ <i>Parabacteroides</i> (UC)	↓ <i>Coriobacteriaceae</i> (CD)	↓ <i>Burkholderiales</i> (UC)	[173]
Pediatric	221	447	0	Biopsy Stool	16S rRNA NGS Metagenomic sequencing	↓ <i>Clostridiales</i> , <i>Erysipelotrichales</i> , <i>Faecalibacterium</i> , <i>Eubacterium</i> , and <i>Roseburia</i> (CD)	↓ <i>Bacteroidales</i> (CD)	-----	↑ <i>Enterobacteriaceae</i> (CD)	[81]
Pediatric	7	12	7	Biopsy	Conventional culturing qPCR	-----	↓ <i>Bacteroides vulgatus</i> (IBD) ↑ <i>Prevotella</i> (UC) ↓ <i>B.</i>	-----	-----	[174]

1.9. The sample types used for studying intestinal microbiota

Most microbiota studies are conducted using either stool or mucosal biopsies [177-182]. Fecal samples are widely used because they are non-invasive and easily collected. However, it is crucial to comprehend that the microbial communities within stool samples differ significantly from those colonizing the intestinal mucosal surface [81, 129, 136]. Generally, fecal samples have a higher abundance of Firmicutes than Bacteroidetes, the two major phyla of intestinal microbiota [131, 183-185]. On the other hand, mucosal biopsies show a relatively equal abundance of these two phyla [173, 186]. Importantly, similar results were observed when comparing stool samples and biopsies from the same individuals [129, 137, 187]. Because the mucosal microbiota is thought to be responsible for the modulation and maturation of the mucosal immune system [138], it is crucial to focus on this microbial subset in microbiota studies of inflammatory disorders such as IBD.

1.10. Other factors that contribute to functional redundancy

Alterations in the function of the intestinal microbiota can be a consequence of the disappearance of certain members of the microbiota, a shift in their relative abundance, or genetic exchange [188]. As previously described, age promotes the maturation of infant microbiota, which results in the disappearance of some bacterial taxa as one ages [189]. Moreover, shift in microbial abundance can be a result of a change in diet, the consumption of antibiotics, and immune system abnormalities and infections [97, 190]. Functional gene exchange can also be attained through lateral gene transfer (LGT) [191]. As a result of LGT, the functional alteration of the

microbiota may increase the fitness of the commensal microbiota or pathogen microbes toward environmental change(s) and may stabilize the ecosystem.

1.11. LGT and functional redundancy

LGT is an essential element of microbial evolution that allows the microbiota to adapt to their microenvironments. For instance, LGT influence behavior traits of Bacteroidetes, such as nutrient utilization, the composition of its cell surface and the sensing of its environment [192], enhancing the capacity of this group of microbes to survive and thrive in the gut environment. Also, the phylogenetic clustering of the microbial genes involved in carbohydrate metabolism, which was obtained from the metagenome of 36 human gut samples, revealed a high rate of LGT in the gut microbial community [193]. The acquisition of functional genes through LGT may benefit not only the bacterial host but also the human host. For instance, the acquisition of the carbohydrate-active enzymes porphyranases from members of the marine red algae Bacteroidetes to the Japanese gut *Bacteroides plebeius* was mainly attributed to LGT [194]. This acquisition aids in the digestion of resistant polysaccharides present in seaweeds, which is an essential part of the Japanese diet. Interestingly, a comparative metagenomic study of the Japanese gut metagenome to that of the North American gut revealed that this function is unique to this Japanese cohort [194]. Moreover, studies of the genetic structure of the microbial genes involved in butyrate synthesis, the major energy source for colonocytes, strongly support that LGT is responsible for the transfer of these genes between bacterial lineages [195]. Using DNA sequencing and degenerate PCR primers based approach, Louis *et al.* found high level of sequence relationships of

the gene β -hydroxybutyryl-CoA dehydrogenase that is involved in butyrate synthesis between different species of Gram-positive human gut bacteria [195]. The acquisition of MGEs may enhance the microbiota ability to colonize the human gut and overcome the colonization obstacles. For example, plasmids and conjugative transposons are responsible for widespread antibiotic resistance genes (especially erythromycin and tetracycline resistance) in human gut microbiota [196-203]. The total microbial plasmids content known as “the plasmidome” in the intestinal tract also carry other genes that are known to contribute to bile acid resistance, redox balance, adhesion to host cells and bacteriocins synthesis [204-212].

LGT can also contribute to microbiota dysbiosis. For instance, the genus of *Salmonella enterica* harbors a myriad of mobile genetic elements (MGEs), such as plasmids, prophages, transposons, and genomic islands that encode for different fitness and virulence factors [188]. One factor is SopE, a T3SS protein that acts as a guanine nucleotide exchange factor and is encoded on a prophage [213-215]. During infection, SopE enhances host cell invasion [216] and participates in gut inflammation [217, 218]. The inflammation induced by *S. enterica* is required for its successful infection because a mutant strain that lacked the ability to trigger the inflammatory response was overtaken by the gut microbiota in a murine model [219]. The inflammation induced by the wild type *S. enterica* was accompanied by the attenuation of the gut microbiota growth, which is characterized by the reduced relative abundance of *Clostridium*, *Bacteroides*, and *Lactobacillus* [219]. SopE also induces host-derived nitrate, which is subsequently used by the bacteria for

respiration [220]. Eventually, this results in an *S. enterica* bloom that is accompanied by intestinal microbiota dysbiosis.

LGT can also cause chronic diseases. *E. coli* represents the most abundant facultative anaerobic bacterium in the gut ecosystem and harbors a myriad of MGEs [188]. In fact, up to 20% of the *E. coli* genome is composed of various MGEs, including plasmids, prophages, transposons, and genomic islands [221]. These properties allow *E. coli* to be highly adaptable to changes in the gut environment and allow it to bloom following any form of disturbance, including infections, inflammation, or antibiotic usage [188, 222]. In a murine inflammation model, the abundance of *E. coli* was 100 fold higher in $\text{Il10}^{-/-}$ mice than in wild-type mice [223]. The *E. coli* strain carried a polyketide synthase (PKS) genotoxic island that coded for colibactin, a genotoxic compound that induces DNA damage [223]. The abundance of PKS carrying *E. coli* was significantly higher in samples from human IBD and colorectal cancer patients than from healthy controls [223]. Furthermore, the monocolonization of azoxymethane-treated $\text{Il10}^{-/-}$ mice with PKS-carrying *E. coli* significantly increased the risk of invasive carcinoma as compared to the monocolonization with PKS-deficient *E. coli* [223]. Altogether, this highlights the involvement of LGT in chronic diseases.

LGT can be achieved through various ways. First, naked DNA can be acquired from the environment, integrated into the host genome, and expressed; this process is known as transformation [188]. Although transformation is observed in different environments, its frequency in the human gut is very limited due to high level of free-DNA degrading enzymes in this environment [188]. The second

mechanism of LGT is the conjugation of transposons or plasmids. During this process, microbial cells are required to have physical contact through conjugative pilus [224]. In the intestinal tract, this form of genetic material exchange seems to be the most efficient, because the gut is one of the most densely colonized ecosystems [225]. The third mechanism for LGT is phage-mediated transduction, which does not require physical contact between the donor and acceptor microorganisms [226]. Because plasmid conjugation and bacteriophage transduction are more commonly seen in the gut, the following sections will focus on those two components of the MGEs.

1.12. Plasmid contribution to intestinal functions in health and disease

Plasmids are extrachromosomal DNA that can carry functional genes and can be mobilized between different bacteria species, mainly through conjugation [227]. Plasmids enhance bacterial adaptation to new or changing environments [228].

Plasmids often carry functional genes involved in various cellular processes. For example, *Providencia stuartii* isolated from human urine samples carries plasmids involved in nutrient utilization, including urea, lactose, and sucrose [229]. Similarly, *Lactobacillus salivarius*, a known probiotic bacteria that naturally lives in the human gut, carries plasmids that code for carbohydrate utilization, cell wall biosynthesis, amino acid metabolism, and bile salt hydrolase [211]. Furthermore, plasmids can also carry genes that are involved in DNA repair and heavy metal resistance [230].

Plasmids also encode functions that enhance bacterial colonization and survival within the host. Human gut commensal and pathogenic bacteria carry

plasmids that code for antibiotic resistance and adhesion to epithelial cells [205, 231]. For example, enteroaggregative *Escherichia coli* (EAEC), which is responsible for most diarrheal infections in developing countries, carries the virulence plasmid encoding for fimbriae required for its distinctive aggregative adherence to intestinal cells [232]. The EAEC O104:H4 strain also contains plasmid coding for beta-lactamase [233], rendering the bacteria resistant to β -lactam antibiotics. Fimbriae and antibiotic resistance coding plasmids were also identified in the commensal *E. coli* strain SE11 (O152:H28), which was isolated from healthy human stool samples [205]. The presence of this plasmid in commensal bacteria improves their colonization efficiency in the human gut. More interestingly, Stecher *et al.* [234] used a murine colitis model to study the transmission of plasmids between members of the *Enterobacteriaceae* family. Upon the induction of inflammation by *Salmonella enterica* serovar Typhimurium infection, the authors found an increased abundance of *Salmonella* species as well as the gut commensal *E. coli*. Remarkably, this observation was coupled with increased transfer of the colicin-plasmid p2 from the *Salmonella* species to *E. coli*. However, in the absence of intestinal inflammation, gut microbiota minimized LGT [234]. This suggests that inflammation triggers the transfer of plasmids from pathogenic to commensal bacteria *in vivo*, which could not only shift the balance of intestinal microbiota but also spread plasmid-encoded functions such as antibiotic resistance, virulence, and fitness genes. It may also suggest plasmid reassortment in other inflammatory conditions such as inflammatory bowel disease. Together, plasmid gene products can enhance the

colonization and survival of pathogenic and non-pathogenic bacteria in the human gut, which may be a double-edged sword for gut-originated plasmids.

Supporting this hypothesis, human studies of the total gut plasmids content “the plasmidome” have shown a potential link between the gut plasmidome and IBD. A study of the abundance of gut-specific plasmids in the metagenomic data from the human intestinal tract (MetaHIT) database [119] revealed a high abundance of three plasmids in the gut metagenomic sequences of UC and CD patients [191, 235]. This observation strongly suggests an important link between the gut plasmidome and IBD. Because the work of Jones *et al.* [191, 235] represents the only study on the human gut plasmidome, the exact link between IBD and the gut plasmidome is still yet to be determined.

1.13. The influence of bacteriophages on microbiota diversity and functional redundancy

Bacteriophage elements are widely spread in bacterial genomes and are capable of LGT as well as modifying the microbiota composition and functionalities. Bacteriophages are viruses that infect bacteria and integrate their genomes into the host bacterial DNA. Their genetic elements (both functional and non-functional) are commonly found in sequenced bacterial genomes and constitute up to 20% of the genomes of some bacteria [221]. Interestingly, a considerable amount of functional genes within the plasmidome originates from bacteriophages due to their ability to integrate into both bacterial chromosomes and plasmids [221]. Bacteriophages are generally divided into lytic and temperate viruses. The lytic viruses promote the synthesis of new virions, leading to cells lysis and the infection of new cells. On the

other hand, the temperate virus genome replicates along with the infected cell chromosome without expressing harmful genes to the host; however, the lytic genes of temperate viruses can be expressed under certain conditions, such as in response to host cell DNA damage or other environmental stresses [221]. Therefore, bacteriophages are able to modulate the microbial diversity through their selective infection and the lysis of certain bacterial species [236]. Also, bacteriophages contribute to LGT through lysogenic conversion or transduction [188]. During these two processes, host genes or phage accessory genes may transfer into new bacterial species [188]. These mobilized genes carry a range of functions, including antibiotic resistance, nutrient utilization and virulence factors, which could enhance the microbe adaptation to the existing or to a new environment [226].

1.14. The link between bacteriophage and IBD

A number of studies have strongly highlighted the possible link between bacteriophages and IBD. For instance, Lepage *et al.* detected a significantly higher number of phage particles in CD patient biopsies collected from both ulcerated and non-ulcerated areas than in healthy control biopsies [237]. Later, Wagner *et al.* [238] sequenced the bacteriophages isolated from pediatric ileal CD biopsies, CD intestinal washes, and non-IBD control biopsies, utilizing 454-pyrosequencing, interestingly, they found a high abundance of phage DNA related reads in CD intestinal washes and ileal CD biopsies compared to a limited number of phage reads in the controls [238]. Most of these reads belonged to *Caudovirales* bacteriophages. This observation was replicated in a more recent study on the IBD intestinal virome [239]. Norman *et al.* [239] purified and sequenced the virome from

stool samples collected from CD and UC patients as well as from healthy controls. *Caudovirales* bacteriophages were enriched in both CD and UC patients compared to controls. This finding strongly suggests that the virome plays a role not only in IBD pathogenesis but also in the microbiota dysbiosis that is associated with the disease. Knowing that prophages have been detected in bacterial plasmids as well as genomes [226], future IBD plasmidome studies could underscore a possible association between plasmidome-associated prophages and IBD.

1.15. Hypothesis and objectives

Human intestinal microbiotas develop early in life and co-evolve with their hosts. This well-balanced microbial ecosystem serves critical functions and provides the host with many benefits. Thus, any disturbance to intestinal microbiotas could result in either short-term disease or long-term chronic illness such as IBD. Pediatric IBD is significantly different from the adult phenotype because children show more severe clinical symptoms and a different population distribution. Also, pediatric onset is less exposed to environmental factors. Therefore, such patients represent a unique group to study the role of the microbiota in IBD pathogenesis. The overall composition of the intestinal microbiota has been extensively studied in adult and pediatric IBD patients. These studies highlighted a clear dysbiosis in IBD intestinal microbiota that affects several known butyrate-producing bacteria. However, a comprehensive characterization of butyrate producers is still lacking, particularly for the pediatric cohort. Other lines of evidence also underscore a possible link between the gut plasmidome and IBD. Nonetheless, a comprehensive comparative characterization of the gut plasmidomes of IBD and healthy controls is missing.

My PhD study had two main objectives. The first objective was to characterize the diversity of the butyrate-producing bacteria in pediatric non-IBD subjects and to compare it to new onset pediatric Crohn's disease. The second objective was to isolate, characterize, and compare the plasmidomes from non-IBD and IBD (both UC and CD) pediatric subjects. I hypothesize that pediatric Crohn's disease patients have a characteristic dysbiosis of gut butyrate producers as compared to non-IBD controls. I also hypothesize that the composition of the gut plasmidomes of pediatric IBD patients are different than for non-IBD controls and are characteristic of the disease subtypes, i.e., UC and CD.

CHAPTER 2

Diversity of Butyrate-producing Bacteria in Pediatric Crohn's Disease Revealed by Characterizing Butyryl-CoA:acetate CoA-transferase and 16S rRNA Genes

2.1. Introduction

The human intestinal microbiota is composed of approximately 1,000 different bacterial species, most of which remain unculturable, harboring 100 times more genes than humans [119, 240]. These genes have a direct effect on human health by encoding proteins involved in the biosynthesis of beneficial metabolites such as vitamins, lipids, and short-chain fatty acids (SCFAs), [121].

SCFAs are the end product of fermentation of undigested dietary carbohydrates, particularly dietary fibers and resistant starches, by special microbial clusters of the intestinal microbiota [121]. There are three main types of intestinal SCFAs: butyrate, propionate, and acetate [149]. Once produced, 95% of SCFAs are quickly absorbed and utilized by the intestinal cells and other organs, and SCFAs represent about 10% of the total caloric requirements of the human body [150, 151]. Butyrate in particular is the preferred energy source for colonocytes where it is oxidized by the mitochondria and accounts for 10% of the intestinal energy [121, 154]. Butyrate concentrations range from 11-25 mM in human stool [149]. The level of butyrate synthesis from the intestinal bacteria is heavily affected by the

concentrations of the complex carbohydrates in the proximal parts of the intestine [146] and by the environmental pH [241]. For instance, Child *et al.* demonstrated that the level of butyrate was significantly higher at pH 5.5 as compared to pH 6.5 using an *in vitro* system simulating the human colon [241]. Accordingly, the level of three major butyrate producers (*Eubacterium rectale*, *Faecalibacterium prausnitzii*, and *Roseburia*) was significantly higher at pH 5.5 than at pH 6.5 [241]. In humans, the pH of the proximal colon is 5.6 and increases gradually, reaching pH 6.6 in the distal colon [242]. Concordingly, the highest level of butyrate in the human colon is detected at the proximal region [242] constituting the major site for butyrate synthesis [149, 157].

Butyrate is involved in a number of beneficial processes (Figure 2.1.1). For instance, the administration of butyrate restored normal mitochondrial respiration and reduced autophagy in germ-free mice which are characterized by defective mitochondrial respiration and increased colonocytes autophagy [243]. In the human colon cancer cell line HT-29, butyrate increases the level of expression and activity of rhodanese, an essential mitochondrial enzyme involved in hydrogen sulfide detoxification [244]. Butyrate is also crucial for the maintenance of the intestinal barrier. Butyrate stimulates mucin secretion, the first line of defense against intestinal microbes [245]. For instance, intra rectal administration of a known butyrate producer, *Clostridium tyrobutyricum*, was shown to protect mice from dextran sodium sulphate (DSS) induced colitis [246]. Also, the treatment of colon cell lines with butyrate increases the expression of the antimicrobial peptide cathelicidin LL-37 and various tight junction (TJ) proteins, including cingulin,

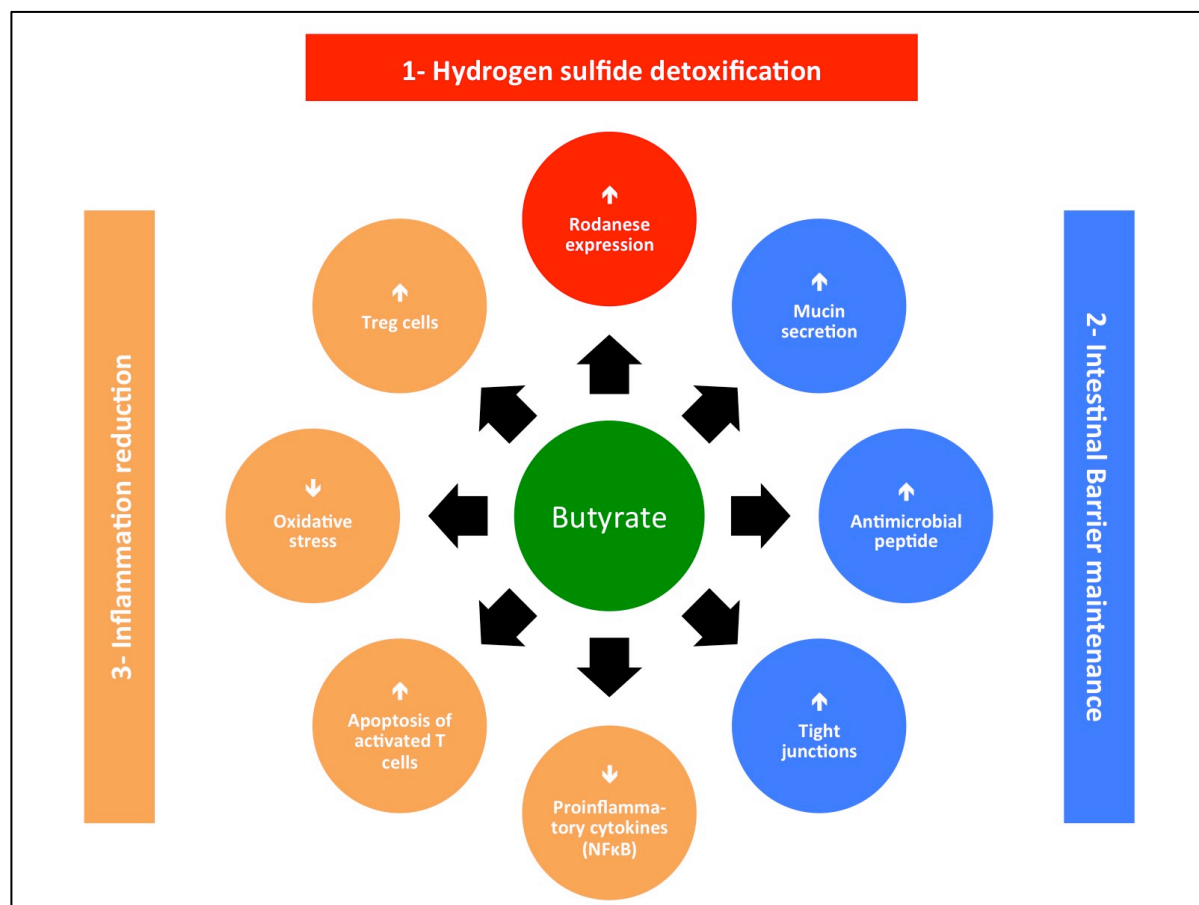


Figure 2.1.1: Summary of beneficial effects of butyrate on intestinal cells.

occludin, ZO-1, and ZO-2, through inhibition of histone deacetylases (HDAC) [247, 248]. Furthermore, butyrate reduces inflammation by many different mechanisms. First, butyrate inhibits the activation of the nuclear factor kappa B (NFκB) [249, 250], which regulates the expression of pro-inflammatory cytokines [251]. Second, butyrate facilitates Fas-mediated apoptosis of T cells through inhibition of HDAC-1 activity, which inhibits INF-γ-induced STAT1 activation and eventually reduces the level of inflammation mediators, including nitric oxide synthase (iNOS) and cyclooxygenase 2 (COX2) [252]. Third, butyrate reduces oxidative stress by increasing the expression of glutathione S-transferase alpha, thus reducing the activation of redox-sensitive NFκB by reactive oxygen species (ROS) [253]. Moreover, the colonization of mice intestines with different butyrate-producing bacteria promoted intestinal immune cell differentiation into the anti-inflammatory T_{reg} cells [254-257]. Finally, two independent studies showed that the microbial metabolite responsible for the T_{reg} differentiation by gut bacteria is butyrate and that the number of colonic T_{reg} cells correlates with butyrate concentration [258, 259]. Thus, butyrate is a crucial modulator of normal intestinal functions.

In agreement with this concept, different studies have reported reduced levels of butyrate-producing bacteria in IBD [78-80, 137]. Using culture-independent shotgun sequencing and qPCR analysis of the 16S rRNA gene, Frank *et al.* demonstrated that OTUs closely related to known butyrate producers were reduced in IBD patients [78]. Using a similar approach on ileal biopsies, Baumgart *et al.* illustrated a significant reduction in *Faecalibacterium*, a major butyrate producer within the Firmicute phylum, relative abundance, in ileal CD (ICD) [79]. On the other

hand, colonic CD (CCD) showed a higher relative abundance of *Faecalibacterium* compared to healthy controls [137]. Using DNA hybridization of 16S rRNA of bacterial DNA extracted from stool samples of 6 control and 6 CD patients, Manichanh *et al.* showed that there was a significant reduction in the *Clostridium coccoides* and *Clostridium leptum* groups, two Clostridial clusters within the Firmicutes phylum mainly dominated by butyrate producers, in CD patients as compared to healthy controls [80]. More recently, Gevers *et al.* illustrated a reduction in the relative abundance of a number of key butyrate producers in treatment-naïve early-onset CD as compared to a control population via next-generation sequencing of the 16S rRNA gene and metagenome [81]. These bacteria included *Eubacterium rectale*, *Faecalibacterium prausnitzii*, and *Roseburia intestinalis*. All together these studies clearly suggest a general reduction of butyrate producers in IBD patients compared to non-IBD controls.

As of today, almost all of the published reports have studied the diversity of butyrate producers using culture-independent techniques involving 16S rRNA sequencing and the quantification of known butyrate producing bacteria. There are nine hypervariable regions (V1–V9) of the 16S rRNA gene that are flanked by phylogenetically conserved regions [260]. The hypervariable regions have significant sequence variability that allows for differentiation between various bacterial species [260]. In the 16S rRNA strategy, oligonucleotide primers can be designed to target a hypervariable region of specific bacterial taxa. Alternatively, the primer could be universal to all or a group of bacteria. Ultimately, the primers are used to amplify and sequence the targeted hypervariable region(s) of the 16S rRNA from metagenomic

DNA. However, characterizing butyrate producers through targeting the 16S rRNA gene has a number of disadvantages. First, the genomes of different bacterial species contain unequal numbers of copies of the 16S rRNA gene (ranging from 1–15 copies per genome), which introduces amplification bias [261]. Moreover, targeting the 16S rRNA gene is unsuitable for identifying novel butyrate producers because it has no metabolic output in butyrate synthesis and requires prior classification of bacteria as butyrate producers. Therefore, the ideal target for such studies would be to characterize a common gene involved in butyrate biosynthesis and that is present only at one copy per bacterial genome.

Butyrate synthesis by intestinal microbiota is a complex process [157]. As illustrated in Figure 2.1.2, butyrate synthesis starts from pyruvate and acetyl-coenzyme A (CoA), which are produced during complex carbohydrate breakdown [262]. Subsequently, two molecules of acetyl-CoA are condensed into butyryl-CoA [262]. Finally butyrate is produced from butyryl-CoA through two different pathways. The first pathway involves transferring the CoA moiety to the acetate molecule by the enzyme butyryl-CoA:acetate CoA-Transferase (BCoAT). In the second pathway, butyryl-CoA is converted into butyryl-phosphate by the enzyme phosphate butyryltransferase, and then to butyrate by the enzyme butyrate-kinase (BUK) [157]. Interestingly, the screening of 38 butyrate producers (isolated from healthy human stool samples) using quantitative polymerase chain reaction (qPCR) and enzymatic assay revealed that the BCoAT pathway is far more prevalent than the BUK pathway for butyrate synthesis [156]. The BCoAT pathway was detected in all isolates and only four were positive for the BUK pathway [156]. More recently, the

screening of the metagenomic sequences of the stool samples from 15 healthy subjects derived from the Human Microbiome Project (HMP) consortium for potential butyrate producers using the hidden Markov model (HMM) revealed that the BCoAT and BUK represent about 78% and 22% of the butyrate synthesis pathways, respectively [262]. Importantly, conserved regions in the BCoAT and BUK gene sequences that flank a hyper variable region of the gene have been identified enabling the design of degenerate universal primers specific to these conserved regions [155, 263]. In turn, these BCoAT universal primers could be used to determine the amount and diversity of butyrate producers from human microbial communities [155, 263]. Because both the BCoAT and BUK genes have one copy per genome and direct metabolic output in butyrate synthesis, this strategy overcomes the limitations of the 16S rRNA strategy.

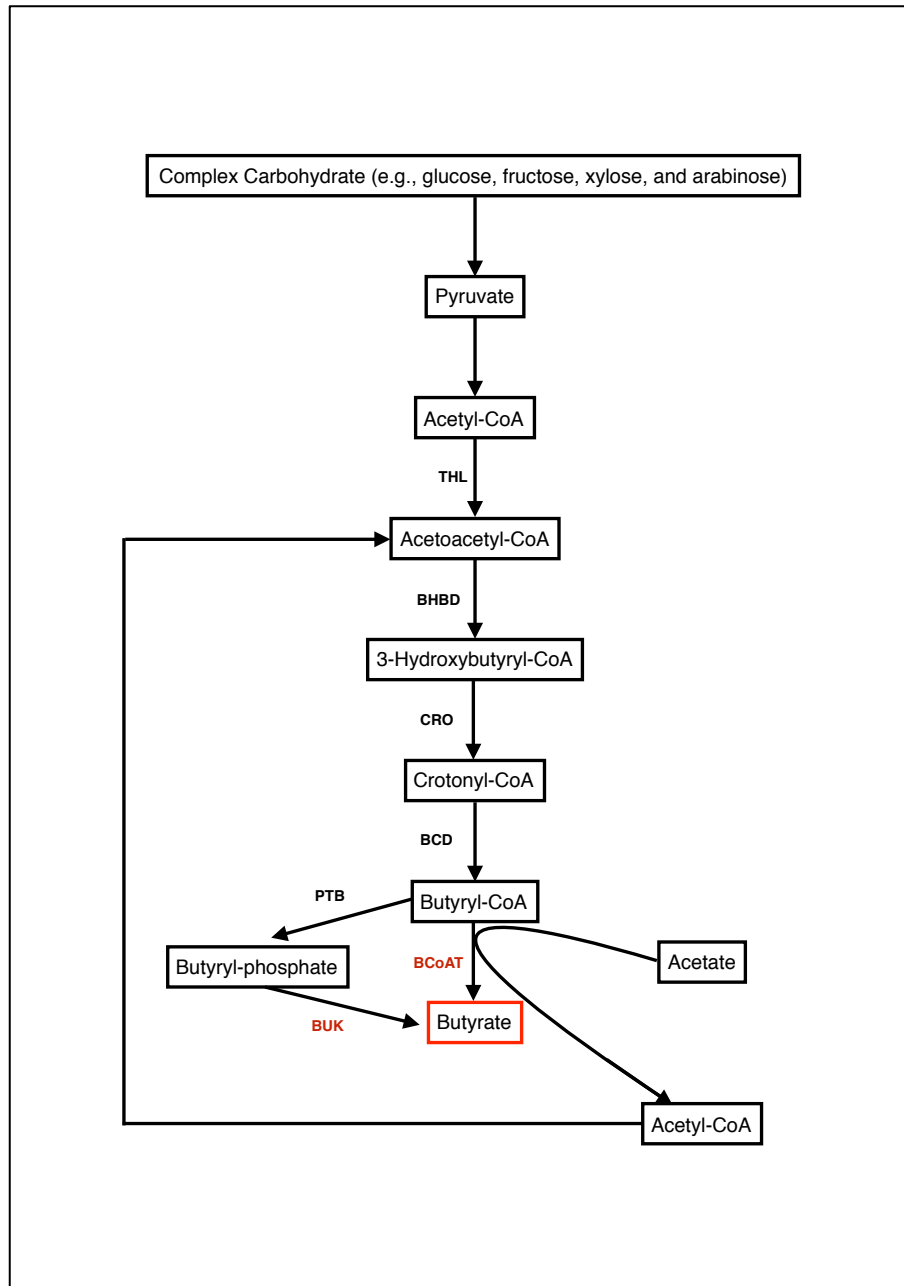


Figure 2.1.2: The simplified overall pathway for butyrate synthesis by gut butyrate producers, highlighting the two main pathways at the terminal step [157]. The genes involved in each step (protein names) are illustrated. The two main pathways at the terminal end are highlighted in red. THL, thiolase; BHBD, beta-hydroxybutyryl-CoA dehydrogenase; CRO, crotonase; BCD, butyryl-CoA dehydrogenase; BCoAT, butyryl-CoA:acetate CoA transferase; PTB, phosphate butyryltransferase; and BUK, butyrate kinase.

To date, there are only two studies that characterized adult human gut butyrate producers by targeting functional genes [155, 263, 264]. Louis *et al.* determined the diversity of butyrate producers that utilize the BCoAT pathway from fecal samples provided by 10 healthy adult volunteers before and after the consumption of inulin [155]. Using cloning and sequencing, they sequenced a total of 1,718 clones and identified 12 known butyrate producers and 20 unclassified OTUs that potentially belong to novel butyrate producers [155]. On the other hand, Vital *et al.* [263] determined the level and diversity of BCoAT and BUK butyrate producers using qPCR and 454 pyrosequencing in the stool aspirate of adult UC patients following colectomies and compared them to healthy controls. Using a higher sequencing depth than Louis *et al.* [155], Vital *et al.* [263] was able to detect 34 BCoAT and 31 BUK butyrate producers. Mariat *et al.* and Agans *et al.* have shown that children carry a different proportion of Firmicutes (the major phylum for butyrate producers) as compared to adults [265, 266], suggesting a community of butyrate producers specific to children.

Here, we sought to characterize this community using a cohort of treatment naïve children with CD. To accomplish this, we collected the mucosa-luminal interface of the ascending colon of children with CD and non-IBD control subjects, and identified their butyrate producers by massively parallel sequencing of the BCoAT gene. This study is the first throughout assessment of the butyrate-producing communities in children with CD and non-IBD pediatric individuals.

2.2. Material and methods

2.2.1. Ethics statement

Ethical approval for our study was attained from the Research Ethics Board of the Children's Hospital of Eastern Ontario (CHEO) with informed consent (or assent) being obtained from the patients and/or parents as indicated.

2.2.2. Patient selection

The targeted subjects for this study were children less than 18 years of age scheduled to have their diagnostic endoscopy at the Children Hospital of Eastern Ontario (CHEO). Eligible subjects must pass a number of exclusion selection criteria which are known to alter the gut microbiota composition. These criteria include 1) body mass index (BMI)-for-age greater than 95th percentile [267], 2) patients with diabetes mellitus [268], 3) patients diagnosed with gastrointestinal tract infection within two months before the diagnostic endoscopy [269, 270], 4) consumption of any medications or food supplements that are known to influence the gut microbiota within the last 4 weeks preceding the endoscopy such as antibiotics [271], immunomodulatory drugs [272, 273], probiotics [274], anti-diarrhea medications, and laxatives [275].

Diagnosis of CD was based on standard clinical, endoscopic, histological, and radiologic criteria [276]. Disease phenotype was assessed using endoscopy and clinical disease activity scores. The site of CD involvement, macroscopic activity, and clinical disease activity were determined using the Paris IBD

Classification [277], Simplified-Endoscopy Score-Crohn's disease [278], and the Pediatric Crohn's Disease Activity Index (PCDAI) [279], respectively. Control subjects were patients scheduled for endoscopy due to rectal bleeding, screening for polyps, or abdominal pain but had normal colon upon macroscopically and microscopically evaluation. Also, any control subject diagnosed with known intestinal inflammatory disorder was excluded. Demographics data (age at time of diagnosis, patient age at time of study, gender, diagnosis, and BMI), environmental exposures (diet, previous antibiotic exposure, and tobacco smoking), and clinical feature were recorded for each participant.

2.2.3. Samples collection

Colonoscopic preparation was done one day ahead of the procedure utilizing standard protocol [280]. During colonoscope entry, and upon reaching the desired intestinal location, colonic mucosa was flushed with sterile water to dislodge the mucosal layer. These mucosal-luminal interface (MLI) samples which contain the strongly adherent mucus layer and the microbiota embedded within it; were aspirated through the colonoscope into a sterile container, immediately placed on ice, and transported to the lab. For this study, only MLI samples collected from the right (ascending) colon were used as this region has been reported to be the most active site for butyrate synthesis [149, 157] and is also a common site for inflammation in pediatric CD [5, 8]. Colonic biopsies from affected and unaffected sites were also collected from the right colon of a subset of the subjects following the mucosa washing and immediately placed on dry ice and transported to the laboratory. In addition, fresh stool samples were collected from a subset of patients

at least 24h prior to the colonoscopy clean-out. All samples were stored at -80°C until further use.

2.2.4. Extraction of metagenomic DNA

DNA was extracted from 5 ml aliquots of the MLI samples, biopsies, and stool samples using the FastDNA® Spin Kit (MP Biomedicals, Santa Ana, California) utilizing two mechanical lysis cycles in a FastPrep® Instrument (MP Biomedicals) set at speed of 6.0 for 40 seconds and following the manufacturer's protocol. Extracted DNA was then stored at -20°C until further use.

2.2.5. Preparation of 16S rRNA sequencing libraries from MLI samples metagenomic DNA for deep sequencing

The 16S rRNA sequencing library was constructed as described previously with few modifications [103]. The V6 hypervariable region of 16S rRNA was amplified and the required adapters and barcodes were integrated into the product using a two-step PCR strategy (illustrated in Figure 2.2.1). The primers for the first PCR were designed to contain, from the 5' to 3', the Illumina paired-end sequencing adapters, 4 to 6 nucleotide barcodes, and the universal V6 primers pair adapted from Sundquist *et al.* [281](Table 2.2.1). Fifty ng of extracted DNA was used in a 50 µl PCR reaction that contains 1X Phusion® High-Fidelity PCR Master Mix (Thermo Fisher Scientific Inc., Waltham, MA) and 0.5 µM of each primer. The amplification condition was as follows: 30 sec initial enzyme activation step at 98°C, followed by 10 cycles of 98°C for 5 sec, 61°C for 15 sec with 1°C drop each cycle and 72°C for 15 sec. Then the amplification was continued with additional 15 cycles using 51°C

as an annealing temperature and ended with a final extension at 72°C for 5 min. The second PCR was conducted using 10 µl of the 1st PCR product, 1X Phusion® High-Fidelity PCR Master Mix (Thermo Fisher Scientific Inc., Waltham, MA), and 0.5 µM of the primers PCRFWD1/PCRRVS1 in a 50 µl PCR reaction. The primers PCRFWD1/PCRRVS1 contain from the 5' to 3', the Illumina flow cell adapters and the Illumina paired-end sequencing adapters (Table 2.2.1). A total of 10 amplification cycles were performed for the second PCR after an initial activation step at 98°C for 30 sec. Denaturation at 98°C for 5 sec, annealing at 65°C 30 sec, and extension at 72°C for 30 sec and 5 min final extension at 72°C were used as the amplification conditions. Each PCR amplicon was inspected on a 1.5% agarose gel, purified using the Montage PCR96 Cleanup Kit (Millipore, Billerica, MA), and the DNA concentration was measured using the Qubit® dsDNA BR Assay Kit (Invitrogen, Carlsbad, CA). Next, 50 ng of amplicon from each sample was pooled and gel purified using the QIAquick Gel Extraction Kit (Qiagen). Finally, the purified pool was sequenced at The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children in Toronto, Canada, on one lane of the Illumina HiSeq 2500 platform.

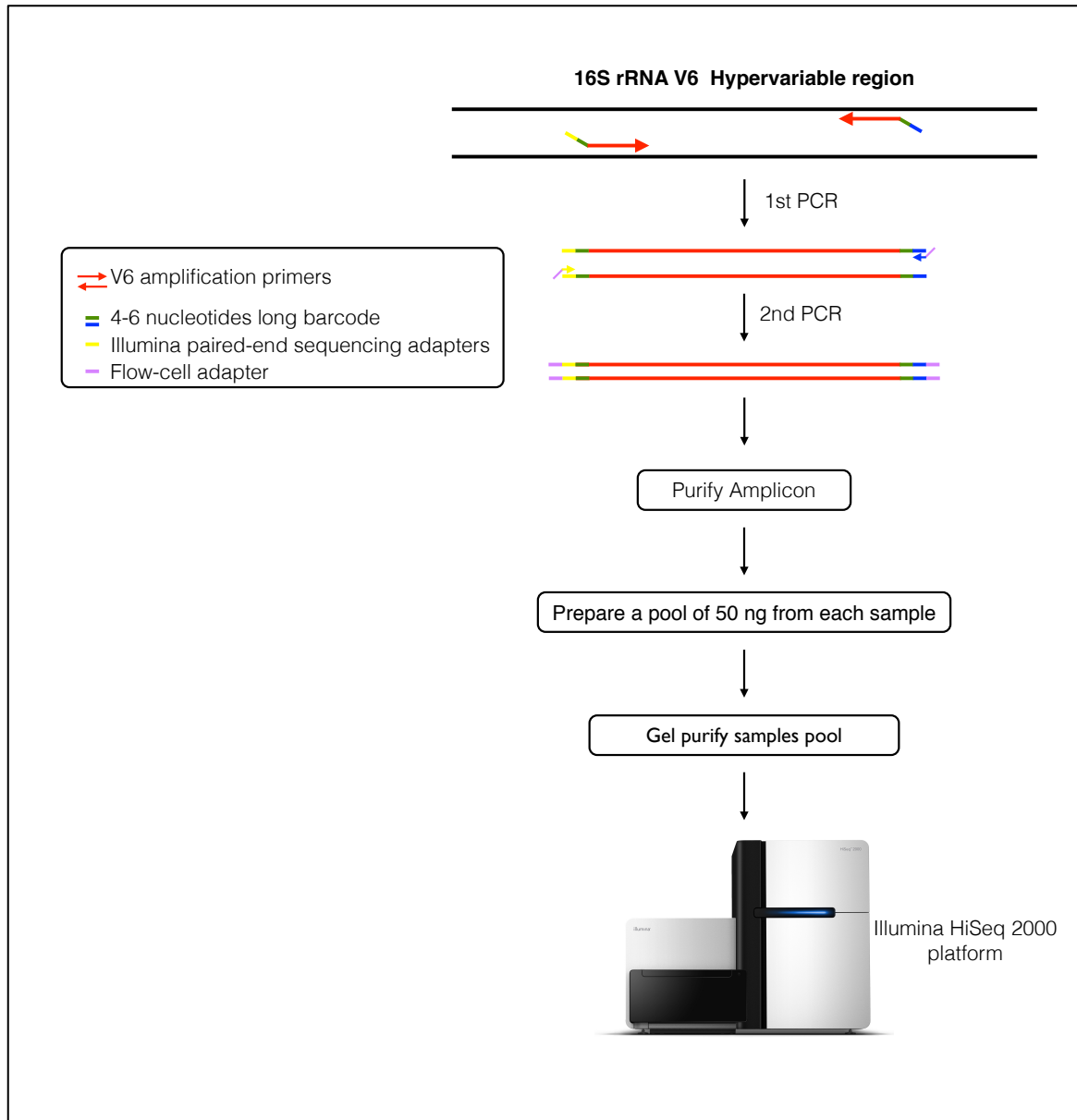


Figure 2.2.1: Illustration of 16S rRNA library construction strategy.

Table 2.2.1: Primers used for 16S rRNA library construction. Primers from MF1 to MF12 and MR1 to MR12 were used in the 1st PCR step. Primers pair PCR-FWD1/PCR-RVS2 was used in the 2nd PCR step. Red nucleotides, universal V6 primers pair adapted from Sundquist *et al.* [281]; green nucleotides, 4-6 nucleotides long barcodes; yellow and blue nucleotides, Illumina paired-end sequencing adapters; and purple nucleotides, Illumina flow-cell adapter.

Primer name	Primer Sequence (5'-3')
MF1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGCGAAACTCAAAGAATTGACGG
MF2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGGTAAACTCAAAGAATTGACGG
MF3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTCATAAACTCAAAGAATTGACGG
MF4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCGTAAACTCAAAGAATTGACGG
MF5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCCGTAAACTCAAAGAATTGACGG
MF6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGTCAAACCTCAAAGAATTGACGG
MF7	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACGTAAACTCAAAGAATTGACGG
MF8	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACGAAACTCAAAGAATTGACGG
MF9	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGACAAACTCAAAGAATTGACGG
MF10	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGAAACTCAAAGAATTGACGG
MF11	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCATAAACTCAAAGAATTGACGG
MF12	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTAAACTCAAAGAATTGACGG
MR1	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTATAGCGAACGAGCTGACGACARCCATG
MR2	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTAGGGTAACGAGCTGACGACARCCATG
MR3	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTTTCATAACGAGCTGACGACARCCATG
MR4	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTGATCGTAACGAGCTGACGACARCCATG
MR5	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTGCCCGTAACGAGCTGACGACARCCATG
MR6	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTCTGTCAAACGAGCTGACGACARCCATG
MR7	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTCACGTAAACGAGCTGACGACARCCATG
MR8	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTCGTACGAACGAGCTGACGACARCCATG
MR9	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTGGACAAACGAGCTGACGACARCCATG
MR10	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTTAGAAACGAGCTGACGACARCCATG
MR11	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTTCATAACGAGCTGACGACARCCATG
MR12	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTACTTACGAGCTGACGACARCCATG
PCR-FWD1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
PCR-RVS1	CAAGCAGAAGACGCCATACGAGATCGGTCTCGGCATTCTGCTGAACCGCTCTTCCGATCT

2.2.6. 16S rRNA sequencing data analysis

The 16S rRNA sequencing data analysis was done as described previously [103]. The paired-end sequences were merged using the Fast Length Adjustment of SHort reads (FLASH) software [282], and uncombined reads were discarded from further analysis. Subsequently, the Novobarcode command, from Novocraft Technologies (www.novocraft.com), was used to demultiplex merged reads according to the 5' and 3' barcode sequences and to trim the barcode sequence from the classified reads. Reads with a minimum quality score of 20 were selected for further analysis using the `fastq_quality_filter` command line from the Fastx-toolkit v0.0.13 (<http://hannonlab.cshl.edu/>). Taxonomy assignment to the genus level was done by aligning each read against the Greengenes database (release 4Feb2011) using the UCLUST Reference-based Operational Taxonomic Units (OTUs) picking method at 97% sequence identity implemented within the Quantitative Insights Into Microbial Ecology (QIIME) software package v1.5.0, [283]. The identified OTUs were then summarized in an OTU table using QIIME, and singletons and doubletons OTUs were discarded from further analysis. Subsequently, the BCoAT butyrate producers were manually selected from the overall microbiota according to the BCoAT butyrate producers list published by Vital *et al.* [263]. The resulting OTU tables of the overall microbiota and BCoAT butyrate producers were used to assess the alpha and beta diversity within and between samples utilizing QIIME. Next, the copy number of the 16S rRNA gene of the BCoAT butyrate producers was normalized to 1 per genome by dividing the number of reads of a given genus by its average 16S rRNA copy number obtained from the ribosomal RNA operons

database (rrnDB) [284]. Then, the normalized relative abundance of the identified BCoAT butyrate producers was compared between control and CD, or control and CD macroscopic inflammation status (normal and inflamed) using Mann-Whitney U test. Finally, the fold change of differential taxa to the control group was calculated by dividing the normalized relative abundance of a given taxa or OTU from each sample by the mean normalized relative abundance of the same taxa from the control group.

2.2.7. Preparation of BCoAT gene libraries from MLI samples metagenomic DNA for deep sequencing

BCoAT library construction was carried out using a two-step PCR strategy (Illustrated in Figure 2.2.2). In the 1st step, 50 ng of metagenomic DNA was used in a 50 μ l PCR reaction containing 1.5 mM $MgCl_2$, 2.5 μ M of the primers pair BCoATscrF/BCoATscrR that were designed to anneal to conserved regions of BCoAT gene from phylogenetically diverse human intestinal bacteria and amplifies 530 nucleotides of the gene (Tabel 2.2.2), 0.2 mM dNTPs, and 2.5U HotStarTaq DNA polymerase (Qiagen, Valencia, California). The mixture was first heated to 95°C for 15 min. Then amplification was carried out for 25 cycles using 94°C for 45 sec, 53°C for 45 sec, and 72°C for 1 min. Finally, the amplification mixture was held at 72°C for 10 min. For the second PCR, 13 fusion primers were designed (12 forward and one reverse) following Roche's Amplicon Fusion Primer Design Guidelines for the GS FLX Titanium Series Lib-L Chemistry. Briefly, the forward primers contain (from 5'-3') the GS FLX Titanium Primer A, a four-base library key, 12 different Multiplex Identifiers (MIDs) barcode sequence, and a BCoATscrF primer

sequence. The reverse primer contains (from 5'-3') the GS FLX Titanium Primer B, a four-base library key, and a BCoATscrR primer sequence. Ten µl of product from the 1st PCR was utilized in a second PCR reaction of 50 µl using the same concentration of PCR component as the 1st PCR. A unique Multiplex Identifier (MID) barcode fusion primer was used for each sample. A total of 15 amplification cycles were performed utilizing the same amplification conditions as the first PCR. For each sample, a total of 5 independent amplification reactions were prepared. Following amplification, PCR products from the same sample were pooled together, visualized on 1.5% agarose gel, and purified using the Montage PCR96 Cleanup Kit (Millipore, Billerica, Massachusetts). Finally, an equimolar amount of samples were pooled together and sequenced on a Roche 454 platform using ¾ plate of GS FLX Titanium chemistry at The McGill University and Génome Quebec Innovation Centre, Montreal, QC, Canada. The sequencing run generated demultiplexed files containing read base calls and per base quality scores for each sample.

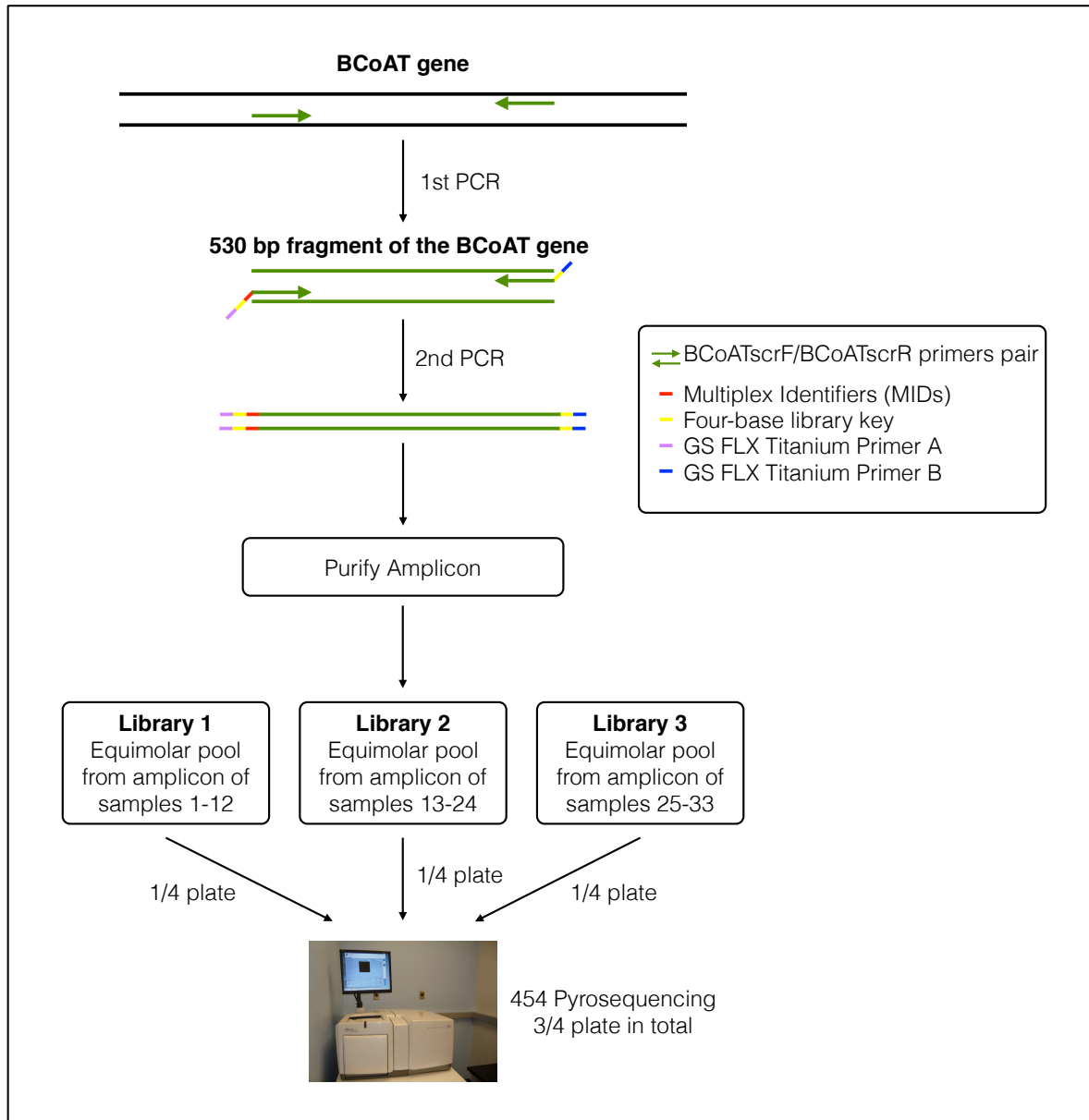


Figure 2.2.2: Illustration of BCoAT library construction strategy.

Table 2.2.2: Primers used for BCoAT library construction. Primers pair BCoATscrF/ BCoATscrR was used for the 1st PCR step of the BCoAT sequencing library. Primers BCoAT-F1 to BCoAT-F14 and BCoAT-R were used for the 2nd PCR step of the BCoAT sequencing library. Purple nucleotides, GS FLX Titanium Primer A; blue nucleotides, GS FLX Titanium Primer B; yellow nucleotides, four-base library key; red nucleotides, Multiplex Identifiers (MIDs); and green nucleotides, BCoATscrF/BCoATscrR primers pair.

Primer name	Primer Sequence 5'-3'	Reference
BCoATscrF	GCIGAI CATTTCACITGGAAYWSITGGCAYATG	[264]
BCoATscrR	CCTGCCTTTGCAATRTCIA CRAANGC	[264]
BCoAT-F1	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGC GTG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F2	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCTCGACAG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F3	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGACGCACTCG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F4	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCACTGTAGG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F5	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGACACGG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F6	CCATCTCATCCCTGCGTGTCTCCGACTCAGATATCGCGAGG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F7	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTGTCTCTAG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F8	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCGCGTGTGCG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F10	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTCTATGCGG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F11	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGATACGTCTG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F13	CCATCTCATCCCTGCGTGTCTCCGACTCAGCATAGTAGTGG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-F14	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGAGATACG CIGAI CATTTCACITGGAAYWSITGGCAYATG	This study
BCoAT-R	CCTATCCCCTGTGTGCCTTGGCAGTCTCAGCCTGCCTTTGC AATRTCIA CRAANGC	This study

2.2.8. BCoAT sequencing data analysis

Demultiplexed reads from each sample were quality filtered using RDP's Pyrosequencing Pipeline (<https://rdp.cme.msu.edu>) based on a minimum quality score of 20 and 200 nucleotides read length cutoff. OTUs clustering at 95% sequence similarity was achieved using the de novo UCLUST algorithm integrated in QIIME V 1.7.0 [283]. Subsequently, singleton and doubletons OTUs were removed. The longest sequence from each OTU was then selected and checked for the presence of chimeric sequences using the UCHIME algorithm [285] integrated in the RDP FunGene Pipeline [286]. Chimeric sequences were removed from subsequent analysis. Then, alpha and beta diversity were computed within and between samples utilizing QIIME. Taxonomy assignment was done as described previously [263]. Briefly, the chimera-free representative sequences were aligned against the *but* database [263], which contains all predicted BCoAT gene sequences from all sequenced bacterial genomes using the RDP FrameBot tool [287] with default settings. At this step, frameshifts introduced by sequencing error were corrected. Then, OTUs with $\geq 75\%$ sequence identity to the closest matches in the *but* reference database [263] were assigned to their representative taxa. OTUs with less than 75% sequence identity to the reference sequences were considered unclassified OTUs. Subsequently, the relative abundance of OTUs was computed by QIIME. To identify differential taxa and OTUs between groups, the relative abundance of each taxa and unclassified OTUs was compared between control and CD, or control and CD macroscopic inflammation status (normal and inflamed) using Mann-Whitney U test. Finally, fold change of differential taxa and OTUs to the

control group was calculated by dividing the relative abundance of a given taxa or OTU from each sample by the mean relative abundance of the same taxa from the control group.

Phylogenetic analysis of the unclassified OTUs was done by aligning the nucleotide sequences of unclassified OTUs to the *but* nucleotide database (downloaded from the Functional Gene Pipeline/Repository, <http://fungene.cme.msu.edu>) utilizing MUSCLE [288]. Then, a phylogenetic tree of the aligned sequences was constructed using maximum-likelihood algorithm with the FastTree tool integrated in QIIME. Visual display of the rooted tree was achieved using the Interactive Tree Of Life (iTOL) tool [289].

2.2.9. Relative quantification of butyrate producers and confirmation of BCoAT gene sequencing results using qPCR

The abundance of butyrate-producing bacteria was determined by relatively quantifying the amount of the BCoAT and BUK genes, utilizing the primers BCoATscrF/BCoATscrR and G_buk_F/G_buk_R, respectively, as previously described [263, 264] (Table 2.2.3 for primer sequences). BCoAT gene was amplified from 50 ng of metagenomic DNA in a 25 µl qPCR reaction containing 1X QuantiTect SYBR Green PCR master mix (Qiagen) and 2.5 µM of BCoATscrF/BCoATscrR primers. The amplification conditions were as follows: 1 cycle of 95°C for 15 min; 40 cycles of 94°C for 15 sec, 53°C, and 72°C each for 30 sec with data acquisition at 72°C. On the other hand, BUK gene was amplified from 50 ng metagenomic DNA in 25 µl qPCR reaction containing 1X QuantiFast SYBR Green PCR master mix (Qiagen) and 0.83 µM of G_buk_F/G_buk_R. BUK gene qPCR amplification started

at 95°C for 5 min followed by 40 cycles of 95°C for 5 sec and 60°C for 10 sec with data acquisition at 60°C. The 16S rRNA gene was amplified simultaneously with BCoAT and BUK genes using the universal primers UniF/UniR adapted from Walker *et al.* [290]. Two qPCR reactions were performed for each sample. Delta-Ct (Δ Ct) for each target (BCoAT or BUK genes) was calculated by subtracting the Ct value of the 16S rRNA from the target's Ct. Then, the Δ Ct values for BCoAT and BUK genes were compared between groups.

In order to validate the BCoAT sequencing results, primers specific to the BCoAT genes of *E. rectale* and *F. prausnitzii* were used in qPCR. In addition, primers targeting the 16S rRNA gene of *E. rectale*/ Clostridium coccoides group (XIVa) and *F. prausnitzii* were used to validate the 16S rRNA based sequencing data. Primer sequences are listed in Table 2.2.3. Validation qPCR was done using QuantiFast reagent (Qiagen) applying similar settings as in the BUK gene assay and using 0.5 μ M of each primer. Δ Ct calculation and comparison was done as described for the BCoAT and BUK genes.

Table 2.2.3: Primers used for qPCR.

Target	Primer Name	Primer Sequence	Ref.
Universal 16S rRNA	UniF	GTGSTGCAYGGYYGTCGTCA	[290]
	UniR	ACGTCRTCCMCNCCTTCCTC	
BCoAT gene	BCoATscrF	GCIGAICATTTACITGGAAYWSIT GGCAYATG	[264]
	BCoATscrR	CCTGCCTTTGCAATRTCIACRAANG C	
BUK gene	G_buk_F	TGCTGTWGTGGWAGAGGYGGA GCAACIGCYTTTTGATTTAATGCAT	[263]
	G_buk_R	GG	
<i>E. rectale</i> / <i>Clostridium</i> <i>coccoides</i> group 16S rRNA gene	UniF338 C.cocR491	ACTCCTACGGGAGGCAGC GCTTCTTAGTCAGGTACCGTCAT	[269]
<i>F. prausnitzii</i> 16S rRNA gene	Fprau 07	CCATGAATTGCCTTCAAACTGTT GAGCCTCAGCGTCAGTTGGT	[291]
	Fprau 02		
<i>E. rectale</i> BCoAT gene	RosEub_F	TCAAATCMGGIGACTGGGTWGA TCATAACCGCCCATATGCCATGAG	[263]
	Eub_R		
<i>F. prausnitzii</i> BCoAT gene	Fprsn_F	GACAAGGGCCGTCAGGTCTA	[263]
	Fprsn_R	GGACAGGCAGATRAAGCTCTTGC	

2.2.10. Statistical Analysis

The qPCR results were analyzed by two-tailed Mann-Whitney U test when comparing control to CD or Kruskal-Wallis test with either a Dunn's post-hoc or Bonferroni correction for multiple-comparison (e.g. control against inflammation status (normal or inflamed) and control against CD patients with mild, moderate, or severe inflammation as determined using the PCDAI). For sequencing data, significant taxa or unclassified OTUs were identified by comparing the relative abundance of each taxa and unclassified OTUs between control and CD or control and CD inflammation status (normal or inflamed) using two-tailed Mann-Whitney U test. *P* values ≤ 0.05 were considered significant. Differential taxa and unclassified OTUs were presented as fold change to the control group calculated by dividing the relative abundance of a given taxa or OTU from each sample by the mean relative abundance of the same taxa from the control group. Correlation between BCoAT and 16S rRNA sequencing was done by calculating the Spearman's rank correlation coefficient (*r*) of paired relative abundance of bacterial taxa identified by the two approaches. Statistical tests were performed using SPSS V.22.0 (IBM Corporation, Armonk, NY), XLSTAT 2014 (Addinsoft, NY), and GraphPad Prism version 6.0c (GraphPad, La Jolla, CA).

2.3. Results

2.3.1. Patient cohort and disease characteristics

As shown in Table 2.3.1 and Appendix I, a total of 119 patients were included in this study (51 control subjects and 68 CD patients). The diagnosis of the participating patients was confirmed by histological examination. Control subjects included patients undergoing endoscopy for rectal bleeding, screening for polyps, or abdominal pain. All controls had both a visually normal macroscopic appearance of the mucosa and normal histology of the mucosal biopsies. CD samples included 30 patients with normal macroscopic appearance of the right colon and 38 with an inflamed right colon. Among the CD patients, there were different levels of disease severity according to PCDAI index including inactive (n=1), mild (n=12), moderate (n=12), and severe (n=43). The site of the macroscopic inflammation of CD was mostly ileocolonic (n=32) followed by isolated ileal CD (n=22) and a minority with colonic only CD (n=13). The average age of participants was 13.12 ± 3.81 for the control group and 13.0 ± 2.64 for the CD patients. The number of male:female patients was 24:27 and 45:23 for control subjects and CD patients, respectively. MLI samples were collected from the right colon of all 119 patients at the time of diagnosis (inception cohort). Ten mucosal biopsies were also collected from control subjects and 22 from CD patients (CD patients biopsies were collected from non-affected (n=12) and affected (n=10) sites). Furthermore, fresh stool samples were collected one day before the colonoscopy cleanout from 5 control subjects and 10

CD patients (6 CD patients with normal macroscopic appearance of the right colon and 4 with inflamed right colon) (Table 2.3.1; Appendix I).

Table 2.3.1: Cohort characteristics.

	Crohn's disease	Controls
Number of Patients	68	51
Age at Colonoscopy (years)	13±3.81	13.1±2.64
Gender		
Male	45	24
Female	23	27
Disease Severity*		
Inactive	1	
Mild	12	
Moderate	13	
Severe	42	
Macroscopic Mucosal Appearance (Normal: Inflamed)	30:38	51:0
Site of Disease		
CD: Upper Disease	1	
CD: Ileal	22	
CD: Ileocolonic	32	
CD: Colonic only	13	
Number of Biopsies (Non-affected: affected)	12:10	10:0
Number of Fresh Stool Samples (Normal right colon: Inflamed right colon)	6:4	5:0

**Disease severity as determined by PCDAI: Pediatric Crohn's Disease Activity Index [292].*

2.3.2. BCoAT but not BUK butyrate producers are reduced in CD patients with colonic inflammation

To determine if the level of butyrate producers is reduced in CD patients as compared to controls, the relative abundance of the BUK and BCoAT genes from right colon MLI and biopsy samples was analyzed by qPCR. The relative abundance of the BCoAT gene, but not of the BUK gene, was found to be decreased at the MLI of CD patients as compared to control subjects (Figure 2.3.1 E). The relative abundance of the BUK and BCoAT genes showed no significant difference when the relative abundance of both genes was compared as a function of gender (Figure 2.3.1 D and H). Further grouping of the CD subjects according to macroscopic inflammation showed that CD patients with inflamed colon had reduced level of BCoAT gene ($P<0.05$), as indicated by a higher ΔC_t value compared to control subjects (Figure 2.3.1 F). The ΔC_t values for the CD patients with normal colon were intermediate between the controls and CD subjects with inflamed colon respectively (Figure 2.3.1 F). This trend suggests a gradual loss of BCoAT butyrate producers in CD subjects together with the increase in macroscopic inflammation status. Stratification of CD patients according to disease severity (as determined by the PCDAI score) revealed a significant reduction in the relative abundance of the BCoAT gene in the MLI samples from patients with mild inflammation as compared to the controls (Figure 2.3.1 G). No statistically significant reduction was observed at the MLI of patients with moderate or severe clinical features of inflammation. Together, this indicates that the reduction in BCoAT butyrate producers correlates with colonic inflammation. The relative abundances of the BUK and BCoAT genes

were analyzed using the mucosal biopsies from non-affected (n=12) and affected (n=10) sites of the right colon of CD patients as well as from control subjects (n=10). No statistically significant difference could be observed (Figure 2.3.2), indicating that only the butyrate producers from the MLI that utilize the BCoAT pathway are reduced in children with CD.

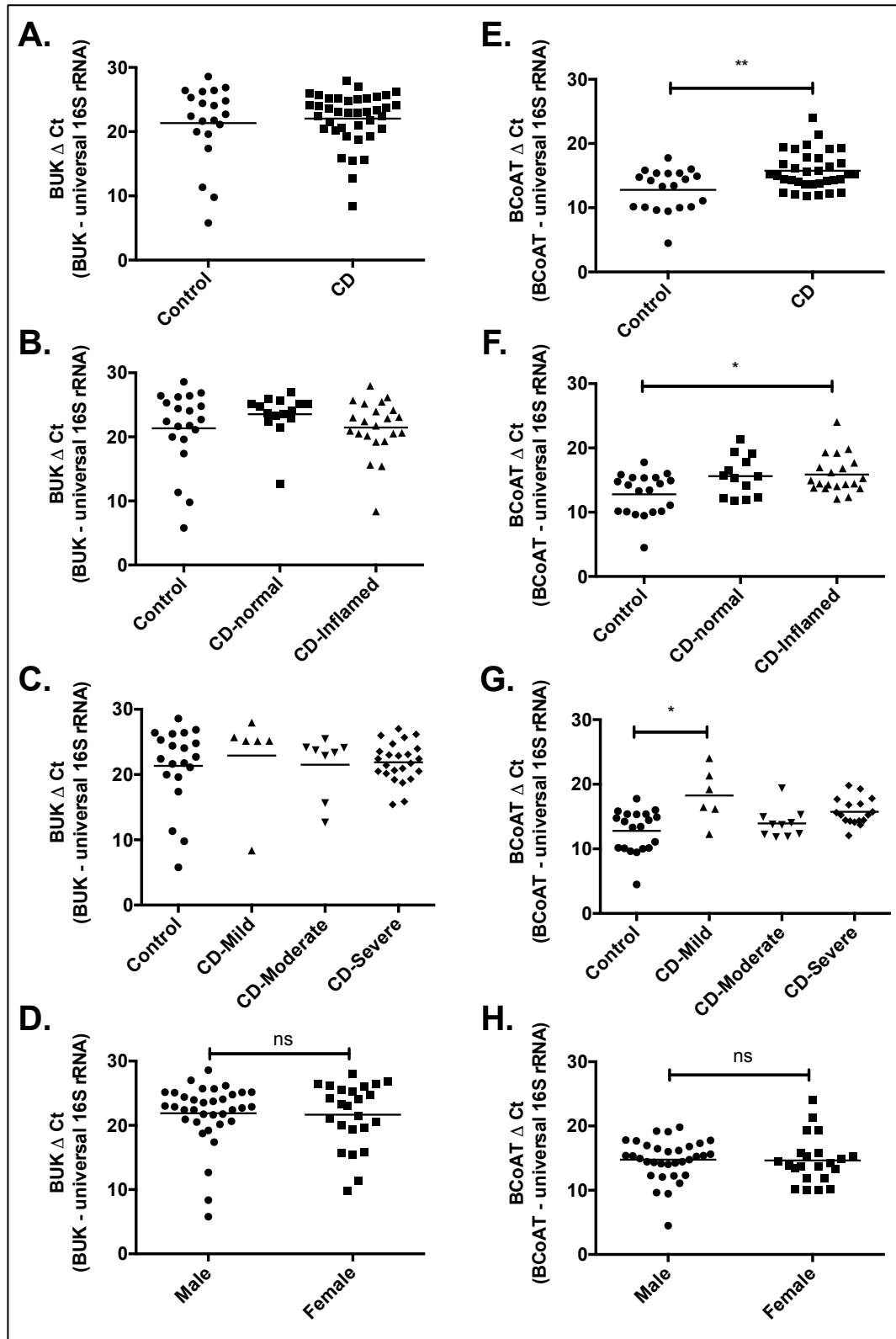


Figure 2.3.1: Relative quantification of butyrate kinase (BUK) and butyryl-CoA:CoA transferase (BCoAT) genes to total bacteria 16S rRNA from right

colon MLI samples. (A and E) BUK and BCoAT genes quantified from control and CD samples. (B and F) CD samples were divided according to colon appearance during sample collection into normal and inflamed. (C and G) CD samples were subclassified according to disease severity into mild, moderate, and severe based on the PCDAI: Pediatric Crohn's Disease Activity Index [292]. D and H represent the level of BUK and BCoAT genes, respectively, according to patients gender. Result is expressed as ΔC_t to total bacteria 16S rRNA. Left panels (A-D) represent the data for the BUK gene, and right panels (E-H) represent data for BCoAT gene. ns, non-significant; *, $P < 0.05$; **, $P < 0.01$.

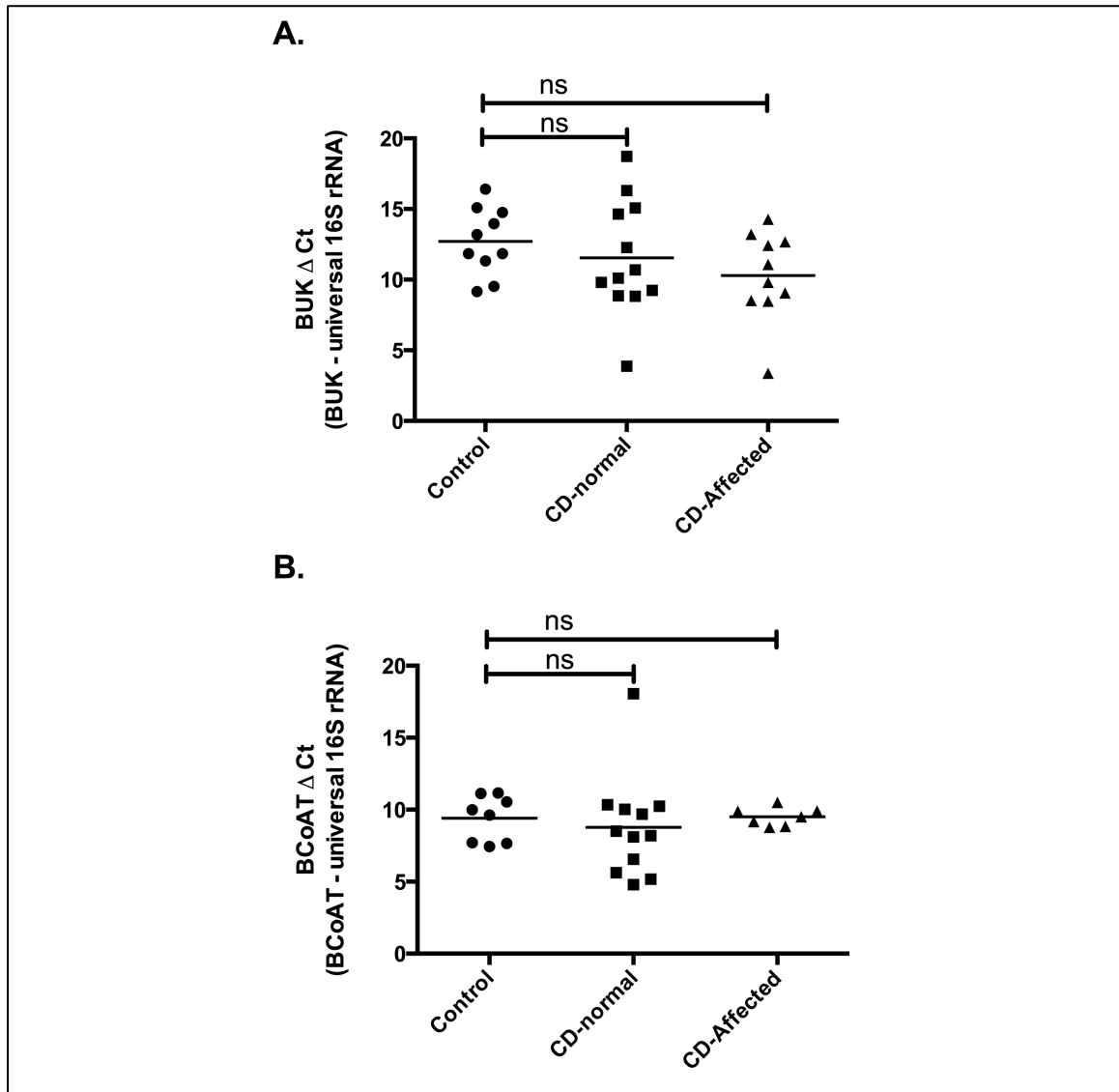


Figure 2.3.2: Relative quantification of butyrate kinase (BUK) and butyryl-CoA:CoA transferase (BCoAT) genes to total bacteria 16S rRNA from biopsies. Biopsies were collected from right colon of control subjects as well as from the normal and inflamed right colon areas of CD patients. Result is expressed as Δ Ct to total bacteria 16S rRNA. (A) Data for the BUK gene; (B) Data for BCoAT gene. ns; $P > 0.05$.

2.3.3. The diversity of BCoAT butyrate producers, revealed by 16S rRNA sequencing, is different in CD patients as compared to the control subjects

The diversity of the overall microbiota and BCoAT butyrate producers was assessed by sequencing the V6 hypervariable region of the 16S rRNA gene as described in section 2.2.5. The total number of identified OTUs (97% sequence identity) for the overall microbiota was 3,288 OTUs. Among these OTUs, 1,970 were shared between control and CD samples, 393 were unique to controls, and 925 were exclusive to CD samples (Appendix II). Chao1 and Shannon indices using equalized data sets to 10,930 reads showed no differences in OTUs evenness and richness between CD samples and controls (Appendix II). Furthermore, the control and CD samples were not separated by multidimensional scaling analysis of UniFrac metrics, presented by principal coordinates analysis (PCoA) plots of both unweighted and weighted UniFrac distances (Appendix II). The BCoAT butyrate producers identified by the 16S rRNA sequencing were represented by 278 OTUs. The control and CD samples shared 191 OTUs, and 39 and 48 OTUs were unique to the control and CD samples, respectively (Figure 2.3.3 A). Interestingly, both Chao1 and Shannon indices of BCoAT butyrate producers were significantly reduced in CD patients as compared to the control subjects (Figure 2.3.3 B). This result indicates that the number of OTUs (richness) and the distribution of sequences within the OTUs (evenness) in the CD samples are lower than in the control samples. Moreover, multidimensional scaling analysis of unweighted UniFrac metrics, presented by PCoA plot, seems to separate the CD patients from the controls with 23.85% of the variance accounted for by coordinate 1 (PCoA1) and

19.24% of the variance attributable to PCoA2 (Figure 2.3.3 C). No clustering of CD and control samples was observed upon using the weighted UniFrac distances in PCoA plot (Figure 2.3.3 C).

A total of 10 genera of butyrate producers were identified with the 16S rRNA approach. The majority of reads were assigned to 4 genera that belong to the Firmicutes phylum (arranged from highest to lowest abundance): *Roseburia* (34.52%±28.66), *Coprococcus* (25.25%±22.01), *Eubacterium* (21.66%±22.95), and *Faecalibacterium* (18.09%±21.13) (Figure 2.3.4). Furthermore, the 16S rRNA approach identified 6 low abundant butyrate producers (total abundance of the 6 genera is less than 1%): *Peptoniphilus*, *Anaerofustis*, *Anaerostipes*, *Butyrivibrio*, *Megasphaera*, and *Treponema* (Figure 2.3.4).

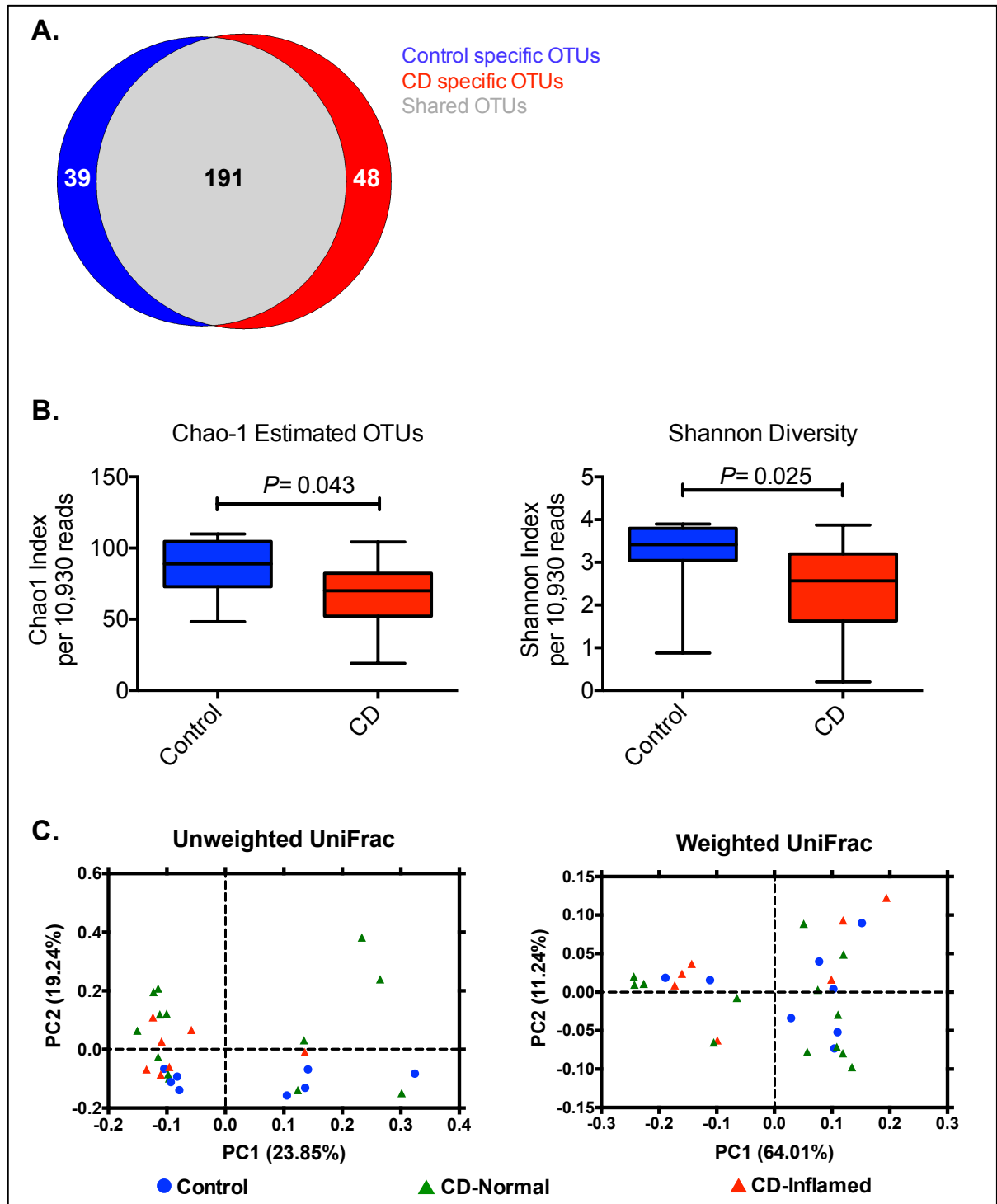


Figure 2.3.3: Diversity of butyrate-producing bacteria revealed by 16S rRNA gene sequencing. (A) Number of Observed Operational Taxonomic Units (OTUs) at 97% sequence similarity. (B) Alpha diversity represented by Chao1 estimated

OTUs (left panel) and Shannon diversity index (right panel). Number of reads was equalized between samples at 10,930 reads. Control, blue bar; CD, red bar. (C) Beta diversity presented by two-dimensional principal coordinates analysis (PCoA) plot of unweighted (left panel) and weighted (right panel) UniFrac distances. Percentage of variance explained by each component is presented for each axis. Control samples, blue circles; CD with normal right colon samples, green triangles; CD with inflamed right colon samples, red triangles.

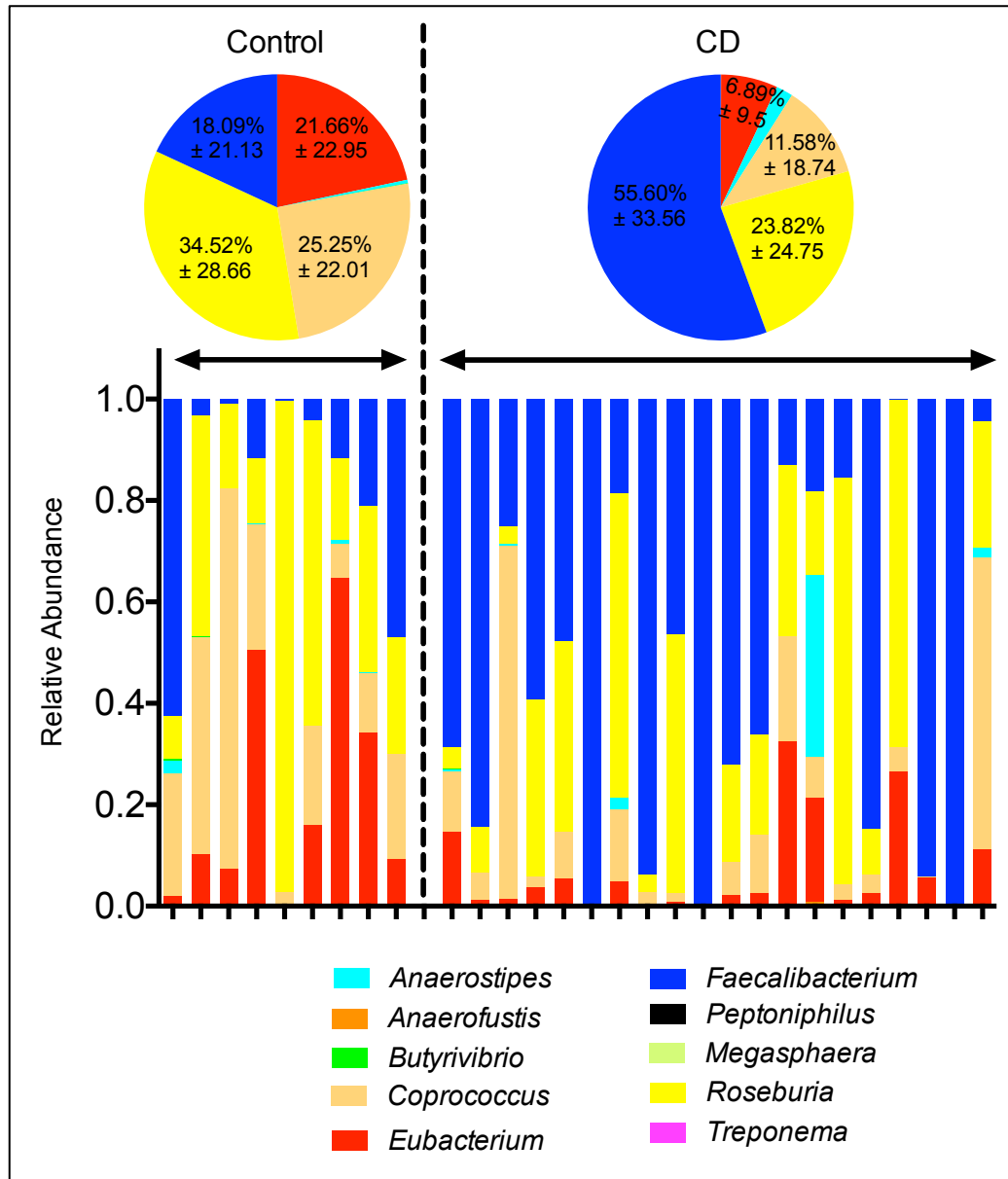


Figure 2.3.4: Butyrate producers, at the genus level, identified by V6 hypervariable region of 16S rRNA sequencing. Pie charts represent the relative abundance of butyrate producers in each group. Stacked bar represent the relative abundance of butyrate producers in individual samples.

2.3.4. The diversity of BCoAT butyrate producers is different in CD patients as compared to the control subjects

To study the diversity of BCoAT butyrate producers, we sequenced a hyper-variable region of the BCoAT functional gene as described in section 2.2.7. A total of 522,370 high quality reads were generated from 33 samples (13 controls and 20 CD; Appendix I) with an average of 15,830 reads per sample (range 44,092-2,919) and an average read length of 465 nucleotides (summarized in Table 2.3.2). Clustering reads at 95% sequence similarity resulted in a total of 802 OTUs with a total OTU number of 649 for controls and 736 for CD. The majority of the observed OTUs were shared between the two groups (583 OTUs) (Figure 2.3.5 A). The Chao1 and Shannon diversity indices were obtained by equalizing the data sets to 2,910 reads. Interestingly, CD samples exhibited a significantly lower Chao1 as compared to the controls ($P=0.03$). In contrast, no difference was observed in the Shannon index (Figure 2.3.5 B). This result indicates that the distribution of sequences within the OTUs in CD subjects is similar to control (evenness), but that the CD samples display lower number of OTUs (richness).

Furthermore, multidimensional scaling analysis of UniFrac metrics, presented by principal coordinates analysis (PCoA) plot, showed that the butyrate producing microbiota of the CD patients and the control subjects were different (Figure 2.3.5 C). Using weighted UniFrac clustering, most control and CD subjects were grouped into two distinct clusters with 32.25% of the variance accounted for by coordinate 1 (PCoA1) and an additional 15.23% of variance attributable to PCoA2. On the other hand, the CD samples were poorly separated from the controls in the unweighted

UniFrac analysis (Figure 2.3.5 C). This suggests that it is the species weight that distinguishes the two groups not the presence of unique bacterial species.

Table 2.3.2: Summary of BCoAT pyrosequencing data.

Sample	Group	Colonoscopy visual appearance of right colon	Number of high quality reads	OTUs at 0.05	Chao-1 at 0.05	Number of Classified Taxa
Subject 1	Control	Normal	12,444	299.9	428.3	7
Subject 2	Control	Normal	15,606	248.2	388.5	9
Subject 3	Control	Normal	5,953	21.3	22.3	3
Subject 4	Control	Normal	7,949	139.4	216.3	7
Subject 5	Control	Normal	14,269	146.3	192.6	6
Subject 6	Control	Normal	44,092	132.9	173.6	7
Subject 7	Control	Normal	15,566	146.9	202.3	5
Subject 8	Control	Normal	14,769	125.1	156.5	5
Subject 9	Control	Normal	19,747	216.2	288.4	7
Subject 10	Control	Normal	30,680	114	162.0	5
Subject 11	Control	Normal	2,919	64.1	142.2	6
Subject 12	Control	Normal	6,993	148.8	248.1	10
Subject 13	Control	Normal	28,212	285.2	425.9	11
Subject 14	CD	Normal	22,687	126.6	163.7	7
Subject 15	CD	Inflamed	5,713	54.4	71.5	6
Subject 16	CD	Normal	16,912	141.1	202.5	6
Subject 17	CD	Inflamed	3,816	54.7	87.5	6
Subject 18	CD	Inflamed	15,512	137	180.0	5
Subject 19	CD	Normal	21,122	144.8	205.7	7
Subject 20	CD	Normal	19,264	162.2	233.1	8
Subject 21	CD	Normal	5,092	55.3	84.2	6
Subject 22	CD	Inflamed	16,557	250.2	313.1	8
Subject 23	CD	Normal	14,178	82.7	97.9	8
Subject 24	CD	Inflamed	14,992	139	214.6	9
Subject 25	CD	Inflamed	4,742	90.6	144.7	8
Subject 26	CD	Normal	13,194	68.3	78.9	9
Subject 27	CD	Inflamed	10,260	148	213.3	7
Subject 28	CD	Normal	16,539	198.5	308.2	7
Subject 29	CD	Normal	20,768	157.5	229.7	8
Subject 30	CD	Normal	29,275	97.6	137.5	6
Subject 31	CD	Normal	28,242	68.4	72.9	8
Subject 32	CD	Normal	5,252	63.4	92.0	5
Subject 33	CD	Normal	19,054	102.5	148.7	7

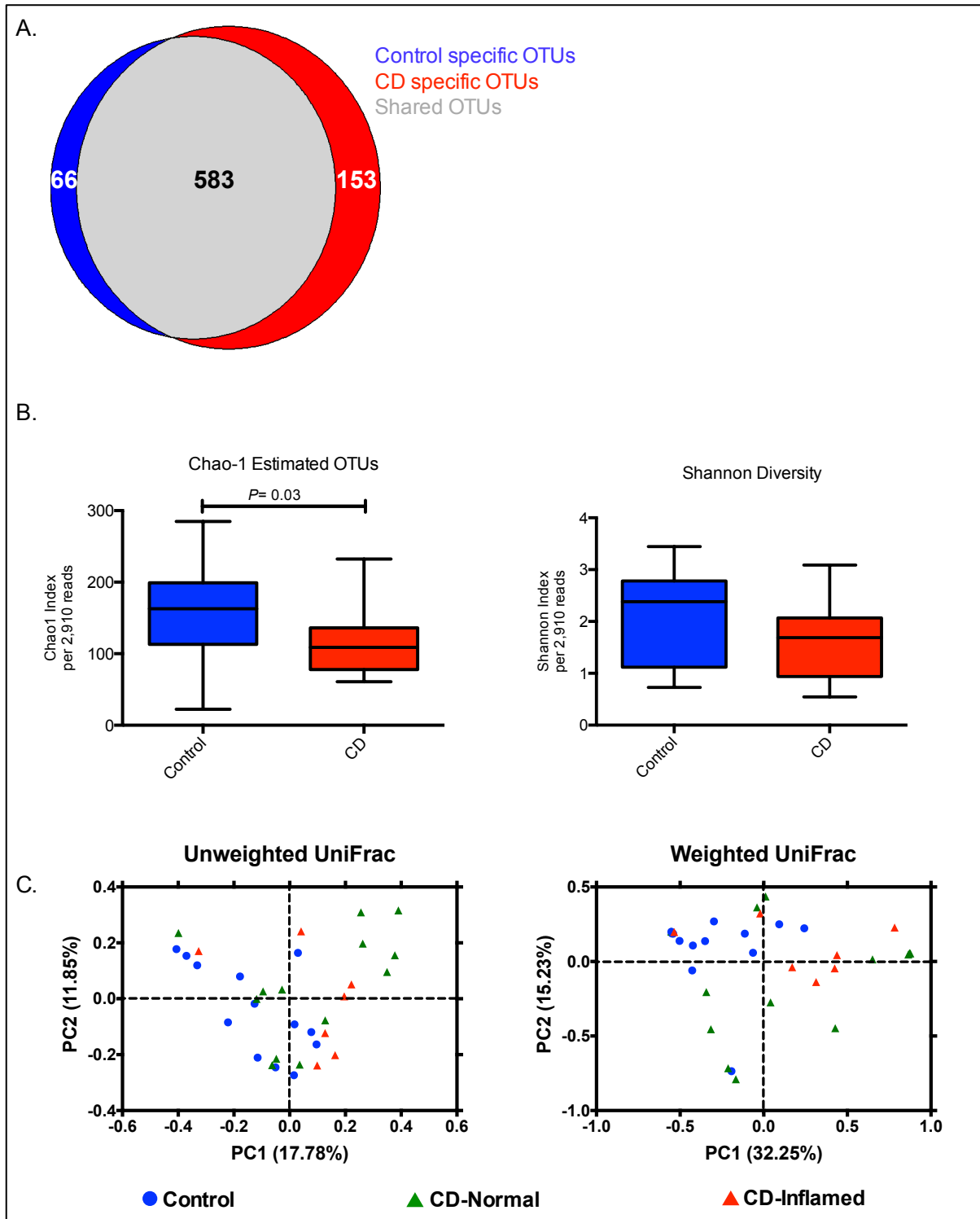


Figure 2.3.5: Diversity of butyrate-producing bacteria revealed by BCoAT gene pyrosequencing. (A) Number of Observed Operational Taxonomic Units (OTUs) at 95% sequence similarity. (B) Alpha diversity represented by Chao1 estimated OTUs (left panel) and Shannon diversity index (right panel). Number of reads was equalized between samples at 2,910 reads. Control, blue bar; CD, red bar. (C) Beta

diversity presented by two-dimensional principal coordinates analysis (PCoA) plot of unweighted (left panel) and weighted (right panel) UniFrac distances. Percentage of variance explained by each component is presented for each axis. Control samples, blue circles; CD with normal right colon samples, green triangles; CD with inflamed right colon samples, red triangles.

2.3.5. Composition of butyrate-producing bacteria revealed by BCoAT sequencing

The composition of BCoAT butyrate producers was also characterized by sequencing the BCoAT gene from control and CD samples. Analyzing the diversity of butyrate producers by this functional gene sequencing approach revealed similar results to the 16S rRNA sequencing with minor differences. Of the total OTUs, 92.5% of the OTUs could be assigned to 12 known butyrate producers while 7.5% of the OTUs were unclassified at the phylum level (Figure 2.3.6).

In the control group, the majority of butyrate-producing bacteria belonged to the Clostridium cluster XIVa. These bacteria included *Eubacterium rectale* (58.7%±29.6), *Eubacterium hallii* (8.5%±8.9), *Roseburia inulinivorans* (2.7%±7.2), *Roseburia hominis* (1%±2.9), *Coprococcus catus* (0.18%±0.65), and *Roseburia intestinalis* (0.05%±0.14). The control group butyrate producers' consortium also harbors *Faecalibacterium prausnitzii*, a member of the Clostridium cluster IV, with a relative abundance of 14.3%±17.4. Other members of the control group butyrate producers included *Clostridium* sp. SS2/1 (11.2%±22.1), butyrate-producing bacterium SS3/4 (0.03%±0.08), *Clostridium* sp. M62/1 (0.018%±0.067), and *Ruminococcus* bacterium D16 ($2.7 \times 10^{-6}\%$ ± $9.8 \times 10^{-6}\%$) (Figure 2.3.6). Also, there were 47 unclassified OTUs making up 3.1%±4.4 of all BCoAT butyrate producers in the control group. Calculating paired Spearman's rank correlation coefficient (r) for the relative abundance of butyrate producers identified by 16S rRNA and BCoAT sequencing revealed a strong correlation between the two datasets ($r=0.74$).

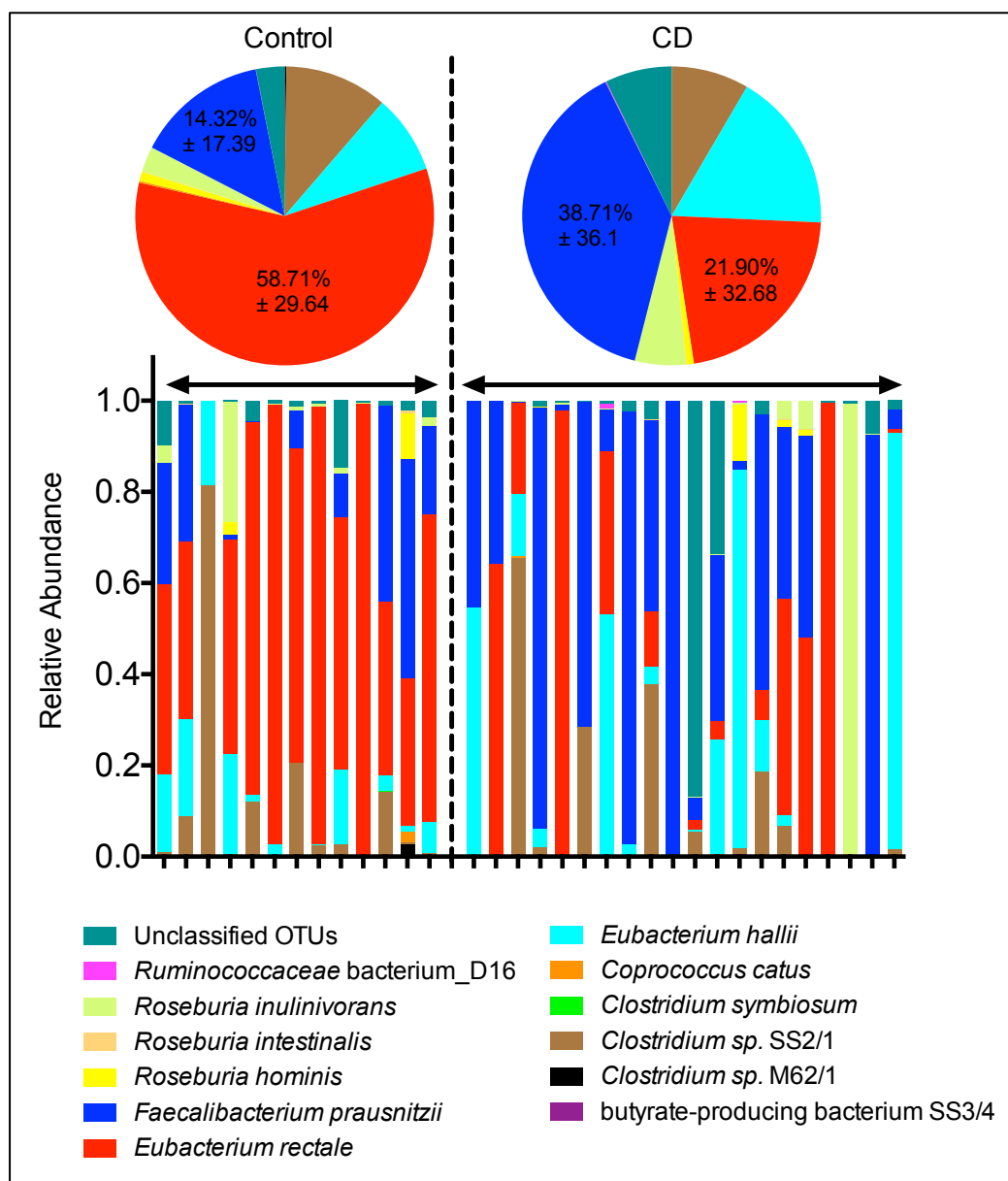


Figure 2.3.6: Identified butyrate-producing bacteria at the species level using BCoAT sequencing. Pie charts illustrate the relative abundance of butyrate producers in each group. Stacked bars represent relative abundance of butyrate producers in individual sample.

2.3.6. *Eubacterium rectale* and *Coprococcus* are depleted and *Faecalibacterium prausnitzii* thrives in the microbiota of the CD patients as compared to control subjects

In order to identify differentially abundant butyrate producers, we compared the relative abundance of the BCoAT butyrate producers, assigned by the functional or 16S rRNA gene approaches, between the CD patients and the control subjects. Using the functional gene approach, we observed a 2.7-fold decrease in the relative abundance of *E. rectale* in children with CD as compared to control subjects (Figure 2.3.7 A). This reduction was significant in patients with macroscopically non-inflamed colon ($P=0.003$) and near significant in children with inflamed colons ($P=0.07$, Figure 2.3.7 B). In contrast, *F. prausnitzii* was increased in CD patients as compared to controls (Figure 2.3.7 A-B). A total of 58 unclassified OTUs were identified in the CD group that represent from 7.2%-20.2% of all the reads of each sample. Although the overall relative abundance of unclassified OTUs was similar between CD patients and controls, five of the unclassified OTUs were significantly reduced in CD ($P<0.05$; Figure 2.3.7 A). Two unclassified OTUs were significantly reduced in CD samples with either normal or inflamed colon (namely, OTU_22 and OTU_55) and three OTUs (OTU_15, OTU_44, and OTU_60) were reduced in CD samples with normal colon only (Figure 2.3.7 B). No significant difference in the relative abundance of the identified butyrate producers or the unclassified OTUs between CD samples with normal or inflamed colon was observed (data not shown).

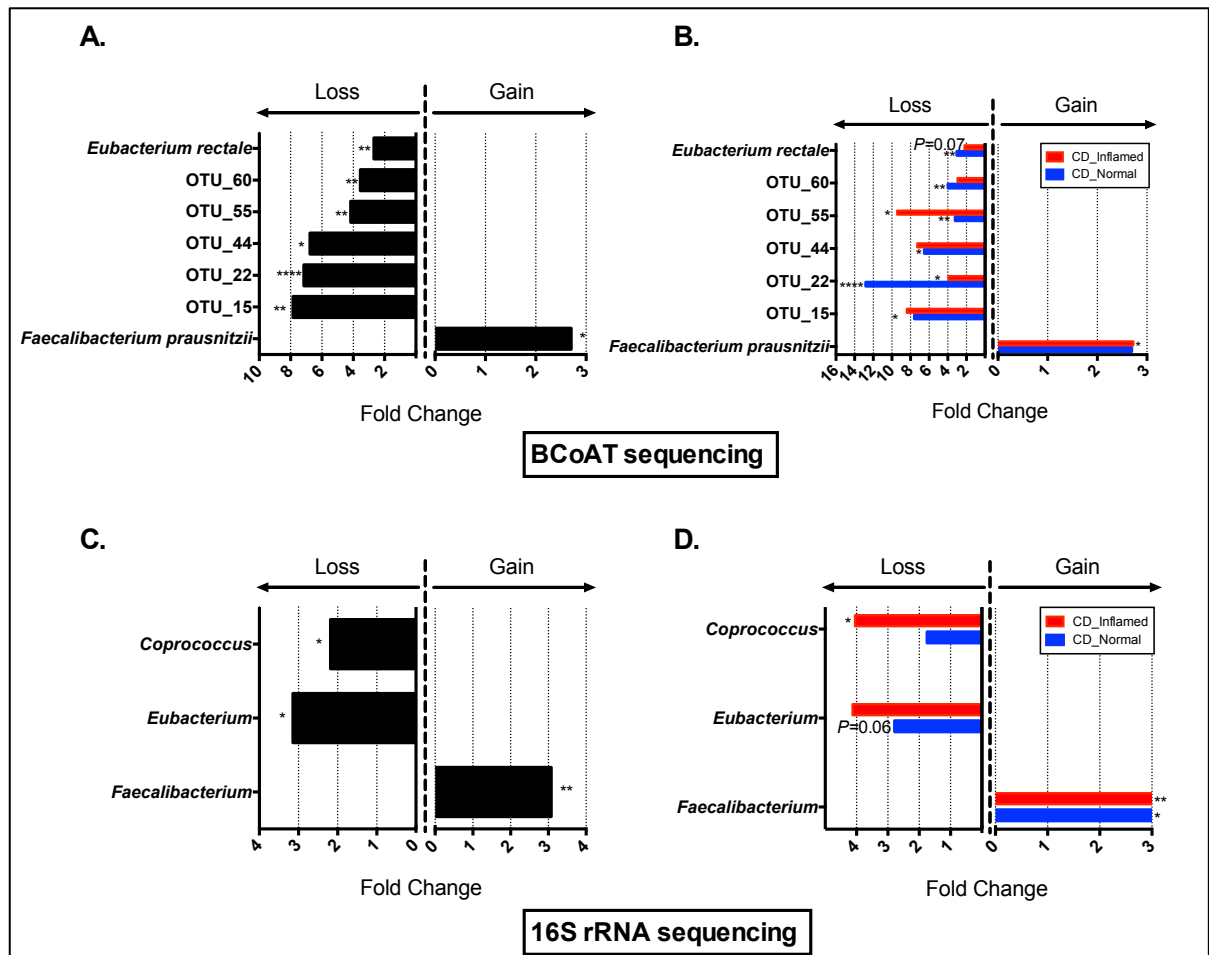


Figure 2.3.7: BCoAT butyrate-producers with differential abundance in CD. Significant taxa or unclassified OTUs were identified by comparing their relative abundance between the control and CD or the control and CD inflammation status (normal or inflamed) using two-tailed Mann-Whitney U test. P values ≤ 0.05 were considered significant. Differential taxa and unclassified OTUs were presented as fold change calculated by dividing the relative abundance of a subject by the mean relative abundance of the control for a given taxon. (A and C) Fold change of differential taxa and unclassified OTUs of CD samples to the controls. (B and D) Fold change of differential taxa and unclassified OTUs of the CD samples stratified according to colon appearance during sample collection (i.e. normal or inflamed) to the controls. (A and B) Differential taxa and unclassified OTUs identified by BCoAT gene sequencing. (C and D) Differential taxa and unclassified OTUs identified by 16S rRNA gene sequencing. Each bar represents the average fold change of a given bacteria. *, $P < 0.05$; **, $P < 0.01$; ****, $P < 0.0001$.

In accordance with the functional gene findings, the 16S rRNA approach highlighted a significant reduction in the relative abundance of *Eubacterium* in CD patients as compared to the control subjects ($P=0.03$) and was near significant in CD patients with macroscopically non-inflamed colon ($P=0.06$). Furthermore, *Faecalibacterium* was higher in CD samples compared to controls ($P=0.005$). This increase was significant in CD patients with inflamed and non-inflamed colons ($P<0.05$; Figure 2.3.7 C-D). Importantly, the relative abundance of *Coprococcus* was decreased in CD patients with inflamed colon ($P=0.01$; Figure 2.3.7 C-D). None of the identified butyrate producers by the 16S rRNA approach showed significant difference between CD samples with normal or inflamed colon (data not shown).

2.3.7. Phylogenetic analysis of unclassified OTUs

In order to assess whether the unclassified OTUs could represent novel butyrate producers, these sequences were aligned to the *but* nucleotide database downloaded from the Functional Gene Pipeline/Repository (<http://fungene.cme.msu.edu>) using MUSCLE. The *but* database contains all nucleotide sequences of probable BCoAT genes identified by Hidden Markov Model searches of the NCBI protein database [263]. Subsequently, the aligned sequences were subjected to phylogenetic tree construction. Forty-five of the 60 unclassified OTUs clustered with known butyrate-producing bacteria. The remaining unclassified OTUs clustered with BCoAT coding sequences from unclassified uncultured bacteria (Figure 2.3.8), suggesting that these 15 unclassified OTUs might belong to novel butyrate producers. Interestingly, the five OTUs exhibiting differential abundance in CD patients as compared to the controls (Figure 2.3.7 A) were grouped in one

branch that clustered with sequences obtained from unclassified uncultured bacteria (Figure 2.3.8).

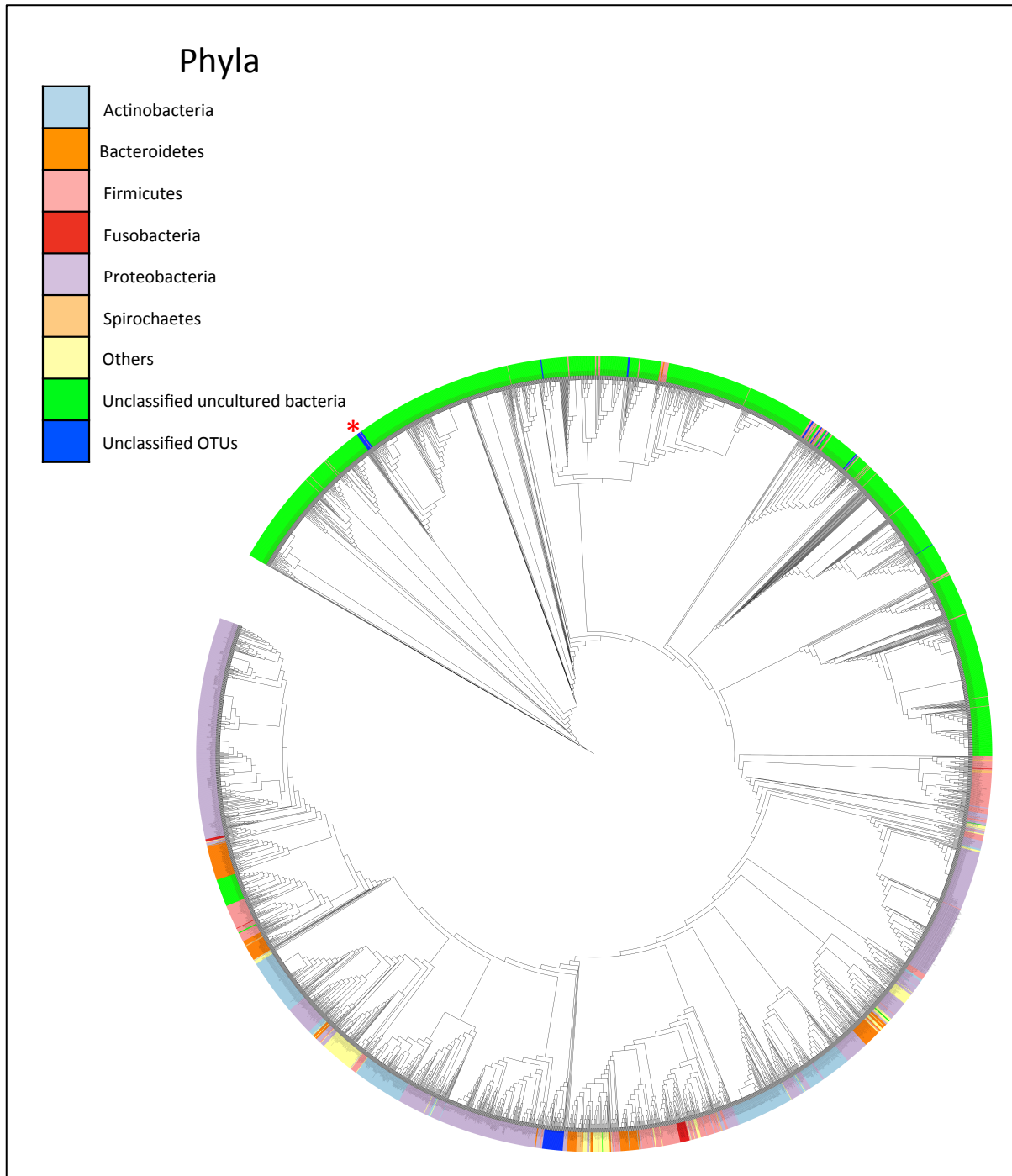


Figure 2.3.8: Rooted phylogenetic analysis of *but* database sequences with the unclassified Operational Taxonomic Units (OTUs) obtained from BCoAT 454 pyrosequencing. Tree is colored by phyla. Red asterisks highlight the branch that contains the 5 unclassified OTUs (OTU_15, OTU_22, OTU_44, OTU_55, and OTU_60) with reduced relative abundance in CD subject.

2.3.8. Relative quantification of key butyrate producers revealed by qPCR

The relative abundance of *E. rectale* and *F. prausnitzii* was further validated by qPCR on 96 samples (39 control and 57 CD; Appendix I) by using specific BCoAT and 16S rRNA primer sets. The relative abundance of *E. rectale* BCoAT gene was found to be reduced at the MLI of CD patients as compared to control subjects ($P<0.05$; Figure 2.3.9 A). Moreover, targeting the 16S rRNA gene of *Eubacterium rectale*/Clostridium coccoides group (Clostridium cluster XIVa), which is dominated by *E. rectale*, by qPCR showed a decreased relative abundance of this cluster at the MLI of CD patients as compared to control subjects (Figure 2.3.9 B). In contrast, the assessment of the level of *F. prausnitzii* by targeting its BCoAT and 16S rRNA genes by qPCR showed that its relative abundance was increased in CD patients (with a greater degree in patients with inflamed right colon) as compared to control subjects ($P<0.05$; Figure 2.3.9 C-D).

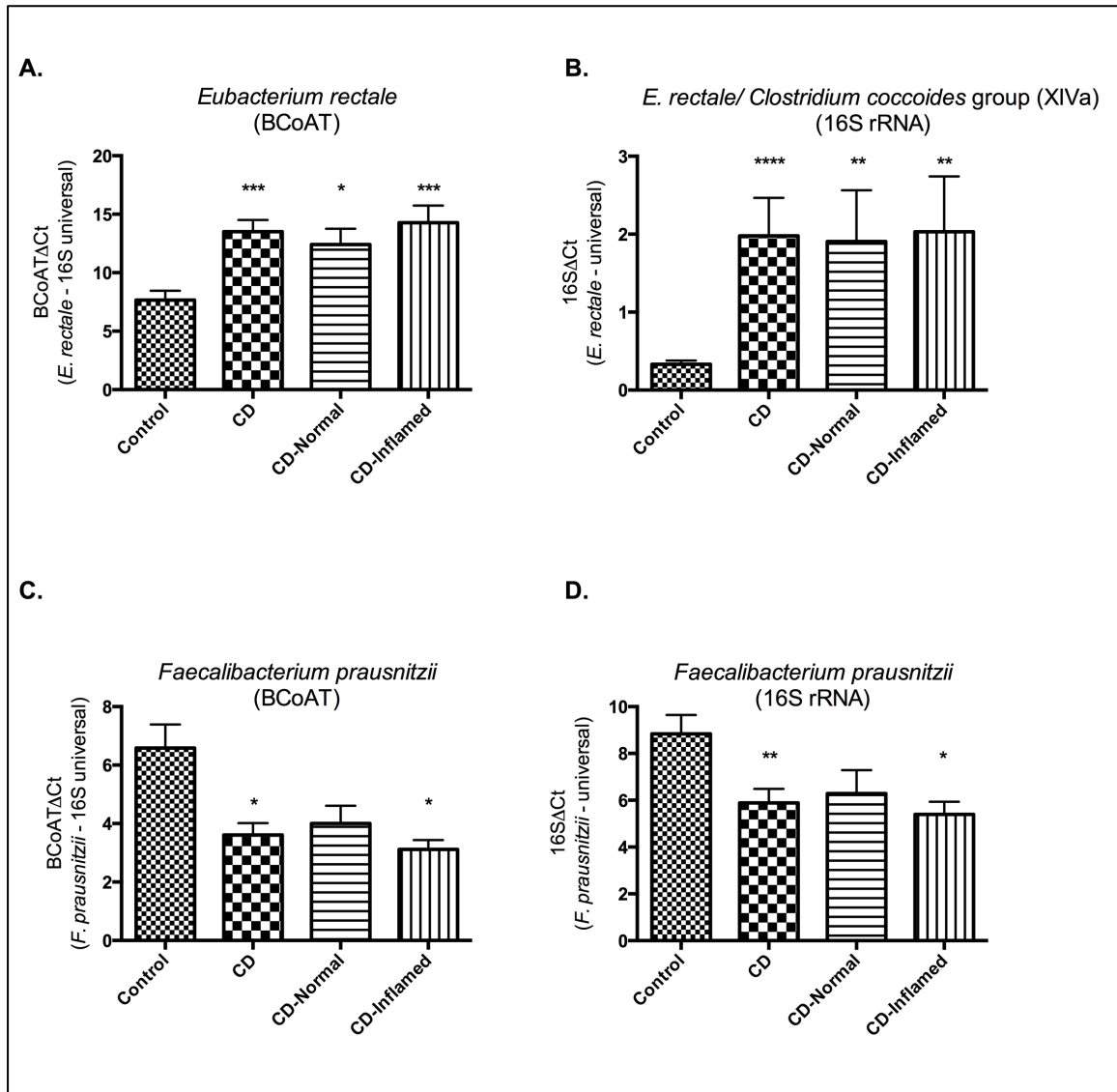


Figure 2.3.9: Quantitative PCR analysis of key butyrate producers. *Eubacterium rectale* and *Faecalibacterium prausnitzii* were quantified using BCoAT and 16S rRNA primers. (A-D) represent the Δ Ct of targeted butyrate producers relative to total bacteria 16S rRNA. Each bar represents the mean Δ Ct of a given bacteria in one group. Error bars represent the standard error of the mean. CD-Normal and CD-Inflamed represent sets of CD samples categorized according to the macroscopic appearance of right colon during samples collection. *, $P < 0.05$; **, $P < 0.01$; *, $P < 0.001$; ****, $P < 0.0001$.**

It is well known that the microbial communities within stools differ significantly from those colonizing the intestinal mucosal surface [81, 129, 136]. Stools have a higher relative abundance of Firmicutes than Bacteroidetes [131, 183-185] while the relative abundance of these two phyla is relatively equal in the mucosa layer [173, 186]. Based on these observations we sought to investigate the relative of *F. prausnitzii* in stool samples. Stools were collected from 5 control and 10 CD patients (6 with normal and 4 with inflamed colons) and extracted metagenomic DNA were subjected to qPCR using 16S rRNA primers specific to *F. prausnitzii*. In contrast to the MLI sample findings, the relative abundance of *F. prausnitzii* in stool samples was similar in CD patients as compared to controls and was unaffected by the macroscopic appearance (normal vs. inflamed) as shown in Figure 2.3.10.

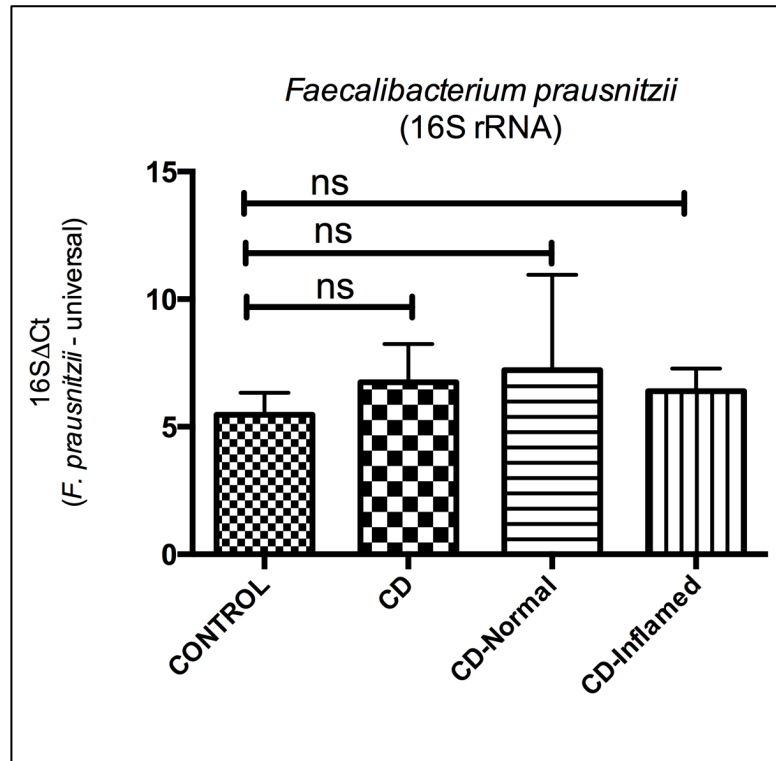


Figure 2.3.10: Quantitative PCR analysis of *Faecalibacterium prausnitzii* from stool samples. *Faecalibacterium prausnitzii* was quantified using specific 16S rRNA primers. Result represents the ΔC_t of *F. prausnitzii* relative to total bacteria 16S rRNA. Error bars represent the standard error of the mean. ns; $P > 0.05$.

2.4. Discussion

In this study, I present a comprehensive characterization of butyrate producers from pediatric non-IBD control and CD subjects. Targeting BCoAT and BUK genes in qPCR, which represent the two main known pathways for butyrate biosynthesis by the gut microbiota, I report a decrease in the level of butyrate producers that utilize the BCoAT pathway in CD patients while there is no change in the level of butyrate producers that use the BUK route. This alteration is observed at the MLI only as no change is noticed at the mucosal layer. The composition of BCoAT butyrate producers was further characterized using two different approaches involving the BCoAT and 16S rRNA genes sequencing. These approaches identified a significant difference in the relative abundance of three major butyrate producers and five potential novel butyrate producers in pediatric CD patients as compared to controls. The sequencing findings were further validated by qPCR which confirmed the sequencing results.

Butyrate production by the gut microbiota is important for the health of the intestinal tract [149]. Butyrate is involved in hydrogen sulfide detoxification [244] and maintenance of the intestinal barrier [245-248]. Moreover, butyrate exhibits anti-inflammatory properties by inhibiting the activation of NF κ B [249-251], inducing Fas-mediated apoptosis of T cells [252], and reducing oxidative stress [253]. Also, butyrate-producing bacteria promotes immune cell differentiation into T_{reg} cells [258, 259]. In fact, reduction in the number of butyrate producers in active CD has been reported in several studies [78-81]. In this study, we demonstrated for the first time that BCoAT butyrate producers, but not BUK butyrate producers, are reduced at the

MLI of the right colon of new early onset CD patients as compared to controls. Although there was a slight decrease in the amount of butyrate producers in CD patients with normal colon as compared to control, the reduction was more pronounced in the inflamed sites. In fact, BCoAT butyrate producers represent the majority of the gut butyrate producers and mainly belong to the Firmicutes phylum which are potent butyrate producers [156]. These butyrate producers are strict anaerobes which by definition are highly sensitive to low levels of oxygen [157]. In healthy human gut, the oxygen tension at the mucosal surface is lower than in intestinal tissue and reaches extremely low levels in the lumen, which favors the growth of strict anaerobic bacteria [293]. Increasing the luminal oxygenation level in the gut of mice resulted in disturbing the gut microbiota, which exhibited a significant reduction in the relative abundance of Firmicutes as compared to control mice [293]. Therefore, the decrease in the level of these bacteria at the inflamed areas might be a consequence of higher oxygen concentration as a result of the inflammatory state [294]. Reduced abundance of butyrate producers is usually coupled with lower concentration of butyrate in the gut [136], resulting in detrimental effect on the integrity of the colonic tissue which will contribute to the inflammation.

In contrast to the results using MLI samples, the relative abundance of the butyrate producers, as measured by quantitative PCR of the BCoAT gene and using the mucosal biopsies, was not affected in the CD patients as compared to the controls. The discrepancy between the results obtained by using the MLI and biopsy samples can be attributed to a number of factors. 1) Colon biopsies represent a very localized small area of the intestine not necessarily representative of the entire colon

and providing limited biological material in contrast to the MLI samples which cover a wider colonic area 2) The biopsy collection protocol involves extensive washing of the mucosal layer which removes the outer loose mucus layer containing the mucosal microbiota [17]. In contrast the MLI samples mainly contain the microbiota rich layer at the MLI. 3) Unpublished work from our laboratory demonstrates that there is a significant variation at the microbiota composition between biopsies and MLI samples with the later being enriched with Firmicutes, the phylum that represents butyrate producers. Altogether, these observations highlight the advantage of using MLI samples over biopsies in studying the butyrate-producing microbiota diversity and explain the observed discrepancies in the results obtained by using the MLI or biopsy samples.

There were few differences in the results obtained from the functional gene and 16S rRNA sequencing data. At the OTU level, the 16S rRNA approach showed significant reduction in the number (richness) and evenness of OTUs that belong to the BCoAT butyrate producers in CD patients as compared to controls. In the contrary, the BCoAT sequencing approach revealed a significant reduction in evenness only in CD patients as compared to controls. Principal coordinate analysis of the unweighted UniFrac distance metric, which analyzes the phylogenetic relationship between the identified OTUs, separated the CD from control samples in the 16S rRNA approach. On the other hand, PCoA analysis of the weighted UniFrac distance metric, which takes into account the relative abundance in addition to the phylogenetic relationship of the identified OTUs, demonstrated a difference between CD and controls in the BCoAT sequencing approach. This indicates that the 16S

rRNA approach mainly attributes the difference between the control and CD samples to the presence of a lower number of BCoAT butyrate producers' OTUs in the CD samples. On the contrary, the BCoAT approach shows that the control and CD samples has similar numbers of OTUs, and the difference between the control and CD samples is mainly attributed to the relative abundance of the identified OTUs. At the taxonomic classification level, *Roseburia* and *Coproccoccus* were the most abundant bacteria in the 16S rRNA approach, while the most abundant bacteria was *Eubacterium rectale* followed by *Faecalibacterium prausnitzii* in the BCoAT gene approach. Finding higher levels of *Roseburia* 16S rRNA than BCoAT genes supports a previous hypothesis that butyrate production is restricted to certain members of the same genus [155, 263]. In the case of *Coproccoccus*, bacteria from this genus can produce butyrate not only via the BCoAT pathway but also through the BUK pathway [263] which could explain the reduced abundance of *Coproccoccus* in the BCoAT data. The functional gene approach missed low abundant bacteria compared to the 16S rRNA approach. This could be attributed to the higher sequencing depth of the HiSeq2500 Illumina platform used for 16S rRNA that generate up to 50 times more reads than the 454 pyrosequencing used for the BCoAT characterization. On the contrary, lack of species level resolution by the 16S rRNA data made it impossible to detect some important butyrate producers that were identified by the functional gene approach such as *Clostridium* sp. M62/1, *Clostridium* sp. SS2/1, and *Clostridium symbiosum*. Therefore, the use of both the functional and 16S rRNA genes approaches are required to generate a comprehensive picture of the butyrate producers composition.

The composition of the BCoAT butyrate producers was affected in pediatric CD patients. For instance, we observed a reduced abundance of *E. rectale* in CD patients as compared to the controls, consistent with other studies conducted on adult and pediatric CD patients [81, 295, 296]. *E. rectale* is one of the most abundant butyrate producers in healthy human gut and produces high levels of butyrate [156, 157]. The functional gene sequencing and qPCR results revealed a reduction in the relative abundance of *E. rectale* at the MLI of non-inflamed and inflamed colons of CD patients as compared to the controls. Thus, reduction of *E. rectale* in CD could not be a consequence of inflammation since the decline was observed in both inflamed and non-inflamed colons.

We have found an increased abundance of *F. prausnitzii* in CD subjects with inflamed colons as compared to controls. This finding is consistent with the study from Reiff *et al.* which similarly reported an increased abundance of *F. prausnitzii* in mucosal biopsies of pediatric CD patients [173]. However, our results are in conflict with most previous studies. Indeed, the relative abundance of *F. prausnitzii* was found to be decreased in adult CD patients by numerous studies (reviewed elsewhere [297]) and more recently in a large cohort of treatment-naïve pediatric CD patients [81]. A number of factors could contribute to these discrepancies. First, increasing evidence illustrates that the composition of the gut microbiota varies along the gastrointestinal tract and differ between colonic and ileal CD [179, 298]. Although, ileal (ICD) or colonic (CCD) classification is not available for most CD studies, one study demonstrated that *F. prausnitzii* was decreased in ICD but not CCD [298]. It is worth noting that the CD subjects with inflamed colons in our study

are either CCD or with ileocolonic involvement. In general, pediatric CD has more extensive colonic and ileocolonic involvement than adults [5, 8]. In the study from Gevers *et al.*, the decline in the relative abundance of *F. prausnitzii* was observed by using terminal ileum and rectal biopsies, which are different gastrointestinal locations than those used in our study (right colon) or the study from Reiff *et al.* (distal colon). Also, the increased abundance that we, and Reiff *et al.*, are reporting is mainly observed in the inflamed areas of the right colon [173], and it is unclear from the study of Gevers *et al.* if the inflammation in their cohort is only restricted to the terminal ileum and rectum or if it is also found in the right and distal colon [81]. This indicates that the different sites investigated and the inflammation status in our study could contribute to the observed differences in *F. prausnitzii* abundance reported by Gevers *et al.* [81]. We also quantified *F. prausnitzii* in a subset of controls and CD samples (with either inflamed and non-inflamed colon) using stool samples. Although MLI samples from the same CD patients with an inflamed colon showed higher level of *F. prausnitzii* compared to controls, there was no difference in *F. prausnitzii* relative abundance when using stool samples. Therefore, sample type (stool versus MLI samples or biopsies) may also explain the observed discrepancies. Wang *et al.* recently showed that the abundance of *F. prausnitzii* is markedly reduced in stool samples compared to biopsies from the same subject with active UC and to a lesser extent in active CD [299]. In another study, no difference in abundance of *F. prausnitzii* was observed between pediatric CD patients and controls using stool samples [173, 300]. Furthermore, it is known that the composition of microbiota within stool samples is significantly different than those

colonizing the intestinal mucosal surface [81, 129, 136]. Normally, fecal samples have a higher abundance of Firmicutes than Bacteroidetes [131, 183-185]. On the other hand, mucosal samples show a relatively equal abundance of Firmicutes and Bacteroidetes [173, 186]. Knowing that *F. prausnitzii* belongs to the Firmicutes phylum [299], we would expect to see a different trend in the relative abundance of this bacteria between stool and MLI samples.

The increased abundance of *F. prausnitzii* in CD subjects with an inflamed colon might suggest that this bacterium is equipped with stress response factors that allow it to thrive under the inflammatory conditions. *F. prausnitzii* was previously thought to be a strict anaerobic bacteria. However, a recently isolated strain from human stool was found to use a low level of oxygen through an extracellular electron shuttle system [301]. This could give *F. prausnitzii* the ability to survive under the reduced anoxic environment as a result of inflammation. It is important to note that higher abundance of *F. prausnitzii* does not necessarily mean higher butyrate production in the gut since butyrate synthesis is highly affected by pH [302], and it is known that patients with IBD have an abnormally high or low pH in the gut lumen [303]. Currently, efforts are underway to isolate *F. prausnitzii* from our samples in order to extensively study this isolate and gain a better understanding of its role in CD.

Our study also identified 60 OTUs, representing 3.1% of reads from control and 7.2% of reads from CD samples, that share low percentage (<75%) of sequence identity to known butyrate producers. These sequences could belong to known butyrate producing bacteria, but due to the high level of sequence variability

in the BCoAT gene, which is observed not only at the species level but also at the strain level [155], they were below our similarity assignment cutoff. The other possibility is that they belong to novel species of butyrate producers. For instance, phylogenetic analysis supports the second possibility for 15 OTUs out of the 60 unclassified OTUs since they were clustered with sequences obtained from uncultured unclassified bacteria isolated from human stool. Interestingly, the 5 unclassified OTUs that were reduced in CD are found in the same clade within the unclassified uncultured bacterium, which indicate that they could belong to one novel butyrate producer or closely related novel butyrate-producing bacteria.

In conclusion, we were able to characterize the diversity of butyrate producers in non-IBD pediatric and CD subjects by targeting the 16S rRNA and the functional BCoAT genes. Using the functional gene approach, we were able to identify differences in the amount and diversity of butyrate producers between pediatric CD and non-CD subjects in the right colon at the MLI. BCoAT sequencing results were concordant with the 16S rRNA sequencing data with the additional advantage of identifying the species level for the taxa assignments. In addition, we were able to identify potential novel butyrate producing bacteria. Future isolation of the key butyrate producers identified here could broaden our view about this important functional group and their involvement in human health and protection against inflammation.

CHAPTER 3

GUT PLASMIDOME IN EARLY ONSET PEDIATRIC INFLAMMATORY BOWEL DISEASE

3.1. Introduction

The human gut is colonized with a complex and dense microbial ecosystem that harbors 10 times more cells than the number of human cells. The colon alone is colonized with 10^{11} - 10^{12} microbial cells/ml [304], which is composed of more than 1,000 different bacterial species [119, 240, 305]. This newly recognized organ co-evolves with its host and shifts its structure in response to host-derived selective pressures or changes in its microenvironment [306]. As a result, either a stable function is maintained (functional redundancy) or a new function is introduced to cope with the environmental changes [306]. It is now evident that the microbiota functional activities not only affect the microbial ecosystem but also the host, which can be beneficial or harmful. In fact, different metabolites produced by intestinal microbiota are implicated in a number of diseases, including obesity, irritable bowel syndrome (IBS), inflammatory bowel disease (IBD), colorectal cancer and inflammation [223, 307-318]. Therefore, recent microbiota studies of human illnesses include the assessment of the functionalities of gut microbiota in addition to their phylogenetic characterization [81, 319, 320].

As discussed in Chapter 1, microbial plasmids are known to harbor functional genes involved in intestinal colonization, fitness and virulence genes [205, 231-234]. To date, there is only one study that characterized the total gut plasmids content “the plasmidome” in human [321]. Jones *et al.* [321] isolated the gut plasmidome from a healthy adult human volunteer using transposon-aided capture (TRACA). In total, they sequenced 21 plasmids using primer walking [235, 321]. These plasmids mainly belonged to Gram-positive bacteria, which are members of the Firmicutes phylum, as well as potentially novel plasmids with no sequence homology to the reference database. The functional annotation of the genes encoded by these plasmids included plasmid replication, plasmid mobilization, segregation stability, site-specific integration or recombination, and the RelBE toxin/anti-toxin (TA) addiction module. Next, Jones *et al.* aligned the nucleotide sequences of the gut plasmids to several databases, including the metagenomic sequences of the human gut microbiota, the murine gut microbiota, the Sargasso Sea, and soil bacteria [191, 235]. Their comparative metagenomic study revealed that two plasmids (namely pTRACA 10 and pTRACA 22) had high nucleotide similarities to metagenomic sequences originating from the gut microbiota of humans across the world, including America, Europe, and Japan. Interestingly, however, no nucleotide similarities to these plasmids were found in the metagenomic databases of the murine gut, marine, and soil microbiota [191, 235]. This finding suggests the existence of unique human gut plasmidome. Plasmids pTRACA 10 and pTRACA 22 carry open reading frames (ORFs) that encode plasmid replication proteins, TA addiction modules, phosphoesterases/phosphohydrolases, and ORFs of unknown function [191, 235].

Most importantly, aligning the plasmidome sequences to the metagenomics human intestinal tract (MetaHIT) database identified three plasmids that are more abundant in adult IBD patients (both UC and CD) compared to healthy individuals [191, 235]. Taken together, these observations highlight a potential association of plasmid-derived gene products with IBD pathogenesis; however, a comprehensive comparative study of non-IBD and IBD gut plasmidome is still necessary. Indeed, the characterization of the gut plasmidome from healthy and IBD subjects would enable a better understanding of the intestinal microbiota, their interactions with the host and their potential involvement in intestinal diseases.

Several approaches have been developed to study the microbiota plasmidome. For instance, metagenomic sequencing has proven to be useful in characterizing microbial functional genes, including the plasmidome; however, metagenomic sequencing only identifies “putative” plasmid sequences, as this approach cannot distinguish chromosomal from plasmid DNA. Moreover, metagenomic studies have underestimated the plasmidome component due to low sequencing coverage and bioinformatic challenges. These obstacles have been mitigated by more recent strategies that rely on the isolation of the plasmidome from metagenomic DNA followed by deep sequencing. To date, there are six reports on the plasmidome of microbial communities; these studies have characterized the plasmidome from healthy adult human gut microbiota [321], bovine rumen [322], human dental plaques [323], activated sludge [324, 325] and wastewater microbiotas [326]. The isolation of the plasmidome in these studies involved traditional and advanced strategies: (1) the chemical lysis of bacteria followed by

plasmid purification [325, 326]; (2) the purification of plasmids from metagenomic DNA by the digestion of linear DNA followed by plasmid DNA amplification [322]; and (3) transposon aided capture (TRACA) of plasmids [321, 323, 324]. The first strategy was used to study the plasmidome of wastewater and activated sludge [325, 326]. Szczepanowski *et al.* plated their wastewater sample in a series of LB agar plates supplemented with different antibiotics. Then, the plasmidome was purified from the growing bacteria using the alkaline lysis method and cesium chloride density gradient centrifugation [326]. Although this study provided valuable information about the plasmidome of wastewater, it excluded the plasmidome of bacteria susceptible to antibiotics and unculturable bacteria. More recently, Sentchilo *et al.* used a culture-independent chemical lysis of bacteria followed by plasmid purification from activated sludge samples [325]. First, the sludge biomass was purified from the activated sludge samples by a series of centrifugations. Then, the plasmidome was purified from the sediment utilizing the alkaline lysis method in combination with acid phenol–chloroform treatment [325]. Although this study surpassed the need for culturing bacteria before plasmid isolation, the purity of the plasmidome was affected by contamination with chromosomal DNA during plasmid isolation [325]. In another study, the rumen plasmidome from 16 cows was isolated by the direct extraction of metagenomic DNA from the samples followed by eliminating the chromosomal DNA by plasmid-safe DNase, leaving only closed circular DNA [322]. The circular DNA was subsequently amplified using Φ 29 DNA polymerase to enrich low abundant plasmids [322]. However, the entire genome amplification step used for the plasmidome preparation in this approach usually

favors small un-nicked DNA, which introduces amplification bias. Finally, the TRACA strategy involved the isolation of metagenomic DNA from the samples and the removal of chromosomal DNA using a plasmid-safe DNase, similar to a study by Brown *et al.* [321, 322]. Then, an EZ-Tn5 transposon that contained an *E. coli* origin of replication and a kanamycin resistance cassette was introduced into the remaining closed circular DNA using an in vitro transposition reaction. Ultimately, the product was cloned in *E. coli* followed by plasmid isolation from the transformants [321]. The TRACA approach in particular has a number of advantages over other methods. It is culture-independent, it minimizes the contamination with chromosomal DNA, and it acquires and maintains plasmids from Gram negative and positive bacteria regardless of the plasmidome encoded functions; however, the TRACA strategy has some disadvantages, including the inactivation of the plasmidome genes by the transposon, difficulties in isolating large plasmids and the inability to identify the original host of the plasmids.

In this section of the thesis, the characterization of pediatric gut plasmidome will be discussed. The gut plasmidome was isolated from pediatric IBD patients (including UC and CD) in an inception cohort of treatment naïve children undergoing diagnostic colonoscopies and a control group of children without IBD using the TRACA strategy. The isolated plasmidome was then characterized using Illumina high-throughput sequencing. The comparison of the gut plasmidome composition of IBD patients to non-IBD subjects showed significant differences that not only differentiated between the control subjects and IBD but also separated UC from CD.

3.2. Material and methods

3.2.1. Patient cohort

This study was reviewed and approved by the ethical board of the CHEO. Patient selection and sample collection were performed as described previously in chapter 2. For the CD patients, the site of the disease, the endoscopic activity and the clinical activity were determined as described in chapter 2. For the UC patients, the disease site was determined using the Paris Modification of the Montreal Classification for IBD [277]. The endoscopic activity was recorded using the Mayo Score Flexible Proctosigmoidoscopy Assessment for ulcerative colitis, and the clinical activity was determined using the Pediatric Ulcerative Colitis Activity Index (PUCAI) [327]. For this study, the right colon mucosal-luminal interface (MLI) samples were collected from 12 patients (4 control subjects, and 4 UC and 4 CD pediatric patients).

3.2.2. Plasmidome isolation

The closed circular DNA was captured from the samples as described previously [321] and as illustrated in figure 3.2.1. Briefly, the metagenomic DNA was extracted from each sample (described in chapter 2). Next, 1 µg of the extracted metagenomic DNA was digested overnight at 37°C with 10 U of a Plasmid-Safe™ ATP-Dependent DNase (Epicentre, Philadelphia, Pennsylvania) to remove sheared and linear DNA. The digestion reaction was stopped by incubating the mixture at 70°C for 30 min, and then it was cooled on ice for 5 min.

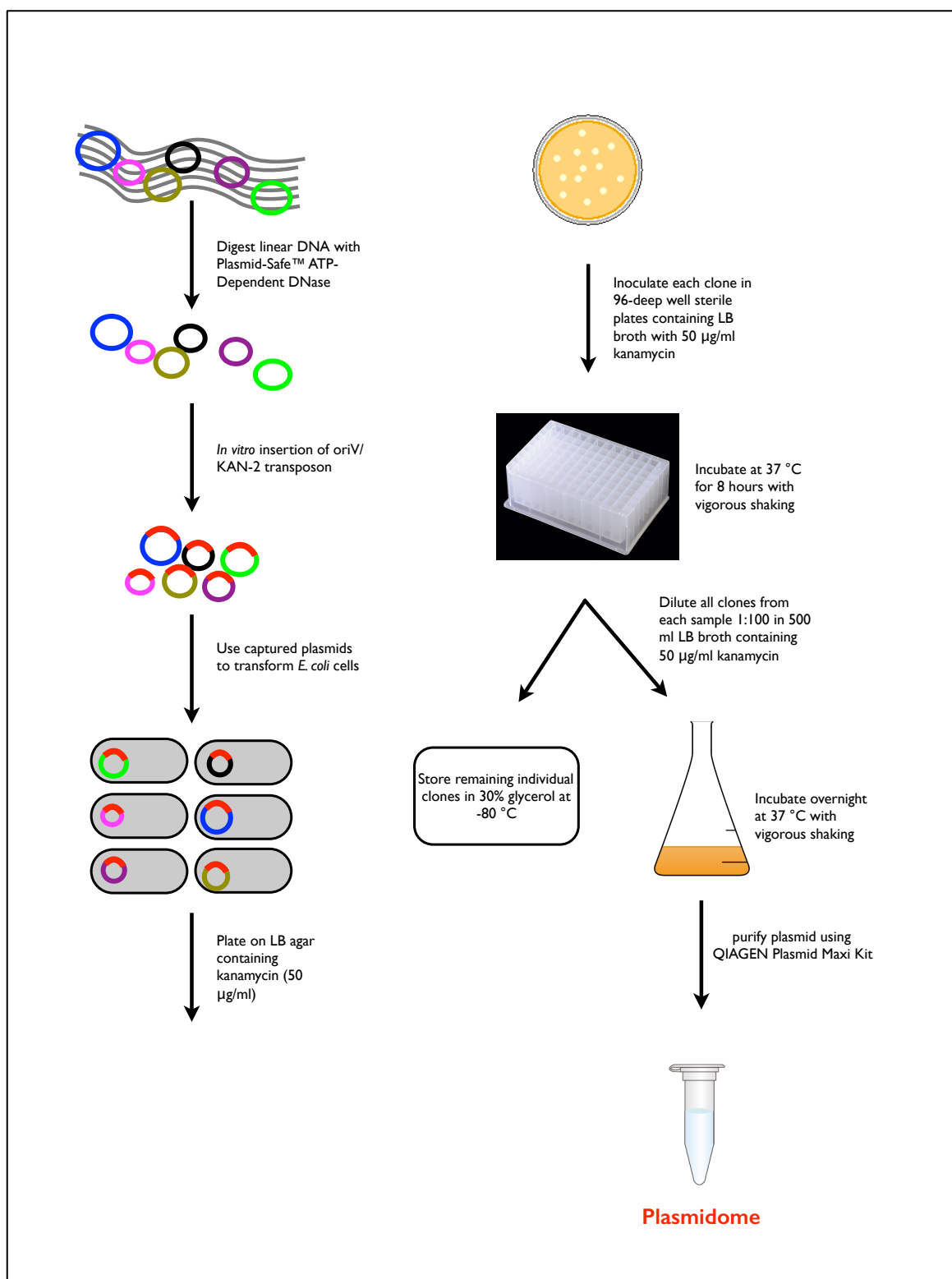


Figure 3.2.1: Illustration of plasmidome capture and its isolation from metagenomic DNA.

Without any further processing, the digested DNA was subsequently used for an *in vitro* transposon reaction utilizing the EZ-Tn5™ <oriV/KAN-2> Insertion Kit (Epicentre) and following the manufacturer's protocol. Thirty-five µl of the digested DNA was used in a 50 µl reaction mixture that contained a 1X EZ-Tn5 reaction buffer, a 0.1 pmol of EZ-Tn5 <KAN-2> Transposon (which harbors a kanamycin resistance cassette and the *E. coli* origin of the plasmid replication (oriV)) and 2 U of the EZ-Tn5 Transposase. The transposon insertion reaction mixture was incubated for 2 hrs at 37°C and the reaction stopped by adding 5 µl of 10X EZ-Tn5 Stop Solution. Then, the mixture was incubated at 70°C for 15 min. Next, the reaction was diluted with 450 µl of deionized sterile water, purified using a Microcon Ultracel YM-100 Centrifugal Filter (Millipore), and concentrated to 10 µl using an Eppendorf Vacufuge Concentrator 5301 (Eppendorf) following standard protocols. Five µl of the purified plasmid was then mixed with 100 µl of TransforMax EPI300 Electrocompetent *E. coli* cells (Epicentre) and incubated on ice for 5 min in a sterile 0.2 cm electroporation cuvette. Then, the electroporation was carried out utilizing an Eppendorf Electroporator 2510 (Eppendorf) set to 2.5 kV. Immediately following electroporation, 500 µl of Super Optimal broth with Catabolite repression (SOC medium) was added to the electroporated cells and then transferred into a sterile 1.5 ml tube and incubated at 37°C with shaking at 225 rpm for 1 hour. One-hundred µl of the bacterial culture was subsequently plated on Luria broth (LB) agar plates (a total of 5 plates per sample) containing 50 µg/ml kanamycin and the plates were incubated overnight at 37°C. The next day, each clone was picked from the LB plates and inoculated individually into 1 ml LB broth containing kanamycin (50

µg/ml) in 96 Deep-Well Plates (VWR). At this step, the number of isolated clones per sample ranged from 196-1,644 clones (Appendix III). The plates were subsequently sealed with sterile AlumaSeal II Sealing Films (Excel Scientific, Inc.) and incubated at 37°C for 8 hours with vigorous shaking. After 8 hours, 50 µl of the 1 ml bacterial culture in the deep-well plates was pooled into sterile flasks and mixed properly. At this step, the bacterial clone cultures that were prepared from the same MLI sample were pooled into one flask. The total number of clones from each group (i.e. control, UC and CD) was adjusted to the lowest number of clones isolated from any of the three groups (i.e. 2,500 clones, Appendix III). The remaining volume of the clone cultures in the deep-well plates was supplemented with 30% sterile glycerol and stored at -80°C. Then, the clone mix was diluted 1:100 in one volume of 500 ml fresh LB broth containing kanamycin (50 µg/ml) and incubated overnight at 37°C. The next day, the cultures were spun down at 4,000 X g for 30 min in a refrigerated centrifuge, and the plasmid DNA was purified from the bacterial pellet using the QIAGEN Plasmid Maxi Kit according to the manufacturer's instructions (QIAGEN). Finally, the DNA concentration of the purified plasmidome was quantified using Qubit® dsDNA BR Assay Kit (Invitrogen) according to the manufacturer's instructions and then stored at -20°C until further use.

3.2.3. Restriction digestion screening of plasmid metagenome

One µg of plasmidome from each sample was used in a digestion reaction containing BamHI, XhoI and HindIII enzymes (5 U each; Invitrogen) and a 1X React® 2 buffer (Invitrogen) in a final volume of 35 µl. Following overnight incubation

at 37°C, the digests were inspected on 2% agarose gel, loading each lane with 10 µl of the digestion products (~285 ng of digested DNA).

3.2.4. Construction of the plasmidome Illumina library

The construction of the sequencing library was carried out as illustrated in Figure 3.2.2. Ten µg of plasmidome from each sample in 250 µl sterile deionized water was sonicated 6X at 7-8 MHz for 20 sec each with at least 1 min cooling on ice between cycles. Following sonication, the DNA concentration of the sonicated plasmidome was quantified using the Qubit® dsDNA BR Assay Kit (Invitrogen) according to the manufacturer's instructions. Three µg of fragmented DNA was subsequently used for terminal end-repair, dA-tailing of repaired DNA and adaptor ligation using the NEBNext® DNA Sample Prep Master Mix Set 1 kit (NEB) according to the manufacturer's instructions. In total, 12 forward and reverse adapters were designed. Each adapter was composed of forward and reverse oligonucleotides that upon annealing to each other formed an adapter that contained (from 5' to 3' end) the Illumina PE Read 1 Sequencing Primer (for forward adapters) or PE Read 2 Sequencing Primer (for reverse adapters), the 4 nucleotide barcode and an overhanging T (oligonucleotide sequences are listed in table 3.2.1). A total of 12 unique barcodes were designed to sequence 12 samples at once in a single lane. Following adapter ligation, the fragments were purified using the PureLink® PCR Purification Kit according to the manufacturer's instructions and then eluted in 30 µl deionized sterile water. Next, 1 µl of the purified fragments was used in 50 µl PCR-enrichment reaction that contained 0.5 µM of the primers PCRFWD1 and PCRRVS1 (which contains an oligonucleotide sequence complimentary to the flow

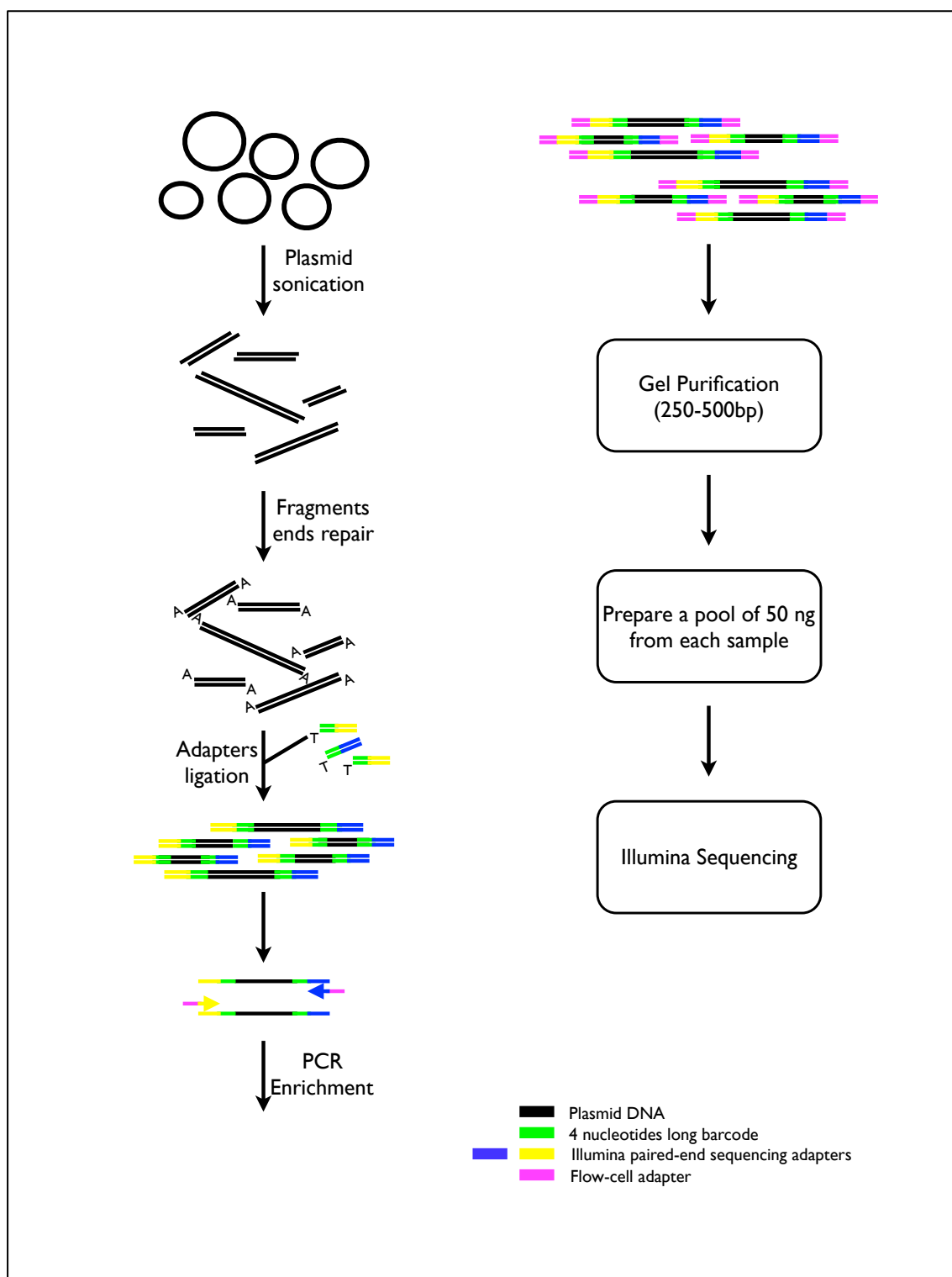


Figure 3.2.2: Construction of Illumina library from plasmidome DNA.

Table 3.2.1: Oligonucleotide sequences used for plasmidome library construction. Oligonucleotides A1BCF and A1BCR with the same barcode sequence were annealed to each other to form the forward adapters. The annealing of A2BCF and A2BCR formed the reverse adapters. The italic nucleotides represent the Illumina paired-end sequencing adapters; the underlined nucleotides represent 4 nucleotides long barcodes; bold T is the overhanging T; and the un-highlighted nucleotides are Illumina flow-cell adapters.

Name	Oligonucleotide Sequence 5'-3'
A1BCF1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATAG T
A1BCF2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGG T
A1BCF3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTCA T
A1BCF4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATC T
A1BCF5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCC T
A1BCF6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGT T
A1BCF7	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACG T
A1BCF8	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTA T
A1BCF9	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAC T
A1BCF10	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGA T
A1BCF11	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAT T
A1BCF12	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTT T
A1BCR1	CTATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR2	CCCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR3	TGAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR4	GATCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR5	GGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR6	ACAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR7	CGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR8	TACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR9	GTCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR10	TCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR11	ATGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A1BCR12	AAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
A2BCF1	GATAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF2	GGGAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF3	ACTTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF4	CTAGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF5	CCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF6	TGTCAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF7	GCACAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF8	ATGCAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF9	CAGGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF10	AGATAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF11	TACTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCF12	TTCAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
A2BCR1	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTTATC T
A2BCR2	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTTCC T
A2BCR3	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTAAGT T
A2BCR4	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTCTAG T
A2BCR5	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTCGGG T
A2BCR6	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTGACA T

A2BCR7	<i>CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGTGCT</i> T
A2BCR8	<i>CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGCAT</i> T
A2BCR9	<i>CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTCCTG</i> T
A2BCR10	<i>CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTATCT</i> T
A2BCR11	<i>CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTAGTA</i> T
A2BCR12	<i>CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTGAAT</i> T
PCRFRWD1	<i>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT</i>
PCRRVS1	<i>CAAGCAGAAGACGCCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT</i>

cell at the 5' end and the Illumina sequencing adapter at the 3' end, see Table 3.1), a 1X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, Tewksbury, MA) and a final adjustment of volume using sterile deionized water. The enrichment reaction was initiated by heating the mixture at 98°C for 30 sec followed by 10 cycles of 98°C for 10 sec, 65°C for 30 sec, 72°C for 30 sec and a final extension step at 72°C for 5 min. The PCR products were subsequently inspected on a 2% agarose gel, and the DNA bands ranging from 250-500 bp were purified from the gel using the QIAquick Gel Extraction Kit (Qiagen). Next, the purified DNA was quantified using the Qubit[®] dsDNA BR Assay Kit (Invitrogen), and an equimolar pool of the purified DNA bands (50 ng from each sample) was prepared. The pool was sequenced in two lanes of the Illumina HiSeq2500 platform (at the Centre for Applied Genomics (TCAG), Toronto) to generate 2X100bp and 2X150bp paired-end sequences.

3.2.5. Data analysis

First, the reads were demultiplexed using an in-house Python script by matching the first four nucleotides of each read to the expected barcodes which were then trimmed. To remove reads that belonged to the cloning host (*Escherichia coli* strain K-12 substrain DH10B) and the transposon (EZ_Tn5_OriV_KAN2), the reads were aligned to the reference sequences using the Burrows-Wheeler Aligner's Smith-Waterman Alignment (BWA-SW v0.7.10) [328]. The unaligned reads were then extracted from the resulting bam files with SAMtools (V1.1) [329]. Finally, the reads were grouped back into pairs using a Python script (<https://github.com/sebhtml/NGS-Pipelines>). The demultiplexing and sequence

filtration processes were conducted in collaboration with Dr. Jacques Corbeil (University of Laval). Next, the nucleotides with a Phred quality score below 20 were trimmed from both ends of the filtered reads using the reformat.sh script of the BBMap software V34.00 (Bushnell, B. BBMap short read aligner, and other bioinformatic tools <http://sourceforge.net/projects/bbmap> (2014)), setting the minimum length of reads after trimming to 70 nucleotides. Following trimming, reads with a Phred quality score of less than 20 over 90% of the read length were discarded from further analysis using the fastq_quality_filter command line from the Fastx-toolkit V0.0.13 (<http://hannonlab.cshl.edu/>). Subsequently, high quality reads from the first and second sequencing runs for each sample were combined. For functional classification, the reads were aligned to the following reference protein databases utilizing the DIAMOND algorithm [330]: Clusters of Orthologous Groups (COGs) Database (<http://www.ncbi.nlm.nih.gov/COG/>), NCBI's Non-Redundant proteins database (NCBI-NR; <http://www.ncbi.nlm.nih.gov>), Human Microbiome Project Gastrointestinal_tract reference sequence (HMP-GIT; <http://www.hmpdacc.org/HMRGD/all/>), A CLAssification of Mobile genetic Elements database V0.4 (ACLAME; <http://aclame.ulb.ac.be>) and Virulence Factors Database (VFDB) (<http://www.mgc.ac.cn/VFs/>). The DIAMOND alignment results were then filtered according to the alignment cutoff set to a 75% sequence identity, an E-value of 1×10^{-5} and Bit-score of 35 using the skimmer.pl script (<http://alrllab.research.pdx.edu>). Next, only the best hit for each read was selected using the blast_best.pl script (<http://alrllab.research.pdx.edu>). The selected hits were summarized for each database in a tabular format that showed a list of unique

database IDs, the number of reads assigned to each ID from each sample and the annotation of each ID using the `blast_summary.pl` script (<http://alrllab.research.pdx.edu>). The rarefaction analysis based on the NCBI-NR database assigned genes was computed by QIIME v1.9.0 utilizing equalized datasets at 5,655,160 reads/sample. The alpha diversity (represented by Chao1 and Shannon indices) and beta diversity for the assigned genes from the various databases were computed using QIIME and compared between groups. Finally, heat maps of $\log_{10}$ relative abundances were generated using the CIMminer tool (<http://discover.nci.nih.gov/cimminer/>) following its default parameters. For the taxonomy assignment, a custom Python script was used to fetch the taxonomy for the assigned NCBI-NR Genbank identifiers (GIs).

3.2.6. 16S rRNA sequencing data analysis

The 16S rRNA-V6 region from 10 out of the 12 samples (3 control subjects, and 4 UC and 3 CD patients) was sequenced and analyzed using a 16S rRNA-based approach as described previously in chapter 2. Then, the 16S rRNA assigned taxa at the phylum and genus levels were compared to that of the plasmidome assigned taxa through the NCBI-NR database. The identified OTUs of the assigned taxa by the 16S rRNA sequencing were also used to predict the COGs functional categories of the microbiota metagenome using the PICRUST algorithm [331]. Finally, the predicted COGs functional categories of the microbiota metagenome were compared to those obtained from the plasmidome analysis using the Wilcoxon signed-rank test.

3.2.7. Statistical analysis

The relative abundance of genes identified by querying the various databases was compared between the control and the IBD samples using the Mann-Whitney U test to identify significantly enriched or depleted genes in IBD patients. Also, the identification of differentially abundant genes as a function of disease status (control vs. UC vs. CD) was carried out using a Kruskal-Wallis test followed by the Dunn's correction to account for multiple comparisons. Next, the microbial communities were compared using principal coordinate analysis (PCoA) of the differentially abundant genes based on the Bray-Curtis distance and using the QIIME software V1.9.0. The correlation between the plasmidome and the 16S rRNA assigned taxa was determined by calculating the Spearman's rank correlation coefficient (r) of the paired relative abundance of the bacterial phyla identified by the two approaches. Compare the relative abundance of the COGs functional categories between the plasmidome and the microbiota metagenome predicted by the PICRUST algorithm was done using the Wilcoxon signed-rank test. The statistical tests were performed using SPSS V.22.0 (IBM Corporation, Armonk, NY), XLSTAT 2014 (Addinsoft, NY) and GraphPad Prism version 6.0c (GraphPad, La Jolla, CA).

3.3. Results

3.3.1. Patient cohort and disease characteristics

A total of 12 patients were included in this study (4 control subjects, and 4 UC and 4 CD patients; Table 3.3.1 and Appendix IV). Histological examinations were performed to confirm the diagnosis of the participating patients. The control subjects included patients undergoing endoscopies for rectal bleeding, screenings for polyps or experiencing abdominal pain. Each control subject had both a visually normal macroscopic appearance of the mucosa and a normal histology of the mucosal biopsies. For the UC patients, 3 subjects had moderate inflammation, and one had severe inflammation according to the PUCAI index. The site of the UC inflammation was mostly left-sided colitis (n=3) in addition to one case with pancolitis. All CD patients had severe inflammation according to the PCDAI index. The site of the macroscopic inflammation of the CD patients was either ileocolonic (n=2) or ileal (n=2). The average age of the participants in years was 11.75 ± 4.2 for the control subjects, 14.75 ± 2.2 for the UC patients and 13.25 ± 1.3 for the CD patients. The control group included 2 male and 2 female subjects. All of the UC and CD patients were males. MLI samples were collected from the right colon of all 12 patients at the time of diagnosis (inception cohort).

Table 3.3.1: Participating patients and disease characteristics.

	Control	UC	CD
Number	4	4	4
Age at Colonoscopy (years)	11.75±4.2	14.75±2.2	13.25±1.3
Gender (M : F)	2:2	4:0	4:0
Disease Severity			
(Moderate : Severe)	NA	3:1	0:4
Macroscopic Mucosal			
Appearance of Right Colon			
(Normal : Inflamed)	4:0	4:0	3:1
Site of Disease			
Ileal : Ileocolonic CD	NA	NA	2:2
Colon Extensive : Pancolitis	NA	3:1	NA

3.3.2. Restriction digestion profile highlights group specific differences in the plasmidome composition

To assess the differences in the plasmidome composition between the control, UC and CD subjects, the isolated plasmidome from each patient or control subject was digested by an enzyme cocktail composed of BamHI, XhoI and HindIII. As shown in Figure 3.3.1, each plasmidome yielded a different restriction profile. Common or unique bands were observed between the plasmidome that originated from either CD or UC patients and the control subjects. For instance, the plasmidome digests from two of the control samples contained a ~ 3,000 bp band that was absent in all CD or UC samples. On the other hand, the UC plasmidome digests contained a ~ 2,200 bp band that was unique to this group. Also, a ~ 500 bp band was observed in 3 of the CD and one of the UC plasmidome digests but never in the control plasmidome. Finally, a ~ 1,500 bp band was observed in 2 of the UC and 2 of the CD plasmidome digests but not in the control plasmidome. Altogether, these results indicate a different plasmidome composition between the three groups. To note, the restriction digestion band of 650 bp in all samples corresponds to the digestion of the oriV/KAN-2 transposon with BamHI (Figure 3.3.1).

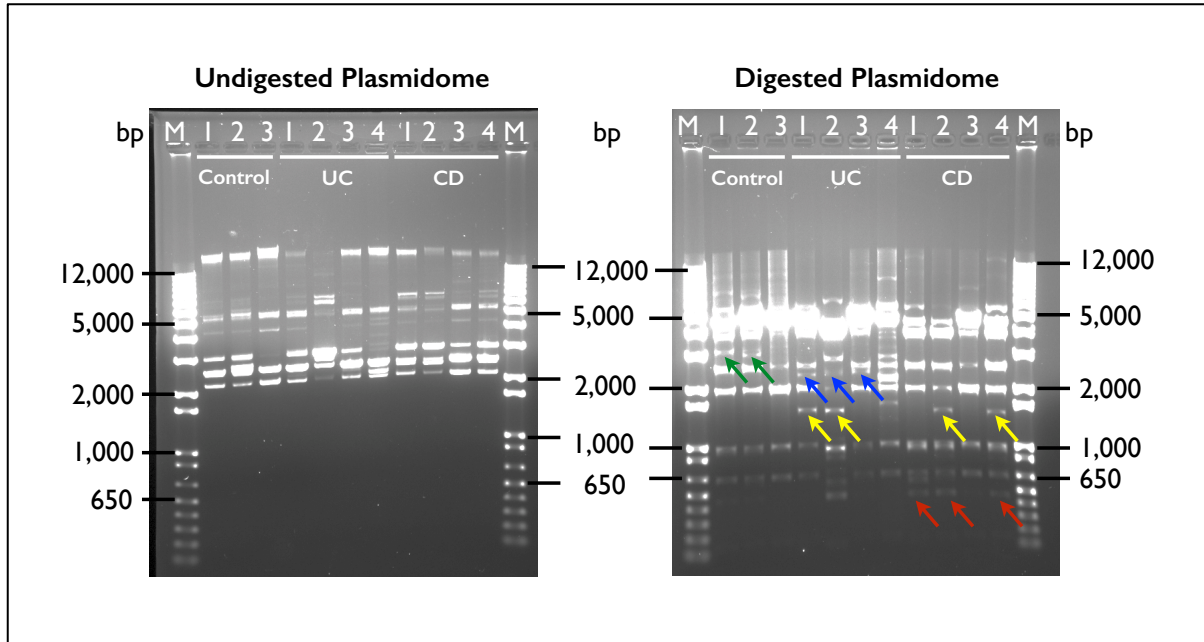


Figure 3.3.1: Plasmidome restriction digest profile. One μg of plasmidome from 3 control, 4 ulcerative colitis (UC) and 4 Crohn's disease (CD) samples were digested with a BamHI, XhoI and HindIII enzyme cocktail. Each lane represents the plasmidome from one sample. The left and right gel pictures represent undigested and digested plasmidome, respectively. The green arrows point at bands unique to the controls; the blue arrows point at bands unique to UC; the red arrows point at bands unique to CD; and the yellow arrows point at bands unique to IBD. Each lane was loaded with ~ 285 ng of DNA.

3.3.3. Sequencing of plasmidome DNA

To further characterize the plasmidome composition, the captured plasmid DNA was sequenced utilizing the Illumina HiSeq2500 platform technology. A total of 336,749,886 reads with an average of 28,062,490 reads/sample (range 32,413,858-23,591,962) were generated (Table 3.3.2). The filtration of the low quality reads and the reads belonging to the inserted transposon and the *E. coli* host strain removed $65.13\% \pm 2.85$ of the raw reads and reduced the average number of reads per sample to 9,737,752 (range 11,277,616 – 8,170,620) (Table 3.3.2).

Table 3.3.2: Number of plasmidome reads generated from each sample. The read numbers represent the total number of sequencing reads generated by two lanes from two independent Illumina HiSeq 2500 sequencing runs.

Sample	Number of raw reads	Number of reads post <i>E. coli</i> /transposon filtration step	Remaining reads after quality filtering	% Reads filtered
Control 1	28,747,012	11,675,344	8,902,621	69.03
Control 2	30,952,670	13,276,520	10,575,089	65.83
Control 3	32,413,858	12,409,791	9,941,283	69.33
Control 4	30,448,836	13,595,683	11,277,616	62.96
UC 1	23,591,962	10,679,412	8,925,847	62.17
UC 2	24,085,724	11,031,841	9,259,280	61.56
UC 3	25,208,470	11,552,456	9,717,539	61.45
UC 4	25,580,554	10,092,409	8,170,620	68.06
CD 1	28,814,416	12,206,395	10,461,650	63.69
CD 2	31,407,352	13,286,368	10,243,112	67.39
CD 3	26,706,410	11,508,233	9,472,077	64.53
CD 4	28,792,622	12,521,162	9,906,293	65.59

3.3.4. The plasmidome host taxa is different than the 16S rRNA-identified phylotypes

To identify the potential hosts harboring the identified plasmids, the reads were aligned to the NCBI-NR protein database, and the host identity was retrieved and summarized. Most of the reads were assigned to the bacterial domain ($72.13\% \pm 2.35$) with a minor representation of Eukaryota ($5.3\% \pm 1.1$), Archea ($0.0006\% \pm 0.0002$), Viruses ($0.32\% \pm 0.16$), unclassified sequences ($0.12\% \pm 0.04$) and other sequences ($22.15\% \pm 1.5$) that mainly belonged to unclassified plasmids (Figure 3.3.2 A). The assigned bacteria belonged to 13 bacterial phyla in addition to the unclassified bacteria and the bacteria isolated from environmental samples (Figure 3.3.2 B). The most abundant bacterial phyla were Proteobacteria ($49.84\% \pm 3.7$), Firmicutes ($20.34\% \pm 4.8$), Bacteroidetes ($12.05\% \pm 8.9$), Spirochaetes ($1.26\% \pm 0.21$) and Actinobacteria ($0.96\% \pm 0.27$) in addition to environmental bacteria ($14.13\% \pm 3.9$) and unclassified bacteria ($1.33\% \pm 0.39$) (Figure 3.3.2 B). At the genus level, the plasmidomes mainly originated from *Escherichia* ($25.3\% \pm 3.4$), *Bacteroides* ($12.1\% \pm 9.5$), *Klebsiella* ($6.5\% \pm 1.1$), *Tyzzarella* ($5.7\% \pm 2.3$) and *Salmonella* ($5.5\% \pm 1.2$) (Figure 3.3.2 C). No significant difference between the relative abundances of the assigned plasmidome hosts of the control, UC or CD samples was identified.

To evaluate the correlation between the phylotypes identified by the gut plasmidome analysis and the 16S rRNA-based assessment, the taxonomic data generated by both approaches from the same samples were compared. At the phylum level, the gut microbiota was dominated by Bacteroidetes ($51\% \pm 19.7$),

Firmicutes ($37.2\% \pm 15.6$) and, to a lesser extent, by Proteobacteria ($9.27\% \pm 13.3$) and Actinobacteria ($2.47\% \pm 4.2$) based on the 16S rRNA-based approach (Figure 3.3.2 D). The pairwise correlation at the phylum level between the plasmidome and 16S rRNA data was calculated using a Spearman's rank correlation test that indicated a modest positive correlation between the two datasets with a Spearman's $r = 0.52$ ($P < 0.0001$). At the genus level, the gut microbiota was dominated by *Bacteroides* ($39.3\% \pm 11.4$), *Faecalibacterium* ($5.04\% \pm 6.9$), *Ruminococcus* ($4.8\% \pm 3.1$), *Parabacteroides* ($4.3\% \pm 5.6$), and *Dorea* ($3.4\% \pm 4.3$). Spearman's rank correlation analysis of the identified genera by the plasmidome and 16S rRNA approaches indicated weak positive correlation between the two datasets with a Spearman's $r = 0.31$ ($P < 0.0001$).

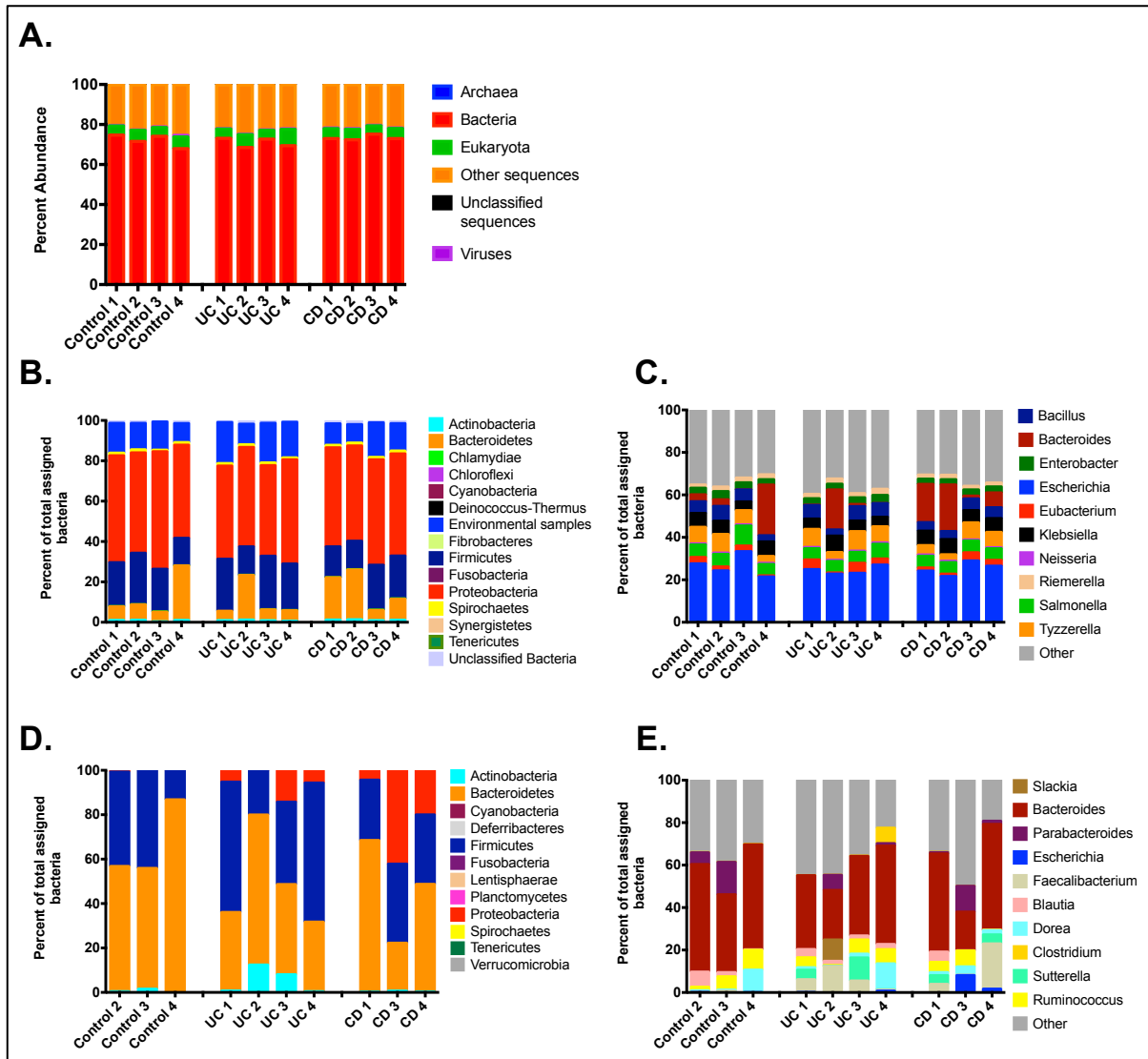


Figure 3.3.2: Taxonomic analysis of the gut plasmidomic data. (A-C) represent the plasmidome taxonomy data generated from the 4 control subjects, and the 4 UC and 4 CD patients using the NCBI-NR database. (D-E) Represents the 16S rRNA-based taxonomy generated from the 3 control subjects, and 4 UC and 3 CD patients using the QIIME software v1.9.0 [283] and the Greengenes database. The stacked bars represent the percent of abundance of the taxa in individual samples. (A) The relative abundance of the assigned taxa at the highest taxonomic rank. (B) The relative abundance of identified bacteria at the phylum level. (C) The relative abundance of identified bacteria at the genus level; the ten most abundant bacterial genera are listed while with the highest abundance and the remaining genera are combined together as “Other”. (D) The relative abundance of the identified bacteria at the phylum level based on the 16S rRNA sequencing. (E) The relative abundance of the identified bacteria at the genus level based on the 16S rRNA sequencing; the ten most abundant bacterial genera are listed while with the highest abundance and the remaining genera are combined together as “Other”.

3.3.5. Functional analysis of the plasmidome data

To characterize the functional properties of the plasmidome, the reads were aligned to the COGs functional database (<http://www.ncbi.nlm.nih.gov/COG/>). The plasmidome exhibited a wide range of functional classifications with 21 COGs categories identified. Four categories represented $99\% \pm 0.8$ of the total assigned functions. These categories were lipid metabolism ($43.04\% \pm 6.1$), amino acid metabolism and transport ($25.9\% \pm 8.2$), translation ($19.8\% \pm 3.5$) and nucleotide metabolism and transport ($10.24\% \pm 4.1$) (Figure 3.3.3). No significant difference between the groups was identified by comparing the COGs functional categories of the control to the IBD or IBD subtypes ($P>0.05$).

To determine if the plasmidome COGs functional categories were different from those of the microbiota metagenome, the metagenome COGs functional categories were predicted from the 16S rRNA sequencing data using PICRUST and compared to the COGs functional categories of the plasmidome. In total, 25 COGs functional categories were predicted from the 16S rRNA sequencing data (Figure 3.3.4). The predicted metagenome COG functional categories were more diverse than the plasmidome COG functional categories (Figure 3.3.4) with relative abundance of each category ranging from 11.4% - 0.005%. The 4 COG functional categories of the predicted metagenome with the highest abundance were general functional prediction only ($11.4\% \pm 0.21$), carbohydrate transport and metabolism ($9.43\% \pm 1.04$), transcription ($8.6\% \pm 0.72$) and amino acid transport and metabolism (7.71 ± 0.26). Comparing the relative abundance of each COG functional category between the plasmidome and the 16S rRNA datasets using the

Wilcoxon signed-rank test revealed that the relative abundance of all identified COGs functional was significantly different between the two datasets ($P > 0.01$; Figure 3.3.5).

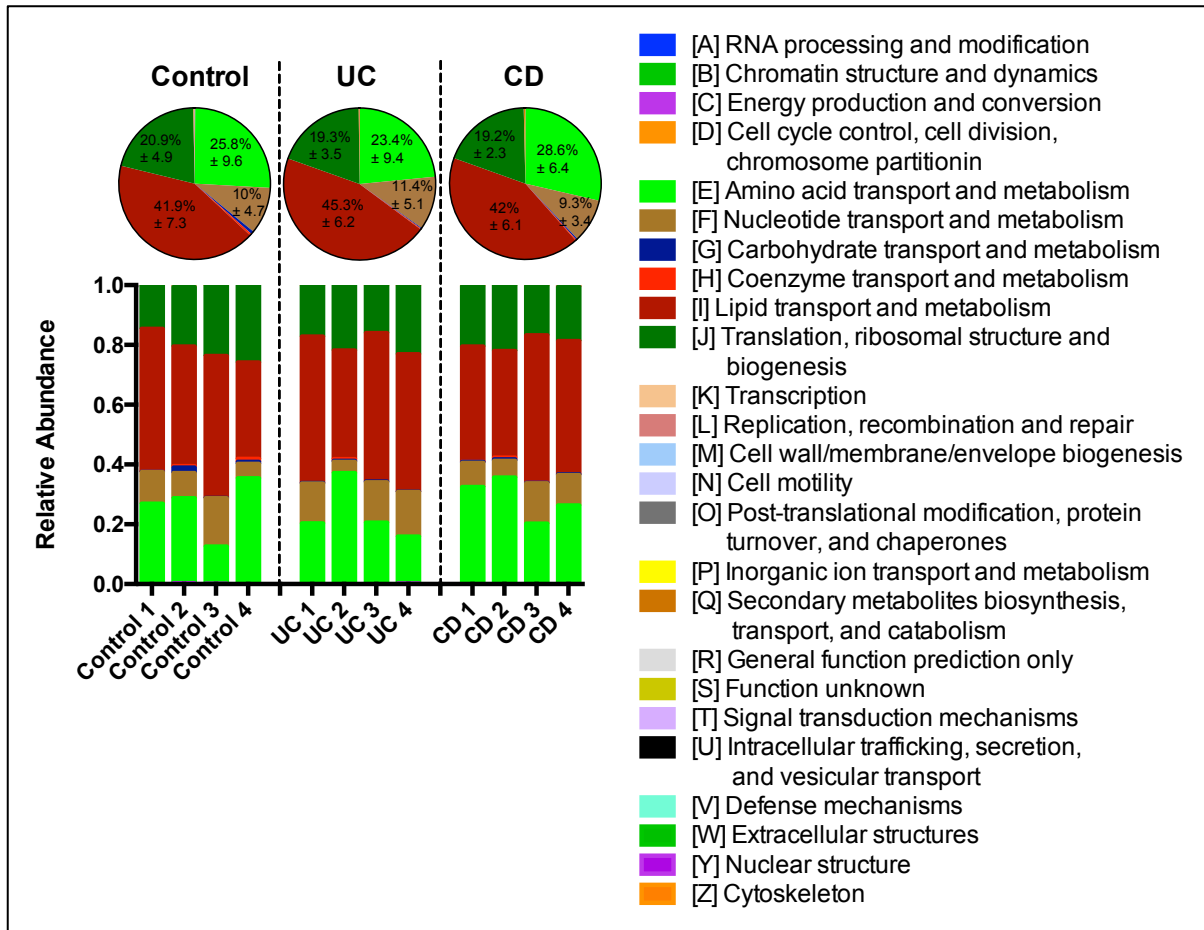


Figure 3.3.3: Plasmidome functional categories in each group. Plasmidome sequences from each sample were aligned against the COGs functional database (<http://www.ncbi.nlm.nih.gov/COG/>). Pie charts represent the average relative abundance of the COGs functional categories in each group. Stacked bars represent the relative abundance of the COGs functional categories in each sample. The comparison of the COGs functional categories between the groups was carried out using a Kruskal-Wallis test followed by a Dunn's post hoc test and showed no significant difference between the groups $P > 0.05$.

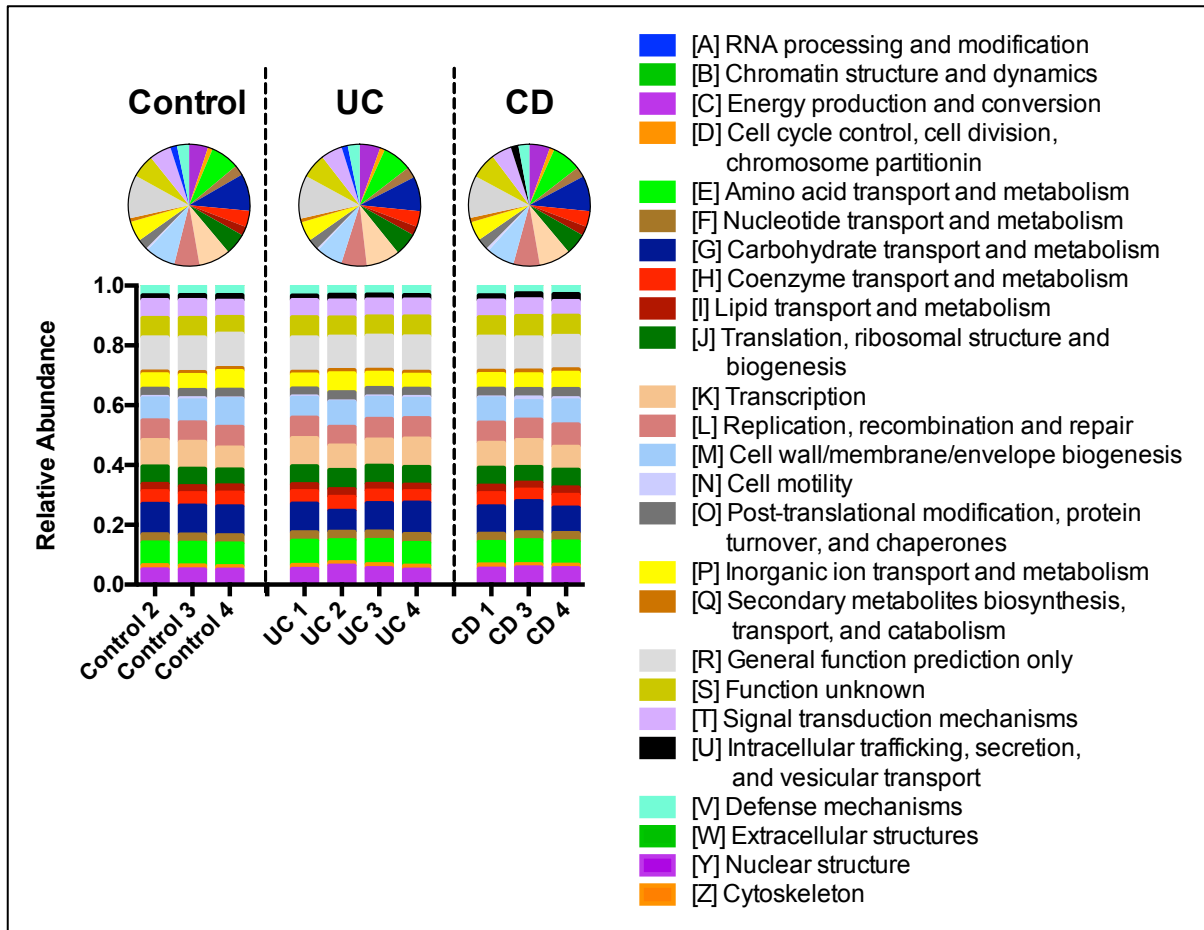


Figure 3.3.4: Predicted metagenome COGs functional categories from the 16S rRNA sequencing data. The relative abundance of the COGs functional categories was predicted from the 16S rRNA sequencing data using the PICRUST algorithm [331]. Pie charts represent the average relative abundance of the COG functional categories in each group. Stacked bars represent the relative abundance of the COG functional categories in each sample.

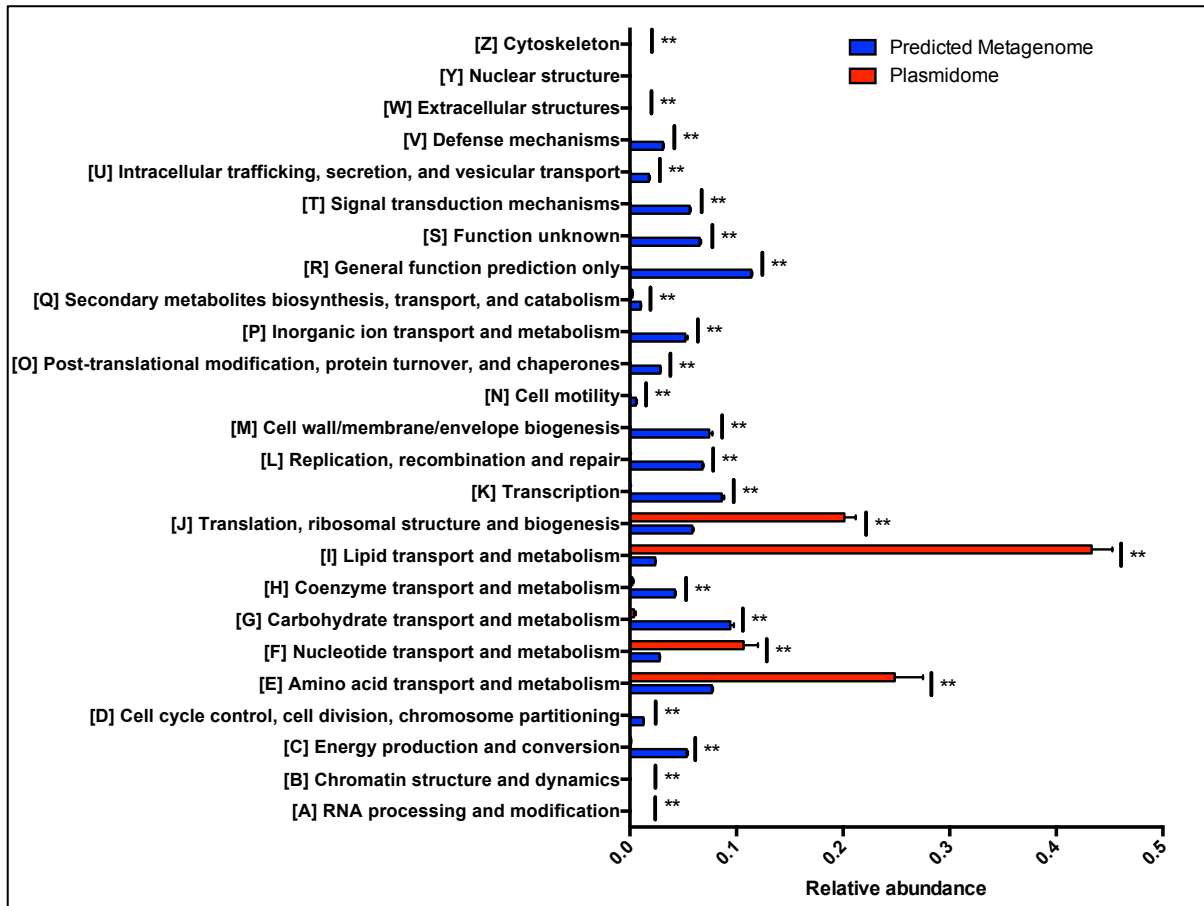


Figure 3.3.5: Comparative analysis of the COGs functional categories of the predicted metagenome vs. the plasmidome. The relative abundance of the COGs functional categories of the predicted metagenome and the plasmidome was compared using the Wilcoxon signed-rank test. **, $P > 0.01$.

3.3.6. Characterizing the plasmidome function using the NCBI-NR reference databases

The plasmidome reads were further aligned to the NCBI-NR database. In total, 7,662 genes were identified from the plasmidome reads, and most of the genes were shared between the groups (Figure 3.3.6 A-B). To estimate whether our sequencing coverage was sufficient to comprehensively identify the plasmidome genes from our samples, we performed a rarefaction analysis. This analysis, which produced a rarefaction curve reaching a plateau (Appendix V), indicates that most plasmidome genes were identified. The evaluation of the alpha diversity based on the assigned genes, as represented by the Chao1 and Shannon diversity indices, showed that neither IBD nor the IBD subtypes exhibited a difference in richness or evenness compared to the control subjects ($P>0.05$; Figure 3.3.6 C-F). Nevertheless, a Mann Whitney test identified 145 differentially abundant genes between the control and IBD patients ($P<0.05$). These differentially abundant genes were grouped into 18 functional annotations (Figure 3.3.7 A-B and Appendix VI). The control group was enriched in genes that code for (arranged from the highest to the lowest abundance) hypothetical proteins, aminoglycoside phosphotransferases, ligases, serine/threonine kinases, replication-associated proteins, acetyl-CoA acetyltransferases, transcriptional regulators, arabinose operon regulatory proteins, DNA methylases, 3-hydroxyacyl-CoA dehydrogenases and DNA-damage repair proteins (Figure 3.3.7 B). In contrast, the IBD group plasmidome was enriched in genes that code for (arranged from the highest to the lowest abundance) beta-lactamases, beta-galactosidases, plasmid mobilization proteins, phage proteins and

transposases (Figure 3.3.7 B). Interestingly, the PCoA analysis based on the abundance of the differential abundant genes clearly separated the control from the IBD samples with 40.67% of the variance accounted for by PCoA1, 24.13% of variance attributable to PCoA2 and 11.54% of variance attributable to PCoA3 (Figure 3.3.8). Accordingly, the IBD plasmidome appears to harbor distinctive functions from that of the control.

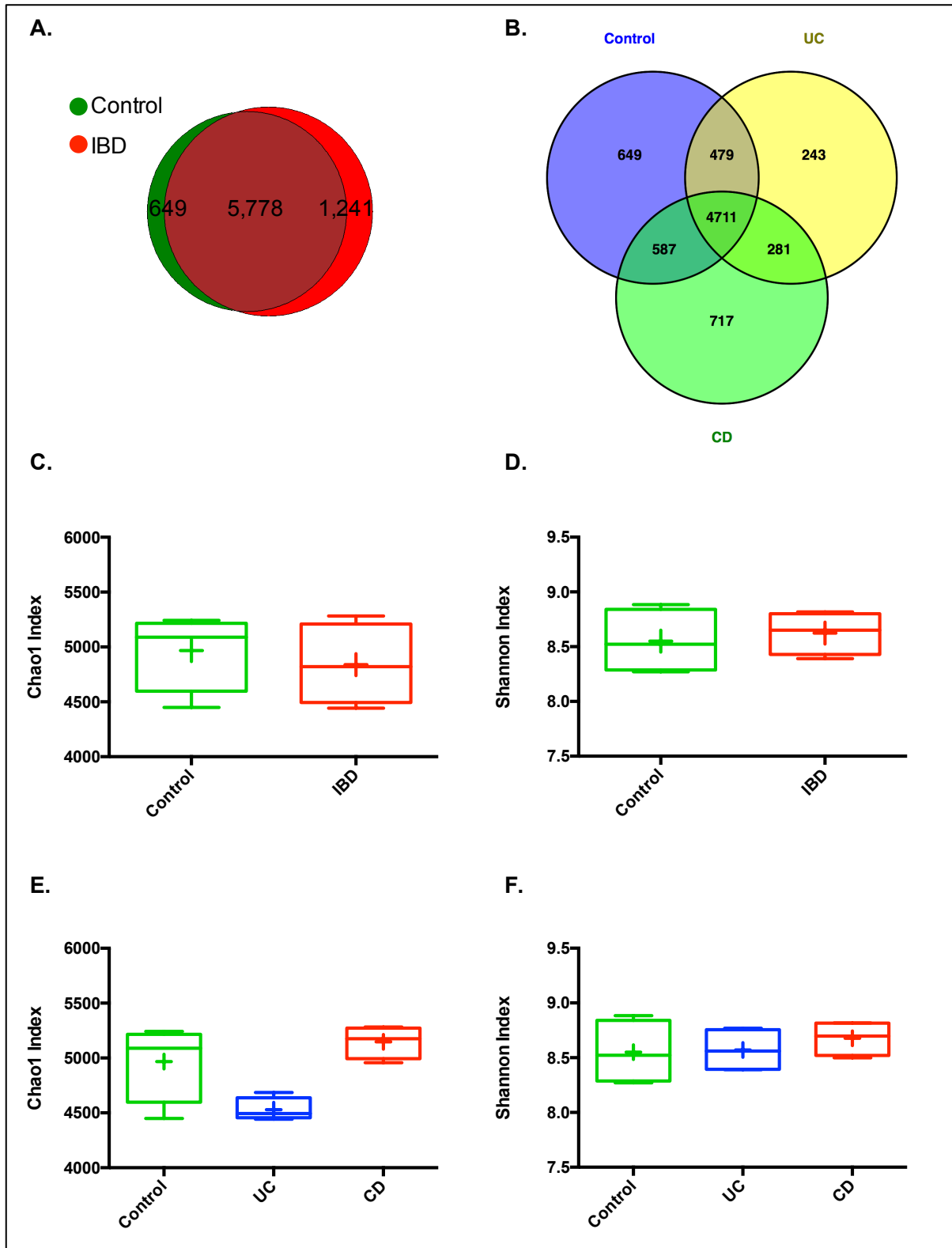


Figure 3.3.6: Diversity of pediatric gut plasmidome. Number of identified NCBI-NR genes using 75% sequence identity, an E-value of 1×10^{-5} and a Bit-score of 35

for cutoff in the control and IBD samples (A) or as a function of IBD subtypes (B). Alpha diversity of identified genes, represented by Chao1 (richness) and Shannon diversity indexes, were calculated using QIIME v1.9.0 and are shown for control and IBD (C-D) or control, UC and CD (E-F). The number of reads was equalized between the samples to 5,655,160 reads. The comparison of Chao1 and Shannon indices between the control and the IBD was carried out using a Mann-Whitney *U* test and a Kruskal-Wallis test followed by a Dunn's post hoc test to compare control vs. UC vs. CD. Both statistical analyses showed no statistical significance between the groups ($P>0.05$).

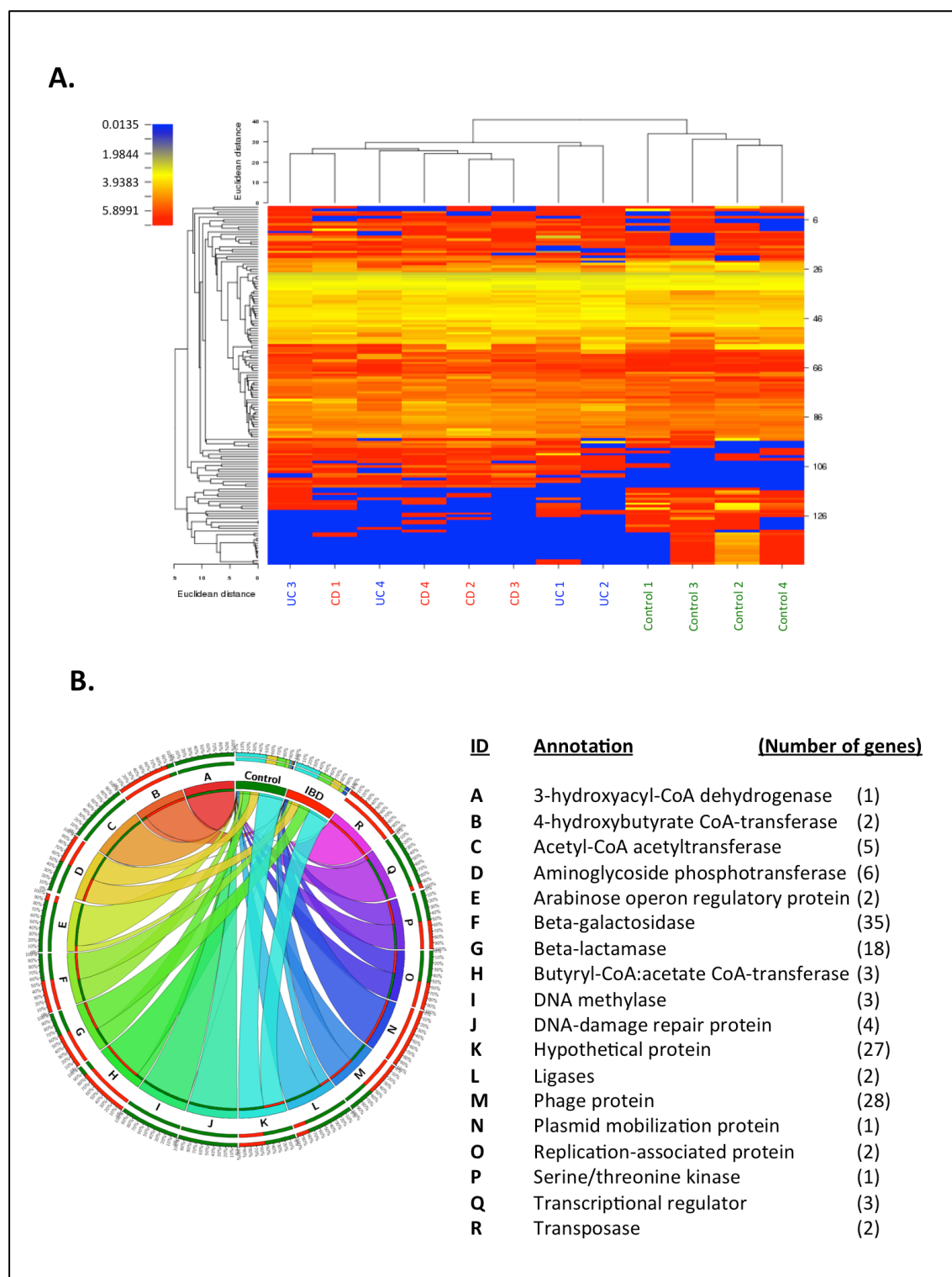


Figure 3.3.7: Genes identified as differentially abundant in the plasmidome of IBD patients. Differentially abundant genes were identified using a Mann-Whitney U test (difference in gene with $P < 0.05$ was considered significant). (A) Heat map constructed using the CIMminer tool (<http://discover.nci.nih.gov/cimminer/>) and the

$\log_{10}$ transformed relative abundance of differentially abundant genes. Rows and columns are clustered using Euclidian distance and the average linkage cluster algorithm. (B) General annotation of the differentially abundant genes. Circos [332] was used to plot the percent of abundance of the differentially abundant genes. The ribbon width is proportional to the percent abundance of genes. The percent abundance of genes with the same annotation is combined, and the number of genes under each annotation category is indicated.

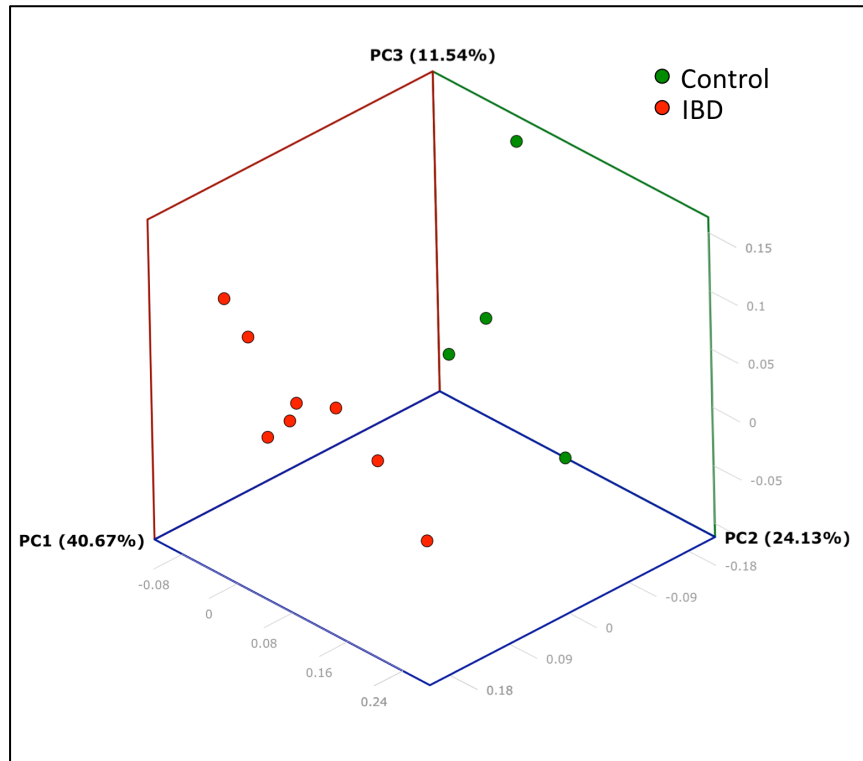


Figure 3.3.8: Beta diversity of the differentially abundant plasmidome genes identified by comparing the relative abundance of the NCBI-NR database assigned genes between the control and the IBD. The relative abundance of the differentially abundant genes was used to calculate the Bray–Curtis dissimilarity measures between the samples. Then, a three-dimensional principal coordinate analysis (PCoA) plot of the Bray–Curtis dissimilarity measures was performed using QIIME v1.9.0. The percentage of variance explained by each component is presented for each axis.

The plasmidome composition was also compared between the controls and the IBD subtypes. Comparing the relative abundance of the NCBI-NR database assigned plasmidome genes between the control, UC and CD samples identified 90 differential genes ($P < 0.05$) that fell under 10 different functional annotations (Figure 3.3.9 A-B and Appendix VI). The UC plasmidome was enriched with aminoglycoside phosphotransferase genes compared to the CD samples. On the other hand, the CD plasmidome had a higher relative abundance of genes coding for hypothetical proteins, beta-lactamases, phage proteins and adenosine-5'-phosphosulfate reductases compared to the control and UC samples (Figure 3.3.9 B). The control, UC and CD samples were well separated on a 3D-PCoA with 64.87% of the variance attributed to the first axis (PCoA1), 18.71% to PCoA2 and 7% to PCoA3 (Figure 3.3.10). Thus, the relative abundance of the NCBI-NR database assigned plasmidome genes demonstrates key differences that not only differentiate between the control and the IBD but also enable IBD subtyping.

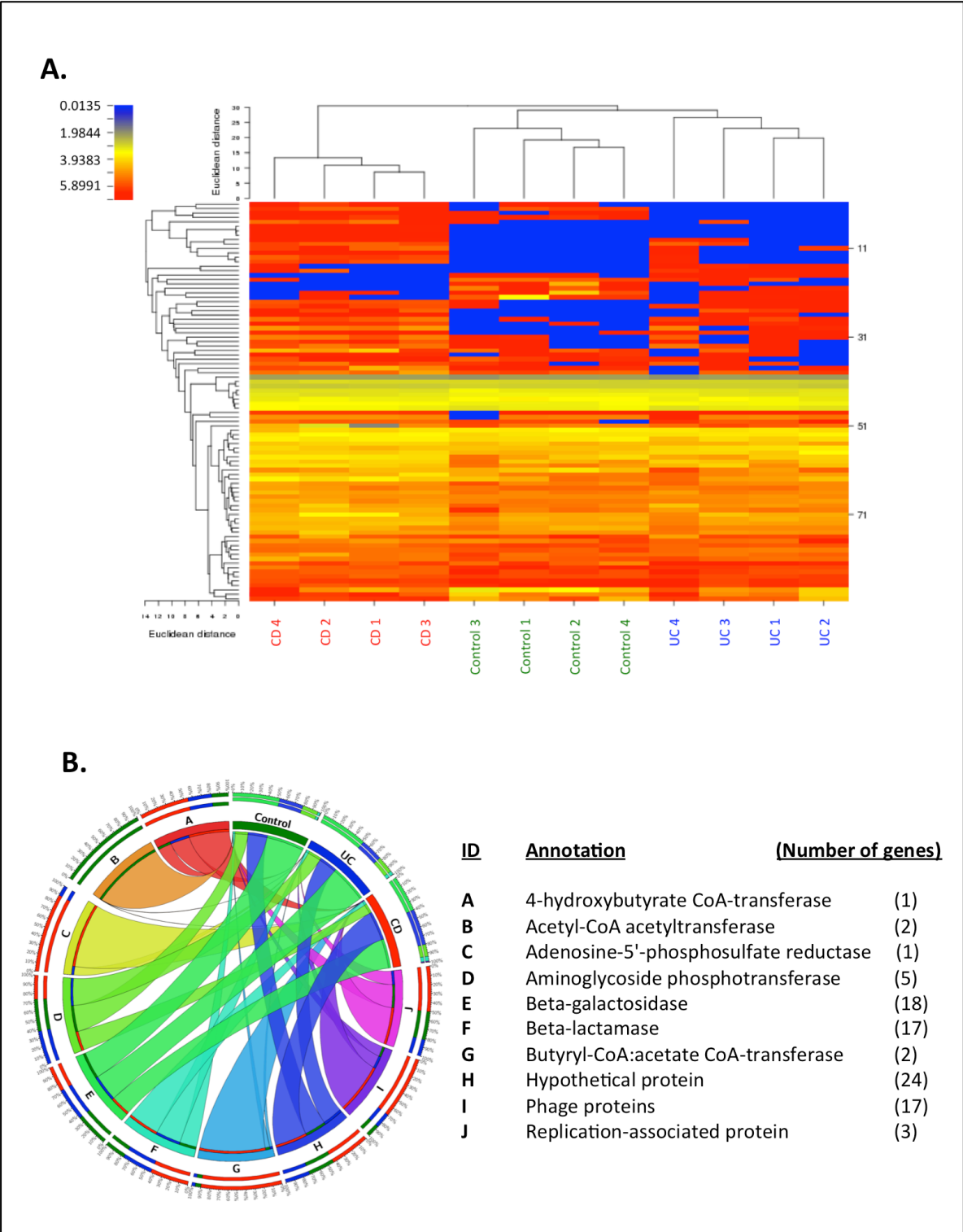


Figure 3.3.9: Plasmidome differentially abundant genes identified by comparing the relative abundance of the NCBI-NR database assigned genes between ulcerative colitis (UC), Crohn's disease (CD) and control samples. Differentially abundant genes were identified using a Kruskal-Wallis test followed by

a Dunn's post hoc test (Dunn's corrected $P < 0.0167$). (A) The relative abundance of the differentially abundant genes was $\log_{10}$ transformed, and a heat map was constructed using the CIMminer tool (<http://discover.nci.nih.gov/cimminer/>). Rows and columns are clustered using the Euclidian distance and the average linkage cluster algorithm. (B) General annotation of the differentially abundant genes. Circos [332] was used to plot the relative abundance of differentially abundant genes. The ribbon width is proportional to the percent of abundance of genes. The relative abundance of genes with the same annotation is combined, and the number of genes under each annotation category is indicated in parentheses.

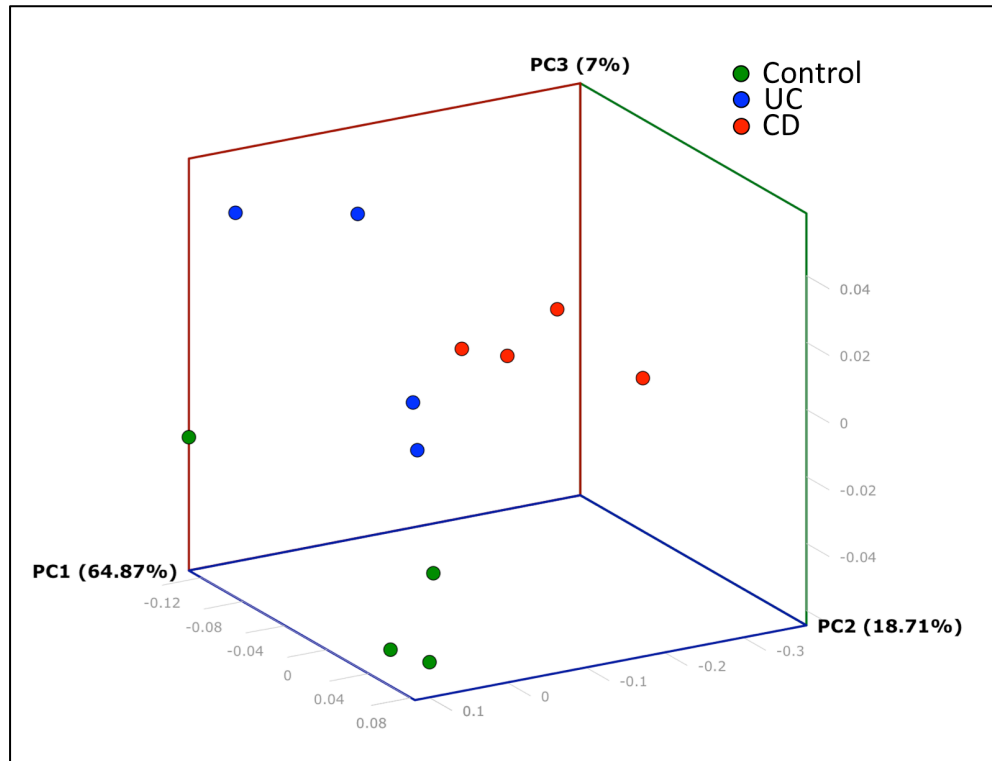


Figure 3.3.10: Beta diversity of the differentially abundant plasmidome genes identified by comparing the relative abundance of the NCBI-NR database assigned genes between ulcerative colitis (UC), Crohn's disease (CD) and control samples. The relative abundance of the differentially abundant genes was used to calculate the Bray–Curtis dissimilarity measures between samples. Then, a three-dimensional principal coordinate analysis (PCoA) plot of the Bray–Curtis dissimilarity measures was performed using QIIME v1.9.0. The percentage of variance explained by each component is presented for each axis.

3.3.7. Characterizing the plasmidome function using other reference databases

The plasmidome reads were also aligned to additional reference databases, including the HMP-GIT, ACLAME and VFDB databases, for a comprehensive characterization of the plasmidome function. Alignment of the plasmidome sequences to these databases confirmed and complemented the results obtained by the NCBI-nr database. For instance, the alignment of the plasmidome data to the Human Microbiome Project Gastrointestinal_tract reference sequence (HMP-GIT) database identified a total of 1,042 genes. The majority of the assigned genes were shared between the control, UC and CD samples (Appendix VII). Similar to the results obtained by the NCBI-NR database, comparing the relative abundance of the HMP-GIT database assigned genes between the control and the IBD samples showed a higher relative abundance of genes that code for 3-hydroxyacyl-CoA dehydrogenases, acetyl-CoA acetyltransferases, DNA-damage repair proteins and hypothetical proteins in the gut plasmidome of controls compared to the IBD samples. Interestingly, the IBD sample plasmidome was enriched with genes that code for phage proteins compared to the controls (Figure 3.3.11 A and Appendix XIII-IX). The relative abundance of phage proteins was significantly higher in UC compared to CD (Dunn's corrected $P < 0.0167$; Figure 3.3.11 B and Appendix XIII-IX).

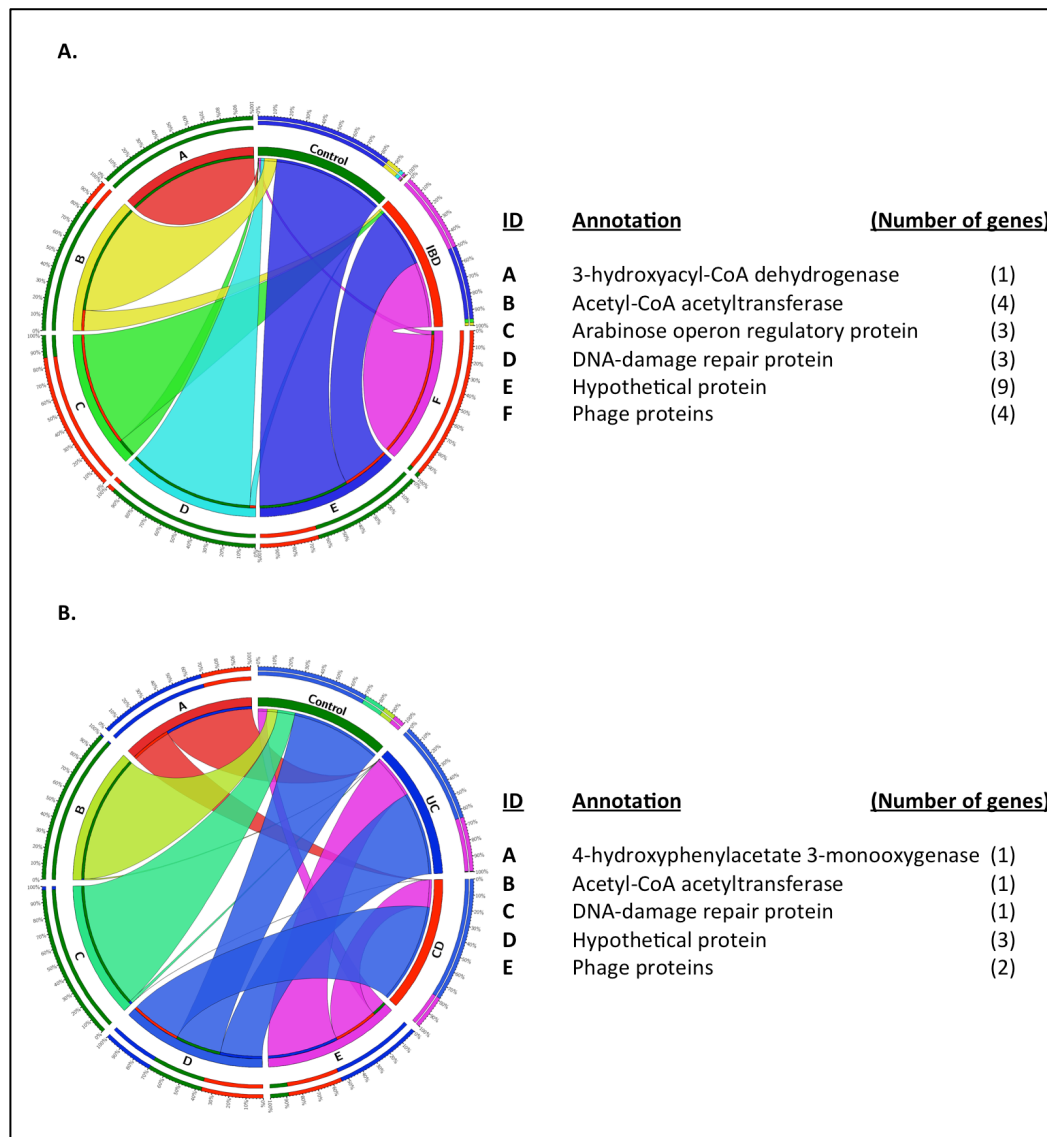


Figure 3.3.11: General annotation of differentially abundant genes identified by comparing the relative abundance of the HMP-GIT database assigned genes between ulcerative colitis (UC), Crohn's disease (CD) and control samples. Differentially abundant genes between the control and the IBD samples were identified using a Mann-Whitney U test ($P < 0.05$). Differentially abundant genes between the control, UC and CD samples were identified using a Kruskal-Wallis test followed by Dunn's post hoc test (Dunn's corrected $P < 0.0167$). Circos [332] was used to plot the relative abundance of the differentially abundant genes; the ribbon width is proportional to the percent abundance of genes. The relative abundance of genes with the same annotation is combined, and the number of genes under each annotation category is indicated in parentheses. (A) Genes identified as differentially abundant between the controls and the IBD patients. (B) Genes identified as differentially abundant between the controls and the IBD subtypes.

The alignment of the plasmidome reads to the A CLAssification of Mobile genetic Elements (ACLAME) database identified 390 genes (Appendix XI). Comparing the relative abundance of the genes assigned to the ACLAME database between the control and the IBD samples showed similar results to the findings obtained using the NCBI-NR database. The control group showed an overall increase in the relative abundance of genes that code for hypothetical proteins, while the IBD patients had an overall enrichment in the abundance of genes that code for phage proteins (Figure 3.3.12 and Appendix XII-XIII). These phage proteins were subdivided into five categories: phage DNA replication, phage function unknown, phage head/capsid internal protein, phage head/capsid major protein and phage prohead/capsid assembly, all of which were increased in the IBD plasmidome except for the phage prohead/capsid assembly related genes, which were increased in the plasmidome of the control (Figure 3.3.12 A and Appendix XII-XIII). The UC plasmidome harbored a higher relative abundance of genes that code for phage head/capsid major protein as compared to the controls, while the CD plasmidome had a higher relative abundance of genes that code for phage DNA replication, phage function unknown and phage head/capsid internal protein.

Finally, using the Virulence Factors (VFDB) database, only a few of the plasmidome reads (0.011% - $2.05 \times 10^{-05}\%$) could be annotated. While this analysis identified 62 genes, the relative abundance of these genes was not affected by the disease status (data not shown).

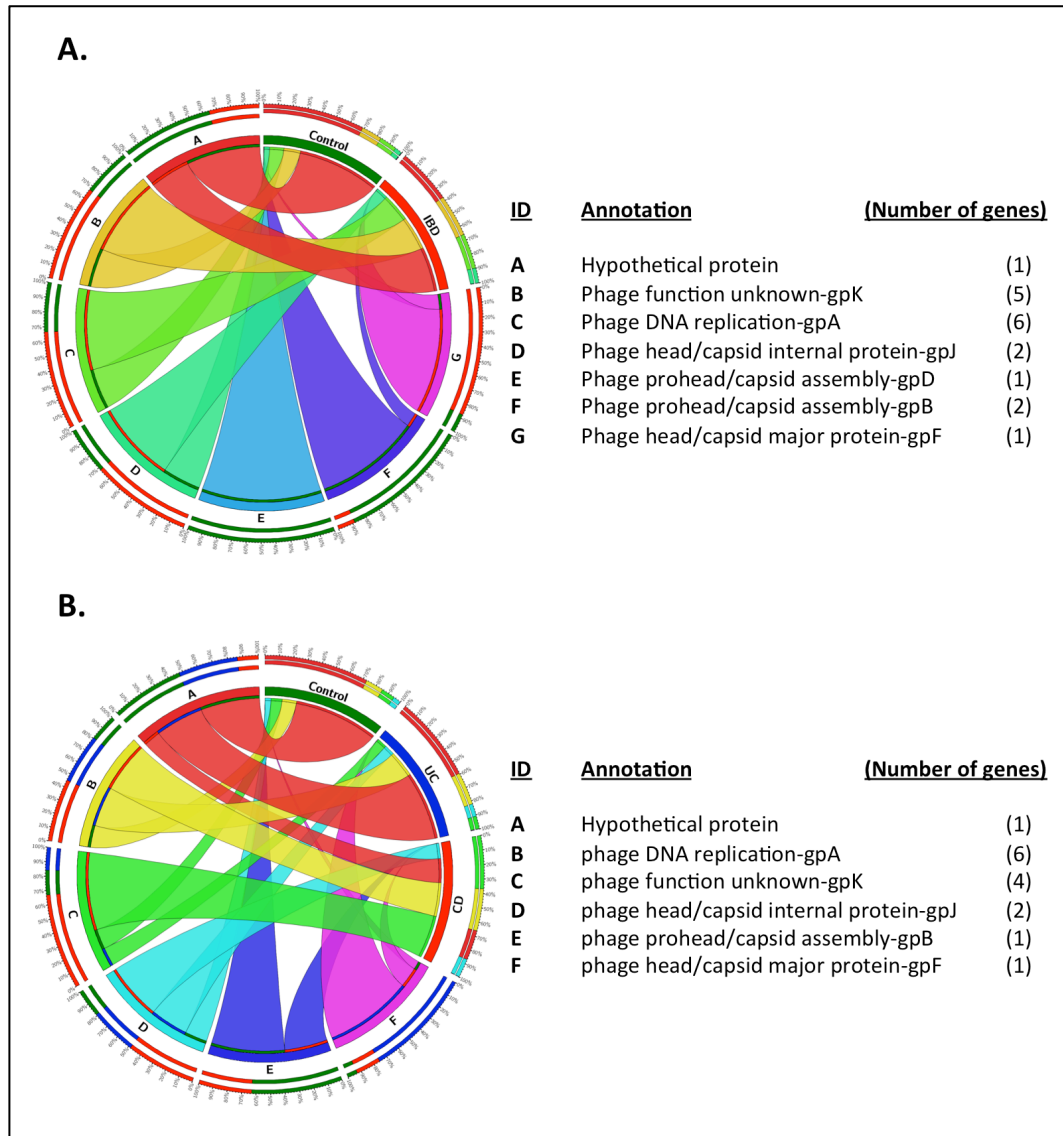


Figure 3.3.12: General annotation of the differentially abundant genes identified by comparing the relative abundance of the ACLAME database assigned genes between ulcerative colitis (UC), Crohn's disease (CD) and control samples. Differentially abundant genes between the control and the IBD samples were identified using a Mann-Whitney U test ($P < 0.05$). Differentially abundant genes between the control, UC and CD samples were identified using a Kruskal-Wallis test followed by Dunn's post hoc test (Dunn's corrected $P < 0.0167$). Circos [332] was used to plot the relative abundance of the differentially abundant genes; the ribbon width is proportional to the percent abundance of genes. The relative abundance of genes with the same annotation is combined, and the number of genes under each annotation category is indicated in parentheses. (A) Genes identified as differentially abundant between the control and IBD samples. (B) Genes identified as differentially abundant between the samples from the control and the IBD subtypes.

3.4. Discussion

This work presents the first study characterizing the gut plasmidome of IBD patients. The gut plasmidome was isolated from mucosal-luminal interface (MLI) samples collected from the right colon of first onset pediatric IBD patients and non-IBD control subjects. Analysis of the plasmidome DNA by restriction enzyme digestion revealed different fragment patterns between the control and the IBD samples. Moreover, the plasmidome restriction profile exhibited visible differences between the UC and the CD samples. Finally, the gut plasmidome sequencing further highlighted the differences between the control, IBD and IBD subtypes plasmidomes by identifying plasmidome-specific genes that correlate either positively or negatively with the disease status.

A comparative metagenomic study of human gut plasmidome strongly suggested a possible association between the gut plasmidome and IBD [191]. Our data further supports this hypothesis in first onset pediatric IBD using a more comprehensive approach. First, the restriction digestion of our gut-plasmidome demonstrated a group-specific restriction profile highlighting a potential difference in the plasmidome composition between the control group, the IBD samples, and the IBD subtypes. This observation prompted us to perform a more detailed evaluation of the isolated plasmidome through next-generation sequencing.

To ensure our sequencing reads are free from sequences belonging to the employed transposon and the cloning strain genomic DNA (introduced during the plasmidome purification from the transformants), we conducted a stringent read

filtration. The filtration resulted in the removal of about 65% of our original reads. Since the human stool plasmidome isolated using the TRACA approach ranged in size from 3-10 kb [321] and the employed transposon size was 1,938 bp, we expected the transposon to account for at least 16.24%-39.25% of the plasmidome reads, assuming a single transposon insertion per plasmid. Consequently, the remaining percentage of the 65% filtered reads (25.75%-48.76%) was expected to correspond to the cloning bacteria chromosomal DNA. Rarefaction analysis indicated that the number of filtered high quality reads generated in our study should be sufficient to characterize the pediatric gut plasmidome.

The hosts of the pediatric gut plasmidome mainly belonged to the bacterial domain ($72.13\% \pm 2.35$). Three bacterial phyla represented up to 80% of the bacterial plasmidome hosts: Proteobacteria ($49.84\% \pm 3.7$), Firmicutes ($20.34\% \pm 4.8$) and Bacteroidetes ($12.05\% \pm 8.9$). The high abundance of sequences originating from Proteobacteria is in agreement with the plasmidomes of other microbial communities, including the plasmidome from the bovine rumen [322]. Notably, the relative abundances of the identified phyla varied significantly from the proportions of the taxa assigned based on the 16S rRNA sequencing from the same samples. Indeed, the 16S rRNA based phylogenetic analysis revealed a significant different phyla distribution dominated by Bacteroidetes ($51\% \pm 19.7$) followed by Firmicutes ($37.2\% \pm 15.6$) and then Proteobacteria ($9.27\% \pm 13.3$). The variations in the phyla proportions between the phylogenetic profiles obtained by the 16S rRNA-based and the TRACA-based approaches suggest

that Proteobacteria harbor more plasmids than the bacteria belonging to the Firmicutes and Bacteroidetes phyla. At the genus level, *Escherichia*, *Bacteroides*, *Klebsiella* and *Salmonella* were the most dominant plasmidome hosts within the Proteobacteria phylum. It is well documented that these genera harbor a wide assortment of plasmids which carry a variety of functions ranging from antibiotic resistance to virulence [333, 334]. Therefore, these genera represent the major contributors to the gut plasmidome; however, the exact host of these plasmids is unknown since the same plasmid can be hosted by phylogenetically different bacteria [335]. To note, the domination of *Escherichia* in the plasmidome is not attributed to contamination from the cloning strain genomic DNA during plasmid purification since none of the reads matched the cloning strain DNA following our multiple quality filtration steps. To note, the TRACA approach is expected to have bias toward Proteobacteria since *E. coli* plasmids doesn't require correct insertion of the transposon leading to more than expected *E. coli* plasmids. Although we did not observe a significant difference in the plasmidome hosts relative abundance between the control, the IBD, and the IBD subtypes, the genetic composition of their plasmidome was significantly different, as discussed below.

The functional analysis of the pediatric gut plasmidome revealed that a number of functional categories were enriched in both the control and the IBD plasmidomes. Four main COG functional categories were identified: lipid metabolism, amino acid metabolism and transport, translation and nucleotide metabolism and transport. The relative abundance of these plasmidome

functional categories was significantly different from those identified by the PICRUST predicted metagenome. This observation implies that the gut plasmidome is enriched with genes that code for specific biological functions. A typical western diet is known to contain high fat, protein, and carbohydrates [336]. Processing of these dietary components involves both human and gut microbiota [153]. This mutualistic relationship benefits the microbiota by generating energy from processing the available nutrient sources from the surrounding environment [153]. Gut microbiota is known to be involved in lipid absorption, storage, and metabolism [153]. For instance, germ-free mice fed a high-fat diet had higher fecal fat and less weight gain than conventional mice [275]. Furthermore, the uptake of fatty acids from the intestine of gnotobiotic zebrafish was stimulated by the gut microbiota, and the number of lipid droplet formations in the intestinal epithelium and liver was positively correlated to the abundance of members of the Firmicutes [337]. The higher abundance of genes involved in amino acid metabolism and transport, translation and nucleotide metabolism and transport is a characteristic of a biosynthetic pathways rich environment [322]. Enrichments in amino acid metabolism was also observed in the plasmidome of cows rumen [322]. Indeed, plasmidome genes are thought to supplement their hosts with additional functions that facilitate the hosts' adaptation to different environments [338].

A number of plasmidome functional genes were enriched in our non-IBD subjects compared to the IBD patients. For instance, the non-IBD plasmidome was enriched with acetyl-CoA acetyltransferase and 3-hydroxyacyl-CoA

dehydrogenase genes, and this result was consistent across the different reference databases. Both genes code for key enzymes involved in the digestion of dietary fibers and butyrate synthesis [157]. Louis *et al.* found a conserved arrangement of six genes involved in butyrate synthesis with a high level of sequence relationships across various species of Gram-positive bacteria isolated from the human gut [195]. This finding suggested that LGT contributed to the spread of butyrate biosynthesis genes between colonic Gram-positive bacteria. In our study, the presence of genes involved in butyrate biosynthesis within the plasmidome further supports this hypothesis. Taking into consideration the anti-inflammatory effects of butyrate [246-253, 258, 259], this may ultimately protect the colonocytes from inflammation; however, this assumption depends on the level of expression of these genes from the plasmidome. Another key finding in the non-IBD plasmidome was the increased relative abundance of serine/threonine kinase genes. The microbial serine/threonine kinase phosphorylates the hydroxyl group of serine or threonine residues upon sensing external stimuli [339]. As a result, it can control the activity of a targeted protein by changing its conformation or through a protein-protein interaction. Consequently, it regulates a wide range of cellular processes, including metabolic activity, mRNA translation, protein and cell wall synthesis, cell division and virulence (reviewed in [339]). A higher abundance of the serine/threonine kinase gene in the non-IBD plasmidome is coupled with the enrichment of other genes that code for regulatory proteins and DNA-damage repair proteins including transcriptional regulators, DNA methylases, and arabinose operon

regulatory proteins. Altogether, the plasmidome of the non-IBD control group had higher abundance of genes involved in external stimuli response, gene regulation, and damage repair than the IBD plasmidome.

Interestingly, a subset of functional genes was enriched in the IBD patient plasmidomes compared to the control samples. Although none of our IBD patients received antibiotic treatment for at least 4 weeks prior to sample collection, the CD plasmidome was enriched with antibiotic resistance genes. The CD plasmidome had an elevated level of beta-lactamases as compared to the controls. These findings suggest that the CD microbiota was exposed, prior diagnosis, to antibiotics to a greater extent than the control non-IBD microbiota, enabling the natural selection of the observed antibiotic resistant genes. Early antibiotic use due to various childhood infections, including middle ear infections, gastrointestinal infections and upper respiratory tract infections, is known to increase the risk of pediatric IBD [63, 91]. Furthermore, pediatric CD patients exposed to antibiotics at the time of diagnosis showed a significant difference in their microbiota composition compared to the CD patients unexposed to antibiotics [81]. Gevers *et al.* showed an increased dysbiosis in the microbiota of the CD patients exposed to antibiotics with about a 10-fold increase in the relative abundance of *Enterobacteriaceae* and *Fusobacteriaceae* and a reduced level of *Pasteurellaceae* and *Veillonellaceae* as compared to microbiota of unexposed CD patients [81]. Hence, the high abundance of antibiotic resistance genes in the plasmidome of our CD patients may indicate earlier exposure to infections and antibiotic treatment. Unfortunately, this data was not collected as

part of our experimental design. The wide spread presence of antibiotic resistance genes in the gut plasmidome of CD patients may have negative consequences for the CD patients. For instance, the spread of antibiotic resistance genes to pathogenic or pro-inflammatory bacteria may render them more resistant to antibiotic treatment. Particularly, IBD patients are at a high risk of infection following immunosuppressive therapy [340-342]. Interestingly, the higher abundance of antibiotic resistance genes in CD plasmidome is coupled with an increased abundance of transposases. It is well documented in the literature that transposons are responsible for the spread of a wide range of antibiotic resistance genes to different bacterial chromosomes and plasmids [343]; however, transposases are also linked to other MGEs, such as bacteriophages.

A considerable amount of pediatric gut plasmidome reads exhibited a high sequence similarity to phage related sequences. Similar to our findings, nucleotide sequences from the human dental plaque plasmidome isolated using the TRACA strategy showed a high sequence similarity to phage metagenomes [323]. In fact, phage elements are frequently found in sequenced plasmids, and a small portion can be found as circular prophages [221]. Since the TRACA approach captures closed circular DNA [321], it is expected to isolate circular prophages as well as bacterial plasmids that could contain phage elements. Plasmidome reads with a high sequence similarity to phage-related genes in the NCBI-NR database were enriched in the CD plasmidome. The same observation was obtained using the HMP-GIT and ACLAME databases. This finding is also in

agreement with a recent report that shows increased phage DNA in ileal washes collected from first onset pediatric CD patients compared to non-IBD controls [238]. Moreover, Lepage *et al.* detected a higher number of phage particles in the mucosa of CD patients compared to healthy control subjects [237]. More recently, an IBD intestinal virome study showed that the abundance of *Caudovirales* bacteriophages is increased in both CD and UC patients as compared to control subjects [239]. The authors of this study suggested that bacteriophages could contribute to the microbiota dysbiosis as well as the intestinal inflammation in IBD [239]. Since prophages have been detected in bacterial plasmids [226], a higher abundance of prophages within the CD plasmidome may accelerate and expand the spread of bacteriophages in the gut of CD patients, which may contribute to the bacterial dysbiosis observed in IBD [239].

Interestingly, the adenosine-5'-phosphosulfate reductase coding gene was found to be significantly enriched in the CD plasmidome as compared to the plasmidome of the control subjects. Adenosine-5'-phosphosulfate reductase is one of the two enzymes that catalyze the oxidation of sulfite into sulfate for energy production in phototrophic sulphur bacteria [344, 345]. Subsequently, sulfate is used in the degradation of organic compounds as a terminal electron acceptor by sulphate-reducing bacteria (SRB), which ultimately produces hydrogen sulfide [344, 345]. At higher concentrations, hydrogen sulfide is toxic to the cells mainly by inhibiting mitochondrial cytochrome c oxidase activity, which is an essential enzyme for mitochondrial respiration [346-348]. Although an

increase in SRB is well documented in UC [349, 350], Morgan *et al.* recently showed that sulfur metabolism is increased in CD patients [351]. Therefore, a higher abundance of adenosine-5'-phosphosulfate reductase genes in CD plasmidome may result in a higher oxidation of sulfite to sulfate. Subsequently, the accumulated sulfate may be used by SRB to generate a higher level of hydrogen sulfide, which could contribute to colonocyte damage.

In conclusion, this work presents the first IBD plasmidome study. The plasmidome was isolated and characterized from treatment naïve early onset IBD patients (both UC and CD) and pediatric subjects without intestinal inflammation. The plasmidome reads were aligned against a panel of reference databases, and the relative abundance of the assigned genes was compared between the controls, the IBD, and the IBD subtypes. The non-IBD plasmidome was enriched with genes involved in the digestion of dietary fibers and butyrate synthesis as well as genes involved in sensing and adapting to environmental changes. On the other hand, the IBD plasmidome was enriched with antibiotic resistance genes and genetic elements enabling the spread of MGEs. The CD plasmidome harbored a higher abundance of genes that code for adenosine-5'-phosphosulfate reductase which could be involved in hydrogen sulfide synthesis. In addition to their potential involvement in the pathogenesis of IBD, the pediatric gut plasmidome genes that are enriched or depleted in IBD and its subtypes could serve as biomarkers for IBD, UC or CD; however, the validity of these potential biomarkers and their potential involvement in IBD and IBD subtypes pathogenesis/protection still requires further confirmation.

CHAPTER 4

GENERAL CONCLUSION

The assessment of butyrate producers in pediatric CD patients and non-IBD subjects revealed significant differences between the two groups. Our results indicate that the right colon of pediatric CD patients has an overall reduction of the major group of butyrate producers that utilize the BCoAT pathway, especially in CD children with an inflamed right colon. The reduction in BCoAT butyrate producers in the inflamed colon might be a result of the known high sensitivity of members of this group to the disturbance of oxygen gradient [157], which is observed with inflammation. On the contrary, the level of butyrate producers that utilize the BUK pathway, which is the second major route for butyrate synthesis, seems to be unaffected in our pediatric CD patients. The decreased level of BCoAT butyrate producers is also accompanied by a shift in the normal balance between their members. In general, *E. rectale*, potent producer of butyrate and the most abundant butyrate producer in the human gut [156, 157], is reduced, regardless of the inflammation status of the right colon in CD children. Other potentially novel butyrate producers are also reduced in the right colons of CD children, and *Coprococcus* is reduced in inflamed CD colons. Considering the beneficial effects of butyrate on colonic health and mucosal immunity [244-253, 258, 259], the high level of butyrate produced by *E. rectale* [156, 157], and the overall

amount of butyrate produced by other members of BCoAT butyrate producers, the loss of *E. rectale* and other BCoAT butyrate producers in pediatric CD could have adverse effects on disease outcome. The reduction in *E. rectale* and other BCoAT butyrate producers' abundances could also open doors for other bacteria to bloom, particularly those that are fit to cope with environmental stress in CD patients' guts. Members of these "stress-coping" bacteria could be other butyrate-producing bacteria. For instance, the abundance of *F. prausnitzii* in the inflamed right colons of CD children is increased (Figure 4.1). Although most butyrate producers are strict anaerobes, strains of *F. prausnitzii* can survive at a low oxygen level [301]. This provides *F. prausnitzii* with a colonization advantage in the reduced anoxic environment in an inflamed gut. However, it is unclear at this stage if the bloom of *F. prausnitzii* in the inflamed right colon of CD children could have any impact on the disease outcome.

Our gut plasmidome characterization results also revealed critical differences between IBD and non-IBD children (both UC and CD). First, non-IBD children's colons are enriched with plasmidome elements with potential beneficial properties, such as genes involved in butyrate synthesis. This observation could be coupled with our butyrate producers finding in which healthy children have a higher level of BCoAT butyrate producers. Together, this could further contribute to colonocyte protection against inflammation. Furthermore, non-IBD microbiota are equipped with a higher level of plasmidome that carries the genes involved in the regulation of different cellular processes and stress responses. A higher level of such plasmidome elements

in the healthy gut microbiota might enhance their adaptation and survival in the gut. On the other hand, IBD plasmidome has a higher abundance of genes involved in antibiotic resistance. Knowing that IBD patients are more likely to be exposed to antibiotics 2–5 years before diagnosis due to various infections [99], prior exposure to antibiotics might be the cause of the observed high level of antibiotic resistance genes in IBD. In addition, the wide spread of these genes in pediatric IBD microbiota might render some members of the IBD microbiota more resilient to antibiotic treatment, which could augment the dysbiosis observed in IBD upon antibiotic treatment [81]. Pediatric IBD plasmidome, particularly CD, is also enriched with phage elements. A higher level of phage elements in IBD in general, and in CD in particular, might contribute to the bacterial dysbiosis through the activation of the phage lytic cycle in the infected bacteria. In addition, they might result in an increased transfer of potentially harmful MGEs that could trigger and/or maintain the disease. Finally, pediatric CD plasmidome has a higher level of Adenosine-5'-phosphosulfate reductase that is involved in hydrogen sulfide synthesis [344, 345]. This in return could contribute to higher levels of hydrogen sulfide in the gut. A higher level of hydrogen sulfide is linked to IBD pathogenesis [352, 353], and an increased hydrogen sulfide metabolism has been demonstrated in CD [351]. At a higher level, hydrogen sulfide interferes with mitochondrial respiration and inhibits butyrate oxidation, which leads to colonocyte damage [346-348]. On the contrary, butyrate is known to increase the expressions of genes involved in hydrogen sulfide detoxification [354]. As a result, the reduced level and

dysbiosis of BCoAT butyrate producers and the higher abundance of the plasmidome adenosine-5'-phosphosulfate reductase gene in pediatric CD might further increase the severity of colonocyte damage and inflammation (Figure 4.1). Altogether, this study represents the first comprehensive characterization of the gut butyrate producers and plasmidome of pediatric subjects that emphasizes a characteristic dysbiosis of butyrate producers in pediatric CD and the potential link of gut plasmidome to IBD pathogenesis.

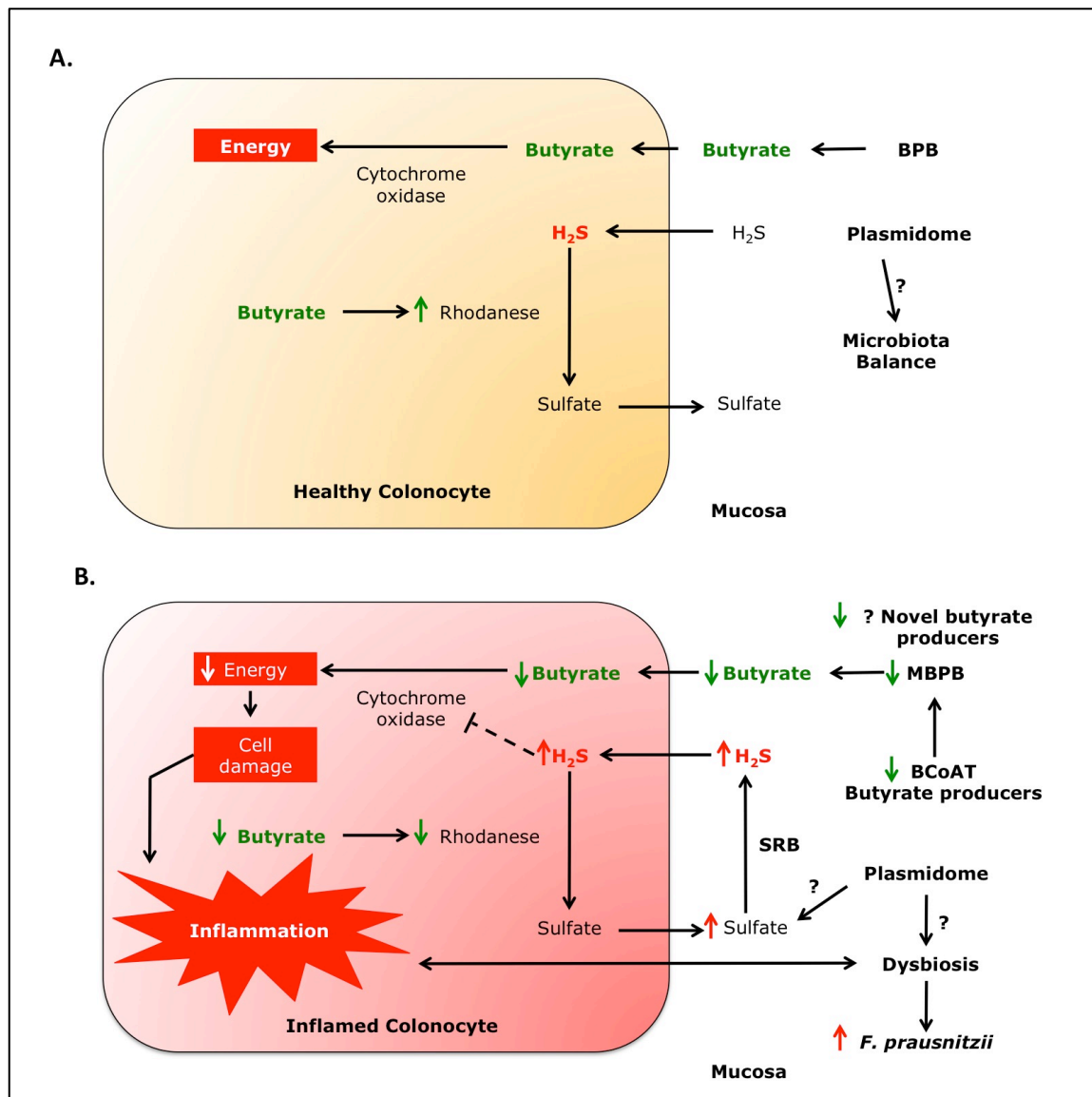


Figure 4.1: Proposed gut butyrate producers and plasmidome interactions in pediatric IBD. (A) In normal healthy colon, both the level and diversity of BCoAT butyrate producing bacteria (BPB) are balanced. This results in normal level of butyrate, which is subsequently oxidised by the mitochondrial cytochrome oxidase in the colonocytes to produce energy. Butyrate also increases the expression of rhodanese, which is involved in hydrogen sulphide (H₂S) detoxification. At the same time, the plasmidome might contribute to the maintenance of microbial balance by containing higher relative abundance of genes involved in external stimuli response, gene regulation, and damage repair than the IBD plasmidome. (B) In IBD, there is a decreased level of BCoAT butyrate producers, especially at the inflamed sites. The decreased level of BCoAT producers is accompanied by distinct dysbiosis characterized by lower levels of major butyrate producing bacteria (MBPB) including *E. rectale* as well as potentially novel butyrate producers. This might result in lower levels of butyrate, which could lower the level

of colonocytes energy. At the same time, lower level of butyrate might reduce the expression of rhodanese, which could result in the accumulation of H_2S that inhibits the activity of mitochondrial cytochrome oxidase. Together, this could result in colonocytes damage and inflammation, which might disrupt the oxygen gradient at the mucosal-luminal interface (MLI). The disrupted anoxic environment at the MLI might result in the disappearance of oxygen-sensitive bacteria, which could open doors for oxygen- tolerant bacterial strains such as *F. prausnitzii* to bloom. The IBD plasmidome might contribute to the maintenance of the observed dysbiosis by containing higher levels of phage and antibiotic resistance genes than non-IBD plasmidome. Also, the CD plasmidome is enriched with adenosine-5'-phosphosulfate reductase coding gene that catalyzes the oxidation of sulfite to sulfate. Higher level of sulfate could result in increased reduction level of sulfate to H_2S by sulphate-reducing bacteria (SRB), which could result in H_2S accumulation and colonocytes damage.

Future Directions

Characterizing the butyrate producers in the right colon of first-onset pediatric CD patients clearly shows that both the level and composition of BCoAT butyrate producers are reduced. This suggests that the level of butyrate at the right colon could also be reduced in CD patients, impacting colonocyte health. However, the quantity and relative abundance of BCoAT butyrate producers may not reflect the state of their metabolic activity. Therefore, measurement of butyrate concentration in the MLI samples using High-performance liquid chromatography (HPLC) and correlation of the butyrate level to the bacterial abundance would clarify this issue. However, butyrate measurement from our samples would be very difficult to interpret, as it is would be difficult to normalize the concentration knowing that the sample volume is different from patient to patient. It is worth noting that the potential anti-inflammatory properties of butyrate producers could be independent of butyrate synthesis, as demonstrated in an earlier study [179]. Alternatively, the transcription level of the BCoAT gene can be measured using qRT-PCR. Although we observed a decreased level of BCoAT butyrate producers, especially in an inflamed colon, it is also crucial to determine whether the reduction is correlated with disease severity. Therefore, more samples are needed at different degrees of severity according to the Pediatric Crohn's Disease Activity Index (PCDAI) [279]. We demonstrated a higher level of *F. prausnitzii* in the inflamed right colon of pediatric CD patients. This highlights that this bacterial species might have an efficient survival mechanism under stress conditions. Although *F. prausnitzii* is generally accepted as beneficial bacteria, this strain

might harbour elements that could have adverse effects on CD outcome. This necessitates the isolation of *F. prausnitzii* strains from our samples. Whole genome sequencing of these isolates, as well as comparative genomic studies involving other *F. prausnitzii* strains, might identify potential stress survival mechanisms and possible roles in CD pathogenesis. Our phylogenetic analysis identified 15 potential novel BCoAT butyrate producers, five of which were reduced in pediatric CD. Isolation of these strains using advanced sorting strategies, such as fluorescence-activated cell sorting (FACS) [355] targeting their BCoAT gene, followed by 16S rRNA gene sequencing, might enable us to identify these bacteria. Our study was conducted on the right colon, as it is thought to be the primary site for butyrate synthesis [149, 157] and a common site for pediatric CD inflammation [5, 8]. However, the left colon is also a major site for butyrate producers, such as *F. prausnitzii* [356], and it is a major site for UC [4]. Therefore, characterizing the butyrate producers at this site in first-onset pediatric UC might show a similar trend to CD patients in the right colon.

Our plasmidome characterization represents the first study that shows a possible link between the plasmidome and IBD pathogenesis. We conducted our study on a low number of samples per group ($n = 4$). Therefore, validation of our result is required on a larger set of samples from each group using qPCR and primers specific to key plasmid elements. Our effort to assemble the plasmidome reads into complete plasmids was unsuccessful due to the short read length and possible high sequence similarities between plasmids. This obstacle can be overcome by utilizing different high-throughput sequencing, such as PacBio

sequencing, which generates reads up to 40 kb in length (Pacific Biosciences of California Inc., 2015). Knowing that our plasmidome clones are stored individually in glycerol at -80°C, we can retrace the plasmids of interest. These plasmids can be tested in a murine colitis model, such as an IL-10^{-/-} mouse model, using *E. coli* that overexpresses the targeted plasmid in humanized mice. The result of such an experiment could further confirm the involvement of plasmidome in intestinal inflammation. However, it is important to note that the transposon could be inserted into the gene of interest during plasmidome capture, which results in their inactivation. In conclusion, my study represents the first comprehensive characterization of pediatric IBD and non-IBD gut butyrate-producing bacteria and plasmidome. The results of my study emphasize a significant variation in the composition of gut butyrate producers and plasmidome between IBD and healthy controls.

CHAPTER 5

REFERENCES

1. Podolsky, D.K., *Inflammatory bowel disease*. N Engl J Med, 2002. **347**(6): p. 417-429.
2. Guindi, M. and R.H. Riddell, *Indeterminate colitis*. Journal of clinical pathology, 2004. **57**(12): p. 1233-1244.
3. Mulder, D.J., et al., *A tale of two diseases: the history of inflammatory bowel disease*. Journal of Crohn's & colitis, 2014. **8**(5): p. 341-348.
4. Hendrickson, B.A., R. Gokhale, and J.H. Cho, *Clinical aspects and pathophysiology of inflammatory bowel disease*. Clinical microbiology ..., 2002.
5. Van Limbergen, J., et al., *Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease*. Gastroenterology, 2008. **135**(4): p. 1114-1122.
6. Ruemmele, F.M., *Pediatric inflammatory bowel diseases: coming of age*. Current Opinion in Gastroenterology, 2010. **26**(4): p. 332-336.
7. Childs, B., et al., *Crohn's disease: influence of age at diagnosis on site and clinical type of disease*. Gastroenterology, 1996.
8. Paul, T., et al., *Distinct phenotype of early childhood inflammatory bowel disease*. Journal of clinical gastroenterology, 2006. **40**(7): p. 583-586.
9. Molodecky, N.A., et al., *Increasing Incidence and Prevalence of the Inflammatory Bowel Diseases With Time, Based on Systematic Review*. YGAST, 2012. **142**(1): p. 46-54.e42.
10. Benchimol, E.I., et al., *Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends*. Inflamm Bowel Dis, 2011. **17**(1): p. 423-439.
11. Rocchi, A., et al., *Inflammatory bowel disease: a Canadian burden of illness review*. Canadian journal of gastroenterology= Journal canadien de gastroenterologie, 2012. **26**(11): p. 811-817.
12. Benchimol, E.I., et al., *Changing Age Demographics of Inflammatory Bowel Disease in Ontario, Canada: A Population-based Cohort Study of Epidemiology Trends*. Inflamm Bowel Dis, 2014. **20**(10): p. 1761-1769.
13. Sonnenberg, A., *Age distribution of IBD hospitalization*. Inflamm Bowel Dis, 2010. **16**(3): p. 452-7.
14. Fiocchi, C., *Inflammatory bowel disease: etiology and pathogenesis*. Gastroenterology, 1998. **115**(1): p. 182-205.
15. Neish, A.S., *Microbes in gastrointestinal health and disease*. Gastroenterology, 2009. **136**(1): p. 65-80.

16. Tsukita, S., M. Furuse, and M. Itoh, *Multifunctional strands in tight junctions*. Nat Rev Mol Cell Biol, 2001. **2**(4): p. 285-93.
17. Hansson, G.C., et al., *The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions*. Proceedings of the National Academy of Sciences, 2011. **108**(Supplement_1): p. 4659-4665.
18. Ganz, T., *Defensins: antimicrobial peptides of innate immunity*. Nat Rev Immunol, 2003. **3**(9): p. 710-20.
19. Harwig, S.S., et al., *Bactericidal properties of murine intestinal phospholipase A2*. J Clin Invest, 1995. **95**(2): p. 603-10.
20. Fagarasan, S. and T. Honjo, *Intestinal IgA synthesis: regulation of front-line body defences*. Nat Rev Immunol, 2003. **3**(1): p. 63-72.
21. Belkaid, Y. and T.W. Hand, *Role of the microbiota in immunity and inflammation*. Cell, 2014. **157**(1): p. 121-41.
22. Lelouard, H., et al., *Peyer's Patch Dendritic Cells Sample Antigens by Extending Dendrites Through M Cell-Specific Transcellular Pores*. Gastroenterology, 2012. **142**(3): p. 592-U269.
23. Man, A.L., M.E. Prieto-Garcia, and C. Nicoletti, *Improving M cell mediated transport across mucosal barriers: do certain bacteria hold the keys?* Immunology, 2004. **113**(1): p. 15-22.
24. Zeuthen, L.H., L.N. Fink, and H. Frokiaer, *Epithelial cells prime the immune response to an array of gut-derived commensals towards a tolerogenic phenotype through distinct actions of thymic stromal lymphopoietin and transforming growth factor-beta*. Immunology, 2008. **123**(2): p. 197-208.
25. Rimoldi, M., et al., *Intestinal immune homeostasis is regulated by the crosstalk between epithelial cells and dendritic cells*. Nature Immunology, 2005. **6**(5): p. 507-514.
26. Taylor, B.C., et al., *TSLP regulates intestinal immunity and inflammation in mouse models of helminth infection and colitis*. J Exp Med, 2009. **206**(3): p. 655-67.
27. Sun, C.M., et al., *Small intestine lamina propria dendritic cells promote de novo generation of Foxp3 T reg cells via retinoic acid*. J Exp Med, 2007. **204**(8): p. 1775-85.
28. Schulz, O., et al., *Intestinal CD103+, but not CX3CR1+, antigen sampling cells migrate in lymph and serve classical dendritic cell functions*. J Exp Med, 2009. **206**(13): p. 3101-14.
29. O'Mahony, C., et al., *Commensal-induced regulatory T cells mediate protection against pathogen-stimulated NF-kappaB activation*. PLoS Pathog, 2008. **4**(8): p. e1000112.
30. Mazmanian, S.K., J.L. Round, and D.L. Kasper, *A microbial symbiosis factor prevents intestinal inflammatory disease*. Nature, 2008. **453**(7195): p. 620-5.
31. Wyatt, J., et al., *Increased gastric and intestinal permeability in patients with Crohn's disease*. Am J Gastroenterol, 1997. **92**(10): p. 1891-6.
32. Teahon, K., et al., *Intestinal permeability in patients with Crohn's disease and their first degree relatives*. Gut, 1992. **33**(3): p. 320-3.
33. Jenkins, R.T., et al., *Small bowel and colonic permeability to 51Cr-EDTA in patients with active inflammatory bowel disease*. Clin Invest Med, 1988. **11**(2): p. 151-5.

34. Hollander, D., et al., *Increased intestinal permeability in patients with Crohn's disease and their relatives. A possible etiologic factor.* Ann Intern Med, 1986. **105**(6): p. 883-5.
35. Jass, J.R. and M.D. Walsh, *Altered mucin expression in the gastrointestinal tract: a review.* J Cell Mol Med, 2001. **5**(3): p. 327-51.
36. Danese, S. and C. Fiocchi, *Ulcerative colitis.* N Engl J Med, 2011. **365**(18): p. 1713-25.
37. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.* Nat Genet, 2010. **42**(12): p. 1118-25.
38. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**(7422): p. 119-124.
39. Simms, L.A., et al., *Reduced alpha-defensin expression is associated with inflammation and not NOD2 mutation status in ileal Crohn's disease.* Gut, 2008. **57**(7): p. 903-910.
40. Gassler, N., et al., *Inflammatory bowel disease is associated with changes of enterocytic junctions.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2001. **281**(1): p. G216-G228.
41. Geremia, A., et al., *Innate and adaptive immunity in inflammatory bowel disease.* Autoimmun Rev, 2014. **13**(1): p. 3-10.
42. Breese, E., et al., *Interleukin-2- and interferon-gamma-secreting T cells in normal and diseased human intestinal mucosa.* Immunology, 1993. **78**(1): p. 127-31.
43. Heller, F., et al., *Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution.* Gastroenterology, 2005. **129**(2): p. 550-564.
44. Fuss, I.J., et al., *Disparate CD4(+) lamina propria (LP) lymphokine secretion profiles in inflammatory bowel disease - Crohn's disease LP cells manifest increased secretion of IFN-gamma, whereas ulcerative colitis LP cells manifest increased secretion of IL-5.* Journal of Immunology, 1996. **157**(3): p. 1261-1270.
45. Fuss, I.J., et al., *Nonclassical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis.* Journal of Clinical Investigation, 2004. **113**(10): p. 1490-1497.
46. Vainer, B., et al., *Colonic expression and synthesis of interleukin 13 and interleukin 15 in inflammatory bowel disease.* Cytokine, 2000. **12**(10): p. 1531-6.
47. Rovedatti, L., et al., *Differential regulation of interleukin 17 and interferon gamma production in inflammatory bowel disease.* Gut, 2009. **58**(12): p. 1629-36.
48. Monteleone, G., et al., *Interleukin-21 enhances T-helper cell type 1 signaling and interferon-gamma production in Crohn's disease.* Gastroenterology, 2005. **128**(3): p. 687-94.
49. Duerr, R.H., et al., *A genome-wide association study identifies IL23R as an inflammatory bowel disease gene.* Science, 2006. **314**(5804): p. 1461-3.
50. McGovern, D.P., et al., *Genome-wide association identifies multiple ulcerative colitis susceptibility loci.* Nat Genet, 2010. **42**(4): p. 332-7.
51. Yao, Z.B., et al., *Herpesvirus Saimiri encodes a new cytokine, IL-17, which binds to a novel cytokine receptor.* Immunity, 1995. **3**(6): p. 811-821.
52. Shalom-Barak, T., J. Quach, and M. Lotz, *Interleukin-17-induced gene expression in articular chondrocytes is associated with activation of mitogen-activated protein*

- kinases and NF-kappa B*. Journal of Biological Chemistry, 1998. **273**(42): p. 27467-27473.
53. Witowski, J., et al., *IL-17 stimulates intraperitoneal neutrophil infiltration through the release of GRO alpha chemokine from mesothelial cells*. Journal of Immunology, 2000. **165**(10): p. 5814-5821.
 54. Jovanovic, D.V., et al., *IL-17 stimulates the production and expression of proinflammatory cytokines, IL-beta and TNF-alpha, by human macrophages*. Journal of Immunology, 1998. **160**(7): p. 3513-3521.
 55. Fossiez, F., et al., *T cell interleukin-17 induces stromal cells to produce proinflammatory and hematopoietic cytokines*. Journal of Experimental Medicine, 1996. **183**(6): p. 2593-2603.
 56. Awane, M., et al., *NF-kappa B-inducing kinase is a common mediator of IL-17-, TNF-alpha-, and IL-1 beta-induced chemokine promoter activation in intestinal epithelial cells*. Journal of Immunology, 1999. **162**(9): p. 5337-5344.
 57. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease*. Nature, 2012. **491**(7422): p. 119-24.
 58. Hugot, J.P., et al., *Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 599-603.
 59. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 603-606.
 60. Inohara, N., et al., *Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease*. J Biol Chem, 2003. **278**(8): p. 5509-12.
 61. Khor, B., A. Gardet, and R.J. Xavier, *Genetics and pathogenesis of inflammatory bowel disease*. Nature, 2011. **474**(7351): p. 307-317.
 62. Lees, C.W., et al., *New IBD genetics: common pathways with other diseases*. Gut, 2011. **60**(12): p. 1739-53.
 63. Aujnarain, A., D.R. Mack, and E.I. Benchimol, *The role of the environment in the development of pediatric inflammatory bowel disease*. Curr Gastroenterol Rep, 2013. **15**(6): p. 326.
 64. Timmer, A., et al., *Oral contraceptive use and smoking are risk factors for relapse in Crohn's disease*. Gastroenterology, 1998. **114**(6): p. 1143-1150.
 65. Nos, P. and E. Domenech, *Management of Crohn's disease in smokers: Is an alternative approach necessary?* World Journal of Gastroenterology, 2011. **17**(31): p. 3567-3574.
 66. Langman, M.J.S. and R.F.A. Logan, *Should Patients with Inflammatory Bowel-Disease Smoke*. British Medical Journal, 1984. **288**(6420): p. 862-862.
 67. Birrenbach, T. and U. Bocker, *Inflammatory bowel disease and smoking - A review of epidemiology, pathophysiology, and therapeutic implications*. Inflammatory Bowel Diseases, 2004. **10**(6): p. 848-859.
 68. Johnson, G.J., J. Cosnes, and J.C. Mansfield, *Review article: smoking cessation as primary therapy to modify the course of Crohn's disease*. Alimentary Pharmacology & Therapeutics, 2005. **21**(8): p. 921-931.
 69. Cosnes, J., et al., *Smoking cessation and the course of Crohn's disease: An intervention study*. Gastroenterology, 2001. **120**(5): p. 1093-1099.

70. Boyko, E.J., et al., *Effects of cigarette smoking on the clinical course of ulcerative colitis*. Scand J Gastroenterol, 1988. **23**(9): p. 1147-52.
71. Beaugerie, L., et al., *Impact of cessation of smoking on the course of ulcerative colitis*. Am J Gastroenterol, 2001. **96**(7): p. 2113-6.
72. Carmody, R.N., et al., *Diet dominates host genotype in shaping the murine gut microbiota*. Cell Host Microbe, 2015. **17**(1): p. 72-84.
73. De Filippo, C., et al., *Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa*. Proc Natl Acad Sci U S A, 2010. **107**(33): p. 14691-14696.
74. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography*. Nature, 2012. **486**(7402): p. 222-227.
75. Wu, G.D., et al., *Linking long-term dietary patterns with gut microbial enterotypes*. Science, 2011. **334**(6052): p. 105-8.
76. David, L.A., et al., *Diet rapidly and reproducibly alters the human gut microbiome*. Nature, 2014. **505**(7484): p. 559-63.
77. Devkota, S., et al., *Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10^{-/-} mice*. Nature, 2012. **487**(7405): p. 104-8.
78. Frank, D.N., et al., *Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases*. Proc Natl Acad Sci U S A, 2007. **104**(34): p. 13780-5.
79. Baumgart, M., et al., *Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum*. The ISME Journal, 2007. **1**(5): p. 403-418.
80. Manichanh, C., et al., *Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach*. Gut, 2006. **55**(2): p. 205-211.
81. Gevers, D., et al., *The Treatment-Naive Microbiome in New-Onset Crohn's Disease*. Cell host & microbe, 2014. **15**(3): p. 382-392.
82. Porter, C.K., et al., *Infectious gastroenteritis and risk of developing inflammatory bowel disease*. Gastroenterology, 2008. **135**(3): p. 781-786.
83. Gradel, K.O., et al., *Increased Short- and Long-Term Risk of Inflammatory Bowel Disease After Salmonella or Campylobacter Gastroenteritis*. Gastroenterology, 2009. **137**(2): p. 495-501.
84. Amre, D.K., et al., *Investigating the hygiene hypothesis as a risk factor in pediatric onset Crohn's disease: A case-control study*. American Journal of Gastroenterology, 2006. **101**(5): p. 1005-1011.
85. Alberto, L., et al., *Acute gastroenteritis is followed by an increased risk of inflammatory bowel disease*. Gastroenterology, 2006. **130**(6): p. 1588-1594.
86. Wurzelmann, J.I., C.M. Lyles, and R.S. Sandler, *Childhood Infections and the Risk of Inflammatory Bowel-Disease*. Digestive Diseases and Sciences, 1994. **39**(3): p. 555-560.
87. Shaw, S.Y., J.F. Blanchard, and C.N. Bernstein, *Association between Early Childhood Otitis Media and Pediatric Inflammatory Bowel Disease: An Exploratory Population-Based Analysis*. Journal of Pediatrics, 2013. **162**(3): p. 510-514.

88. Jakobsen, C., et al., *Environmental factors and risk of developing paediatric inflammatory bowel disease - A population based study 2007-2009*. Journal of Crohns & Colitis, 2013. **7**(1): p. 79-88.
89. Virta, L., et al., *Association of Repeated Exposure to Antibiotics With the Development of Pediatric Crohn's Disease-A Nationwide, Register-based Finnish Case-Control Study*. American Journal of Epidemiology, 2012. **175**(8): p. 775-784.
90. Shaw, S.Y., J.F. Blanchard, and C.N. Bernstein, *Association Between the Use of Antibiotics and New Diagnoses of Crohn's Disease and Ulcerative Colitis*. American Journal of Gastroenterology, 2011. **106**(12): p. 2133-2142.
91. Hviid, A., H. Svanstrom, and M. Frisch, *Antibiotic use and inflammatory bowel diseases in childhood*. Gut, 2011. **60**(1): p. 49-54.
92. Tanaka, S., et al., *Influence of antibiotic exposure in the early postnatal period on the development of intestinal microbiota*. FEMS Immunol Med Microbiol, 2009. **56**(1): p. 80-7.
93. Ng, K.M., et al., *Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens*. Nature, 2013. **502**(7469): p. 96-9.
94. Kamada, N., et al., *Control of pathogens and pathobionts by the gut microbiota*. Nat Immunol, 2013. **14**(7): p. 685-90.
95. Cho, I., et al., *Antibiotics in early life alter the murine colonic microbiome and adiposity*. Nature, 2012. **488**(7413): p. 621-6.
96. Dethlefsen, L., et al., *The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing*. PLoS Biol, 2008. **6**(11): p. e280.
97. Dethlefsen, L. and D.A. Relman, *Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation*. Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4554-61.
98. Fouhy, F., et al., *High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin*. Antimicrob Agents Chemother, 2012. **56**(11): p. 5811-20.
99. Shaw, S.Y., J.F. Blanchard, and C.N. Bernstein, *Association between the use of antibiotics and new diagnoses of Crohn's disease and ulcerative colitis*. Am J Gastroenterol, 2011. **106**(12): p. 2133-42.
100. Shaw, S.Y., J.F. Blanchard, and C.N. Bernstein, *Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease*. Am J Gastroenterol, 2010. **105**(12): p. 2687-92.
101. Kronman, M.P., et al., *Antibiotic exposure and IBD development among children: a population-based cohort study*. Pediatrics, 2012. **130**(4): p. e794-803.
102. Keeney, K.M., et al., *Effects of antibiotics on human microbiota and subsequent disease*. Annu Rev Microbiol, 2014. **68**: p. 217-35.
103. Abujamel, T., et al., *Defining the Vulnerable Period for Re-Establishment of Clostridium difficile Colonization after Treatment of C. difficile Infection with Oral Vancomycin or Metronidazole*. PLoS One, 2013. **8**(10): p. e76269.

104. Sjolund, M., et al., *Long-term persistence of resistant Enterococcus species after antibiotics to eradicate Helicobacter pylori*. Ann Intern Med, 2003. **139**(6): p. 483-7.
105. Sjolund, M., et al., *Persistence of resistant Staphylococcus epidermidis after single course of clarithromycin*. Emerg Infect Dis, 2005. **11**(9): p. 1389-93.
106. Lerebours, E., et al., *Stressful life events as a risk factor for inflammatory bowel disease onset: A population-based case-control study*. American Journal of Gastroenterology, 2007. **102**(1): p. 122-131.
107. Cornish, J.A., et al., *The Risk of Oral Contraceptives in the Etiology of Inflammatory Bowel Disease: A Meta-Analysis*. American Journal of Gastroenterology, 2008. **103**(9): p. 2394-2400.
108. Kefalakes, H., et al., *Exacerbation of inflammatory bowel diseases associated with the use of nonsteroidal anti-inflammatory drugs: myth or reality?* European Journal of Clinical Pharmacology, 2009. **65**(10): p. 963-970.
109. Feagins, L.A. and B.L. Cryer, *Do Non-steroidal Anti-inflammatory Drugs Cause Exacerbations of Inflammatory Bowel Disease?* Digestive Diseases and Sciences, 2010. **55**(2): p. 226-232.
110. Biedermann, L., et al., *Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans*. PLoS One, 2013. **8**(3): p. e59260.
111. Pham, T.A. and T.D. Lawley, *Emerging insights on intestinal dysbiosis during bacterial infections*. Curr Opin Microbiol, 2014. **17**: p. 67-74.
112. Collins, S.M. and P. Bercik, *The Relationship Between Intestinal Microbiota and the Central Nervous System in Normal Gastrointestinal Function and Disease*. Gastroenterology, 2009. **136**(6): p. 2003-2014.
113. Makivuokko, H., et al., *The effect of age and non-steroidal anti-inflammatory drugs on human intestinal microbiota composition*. British Journal of Nutrition, 2010. **103**(2): p. 227-234.
114. Tyakht, A.V., et al., *Human gut microbiota community structures in urban and rural populations in Russia*. Nat Commun, 2013. **4**: p. 2469.
115. Sakamoto, N., et al., *Dietary risk factors for inflammatory bowel disease: a multicenter case-control study in Japan*. Inflamm Bowel Dis, 2005. **11**(2): p. 154-63.
116. Garg, M., et al., *Review article: vitamin D and inflammatory bowel disease--established concepts and future directions*. Aliment Pharmacol Ther, 2012. **36**(4): p. 324-44.
117. Bernstein, C.N., et al., *A population-based case control study of potential risk factors for IBD*. Am J Gastroenterol, 2006. **101**(5): p. 993-1002.
118. Soon, I.S., et al., *The relationship between urban environment and the inflammatory bowel diseases: a systematic review and meta-analysis*. BMC Gastroenterology, 2012. **12**.
119. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing*. Nature, 2010. **464**(7285): p. 59-65.
120. Sekirov, I., et al., *Gut microbiota in health and disease*. Physiol Rev, 2010. **90**(3): p. 859-904.

121. Nicholson, J.K., et al., *Host-gut microbiota metabolic interactions*. Science, 2012. **336**(6086): p. 1262-7.
122. Arrieta, M.C., et al., *The intestinal microbiome in early life: health and disease*. Front Immunol, 2014. **5**: p. 427.
123. Lozupone, C.A., et al., *Meta-analyses of studies of the human microbiota*. Genome Res, 2013. **23**(10): p. 1704-14.
124. Dominguez-Bello, M.G., et al., *Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns*. Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11971-5.
125. Matamoros, S., et al., *Development of intestinal microbiota in infants and its impact on health*. Trends Microbiol, 2013. **21**(4): p. 167-73.
126. Sela, D.A., et al., *The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome*. Proc Natl Acad Sci U S A, 2008. **105**(48): p. 18964-9.
127. Fallani, M., et al., *Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres*. Microbiology, 2011. **157**(Pt 5): p. 1385-92.
128. Martens, E.C., et al., *Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts*. PLoS Biol, 2011. **9**(12): p. e1001221.
129. Eckburg, P.B., et al., *Diversity of the human intestinal microbial flora*. Science, 2005. **308**(5728): p. 1635-8.
130. Biagi, E., et al., *Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians*. PLoS One, 2010. **5**(5): p. e10667.
131. Claesson, M.J., et al., *Composition, variability, and temporal stability of the intestinal microbiota of the elderly*. Proc Natl Acad Sci U S A, 2011. **108** Suppl 1: p. 4586-91.
132. Arumugam, M., et al., *Enterotypes of the human gut microbiome*. Nature, 2011. **473**(7346): p. 174-80.
133. Koren, O., et al., *A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets*. PLoS Comput Biol, 2013. **9**(1): p. e1002863.
134. Knights, D., et al., *Rethinking "enterotypes"*. Cell Host Microbe, 2014. **16**(4): p. 433-7.
135. Zoetendal, E.G., et al., *Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces*. Appl Environ Microbiol, 2002. **68**(7): p. 3401-7.
136. Takaishi, H., et al., *Imbalance in intestinal microflora constitution could be involved in the pathogenesis of inflammatory bowel disease*. Int J Med Microbiol, 2008. **298**(5-6): p. 463-72.
137. Willing, B.P., et al., *A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes*. Gastroenterology, 2010. **139**(6): p. 1844-1854 e1.
138. Hooper, L.V., D.R. Littman, and A.J. Macpherson, *Interactions between the microbiota and the immune system*. Science, 2012. **336**(6086): p. 1268-73.

139. Caporaso, J.G., et al., *Moving pictures of the human microbiome*. Genome Biol, 2011. **12**(5): p. R50.
140. Human Microbiome Project, C., *Structure, function and diversity of the healthy human microbiome*. Nature, 2012. **486**(7402): p. 207-14.
141. Faith, J.J., et al., *The long-term stability of the human gut microbiota*. Science, 2013. **341**(6141): p. 1237439.
142. Schloissnig, S., et al., *Genomic variation landscape of the human gut microbiome*. Nature, 2013. **493**(7430): p. 45-50.
143. Lozupone, C.A., et al., *Diversity, stability and resilience of the human gut microbiota*. Nature, 2012. **489**(7415): p. 220-30.
144. Backhed, F., et al., *Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications*. Cell Host Microbe, 2012. **12**(5): p. 611-22.
145. Dongowski, G., A. Lorenz, and H. Anger, *Degradation of pectins with different degrees of esterification by Bacteroides thetaiotaomicron isolated from human gut flora*. Appl Environ Microbiol, 2000. **66**(4): p. 1321-7.
146. Cummings, J.H. and G.T. Macfarlane, *The control and consequences of bacterial fermentation in the human colon*. J Appl Bacteriol, 1991. **70**(6): p. 443-59.
147. Simpson, H.L. and B.J. Campbell, *Review article: dietary fibre-microbiota interactions*. Aliment Pharmacol Ther, 2015. **42**(2): p. 158-79.
148. Chassard, C. and C. Lacroix, *Carbohydrates and the human gut microbiota*. Curr Opin Clin Nutr Metab Care, 2013. **16**(4): p. 453-60.
149. Hamer, H.M., et al., *Review article: the role of butyrate on colonic function*. Aliment Pharmacol Ther, 2008. **27**(2): p. 104-19.
150. Bergman, E.N., *Energy contributions of volatile fatty acids from the gastrointestinal tract in various species*. Physiol Rev, 1990. **70**(2): p. 567-90.
151. Topping, D.L. and P.M. Clifton, *Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides*. Physiol Rev, 2001. **81**(3): p. 1031-64.
152. Wolever, T.M., P. Spadafora, and H. Eshuis, *Interaction between colonic acetate and propionate in humans*. Am J Clin Nutr, 1991. **53**(3): p. 681-7.
153. Tremaroli, V. and F. Backhed, *Functional interactions between the gut microbiota and host metabolism*. Nature, 2012. **489**(7415): p. 242-9.
154. Roediger, W.E., *Utilization of nutrients by isolated epithelial cells of the rat colon*. Gastroenterology, 1982. **83**(2): p. 424-9.
155. Louis, P., et al., *Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene*. Environ Microbiol, 2010. **12**(2): p. 304-314.
156. Louis, P., et al., *Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon*. J Bacteriol, 2004. **186**(7): p. 2099-2106.
157. Flint, H.J., et al., *Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine*. FEMS microbiology letters, 2009. **294**(1): p. 1-8.

158. Janowitz, H.D., E.C. Croen, and D.B. Sachar, *The role of the fecal stream in Crohn's disease: an historical and analytic review*. Inflamm Bowel Dis, 1998. **4**(1): p. 29-39.
159. Kang, S.S., et al., *An antibiotic-responsive mouse model of fulminant ulcerative colitis*. PLoS Med, 2008. **5**(3): p. e41.
160. Stepankova, R., et al., *Segmented filamentous bacteria in a defined bacterial cocktail induce intestinal inflammation in SCID mice reconstituted with CD45RB^{high} CD4⁺ T cells*. Inflamm Bowel Dis, 2007. **13**(10): p. 1202-11.
161. Sellon, R.K., et al., *Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice*. Infect Immun, 1998. **66**(11): p. 5224-31.
162. Matsumoto, S., et al., *Inflammatory bowel disease-like enteritis and caecitis in a senescence accelerated mouse P1/Yit strain*. Gut, 1998. **43**(1): p. 71-8.
163. Garrett, W.S., et al., *Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system*. Cell, 2007. **131**(1): p. 33-45.
164. Ott, S.J., et al., *Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease*. Gut, 2004. **53**(5): p. 685-693.
165. Lepage, P., et al., *Twin Study Indicates Loss of Interaction Between Microbiota and Mucosa of Patients With Ulcerative Colitis*. YGAST, 2011. **141**(1): p. 227-236.
166. Willing, B.P., et al., *A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes*. Gastroenterology, 2010. **139**(6): p. 1844-1854.e1.
167. Gophna, U., et al., *Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis*. J Clin Microbiol, 2006. **44**(11): p. 4136-41.
168. Kang, S., et al., *Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray*. Inflamm Bowel Dis, 2010. **16**(12): p. 2034-42.
169. Walker, A.W., et al., *High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease*. BMC Microbiol, 2011. **11**: p. 7.
170. Morgan, X.C., et al., *Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment*. Genome Biol, 2012. **13**(9): p. R79.
171. Wang, W., et al., *Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease*. J Clin Microbiol, 2014. **52**(2): p. 398-406.
172. Tong, M., et al., *A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease*. PLoS One, 2013. **8**(11): p. e80702.
173. Reiff, C., et al., *Microbiota of de-novo pediatric IBD: increased Faecalibacterium prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis*. The American journal of gastroenterology, 2012. **107**(12): p. 1913-1922.
174. Conte, M.P., et al., *Gut-associated bacterial microbiota in paediatric patients with inflammatory bowel disease*. Gut, 2006. **55**(12): p. 1760-7.

175. Schwiertz, A., et al., *Microbiota in pediatric inflammatory bowel disease*. J Pediatr, 2010. **157**(2): p. 240-244 e1.
176. Joossens, M., et al., *Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives*. Gut, 2011. **60**(5): p. 631-7.
177. Swidsinski, A., et al., *Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease*. J Clin Microbiol, 2005. **43**(7): p. 3380-9.
178. Sokol, H., et al., *Low counts of Faecalibacterium prausnitzii in colitis microbiota*. Inflamm Bowel Dis, 2009. **15**(8): p. 1183-9.
179. Sokol, H., et al., *Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients*. Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16731-16736.
180. Conte, M.P., et al., *Gut-associated bacterial microbiota in paediatric patients with inflammatory bowel disease*. Gut, 2006. **55**(12): p. 1760-1767.
181. Cucchiara, S., et al., *The microbiota in inflammatory bowel disease in different age groups*. Dig Dis, 2009. **27**(3): p. 252-8.
182. PhD, A.S., et al., *Microbiota in Pediatric Inflammatory Bowel Disease*. The Journal of Pediatrics, 2010. **157**(2): p. 240-244.e1.
183. Li, M., et al., *Symbiotic gut microbes modulate human metabolic phenotypes*. Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2117-22.
184. Zhang, H., et al., *Human gut microbiota in obesity and after gastric bypass*. Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2365-70.
185. Tap, J., et al., *Towards the human intestinal microbiota phylogenetic core*. Environ Microbiol, 2009. **11**(10): p. 2574-84.
186. Hong, P.Y., et al., *Pyrosequencing-based analysis of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-heterogeneity*. PLoS One, 2011. **6**(9): p. e25042.
187. van den Bogert, B., et al., *Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples*. Appl Environ Microbiol, 2011. **77**(6): p. 2071-80.
188. Stecher, B., L. Maier, and W.D. Hardt, *'Blooming' in the gut: how dysbiosis might contribute to pathogen evolution*. Nat Rev Microbiol, 2013. **11**(4): p. 277-84.
189. Sharon, I., et al., *Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization*. Genome Res, 2013. **23**(1): p. 111-20.
190. Walker, A.W., et al., *Dominant and diet-responsive groups of bacteria within the human colonic microbiota*. ISME J, 2011. **5**(2): p. 220-30.
191. Jones, B.V., *The human gut mobile metagenome: a metazoan perspective*. Gut Microbes, 2010. **1**(6): p. 415-431.
192. Xu, J., et al., *Evolution of symbiotic bacteria in the distal human intestine*. PLoS Biol, 2007. **5**(7): p. e156.
193. Lozupone, C.A., et al., *The convergence of carbohydrate active gene repertoires in human gut microbes*. Proc Natl Acad Sci U S A, 2008. **105**(39): p. 15076-81.
194. Hehemann, J.H., et al., *Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota*. Nature, 2010. **464**(7290): p. 908-12.

195. Louis, P., et al., *Organization of butyrate synthetic genes in human colonic bacteria: phylogenetic conservation and horizontal gene transfer*. FEMS Microbiol Lett, 2007. **269**(2): p. 240-7.
196. Shoemaker, N.B., et al., *Evidence for extensive resistance gene transfer among Bacteroides spp. and among Bacteroides and other genera in the human colon*. Appl Environ Microbiol, 2001. **67**(2): p. 561-8.
197. Salyers, A.A., et al., *Conjugative transposons: an unusual and diverse set of integrated gene transfer elements*. Microbiol Rev, 1995. **59**(4): p. 579-90.
198. Salyers, A.A., A. Gupta, and Y. Wang, *Human intestinal bacteria as reservoirs for antibiotic resistance genes*. Trends Microbiol, 2004. **12**(9): p. 412-6.
199. Gupta, A., et al., *A new Bacteroides conjugative transposon that carries an ermB gene*. Appl Environ Microbiol, 2003. **69**(11): p. 6455-63.
200. Scott, K.P., et al., *Occurrence of the new tetracycline resistance gene tet(W) in bacteria from the human gut*. Antimicrob Agents Chemother, 2000. **44**(3): p. 775-7.
201. Melville, C.M., et al., *Novel tetracycline resistance gene, tet(32), in the Clostridium-related human colonic anaerobe K10 and its transmission in vitro to the rumen anaerobe Butyrivibrio fibrisolvens*. Antimicrob Agents Chemother, 2001. **45**(11): p. 3246-9.
202. Melville, C.M., et al., *The Butyrivibrio fibrisolvens tet(W) gene is carried on the novel conjugative transposon TnB1230, which contains duplicated nitroreductase coding sequences*. J Bacteriol, 2004. **186**(11): p. 3656-9.
203. Kazimierczak, K.A., H.J. Flint, and K.P. Scott, *Comparative analysis of sequences flanking tet(W) resistance genes in multiple species of gut bacteria*. Antimicrob Agents Chemother, 2006. **50**(8): p. 2632-9.
204. Jones, B.V. and J.R. Marchesi, *Accessing the mobile metagenome of the human gut microbiota*. Mol Biosyst, 2007. **3**(11): p. 749-758.
205. Oshima, K., et al., *Complete genome sequence and comparative analysis of the wild-type commensal Escherichia coli strain SE11 isolated from a healthy adult*. DNA Res, 2008. **15**(6): p. 375-86.
206. Muriana, P.M. and T.R. Klaenhammer, *Conjugal Transfer of Plasmid-Encoded Determinants for Bacteriocin Production and Immunity in Lactobacillus acidophilus 88*. Appl Environ Microbiol, 1987. **53**(3): p. 553-60.
207. Moon, K., J. Sonnenburg, and A.A. Salyers, *Unexpected effect of a Bacteroides conjugative transposon, CTnDOT, on chromosomal gene expression in its bacterial host*. Mol Microbiol, 2007. **64**(6): p. 1562-71.
208. Millette, M., et al., *Purification and identification of the pediocin produced by Pediococcus acidilactici MM33, a new human intestinal strain*. J Appl Microbiol, 2008. **104**(1): p. 269-75.
209. Li, Y., et al., *Distribution of megaplasms in Lactobacillus salivarius and other lactobacilli*. J Bacteriol, 2007. **189**(17): p. 6128-39.
210. Dudley, E.G., et al., *An IncI1 plasmid contributes to the adherence of the atypical enteroaggregative Escherichia coli strain C1096 to cultured cells and abiotic surfaces*. Infect Immun, 2006. **74**(4): p. 2102-14.
211. Claesson, M.J., et al., *Multireplicon genome architecture of Lactobacillus salivarius*. Proc Natl Acad Sci U S A, 2006. **103**(17): p. 6718-23.

212. Chassy, B.M., E.M. Gibson, and A. Guiffrida, *Evidence for plasmid-associated lactose metabolism in Lactobacillus casei subsp. casei*. Curr Microbiol, 1978. **1**(3): p. 141-4.
213. Mirolid, S., et al., *Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic Salmonella typhimurium strain*. Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9845-50.
214. Mirolid, S., et al., *Salmonella host cell invasion emerged by acquisition of a mosaic of separate genetic elements, including Salmonella pathogenicity island 1 (SPI1), SPI5, and sopE2*. J Bacteriol, 2001. **183**(7): p. 2348-58.
215. Hardt, W.D., H. Urlaub, and J.E. Galan, *A substrate of the centisome 63 type III protein secretion system of Salmonella typhimurium is encoded by a cryptic bacteriophage*. Proc Natl Acad Sci U S A, 1998. **95**(5): p. 2574-9.
216. Hardt, W.D., et al., *S. typhimurium encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells*. Cell, 1998. **93**(5): p. 815-26.
217. Muller, A.J., et al., *The S. Typhimurium effector SopE induces caspase-1 activation in stromal cells to initiate gut inflammation*. Cell Host Microbe, 2009. **6**(2): p. 125-36.
218. Hapfelmeier, S., et al., *Role of the Salmonella pathogenicity island 1 effector proteins SipA, SopB, SopE, and SopE2 in Salmonella enterica subspecies 1 serovar Typhimurium colitis in streptomycin-pretreated mice*. Infect Immun, 2004. **72**(2): p. 795-809.
219. Stecher, B., et al., *Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota*. PLoS Biol, 2007. **5**(10): p. 2177-89.
220. Lopez, C.A., et al., *Phage-mediated acquisition of a type III secreted effector protein boosts growth of salmonella by nitrate respiration*. MBio, 2012. **3**(3).
221. Casjens, S., *Prophages and bacterial genomics: what have we learned so far?* Mol Microbiol, 2003. **49**(2): p. 277-300.
222. Mason, T.G. and G. Richardson, *Escherichia coli and the human gut: some ecological considerations*. J Appl Bacteriol, 1981. **51**(1): p. 1-16.
223. Arthur, J.C., et al., *Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota*. Science, 2012.
224. Frost, L.S., et al., *Mobile genetic elements: the agents of open source evolution*. Nat Rev Microbiol, 2005. **3**(9): p. 722-32.
225. Stotzky, G. and H. Babich, *Survival of, and genetic transfer by, genetically engineered bacteria in natural environments*. Adv Appl Microbiol, 1986. **31**: p. 93-138.
226. Brussow, H., C. Canchaya, and W.D. Hardt, *Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion*. Microbiol Mol Biol Rev, 2004. **68**(3): p. 560-602, table of contents.
227. Walker, A., *Welcome to the plasmidome*. Nature reviews Microbiology, 2012. **10**(6): p. 379-379.
228. Thomas, C.M., *Paradigms of plasmid organization*. Mol Microbiol, 2000. **37**(3): p. 485-91.
229. Mobley, H.L., et al., *Variable phenotypes of Providencia stuartii due to plasmid-encoded traits*. J Clin Microbiol, 1985. **22**(5): p. 851-3.

230. Broaders, E., C.G. Gahan, and J.R. Marchesi, *Mobile genetic elements of the human gastrointestinal tract: potential for spread of antibiotic resistance genes*. Gut Microbes, 2013. **4**(4): p. 271-80.
231. Weintraub, A., *Enteroaggregative Escherichia coli: epidemiology, virulence and detection*. J Med Microbiol, 2007. **56**(Pt 1): p. 4-8.
232. Nataro, J.P., et al., *Aggregative adherence fimbria I expression in enteroaggregative Escherichia coli requires two unlinked plasmid regions*. Infect Immun, 1993. **61**(3): p. 1126-31.
233. Ahmed, S.A., et al., *Genomic comparison of Escherichia coli O104:H4 isolates from 2009 and 2011 reveals plasmid, and prophage heterogeneity, including shiga toxin encoding phage stx2*. PLoS One, 2012. **7**(11): p. e48228.
234. Stecher, B., et al., *Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae*. Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1269-74.
235. Jones, B.V., et al., *Comparative metagenomic analysis of plasmid encoded functions in the human gut microbiome*. BMC Genomics, 2010. **11**: p. 46.
236. Rohwer, F., D. Prangishvili, and D. Lindell, *Roles of viruses in the environment*. Environ Microbiol, 2009. **11**(11): p. 2771-4.
237. Lepage, P., et al., *Dysbiosis in inflammatory bowel disease: a role for bacteriophages?* Gut, 2008. **57**(3): p. 424-5.
238. Wagner, J., et al., *Bacteriophages in gut samples from pediatric Crohn's disease patients: metagenomic analysis using 454 pyrosequencing*. Inflamm Bowel Dis, 2013. **19**(8): p. 1598-608.
239. Norman, J.M., et al., *Disease-specific alterations in the enteric virome in inflammatory bowel disease*. Cell, 2015. **160**(3): p. 447-60.
240. Suau, A., et al., *Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut*. Applied and environmental microbiology, 1999. **65**(11): p. 4799-4807.
241. Child, M.W., et al., *pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon*. Applied and environmental microbiology, 2005. **71**(7): p. 3692-3700.
242. Cummings, J.H., et al., *Short chain fatty acids in human large intestine, portal, hepatic and venous blood*. Gut, 1987. **28**(10): p. 1221-7.
243. Donohoe, D.R., et al., *The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon*. Cell Metab, 2011. **13**(5): p. 517-26.
244. Ramasamy, S., et al., *Sulfide-detoxifying enzymes in the human colon are decreased in cancer and upregulated in differentiation*. Am J Physiol Gastrointest Liver Physiol, 2006. **291**(2): p. G288-G296.
245. Barcelo, A., et al., *Mucin secretion is modulated by luminal factors in the isolated vascularly perfused rat colon*. Gut, 2000. **46**(2): p. 218-224.
246. Hudcovic, T., et al., *Protective effect of Clostridium tyrobutyricum in acute dextran sodium sulphate-induced colitis: differential regulation of tumour necrosis factor-alpha and interleukin-18 in BALB/c and severe combined immunodeficiency mice*. Clin Exp Immunol, 2012. **167**(2): p. 356-65.

247. Schaubert, J., et al., *Expression of the cathelicidin LL-37 is modulated by short chain fatty acids in colonocytes: relevance of signalling pathways*. Gut, 2003. **52**(5): p. 735-741.
248. Bordin, M., et al., *Histone deacetylase inhibitors up-regulate the expression of tight junction proteins*. Mol Cancer Res, 2004. **2**(12): p. 692-701.
249. Luhrs, H., et al., *Butyrate inhibits NF-kappaB activation in lamina propria macrophages of patients with ulcerative colitis*. Scand J Gastroenterol, 2002. **37**(4): p. 458-66.
250. Segain, J.P., et al., *Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease*. Gut, 2000. **47**(3): p. 397-403.
251. Baldwin, A.S., Jr., *The NF-kappa B and I kappa B proteins: new discoveries and insights*. Annu Rev Immunol, 1996. **14**: p. 649-83.
252. Zimmerman, M.A., et al., *Butyrate suppresses colonic inflammation through HDAC1-dependent Fas upregulation and Fas-mediated apoptosis of T cells*. Am J Physiol Gastrointest Liver Physiol, 2012. **302**(12): p. G1405-15.
253. Russo, I., et al., *Butyrate attenuates lipopolysaccharide-induced inflammation in intestinal cells and Crohn's mucosa through modulation of antioxidant defense machinery*. PLoS One, 2012. **7**(3): p. e32841.
254. Lathrop, S.K., et al., *Peripheral education of the immune system by colonic commensal microbiota*. Nature, 2011. **478**(7368): p. 250-4.
255. Ivanov, II, et al., *Induction of intestinal Th17 cells by segmented filamentous bacteria*. Cell, 2009. **139**(3): p. 485-98.
256. Atarashi, K., et al., *Induction of colonic regulatory T cells by indigenous Clostridium species*. Science, 2011. **331**(6015): p. 337-41.
257. Atarashi, K., et al., *Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota*. Nature, 2013. **500**(7461): p. 232-6.
258. Furusawa, Y., et al., *Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells*. Nature, 2013. **504**(7480): p. 446-50.
259. Arpaia, N., et al., *Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation*. Nature, 2013. **504**(7480): p. 451-5.
260. Van de Peer, Y., S. Chapelle, and R. De Wachter, *A quantitative map of nucleotide substitution rates in bacterial rRNA*. Nucleic Acids Res, 1996. **24**(17): p. 3381-91.
261. Větrovský, T. and P. Baldrian, *The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses*. PLoS One, 2013. **8**(2): p. e57923.
262. Vital, M., et al., *Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data*. mBio, 2014. **5**(2): p. e00889-e00889-14.
263. Vital, M., et al., *A gene-targeted approach to investigate the intestinal butyrate-producing bacterial community*. Microbiome, 2013. **1**(1): p. 8.
264. Louis, P. and H.J. Flint, *Development of a semiquantitative degenerate real-time pcr-based assay for estimation of numbers of butyryl-coenzyme A (CoA) CoA transferase genes in complex bacterial samples*. Applied and environmental microbiology, 2007. **73**(6): p. 2009-2012.
265. Mariat, D., et al., *The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age*. BMC Microbiol, 2009. **9**: p. 123.

266. Agans, R., et al., *Distal gut microbiota of adolescent children is different from that of adults*. FEMS Microbiology Ecology, 2011. **77**(2): p. 404-412.
267. Turnbaugh, P.J., et al., *An obesity-associated gut microbiome with increased capacity for energy harvest*. Nature, 2006. **444**(7122): p. 1027-31.
268. Larsen, N., et al., *Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults*. PLoS ONE, 2010. **5**(2): p. e9085.
269. Barman, M., et al., *Enteric salmonellosis disrupts the microbial ecology of the murine gastrointestinal tract*. Infection and immunity, 2008. **76**(3): p. 907-915.
270. Kuehl, C.J., et al., *Colonization of the cecal mucosa by Helicobacter hepaticus impacts the diversity of the indigenous microbiota*. Infection and immunity, 2005. **73**(10): p. 6952-6961.
271. Dethlefsen, L., et al., *The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing*. PLoS biology, 2008. **6**(11): p. e280.
272. Swidsinski, A., et al., *Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora*. Inflamm Bowel Dis, 2008. **14**(2): p. 147-61.
273. Dorffel, Y., et al., *Common biostructure of the colonic microbiota in neuroendocrine tumors and Crohn's disease and the effect of therapy*. Inflamm Bowel Dis, 2012. **18**(9): p. 1663-71.
274. Gerritsen, J., et al., *Intestinal microbiota in human health and disease: the impact of probiotics*. Genes & Nutrition, 2011. **6**(3): p. 209-240.
275. Rogers, M.A. and D.M. Aronoff, *The Influence of Nonsteroidal Anti-Inflammatory Drugs on the Gut Microbiome*. Clin Microbiol Infect, 2015.
276. North American Society for Pediatric Gastroenterology, H., et al., *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*. J Pediatr Gastroenterol Nutr, 2007. **44**(5): p. 653-74.
277. Levine, A., et al., *Pediatric modification of the Montreal classification for inflammatory bowel disease*. Inflammatory Bowel Diseases, 2011. **17**(6): p. 1314-1321.
278. Daperno, M., et al., *Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD*. Gastrointest Endosc, 2004. **60**(4): p. 505-12.
279. Hyams, J., et al., *Evaluation of the pediatric crohn disease activity index: a prospective multicenter experience*. J Pediatr Gastroenterol Nutr, 2005. **41**(4): p. 416-21.
280. Jimenez-Rivera, C., et al., *Comparison of two common outpatient preparations for colonoscopy in children and youth*. Gastroenterol Res Pract, 2009. **2009**: p. 518932.
281. Sundquist, A., et al., *Bacterial flora-typing with targeted, chip-based Pyrosequencing*. BMC Microbiology, 2007. **7**(1): p. 108.
282. Magoc, T. and S.L. Salzberg, *FLASH: fast length adjustment of short reads to improve genome assemblies*. Bioinformatics, 2011. **27**(21): p. 2957-2963.

283. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data*. Nat Methods, 2010. **7**(5): p. 335-336.
284. Lee, Z.M.-P., et al., *rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea*. Nucleic Acids Res, 2009. **37**(Database issue): p. D489-93.
285. Edgar, R.C., et al., *UCHIME improves sensitivity and speed of chimera detection*. Bioinformatics, 2011. **27**(16): p. 2194-200.
286. Fish, J.A., et al., *FunGene: the functional gene pipeline and repository*. Front Microbiol, 2013. **4**: p. 291.
287. Wang, Q., et al., *Ecological patterns of nifH genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool*. MBio, 2013. **4**(5): p. e00592-13.
288. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
289. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation*. Bioinformatics, 2007. **23**(1): p. 127-128.
290. Walker, A.W., et al., *Dominant and diet-responsive groups of bacteria within the human colonic microbiota*. ISME J, 2011. **5**(2): p. 220-230.
291. Vermeiren, J., et al., *Decreased colonization of fecal Clostridium coccoides/Eubacterium rectale species from ulcerative colitis patients in an in vitro dynamic gut model with mucin environment*. FEMS Microbiology Ecology, 2012. **79**(3): p. 685-696.
292. Turner, D., et al., *Appraisal of the Pediatric Crohn's Disease Activity Index on Four Prospectively Collected Datasets: Recommended Cutoff Values and Clinimetric Properties*. The American Journal of Gastroenterology, 2010. **105**(9): p. 2085-2092.
293. Albenberg, L., et al., *Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota*. Gastroenterology, 2014. **147**(5): p. 1055-63 e8.
294. Rigottier-Gois, L., *Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis*. ISME J, 2013. **7**(7): p. 1256-61.
295. Kang, S., et al., *Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray*. Inflammatory Bowel Diseases, 2010. **16**(12): p. 2034-2042.
296. Mondot, S., et al., *Highlighting new phylogenetic specificities of Crohn's disease microbiota*. Inflammatory Bowel Diseases, 2011. **17**(1): p. 185-192.
297. Miquel, S., et al., *Faecalibacterium prausnitzii and human intestinal health*. Current Opinion in Microbiology, 2013. **16**(3): p. 255-261.
298. Willing, B., et al., *Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease*. Inflammatory Bowel Diseases, 2009. **15**(5): p. 653-660.
299. Wang, W., et al., *Increased Proportions of Bifidobacterium and the Lactobacillus Group and Loss of Butyrate-Producing Bacteria in Inflammatory Bowel Disease*. J Clin Microbiol, 2014. **52**(2): p. 398-406.
300. Kaakoush, N.O., et al., *Microbial dysbiosis in pediatric patients with Crohn's disease*. J Clin Microbiol, 2012. **50**(10): p. 3258-3266.

301. Khan, M.T., et al., *The gut anaerobe Faecalibacterium prausnitzii uses an extracellular electron shuttle to grow at oxic-anoxic interphases*. The ISME Journal, 2012. **6**(8): p. 1578-1585.
302. Belenguer, A., et al., *Impact of pH on Lactate Formation and Utilization by Human Fecal Microbial Communities*. Applied and environmental microbiology, 2007. **73**(20): p. 6526-6533.
303. Kumar, D., et al., *Intestinal luminal pH in inflammatory bowel disease: possible determinants and implications for therapy with aminosalicylates and other drugs*. 2001. **48**(4): p. 571-577.
304. Walter, J. and R. Ley, *The human gut microbiome: ecology and recent evolutionary changes*. Annu Rev Microbiol, 2011. **65**: p. 411-29.
305. Goodman, A.L., et al., *Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice*. Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6252-7.
306. Ley, R.E., D.A. Peterson, and J.I. Gordon, *Ecological and evolutionary forces shaping microbial diversity in the human intestine*. Cell, 2006. **124**(4): p. 837-848.
307. Zhu, Q., et al., *The role of gut microbiota in the pathogenesis of colorectal cancer*. Tumour Biol, 2013. **34**(3): p. 1285-300.
308. Wall, R., et al., *Metabolic activity of the enteric microbiota influences the fatty acid composition of murine and porcine liver and adipose tissues*. Am J Clin Nutr, 2009. **89**(5): p. 1393-401.
309. Smith, P.M., et al., *The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis*. Science, 2013. **341**(6145): p. 569-73.
310. Shin, A., et al., *Bowel functions, fecal unconjugated primary and secondary bile acids, and colonic transit in patients with irritable bowel syndrome*. Clin Gastroenterol Hepatol, 2013. **11**(10): p. 1270-1275 e1.
311. Ridaura, V.K., et al., *Gut microbiota from twins discordant for obesity modulate metabolism in mice*. Science, 2013. **341**(6150): p. 1241214.
312. Ou, J., et al., *Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans*. Am J Clin Nutr, 2013. **98**(1): p. 111-20.
313. Le Chatelier, E., et al., *Richness of human gut microbiome correlates with metabolic markers*. Nature, 2013. **500**(7464): p. 541-6.
314. Fanning, S., et al., *Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection*. Proc Natl Acad Sci U S A, 2012. **109**(6): p. 2108-13.
315. Duboc, H., et al., *Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases*. Gut, 2013. **62**(4): p. 531-9.
316. Duboc, H., et al., *Increase in fecal primary bile acids and dysbiosis in patients with diarrhea-predominant irritable bowel syndrome*. Neurogastroenterol Motil, 2012. **24**(6): p. 513-20, e246-7.
317. Cotillard, A., et al., *Dietary intervention impact on gut microbial gene richness*. Nature, 2013. **500**(7464): p. 585-8.
318. Chen, W., et al., *Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer*. PLoS One, 2012. **7**(6): p. e39743.

319. Haberman, Y., et al., *Corrigendum. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature*. J Clin Invest, 2015. **125**(3): p. 1363.
320. Haberman, Y., et al., *Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature*. J Clin Invest, 2014. **124**(8): p. 3617-33.
321. Jones, B.V. and J.R. Marchesi, *Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome*. Nat Methods, 2007. **4**(1): p. 55-61.
322. Brown Kav, A., et al., *Insights into the bovine rumen plasmidome*. Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5452-5457.
323. Warburton, P.J., et al., *Isolation of bacterial extrachromosomal DNA from human dental plaque associated with periodontal disease, using transposon-aided capture (TRACA)*. FEMS Microbiology Ecology, 2011. **78**(2): p. 349-354.
324. Ye, L., T. Zhang, and X.-X. Zhang, *Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge*. PLoS One, 2011. **6**(10): p. e26041.
325. Sentchilo, V., et al., *Community-wide plasmid gene mobilization and selection*. ISME J, 2013. **7**(6): p. 1173-86.
326. Szczepanowski, R., et al., *Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology*. Journal of biotechnology, 2008. **136**(1-2): p. 54-64.
327. Turner, D., et al., *Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study*. Gastroenterology, 2007. **133**(2): p. 423-32.
328. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform*. Bioinformatics, 2010. **26**(5): p. 589-95.
329. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
330. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND*. Nat Methods, 2015. **12**(1): p. 59-60.
331. Langille, M.G., et al., *Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences*. Nat Biotechnol, 2013. **31**(9): p. 814-21.
332. Krzywinski, M., et al., *Circos: an information aesthetic for comparative genomics*. Genome Res, 2009. **19**(9): p. 1639-45.
333. Beceiro, A., M. Tomas, and G. Bou, *Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world?* Clin Microbiol Rev, 2013. **26**(2): p. 185-230.
334. Hacker, J. and J.B. Kaper, *Pathogenicity islands and the evolution of microbes*. Annu Rev Microbiol, 2000. **54**: p. 641-79.
335. Mazodier, P. and J. Davies, *Gene transfer between distantly related bacteria*. Annu Rev Genet, 1991. **25**: p. 147-71.
336. Graf, D., et al., *Contribution of diet to the composition of the human gut microbiota*. Microb Ecol Health Dis, 2015. **26**: p. 26164.
337. Semova, I., et al., *Microbiota regulate intestinal absorption and metabolism of fatty acids in the zebrafish*. Cell Host Microbe, 2012. **12**(3): p. 277-88.

338. Dib, J.R., et al., *Strategies and approaches in plasmidome studies-uncovering plasmid diversity disregarding of linear elements?* Front Microbiol, 2015. **6**: p. 463.
339. Pereira, S.F., L. Goss, and J. Dworkin, *Eukaryote-like serine/threonine kinases and phosphatases in bacteria*. Microbiol Mol Biol Rev, 2011. **75**(1): p. 192-212.
340. Lichtenstein, G.R., et al., *A pooled analysis of infections, malignancy, and mortality in infliximab- and immunomodulator-treated adult patients with inflammatory bowel disease*. Am J Gastroenterol, 2012. **107**(7): p. 1051-63.
341. Ananthakrishnan, A.N. and E.L. McGinley, *Infection-related hospitalizations are associated with increased mortality in patients with inflammatory bowel diseases*. J Crohns Colitis, 2013. **7**(2): p. 107-12.
342. Afif, W. and E.V. Loftus, Jr., *Safety profile of IBD therapeutics: infectious risks*. Gastroenterol Clin North Am, 2009. **38**(4): p. 691-709.
343. Baquero, F., A.P. Tedim, and T.M. Coque, *Antibiotic resistance shaping multi-level population biology of bacteria*. Front Microbiol, 2013. **4**: p. 15.
344. Muyzer, G. and A.J. Stams, *The ecology and biotechnology of sulphate-reducing bacteria*. Nat Rev Microbiol, 2008. **6**(6): p. 441-54.
345. Bruser, T., T. Selmer, and C. Dahl, *"ADP sulfurylase" from Thiobacillus denitrificans is an adenylylsulfate:phosphate adenylyltransferase and belongs to a new family of nucleotidyltransferases*. J Biol Chem, 2000. **275**(3): p. 1691-8.
346. Roediger, W.E., et al., *Sulphide impairment of substrate oxidation in rat colonocytes: a biochemical basis for ulcerative colitis?* Clin Sci (Lond), 1993. **85**(5): p. 623-7.
347. Leschelle, X., et al., *Adaptative metabolic response of human colonic epithelial cells to the adverse effects of the luminal compound sulfide*. Biochim Biophys Acta, 2005. **1725**(2): p. 201-12.
348. Attene-Ramos, M.S., et al., *DNA damage and toxicogenomic analyses of hydrogen sulfide in human intestinal epithelial FHs 74 Int cells*. Environ Mol Mutagen, 2010. **51**(4): p. 304-14.
349. Gibson, G.R., J.H. Cummings, and G.T. Macfarlane, *Growth and activities of sulphate-reducing bacteria in gut contents of healthy subjects and patients with ulcerative colitis*. FEMS microbiology letters, 1991. **86**(2): p. 103-111.
350. Roediger, W. and J. Moore, *Colonic sulfide in pathogenesis and treatment of ulcerative colitis*. Digestive diseases and sciences, 1997: p. 1-9.
351. Morgan, X.C., et al., *Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment*. Genome Biol, 2012. **13**(9): p. R79.
352. Turnbaugh, P.J., et al., *The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice*. Sci Transl Med, 2009. **1**(6): p. 6ra14.
353. Levine, J., et al., *Fecal hydrogen sulfide production in ulcerative colitis*. Am J Gastroenterol, 1998. **93**(1): p. 83-7.
354. Ramasamy, S., et al., *Sulfide-detoxifying enzymes in the human colon are decreased in cancer and upregulated in differentiation*. Am J Physiol Gastrointest Liver Physiol, 2006. **291**(2): p. G288-96.
355. Lay, C., J. Dore, and L. Rigottier-Gois, *Separation of bacteria of the Clostridium leptum subgroup from the human colonic microbiota by fluorescence-activated*

- cell sorting or group-specific PCR using 16S rRNA gene oligonucleotides.* FEMS Microbiol Ecol, 2007. **60**(3): p. 513-20.
356. Ahmed, S., et al., *Mucosa-Associated Bacterial Diversity in Relation to Human Terminal Ileum and Colonic Biopsy Samples.* Applied and environmental microbiology, 2007. **73**(22): p. 7435-7442.

Collaborators Contributions

- Dr. Jacques Corbeil, Dr. Frédéric Raymond, and Pier-Luc Plante (Infectious Diseases Research Center, Faculty of Medicine, University of Laval, Quebec, QC, Canada) demultiplexed the plasmidome reads and removed the reads that belong to the cloning strain and transposon from the demultiplexed reads.
- Dr. Walid Mottawea (Department of Biochemistry, Microbiology and Immunology, University of Ottawa) helped in the construction of 16S rRNA library for Illumina sequencing and 16S rRNA data analysis.
- Dr. Alain Stintzi (Department of Biochemistry, Microbiology and Immunology, University of Ottawa) supervised this project.

Appendix

Appendix I: Clinical data of participating subject in the study. This table also illustrates which subjects were used for BCoAT and 16S rRNA sequencing and qPCR.

Sample	HMC ID	Gender	Age	Group	Colonoscopy			Category	PCDAI/pUCAI	Severity	BCoAT Sequencing	16S rRNA sequencing	Mucosal wash Used for qPCR	Biopsies used for qPCR	Stool used for qPCR
					visual appearance	Paris Classification	of right colon								
Subject 1	18	M	13	Control	Normal	NA	NA	NA	NA	NA	Yes	Yes	Yes	No	No
Subject 2	28	M	13	Control	Normal	NA	NA	NA	NA	NA	Yes	Yes	Yes	No	No
Subject 3	43	F	16	Control	Normal	NA	NA	NA	NA	NA	Yes	Yes	Yes	No	No
Subject 4	55	M	6	Control	Normal	NA	NA	NA	NA	NA	Yes	Yes	Yes	No	No
Subject 5	59	M	14	Control	Normal	NA	NA	NA	NA	NA	Yes	Yes	No	No	No
Subject 6	67	F	17	Control	Normal	NA	NA	NA	NA	NA	Yes	Yes	Yes	No	No
Subject 7	87	F	18	Control	Normal	NA	NA	NA	NA	NA	Yes	Yes	Yes	No	No
Subject 8	98	F	16	Control	Normal	NA	NA	NA	NA	NA	Yes	Yes	Yes	No	No
Subject 9	208	F	16	Control	Normal	NA	NA	NA	NA	NA	Yes	Yes	Yes	No	No
Subject 10	221	F	9	Control	Normal	NA	NA	NA	NA	NA	Yes	No	No	No	No
Subject 11	222	M	10	Control	Normal	NA	NA	NA	NA	NA	Yes	No	No	No	No
Subject 12	232	F	15	Control	Normal	NA	NA	NA	NA	NA	Yes	No	No	No	No
Subject 13	235	M	11	Control	Normal	NA	NA	NA	NA	NA	Yes	No	No	No	No
Subject 14	14	F	14	CD	Normal	A1b, L3, L4a, B1, G1	Ileocolonic	37.5	Moderate	Yes	Yes	Yes	Yes	No	No

Subject 15	17	M	13	CD	Inflamed	A1b, L2, L4a, B1, G1	Colonic	57.5	Severe	Yes	Yes	Yes	No	No
Subject 16	22	M	14	CD	Normal	A1b, L1, B1, G1	Ileal	37.5	Moderate	Yes	Yes	Yes	No	No
Subject 17	29	M	14	CD	Inflamed	A1b, L3, L4a, B1, G0	Ileocolonic	32.5	Moderate	Yes	Yes	Yes	No	No
Subject 18	38	F	16	CD	Inflamed	A1b, L2, B1, G0	Colonic	20	Mild	Yes	Yes	Yes	No	No
Subject 19	39	F	13	CD	Normal	A1b, L1, L4a, B1, G1	Ileal	60	Severe	Yes	Yes	Yes	No	No
Subject 20	41	F	15	CD	Normal	A1b, L3, B1, G0	Ileocolonic	57.5	Severe	Yes	Yes	Yes	No	No
Subject 21	44	F	15	CD	Normal	A1b, L1, B3, G1	Ileal	52.5	Severe	Yes	Yes	No	No	No
Subject 22	61	M	9	CD	Inflamed	A1a, L2, B1, G0	Colonic	32.5	Moderate	Yes	Yes	Yes	No	No
Subject 23	63	M	13	CD	Normal	A1b, L1, L4a, B1, G1	Ileal	65	Severe	Yes	Yes	Yes	No	No
Subject 24	72	M	12	CD	Inflamed	A1b, L3, L4a, B1, G1	Ileocolonic	55	Severe	Yes	Yes	Yes	No	No
Subject 25	75	M	14	CD	Inflamed	A1b, L2, B1, G1	Colonic	50	Severe	Yes	Yes	No	No	No
Subject 26	78	F	15	CD	Normal	A1b, L1, B1, G0	Ileal	45	Severe	Yes	Yes	Yes	No	No
Subject 27	79	M	16	CD	Inflamed	A1b, L3, B1, G0	Ileocolonic	45	Severe	Yes	Yes	Yes	No	No
Subject 28	85	M	16	CD	Normal	A1b, L2, B1, G1	Colonic	62.5	Severe	Yes	Yes	Yes	No	No
Subject 29	86	M	8	CD	Normal	A1a, L1, L4ab, B1, G0	Ileal	22.5	Mild	Yes	Yes	Yes	No	No
Subject 30	94	M	11	CD	Normal	A1b, L1, B1p, G0	Ileal	45	Severe	Yes	Yes	No	No	No
Subject 31	97	M	12	CD	Normal	A1b, L3, L4a, B1, G0	Ileocolonic	40	Severe	Yes	Yes	No	No	No
Subject 32	16	M	13	CD	Normal	A1b, L3, B1, G0	Ileocolonic	20	Mild	Yes	Yes	Yes	No	No
Subject 33	205	M	13	CD	Normal	A1b, L1, L4a, B1p, G1	Ileal	45	Severe	Yes	Yes	Yes	No	No
Subject 34	106	F	16	Control	Normal	NA	NA	NA	NA	No	No	Yes	No	No
Subject 35	109	F	3	Control	Normal	NA	NA	NA	NA	No	No	Yes	No	No
Subject 36	110	F	4	Control	Normal	NA	NA	NA	NA	No	No	Yes	No	No
Subject	112	M	9	Control	Normal	NA	NA	NA	NA	No	No	Yes	No	No

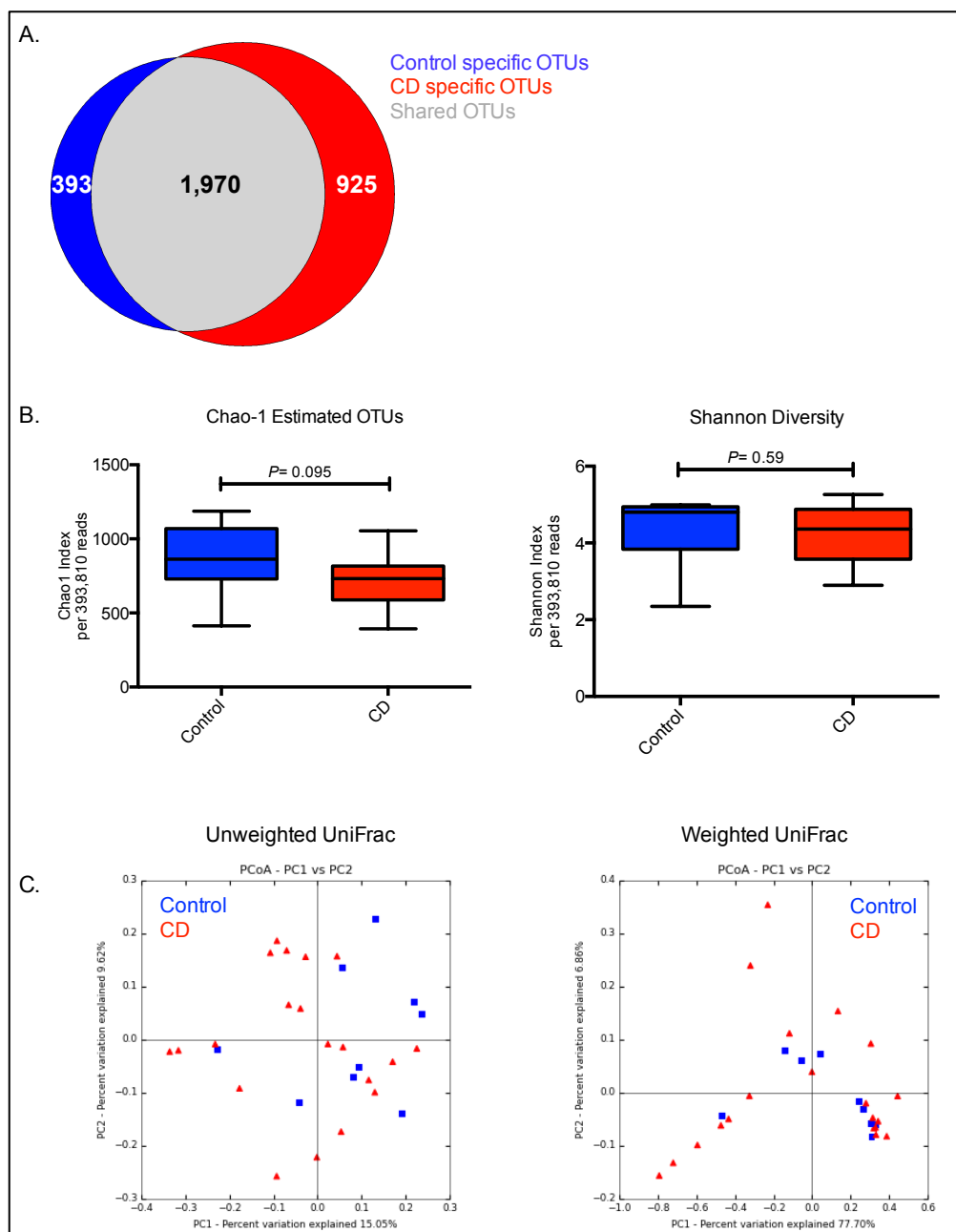
Subject	225	F	13	Control	Normal	NA	NA	NA	NA	No	No	Yes	No	No
Subject 60														
Subject 61	241	F	12	Control	Normal	NA	NA	NA	NA	No	No	Yes	No	No
Subject 62	73	F	16	Control	Normal	NA	NA	NA	NA	No	No	Yes	No	No
Subject 63	321	F	16	Control	Normal	NA	NA	NA	NA	No	No	Yes	Yes	No
Subject 64	324	F	8	Control	Normal	NA	NA	NA	NA	No	No	Yes	Yes	No
Subject 65	368	M	7	Control	Normal	NA	NA	NA	NA	No	No	No	Yes	No
Subject 66	371	M	7	Control	Normal	NA	NA	NA	NA	No	No	No	Yes	No
Subject 67	376	F	10	Control	Normal	NA	NA	NA	NA	No	No	No	Yes	No
Subject 68	387	M	11	Control	Normal	NA	NA	NA	NA	No	No	No	Yes	No
Subject 69	403	F	16	Control	Normal	NA	NA	NA	NA	No	No	No	Yes	No
Subject 70	381	M	17	Control	Normal	NA	NA	NA	NA	No	No	No	Yes	No
Subject 71	397	F	10	Control	Normal	NA	NA	NA	NA	No	No	No	Yes	No
Subject 72	217	F	14	CD	Normal	A1b, L1, L4b, B1, G0	Ileal	55	Severe	No	No	Yes	No	No
Subject 73	231	F	15	CD	Normal	A1b, L1, B1, G0	Ileal	37.5	Moderate	No	No	Yes	No	No
Subject 74	49	M	12	CD	Normal	A1b, L1, B1, G1	Ileal	40	Severe	No	No	Yes	No	No
Subject 75	266	F	10	CD	Normal	A1a, L2, B1, G1	Colonic	27.5	Mild	No	No	Yes	No	Yes
Subject 76	269	M	8	CD	Normal	A1a, L4a, B1, G0	Upper disease	2.5	Inactive	No	No	Yes	No	Yes
Subject 77	278	M	16	CD	Normal	A1a, L1, B1, G0	Ileal	10	Mild	No	No	Yes	No	Yes
Subject 78	298	M	12	CD	Normal	A1b, L1, B1, G0	Ileal	52.5	Severe	No	No	Yes	No	Yes
Subject 79	267	M	10	CD	Normal	A1a, L3, B1, G0	Ileocolonic	10	Mild	No	No	Yes	No	Yes
Subject 80	256	F	10	CD	Normal	A1b, L2, B1, G0	Colonic	30	Moderate	No	No	Yes	No	No
Subject 81	209	F	17	CD	Normal	A1b, L1, L4a, B1, G0	Ileal	35	Moderate	No	No	Yes	No	No
Subject	240	M	10	CD	Normal	A1b, L1, B1, G0	Ileal	65	Severe	No	No	Yes	No	No

Subject 83	319	M	12	CD	Normal	A1b, L1, B1, G1	Ileal	45	Severe	No	No	Yes	No	No
Subject 84	268	M	14	CD	Normal	A1b, L1, B1p, G1	ileal	50	Severe	No	No	Yes	No	No
Subject 85	65	M	15	CD	Normal	A1b, L1, B2, B3, G0	Ileal	50	Severe	No	No	Yes	No	No
Subject 86	261	F	13	CD	Normal	A1b, L3, L4, B1p, G0	Ileocolonic	32.5	Moderate	No	No	Yes	Yes	Yes
Subject 87	262	M	15	CD	Normal	A1b, L1, L4, AB1p, G0	Ileal	27.5	Mild	No	No	Yes	Yes	No
Subject 88	374	M	15	CD	Normal	A1b, L3, B2, G0	Ileocolonic	37.5	Moderate	No	No	No	Yes	No
Subject 89	201	F	11	CD	Inflamed	A1b, L2, B1, G1	Colonic	37.5	Moderate	No	No	Yes	No	No
Subject 90	203	M	11	CD	Inflamed	A1b, L3, B1, G0	Ileocolonic	45	Severe	No	No	Yes	No	No
Subject 91	206	M	14	CD	Inflamed	A1b, L3, B1, G1	Ileocolonic	45	Severe	No	No	Yes	No	No
Subject 92	219	M	14	CD	Inflamed	A1b, L3, B1, G1	Ileocolonic	37.5	Moderate	No	No	Yes	No	No
Subject 93	223	F	9	CD	Inflamed	A1a, L2, B1, G1	Colonic	45	Severe	No	No	Yes	No	No
Subject 94	248	M	10	CD	Inflamed	A1b, L3, B1p, G0	Ileocolonic	42.5	Severe	No	No	Yes	No	Yes
Subject 95	47	M	16	CD	Inflamed	A1b, L3, L4a, B1, G1	Ileocolonic	65	Severe	No	No	Yes	No	No
Subject 96	62	F	15	CD	Inflamed	A1b, L3, L4a, B3, G0	Ileocolonic	57.5	Severe	No	No	Yes	No	No
Subject 97	90	M	16	CD	Inflamed	A1b, L3, L4, B1, G0	Ileocolonic	52.5	Severe	No	No	Yes	No	No
Subject 98	311	F	15	CD	Inflamed	A1b, L1, B1p, G0	Ileal	32.5	Moderate	No	No	Yes	No	Yes
Subject 99	253	F	11	CD	Inflamed	A1b, L4a, L2, B1, G0	Colonic	40	Severe	No	No	Yes	No	No
Subject 100	239	M	3	CD	Inflamed	A1a, L3, B1P, G0	Ileocolonic	27.5	Mild	No	No	Yes	No	No
Subject 101	295	F	14	CD	Inflamed	A1b, L2, B1, G0	Colonic	25	Mild	No	No	Yes	No	No
Subject 102	316	M	14	CD	Inflamed	A1b, L3, L4a, B1, G0	Ileocolonic	27.5	Mild	No	No	Yes	No	No
Subject 103	251	M	8	CD	Inflamed	A1a, L3, L4a, B1, G0	Ileocolonic	47.5	Severe	No	No	Yes	No	No
Subject 104	280	M	15	CD	Inflamed	A1b, L3, B1, G0	Ileocolonic	45	Severe	No	No	Yes	No	No

Subject 105	234	F	13	CD	Inflamed	A1b, L3, B1, G0	Ileocolonic	52.5	Severe	No	No	Yes	No	No
Subject 106	301	F	14	CD	Inflamed	A1b, L3, L4a, B1, G0	Ileocolonic	60	Severe	No	No	Yes	No	No
Subject 107	271	M	15	CD	Inflamed	A1b, L3, B1P, G0	Ileocolonic	67.5	Severe	No	No	Yes	No	No
Subject 108	272	M	12	CD	Inflamed	A1b, L3, B1P, G1	Ileocolonic	65	Severe	No	No	Yes	No	No
Subject 109	294	M	12	CD	Inflamed	A1b, L3, L4a, B1, G0	Ileocolonic	77.5	Severe	No	No	Yes	No	No
Subject 110	254	M	15	CD	Inflamed	A1b, L1, B2, G0	Ileal	27.5	Mild	No	No	Yes	No	No
Subject 111	229	M	10	CD	Inflamed	A1b, L2, L4a, B1, G0	Colonic	57.5	Severe	No	No	Yes	Yes	Yes
Subject 112	260	M	12	CD	Inflamed	A1b, L2, B1p, G0	Colonic	47.5	Severe	No	No	Yes	Yes	Yes
Subject 113	323	F	16	CD	Inflamed	A1b, L3, B1, G0	Ileocolonic	55	Severe	No	No	Yes	Yes	No
Subject 114	338	F	14	CD	Inflamed	A1b, L3, B1, G0	Ileocolonic	47.5	Severe	No	No	No	Yes	No
Subject 115	329	M	12	CD	Inflamed	A1b, L3, B1p, G0	Ileocolonic	52.5	Severe	No	No	No	Yes	No
Subject 116	343	M	16	CD	Inflamed	A1b, L3, B1, G0	Ileocolonic	57.5	Severe	No	No	No	Yes	No
Subject 117	345	M	17	CD	Inflamed	A1b, L3, L4a, B1, G0	Ileocolonic	57.5	Severe	No	No	No	Yes	No
Subject 118	379	F	10	CD	Inflamed	A1b, L1, B1p, G0	Ileal	27.5	Mild	No	No	No	Yes	No
Subject 119	394	M	16	CD	Inflamed	A1b, L3, B1, G0	Ileocolonic	32.5	Moderate	No	No	No	Yes	No

Appendix II: Diversity of all bacteria revealed by 16S rRNA gene sequencing.

(A) Number of Observed Operational Taxonomic Units (OTUs) at 97% sequence similarity. (B) Alpha diversity represented by Chao1 estimated OTUs (left panel) and Shannon diversity index (right panel). Number of reads was equalized between samples at 393,810, reads. Control, blue bar; CD, red bar. (C) Beta diversity presented by two-dimensional principal coordinates analysis (PCoA) plot of unweighted (left panel) and weighted (right panel) UniFrac distances. Percentage of variance explained by each component is presented for each axis. Control samples, blue squares; CD samples, red triangles.



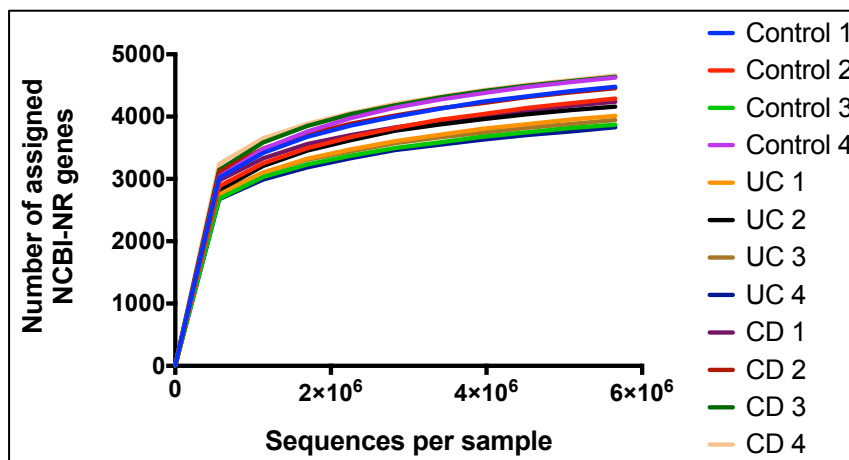
Appendix III: Number of isolated plasmid clones from each sample. Clones from each sample were pooled together before plasmidome DNA purification. Total number of clones per group was equalized to 2,500 clones. UC, ulcerative colitis; and CD, Crohn's disease.

Sample	Number of isolated clones	Sample	Number of isolated clones	Sample	Number of isolated clones
Control 1	362	UC 1	278	CD 1	196
Control 2	489	UC 2	250	CD 2	1,644
Control 3	353	UC 3	1,595	CD 3	374
Control 4	1,296	UC 4	377	CD 4	286
Total Clones	2,500	Total Clones	2,500	Total Clones	2,500

Appendix IV: Clinical data of participating subject in the study. Table also illustrates which subjects were used for 16S rRNA sequencing.

Sample	HMC ID	Gender	Age	Group	Colonoscopy visual appearance of right colon	Paris Classification	Category	PCDAI/PUCAI	Severity	16S rRNA sequencing
Control 1	20	F	12	Control	Normal	NA	NA	NA	NA	No
Control 2	28	M	13	Control	Normal	NA	NA	NA	NA	Yes
Control 3	55	M	6	Control	Normal	NA	NA	NA	NA	Yes
Control 4	43	F	16	Control	Normal	NA	NA	NA	NA	Yes
UC 1	77	M	17	UC	Normal	E3S0	Extensive (hepatic flexure distally); never severe	50	Moderate	Yes
UC 2	92	M	12	UC	Normal	E3S1	Extensive (hepatic flexure distally); ever severe	75	Severe	Yes
UC 3	23	M	14	UC	Normal	E4S1	Pancolitis (proximal to hepatic flexure); ever severe	50	Moderate	Yes
UC 4	24	M	16	UC	Normal	E3S1	Extensive (hepatic flexure distally); ever severe	40	Moderate	Yes
CD 1	90	M	15	CD	Inflamed	Ab1 L3L4a B1 G0	Ileocolonic	52.5	Severe	Yes
CD 2	97	M	13	CD	Normal	A1b L3L4a B1 G0	Ileocolonic	40	Severe	No
CD 3	49	M	12	CD	Normal	A1b L1 B1 G1	Ileal	40	Severe	Yes
CD 4	63	M	13	CD	Normal	A1b L1L4a B1 G1	Ileal with upper disease	65	Severe	Yes

Appendix V: Rarefaction curves calculated for each plasmidome sample based on the NCBI-NR assigned genes. Genes were assigned to the plasmidome reads from each sample using 75% sequence identity, E-value of 1×10^{-5} , and Bit-score of 35 cutoff to the NCBI-NR database utilizing the DIAMOND algorithm [330]. Rarefaction analysis based on the identified genes from each sample was computed using QIIME V1.9.0. The number of reads was equalized between samples at 5,655,160 reads.



Appendix VI: NCBI-NR genes with differential abundance. Mann-Whitney U test was used to identify significant genes between control and IBD ($P < 0.05$). Kruskal-Wallis test followed by Dunn's post hoc test was utilized to identify significant genes between control and IBD subtypes (Dunn's corrected $P < 0.0167$). Values in bold indicate statistical significance.

NCBI-NR ID	$P IBD$	$P Control$	$P UC$	$P CD$
gi 108796630 gb ABG21299.1 Control	0.0481	1	0.117	0.024
gi 11095670 gb AAG29969.1 Control	0.0161	1	0.155	0.004
gi 119394773 gb ABL74511.1 Control	0.1091	1	0.009	0.512
gi 125661553 gb ABN49801.1 Control	0.2141	1	1	0.005
gi 133774810 gb EBA38630.1 Control	0.0731	1	0.433	0.014
gi 134260421 gb ABO65261.1 Control	0.0281	1	0.039	0.141
gi 146260130 gb ABQ14404.1 Control	0.0481	1	0.117	0.05
gi 146743440 gb ABQ43167.1 Control	0.0281	1	0.239	0.008
gi 148873070 gb EDL71205.1 Control	0.0481	1	0.021	0.021
gi 1515098 emb CAA68910.1 Control	0.0281	1	0.17	0.031
gi 152676 gb AAA26475.1 Control	0.0481	1	0.062	0.169
gi 161347206 gb ABX65303.1 Control	0.0161	1	0.141	0.008
gi 16798374 gb AAL29435.1 AF427129_1 Control	0.0481	1	0.096	0.062
gi 169665589 gb ACA63479.1 CD	0.3681	0.031	0.011	1
gi 169665597 gb ACA63483.1 Control	0.0161	1	0.006	0.17
gi 18766963 gb AAL79196.1 AF480833_2 Control	0.0481	1	0.096	0.202
gi 20302879 gb AAM18924.1 AF495873_1 Control	0.0161	1	0.202	0.004
gi 208016 gb AAA73162.1 Control	0.1091	1	0.695	0.011
gi 208459 gb AAA72769.1 CD	0.9331	0.141	0.008	1
gi 208984 gb AAA73221.1 Control	0.0081	1	0.048	0.014
gi 209095 gb AAA72686.1 Control	0.0281	1	0.17	0.031
gi 209869166 emb CAR92282.1 Control	0.1091	1	0.585	0.006
gi 215212 gb AAA32255.1 Control	0.0481	1	0.281	0.014
gi 220683460 gb ACL80551.1 Control	0.0281	1	0.202	0.011
gi 238860059 gb ACR62057.1 Control	0.0481	1	0.096	0.062
gi 241874821 gb ACS70152.1 Control	0.0281	1	0.461	0.141
gi 241875610 gb ACS70546.1 CD	1.0001	0.169	0.014	1
gi 241875740 gb ACS70611.1 Control	0.0481	1	0.053	0.155
gi 241875800 gb ACS70641.1 Control	0.1541	1	0.834	0.014
gi 2440160 emb CAA75108.1 Control	0.1541	1	0.845	0.014
gi 2501755 gb AAC48874.1 Control	0.0481	1	0.492	0.17
gi 262064121 gb ACY07107.1 Control	0.0731	1	0.618	0.008
gi 263233443 gb EEZ19092.1 CD	0.5701	0.05	0.006	1
gi 27573829 pdb 1M0F B Control	0.0731	1	0.008	0.556

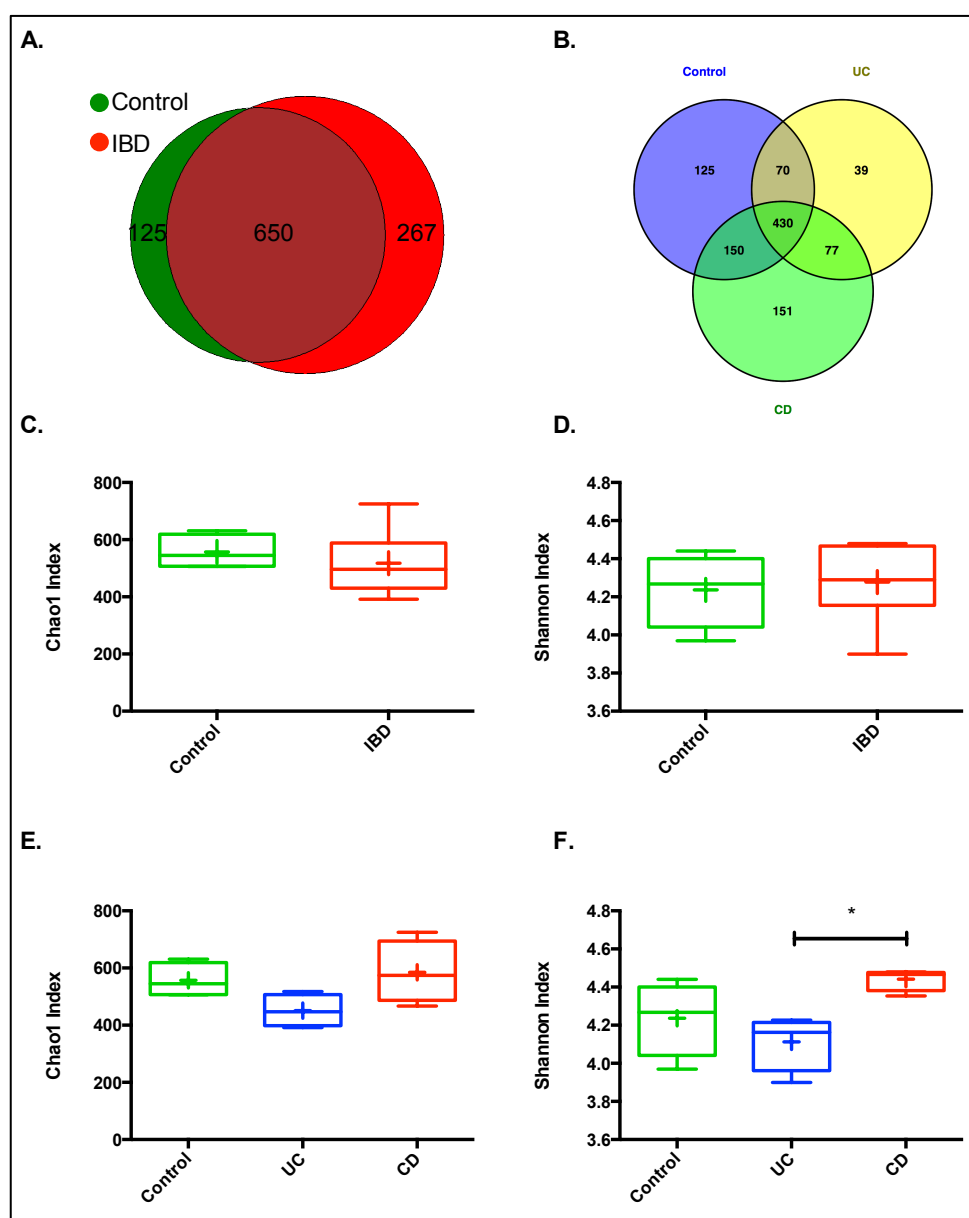
gi 291471672 gb EFF14155.1 Control	0.0481	1	0.327	0.096
gi 292809761 gb EFF68966.1 Control	0.0481	1	0.021	0.021
gi 295115075 emb CBL35922.1 Control	0.0481	1	0.031	0.129
gi 309400374 gb ADO79610.1 Control	0.0281	1	0.062	0.05
gi 315632435 gb ADU53763.1 Control	0.0281	1	0.096	0.117
gi 330752427 emb CBL87378.1 Control	0.0281	1	0.281	0.006
gi 333972755 gb AEG39560.1 Control	0.0481	1	0.28	0.014
gi 375058382 dbj BAL60567.1 Control	0.0281	1	0.327	0.845
gi 37693755 gb AAQ98890.1 Control	0.0481	1	0.845	0.281
gi 378939967 gb AFC75523.1 Control	0.0281	1	0.281	0.062
gi 3808121 emb CAA09780.1 Control	0.0081	1	0.096	0.006
gi 383842778 gb AFH53881.1 Control	0.0081	1	0.239	0.019
gi 3953663 dbj BAA34762.1 Control	0.0481	1	0.096	0.031
gi 402950453 gb EJX68456.1 CD	0.9331	0.127	0.006	1
gi 403017426 gb EJY30174.1 Control	0.0081	1	0.081	0.006
gi 4210487 gb AAD11974.1 CD	0.5701	0.05	0.006	1
gi 43739 emb CAA23889.1 Control	0.0081	1	0.031	0.011
gi 446100002 ref WP_000177857.1 CD	0.3681	0.031	0.011	1
gi 446100092 ref WP_000177947.1 Control	0.0481	1	0.377	0.039
gi 446100096 ref WP_000177951.1 Control	0.0281	1	0.096	0.062
gi 446100119 ref WP_000177974.1 Control	0.0281	1	0.019	0.019
gi 446832557 ref WP_000909813.1 Control	0.0481	1	0.377	0.377
gi 447057677 ref WP_001134933.1 Control	0.0161	1	0.067	0.015
gi 451928584 pdb 3ZDJ A Control	0.0731	1	0.556	0.008
gi 46277646 gb AAS87025.1 CD	0.9331	0.222	0.006	1
gi 46850195 gb AAT02530.1 CD	0.9331	0.117	0.004	1
gi 476116409 gb EMV39178.1 CD	0.1091	0.004	0.016	1
gi 477137054 gb ENE15873.1 Control	0.0481	1	0.023	0.08
gi 481041386 ref WP_001295762.1 Control	0.0481	1	0.295	0.496
gi 485746862 ref WP_001373242.1 Control	0.0081	1	0.078	0.008
gi 485793337 ref WP_001415026.1 Control	0.0041	1	0.093	0.002
gi 485806691 ref WP_001426975.1 Control	0.0281	1	0.281	0.327
gi 485860877 ref WP_001460269.1 Control	0.0481	1	0.326	0.016
gi 485871440 ref WP_001468384.1 Control	0.0481	1	0.555	0.077
gi 487285635 ref WP_001632456.1 Control	0.0481	1	0.189	0.022
gi 487405847 ref WP_001674007.1 Control	0.0161	1	0.141	0.141
gi 487564528 ref WP_001717547.1 Control	0.0281	1	0.039	0.077
gi 487661010 ref WP_001751343.1 CD	0.2141	0.005	0.005	1
gi 487853220 ref WP_001926686.1 CD	0.1091	0.004	0.016	1
gi 487907108 ref WP_001980574.1 Control	0.0081	1	0.139	0.003
gi 487910581 ref WP_001984047.1 Control	0.0481	1	0.556	0.039
gi 488180921 ref WP_002252129.1 Control	0.0731	1	0.489	0.01
gi 488182696 ref WP_002253904.1 Control	0.1541	1	0.845	0.014

gi 488279268 ref WP_002350476.1	Control	0.0281	1	0.008	0.238
gi 489810083 ref WP_003713939.1	Control	0.0481	1	0.433	0.117
gi 490540999 ref WP_004406128.1	Control	0.0481	1	0.065	0.058
gi 490714494 ref WP_004577211.1	Control	0.1541	1	0.765	0.017
gi 490987317 ref WP_004849050.1	Control	0.0161	1	0.094	0.006
gi 490987753 ref WP_004849486.1	Control	0.0481	1	0.126	0.08
gi 491160894 ref WP_005019281.1	Control	0.1541	1	0.845	0.014
gi 491162098 ref WP_005020484.1	Control	0.0481	1	0.295	0.006
gi 491266388 ref WP_005124526.1	Control	0.0481	1	0.029	0.006
gi 492292168 ref WP_005800676.1	Control	0.0161	1	0.072	0.036
gi 492677637 ref WP_005920083.1	Control	0.0081	1	0.062	0.011
gi 493329074 ref WP_006286292.1	Control	0.0281	1	0.202	0.011
gi 494817674 ref WP_007553082.1	Control	0.0281	1	0.17	0.014
gi 495674769 ref WP_008399348.1	Control	0.0481	1	0.08	0.023
gi 495674836 ref WP_008399415.1	Control	0.0161	1	0.093	0.03
gi 498572590 ref WP_010835044.1	CD	1.0001	0.17	0.014	1
gi 499964299 ref WP_011645017.1	Control	0.0481	1	0.008	0.377
gi 500821251 ref WP_012000845.1	Control	0.0281	1	0.202	0.011
gi 501083106 ref WP_012133644.1	Control	0.0281	1	0.281	0.006
gi 502533835 ref WP_012814673.1	Control	0.0481	1	0.096	0.117
gi 503044338 ref WP_013279314.1	Control	0.0081	1	0.014	0.05
gi 503320404 ref WP_013555065.1	Control	0.0281	1	0.039	0.078
gi 505348778 ref WP_015535880.1	Control	0.0481	1	0.021	0.021
gi 505381927 ref WP_015569029.1	Control	0.0161	1	0.093	0.014
gi 507590704 gb EOR83978.1	Control	0.0281	1	0.239	0.239
gi 510927756 ref WP_016248400.1	Control	0.1091	1	0.695	0.011
gi 514663242 gb EPJ19297.1	Control	0.0281	1	0.695	0.17
gi 519784041 gb EPP30712.1	Control	0.0081	1	0.139	0.003
gi 523614619 ref WP_020762354.1	Control	0.0481	1	0.695	0.433
gi 531494529 gb EQH80598.1	Control	0.0481	1	0.277	0.014
gi 531556724 gb EQI39009.1	Control	0.0281	1	0.281	0.845
gi 531626711 gb EQJ05633.1	Control	0.0281	1	0.624	0.031
gi 534787977 gb EQP40445.1	Control	0.0081	1	0.096	0.006
gi 536846312 gb ERC29327.1	Control	0.0161	1	0.239	0.003
gi 536879613 gb ERC61643.1	Control	0.0481	1	0.281	0.17
gi 53756912 gb AAU91203.1	Control	0.0041	1	0.052	0.004
gi 544967051 ref WP_021370747.1	Control	0.0281	1	0.377	0.377
gi 544980124 ref WP_021380190.1	Control	0.0161	1	0.238	0.005
gi 545048784 ref WP_021423038.1	Control	0.0081	1	0.05	0.014
gi 545296843 ref WP_021577058.1	CD	0.5701	0.062	0.011	1
gi 546734476 ref WP_021888857.1	Control	0.0481	1	0.022	0.189
gi 547220975 ref WP_021960263.1	Control	0.0481	1	0.021	0.021
gi 547313703 ref WP_022045363.1	Control	0.0281	1	0.077	0.141

gi 547811649 ref WP_022220766.1 Control	0.0481	1	0.189	0.022
gi 547941520 ref WP_022342824.1 Control	0.0481	1	0.078	0.039
gi 548203089 ref WP_022423198.1 Control	0.0481	1	0.021	0.021
gi 548226340 ref WP_022445215.1 Control	0.0481	1	0.075	0.129
gi 548329222 ref WP_022514978.1 Control	0.0481	1	0.231	0.231
gi 548374420 ref WP_022518728.1 Control	0.0481	1	0.129	0.031
gi 552996 gb AAA32208.1 Control	0.1091	1	0.624	0.014
gi 553928417 ref WP_023140126.1 Control	0.0281	1	0.327	0.17
gi 555421419 ref WP_023263970.1 Control	0.0481	1	0.202	0.007
gi 564587202 gb AHB86974.1 CD	1.0001	0.281	0.014	1
gi 564909311 gb ETE09584.1 Control	0.0481	1	0.023	0.08
gi 568816383 gb AHE38449.1 CD	0.9331	0.281	0.014	1
gi 57157779 dbj BAD83862.1 CD	0.9331	0.277	0.014	1
gi 57157820 dbj BAD83864.1 Control	0.0481	1	0.17	0.031
gi 572727913 gb ETS26555.1 Control	0.0481	1	0.199	0.277
gi 575535143 gb ETX35607.1 CD	0.2141	0.005	0.005	1
gi 585339574 ref WP_024221205.1 Control	0.0481	1	0.05	0.031
gi 585348282 ref WP_024229899.1 Control	0.0481	1	0.202	0.024
gi 596222225 gb EYA67038.1 Control	0.0161	1	0.053	0.019
gi 596233172 gb EYA76433.1 Control	0.0481	1	0.084	0.186
gi 599342835 gb AHN62075.1 CD	0.3681	0.031	0.011	1
gi 603578 emb CAA58446.1 Control	0.0481	1	0.117	0.024
gi 606748092 gb EYX90533.1 Control	0.0081	1	0.019	0.039
gi 607741474 gb EZE23258.1 Control	0.0481	1	0.556	0.141
gi 635649404 ref WP_024296277.1 Control	0.0481	1	0.021	0.021
gi 642975646 ref WP_025188709.1 Control	0.0281	1	0.078	0.141
gi 643562994 ref WP_025232309.1 Control	0.0481	1	0.063	0.08
gi 6457346 gb AAF09485.1 AF191496_1 Control	0.0041	1	0.006	0.05
gi 654789027 ref WP_028243020.1 Control	0.0481	1	0.05	0.006
gi 660355405 gb KEO42207.1 CD	0.5701	0.425	0.005	1
gi 667745630 gb KFA83936.1 CD	0.2831	0.024	0.014	1
gi 671561717 ref WP_031542377.1 Control	0.0041	1	0.016	0.004
gi 671564761 ref WP_031545132.1 Control	0.1091	1	0.624	0.014
gi 690410436 ref WP_031942887.1 Control	0.0481	1	0.141	0.377
gi 691347822 ref WP_032072697.1 Control	0.0081	1	0.05	0.014
gi 692193070 ref WP_032106118.1 Control	0.1091	1	0.624	0.014
gi 692983954 ref WP_032173612.1 Control	0.0481	1	0.209	0.012
gi 693003576 ref WP_032189073.1 CD	1.0001	0.209	0.012	1
gi 693025390 ref WP_032202178.1 Control	0.0281	1	0.769	0.039
gi 693113791 ref WP_032264873.1 Control	0.0041	1	0.096	0.062
gi 693120738 ref WP_032269776.1 Control	0.0481	1	0.141	0.078
gi 693123444 ref WP_032271733.1 Control	0.0081	1	0.078	0.008
gi 693199697 ref WP_032320913.1 Control	0.0161	1	0.155	0.053

gi 693210867 ref WP_032329776.1	Control	0.0281	1	0.377	0.078
gi 694087296 ref WP_032432063.1	CD	0.9331	0.239	0.008	1
gi 695263285 ref WP_032490155.1	CD	0.5701	0.062	0.011	1
gi 695269199 ref WP_032491376.1	CD	0.5701	0.05	0.006	1
gi 695317687 ref WP_032516898.1	Control	0.0481	1	0.016	0.174
gi 695319432 ref WP_032518622.1	Control	0.0481	1	0.141	0.078
gi 695323975 ref WP_032522284.1	Control	0.0481	1	0.038	0.767
gi 695348632 ref WP_032545056.1	Control	0.0481	1	0.096	0.202
gi 695401614 ref WP_032556970.1	Control	0.0481	1	0.17	0.327
gi 695539704 ref WP_032596024.1	Control	0.0481	1	0.086	0.219
gi 71842885 gb AAZ49080.1	Control	0.0481	1	0.202	0.281
gi 71842945 gb AAZ49135.1	Control	0.0481	1	0.556	0.377
gi 727108235 ref WP_033623324.1	Control	0.0281	1	0.09	0.194
gi 727367979 ref WP_033801914.1	Control	0.0481	1	0.845	0.117
gi 727372691 ref WP_033802872.1	Control	0.0281	1	0.169	0.154
gi 727386467 ref WP_033805340.1	Control	0.0081	1	0.141	0.239
gi 729796164 gb KHG91472.1	Control	0.1091	1	0.624	0.014
gi 729951767 gb KHI42278.1	Control	0.0481	1	0.845	0.433
gi 737108003 ref WP_035095871.1	Control	0.0481	1	0.295	0.006
gi 738288061 ref WP_036241831.1	Control	0.0481	1	0.209	0.012
gi 739010638 ref WP_036892165.1	Control	0.0481	1	0.08	0.063
gi 739495614 ref WP_037355341.1	Control	0.0281	1	0.078	0.141
gi 746822997 gb AJD80684.1	Control	0.0161	1	0.17	0.006
gi 749549041 ref WP_040182562.1	Control	0.0281	1	0.327	0.004
gi 749939258 ref WP_040324867.1	CD	0.5701	0.542	0.015	1
gi 971143 dbj BAA09901.1	Control	0.0081	1	0.062	0.024
gi 992929 dbj BAA06208.1	Control	0.0481	1	0.922	0.845

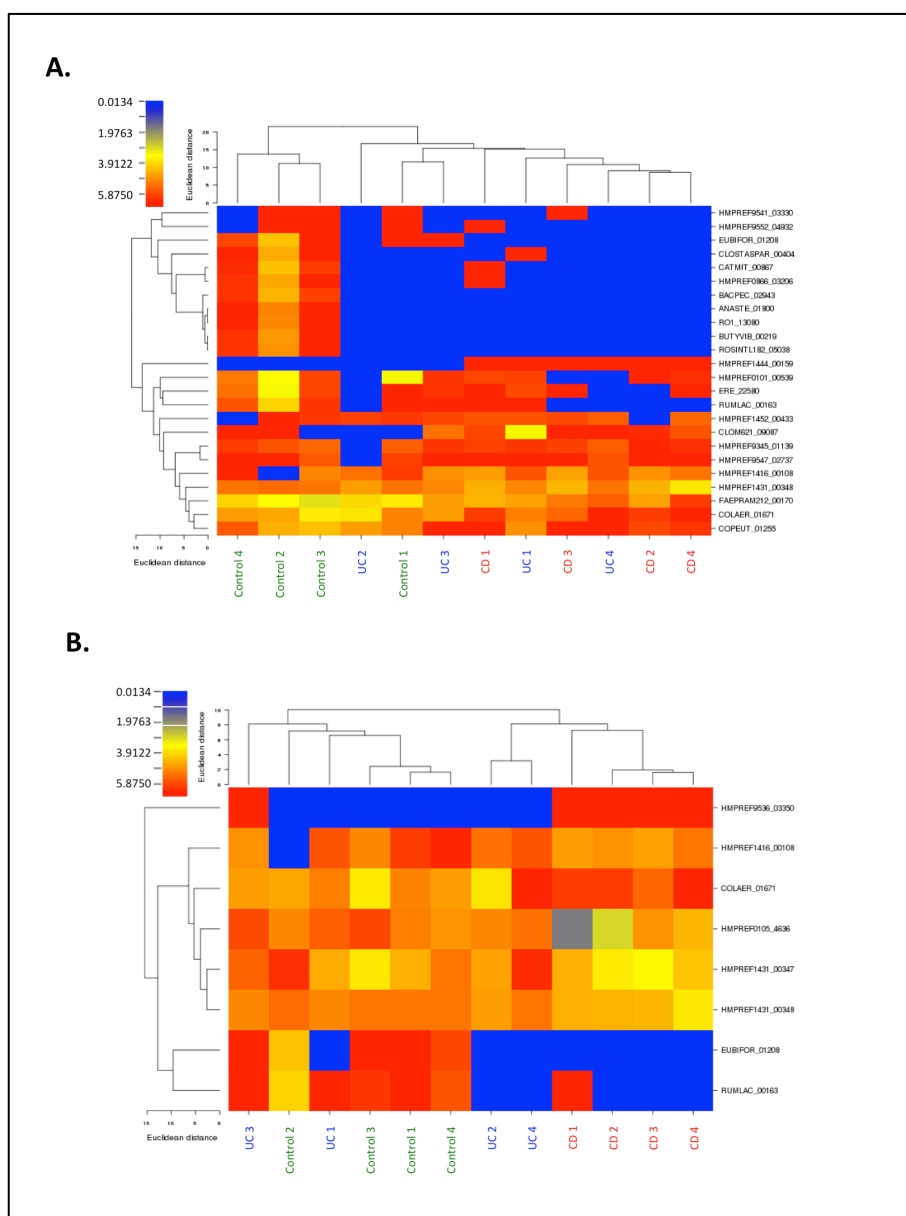
Appendix VII: Alpha diversity of plasmidome revealed by alignment to the HMP-GIT database. Number of identified HMP-GIT genes using 75% sequence identity, E-value of 1×10^{-5} , and Bit-score of 35 cutoff in the control and IBD samples (A) or as a function of IBD subtypes (B). Alpha diversity of identified genes, represented by Chao1 (richness) and Shannon diversity indexes, were calculated using QIIME v1.9.0 and are shown for control and IBD (C-D) or control, UC, and CD (E-F). Number of reads was equalized between samples at 4,787,640 reads. Comparison of Chao1 and Shannon indices between control and IBD was carried out using a Mann-Whitney *U* test and a Kruskal-Wallis test followed by Dunn's post hoc test to compare control vs. UC vs. CD. *, $P < 0.05$.



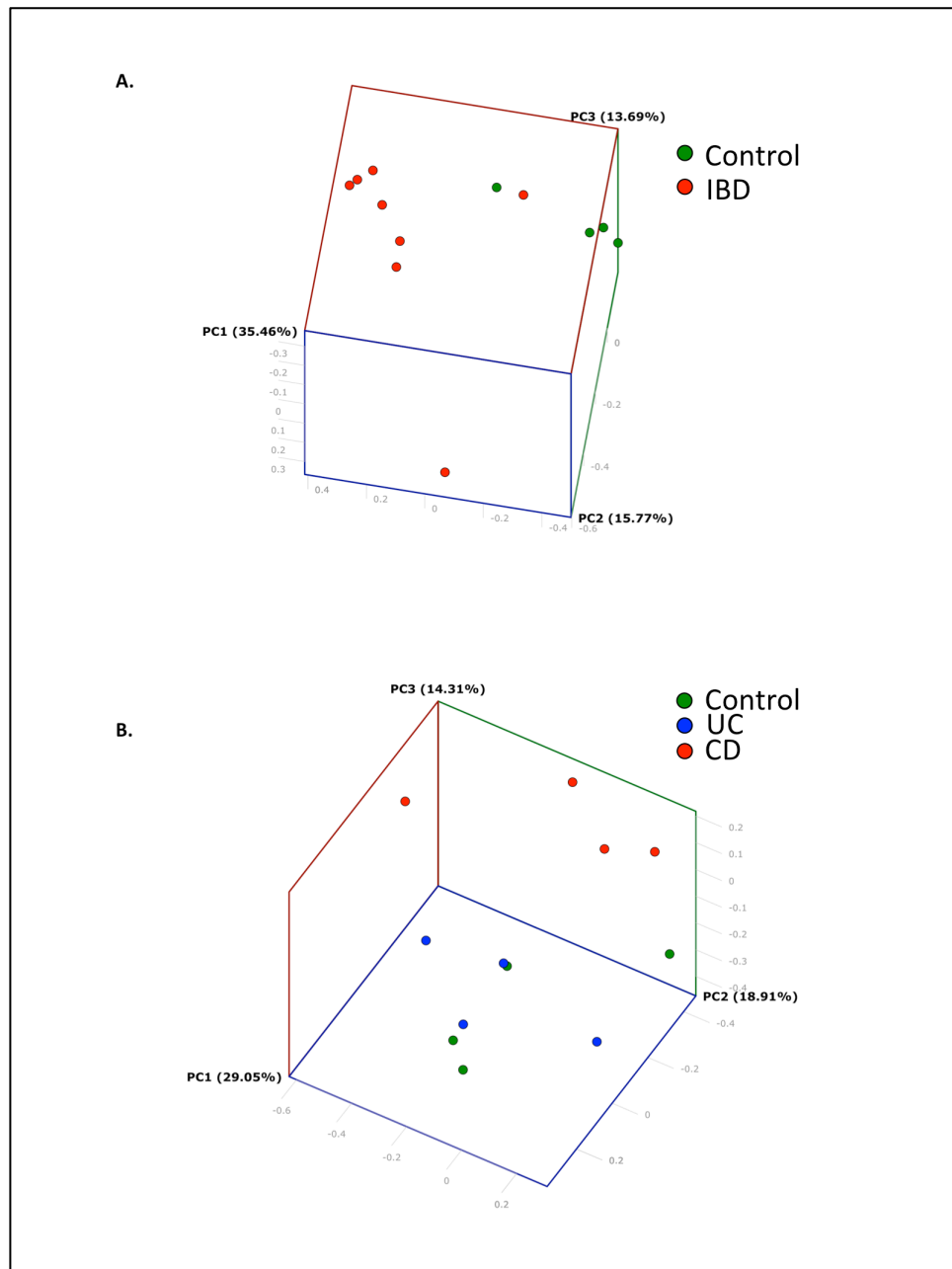
Appendix VIII: HMP-GIT genes with differential abundance. Mann-Whitney *U* test was used to identify significant genes between control and IBD ($P<0.05$). Kruskal-Wallis test followed by Dunn's post hoc test was utilized to identify significant genes between control and IBD subtypes (Dunn's corrected $P<0.0167$). Values in bold indicate statistical significance.

HMP-GIT ID	<i>P</i> IBD	<i>P</i> Control	<i>P</i> UC	<i>P</i> CD
ANASTE_01800 Control	0.0481	1	0.021	0.021
BACPEC_02943 Control	0.0481	1	0.021	0.021
BUTYVIB_00219 Control	0.0481	1	0.021	0.021
CATMIT_00867 Control	0.0481	1	0.023	0.08
CLOM621_09087 Control	0.0481	1	0.075	0.075
CLOSTASPAR_00404 Control	0.0481	1	0.08	0.023
COLAER_01671 Control	0.0481	1	0.281	0.014
COPEUT_01255 Control	0.0481	1	0.141	0.039
ERE_22580 Control	0.0481	1	0.114	0.048
EUBIFOR_01208 Control	0.0041	1	0.016	0.004
FAEPRAM212_00170 Control	0.0081	1	0.039	0.019
HMPREF0101_00539 Control	0.0081	1	0.018	0.038
HMPREF0105_4636 CD	0.5701	0.062	0.011	1
HMPREF0866_03206 Control	0.0481	1	0.023	0.08
HMPREF1416_00108 Control	0.0481	1	0.327	0.011
HMPREF1431_00347 CD	0.6831	0.096	0.014	1
HMPREF1431_00348 Control	0.0081	1	0.17	0.002
HMPREF1444_00159 Control	0.0481	1	0.116	0.028
HMPREF1452_00433 Control	0.0281	1	0.044	0.095
HMPREF9345_01139 Control	0.0481	1	0.078	0.078
HMPREF9536_03350 CD	0.1091	0.004	0.016	1
HMPREF9541_03330 Control	0.0481	1	0.023	0.08
HMPREF9547_02737 Control	0.0481	1	0.078	0.078
HMPREF9552_04932 Control	0.0481	1	0.023	0.08
RO1_13080 Control	0.0481	1	0.021	0.021
ROSINTL182_05038 Control	0.0481	1	0.021	0.021
RUMLAC_00163 Control	0.0081	1	0.042	0.011

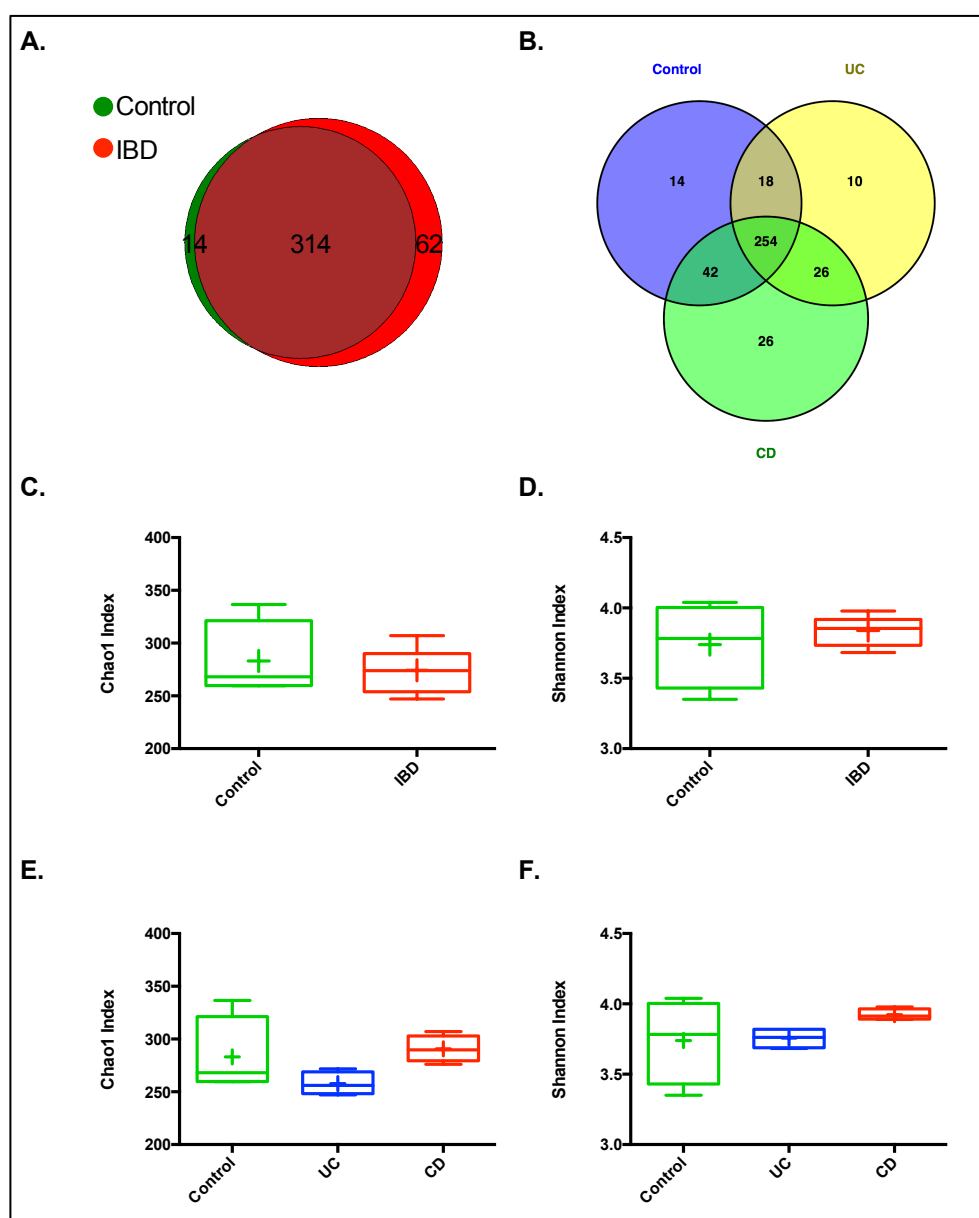
Appendix IX: Plasmidome differentially abundant genes identified in IBD by the HMP-GIT database. Mann-Whitney U test was used to identify significant genes between control and IBD ($P < 0.05$). Kruskal-Wallis test followed by Dunn's post hoc test was utilized to identify significant genes between control and IBD subtypes (Dunn's corrected $P < 0.0167$). The relative abundance of differentially abundant genes was $\log_{10}$ transformed and used to construct the heat maps using CIMminer tool (<http://discover.nci.nih.gov/cimminer/>). The Heat map illustrates the $\log_{10}$ transformed relative abundance of differentially abundant genes. Rows and columns are clustered using Euclidian distance and the average linkage cluster algorithm. (A) The plasmidome differential genes identified between the control and IBD. (B) The plasmidome differential genes identified between the control, UC, and CD.



Appendix X: Beta diversity of the differentially abundant plasmidome genes identified by comparing the relative abundance of the HMP-GIT database assigned genes between groups. The relative abundance of deferentially abundant genes was used to calculate the Bray–Curtis dissimilarity measures between groups. Then, three-dimensional principal coordinates analysis (PCoA) plot of the Bray–Curtis dissimilarity measures between the control and IBD samples (A) or the control, UC, and CD (B). The percentage of variance explained by each component is presented for each axis.



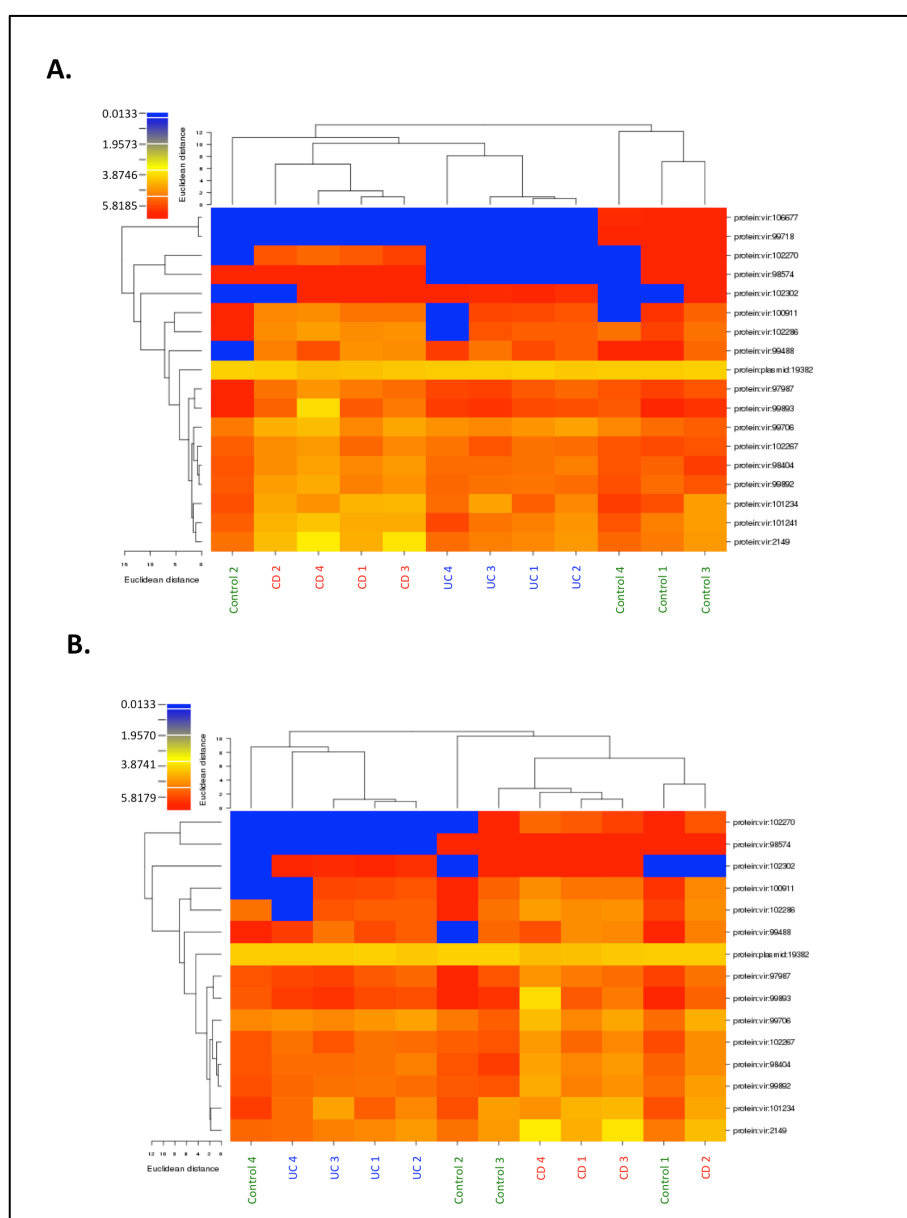
Appendix XI: Alpha diversity of plasmidome revealed by alignment to the ACLAME database. Number of identified ACLAME genes using 75% sequence identity, E-value of 1×10^{-5} , and Bit-score of 35 cutoff in the control and IBD samples (A) or as a function of IBD subtypes (B). Alpha diversity of identified genes, represented by Chao1 (richness) and Shannon diversity indexes, were calculated using QIIME v1.9.0 and are shown for control and IBD (C-D) or control, UC, and CD (E-F). Number of reads was equalized between samples at 4,321,860 reads. Comparison of Chao1 and Shannon indices between control and IBD was carried out using a Mann-Whitney *U* test and a Kruskal-Wallis test followed by Dunn's post hoc test to compare control vs. UC vs. CD. Both statistical analyses showed no statistical significance between groups ($P > 0.05$).



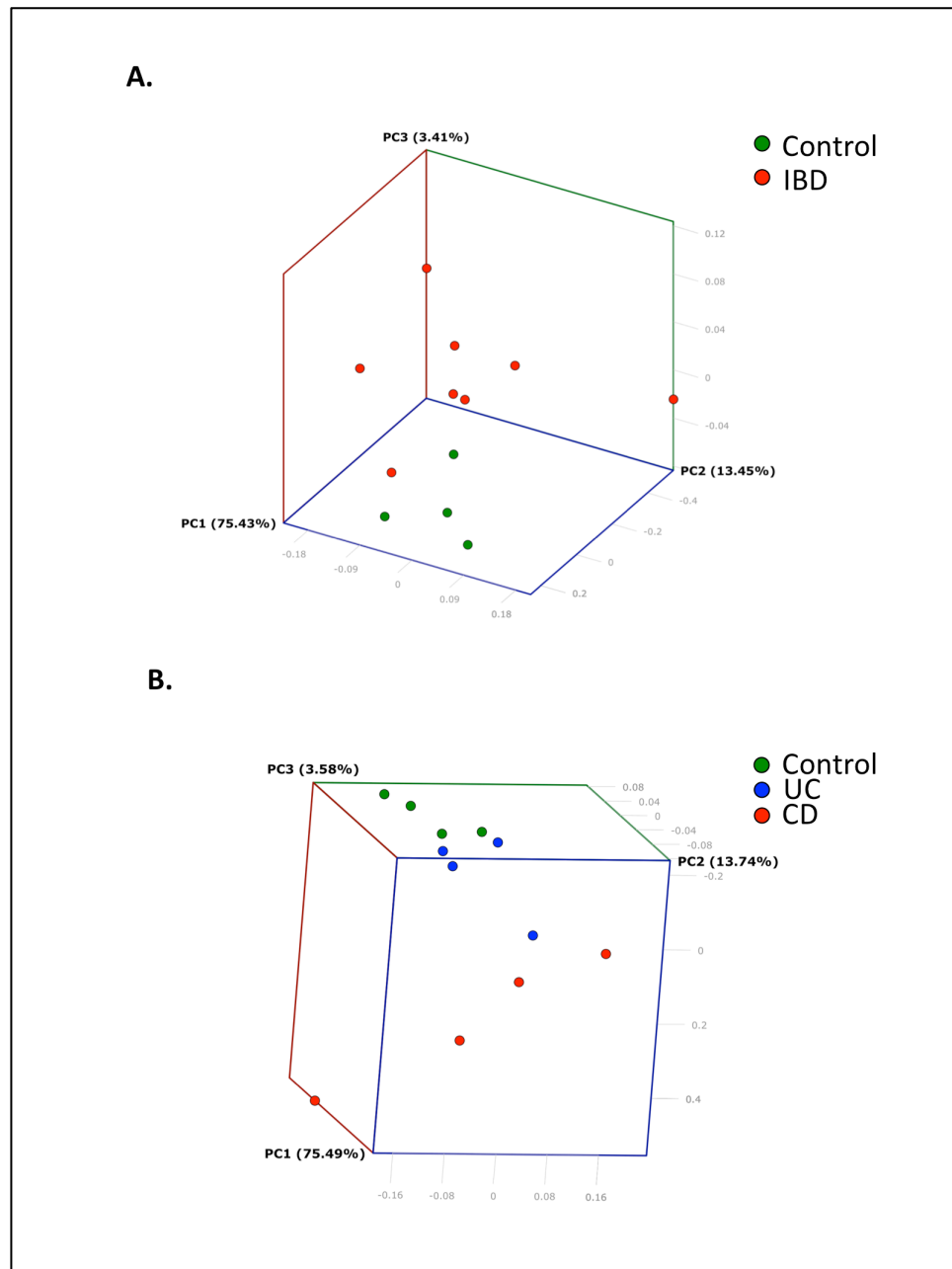
Appendix XII: ACLAME genes with differential abundance. Mann-Whitney *U* test was used to identify significant genes between control and IBD ($P<0.05$). Kruskal-Wallis test followed by Dunn's post hoc test was utilized to identify significant genes between control and IBD subtypes (Dunn's corrected $P<0.0167$). Values in bold indicate statistical significance.

ACLAME ID	<i>P</i> IBD	<i>P</i> Control	<i>P</i> UC	<i>P</i> CD
protein:plasmid:19382 Control	0.035	1	0.624	0.014
protein:vir:100911 Control	0.024	1	0.768	0.012
protein:vir:101234 Control	0.036	1	0.326	0.011
protein:vir:101241 Control	0.025	1	1	0.019
protein:vir:102267 Control	0.022	1	0.096	0.006
protein:vir:102270 CD	0.01	0.036	0.003	1
protein:vir:102286 CD	0.023	0.031	0.011	1
protein:vir:102302 Control	0.038	1	0.011	0.134
protein:vir:106677 Control	0.028	1	0.021	0.021
protein:vir:2149 Control	0.022	1	0.623	0.009
protein:vir:97987 Control	0.018	1	0.433	0.006
protein:vir:98404 Control	0.007	1	0.117	0.002
protein:vir:98574 CD	0.017	0.155	0.004	1
protein:vir:99488 Control	0.049	1	0.281	0.014
protein:vir:99706 Control	0.04	1	0.105	0.012
protein:vir:99718 Control	0.028	1	0.021	0.021
protein:vir:99892 Control	0.01	1	0.17	0.002
protein:vir:99893 Control	0.035	1	0.624	0.014

Appendix XIII: Plasmidome differentially abundant genes identified in IBD by the ACLAME database. Mann-Whitney U test was used to identify significant genes between control and IBD ($P < 0.05$). Kruskal-Wallis test followed by Dunn's post hoc test was utilized to identify significant genes between control and IBD subtypes (Dunn's corrected $P < 0.0167$). The relative abundance of differentially abundant genes was $\log_{10}$ transformed and used to construct the heat maps using CIMminer tool (<http://discover.nci.nih.gov/cimminer/>). The Heat map illustrates the $\log_{10}$ transformed relative abundance of differentially abundant genes. Rows and columns are clustered using Euclidian distance and the average linkage cluster algorithm. (A) The plasmidome differential genes identified between the control and IBD. (B) The plasmidome differential genes identified between the control, UC, and CD.



Appendix XIV: Beta diversity of the differentially abundant plasmidome genes identified by comparing the relative abundance of the ACLAME database assigned genes between groups. The relative abundance of deferentially abundant genes was used to calculate the Bray–Curtis dissimilarity measures between groups. Then, three-dimensional principal coordinates analysis (PCoA) plot of the Bray–Curtis dissimilarity measures between the control and IBD samples (A) or the control, UC, and CD (B). The percentage of variance explained by each component is presented for each axis.



Appendix XV: Recent publication as primary author.

OPEN ACCESS Freely available online



Defining the Vulnerable Period for Re-Establishment of *Clostridium difficile* Colonization after Treatment of *C. difficile* Infection with Oral Vancomycin or Metronidazole

Turki Abujamel¹, Jennifer L. Cadnum², Lucy A. Jury³, Venkata C. K. Sunkesula⁴, Sirisha Kundrapu⁴, Robin L. Jump³, Alain C. Stintzi^{1*}, Curtis J. Donskey^{3,4*}

¹ Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ontario, Canada, ² Research Service, Cleveland Veterans Affairs Medical Center, Cleveland, Ohio, United States of America, ³ Geriatric Research, Education and Clinical Center, Cleveland, Ohio, United States of America, ⁴ Department of Medicine, Infectious Diseases Division, Case Western Reserve University School of Medicine, Cleveland, Ohio, United States of America

Abstract

Background: *Clostridium difficile* is an anaerobic, spore-forming bacterium that is the most common cause of healthcare-associated diarrhea in developed countries. A significant proportion of patients receiving oral vancomycin or metronidazole for treatment of *Clostridium difficile* infection (CDI) develop recurrences. However, the period of vulnerability to re-establishment of colonization by *C. difficile* after therapy is not well defined.

Principal Findings: In a prospective study of CDI patients, we demonstrated that most vancomycin-treated patients maintained inhibitory concentrations of vancomycin in stool for 4 to 5 days after therapy, whereas metronidazole was only detectable during therapy. From the time of elimination of the antibiotics to 14 to 21 days after therapy, a majority of stool suspensions supported growth of *C. difficile* and deep 16S rRNA sequencing demonstrated persistent marked alteration of the indigenous microbiota. By 21 to 28 days after completion of CDI treatment, a majority of stool suspensions inhibited growth of *C. difficile* and there was evidence of some recovery of the microbiota.

Conclusions: These data demonstrate that there is a vulnerable period for re-establishment of *C. difficile* colonization after CDI treatment that begins within a few days after discontinuation of treatment and extends for about 3 weeks in most patients.

Citation: Abujamel T, Cadnum JL, Jury LA, Sunkesula VCK, Kundrapu S, et al. (2013) Defining the Vulnerable Period for Re-Establishment of *Clostridium difficile* Colonization after Treatment of *C. difficile* Infection with Oral Vancomycin or Metronidazole. PLoS ONE 8(10): e76269. doi:10.1371/journal.pone.0076269

Editor: Daniel Paredes-Sabja, Universidad Andres Bello, Chile

Received: May 20, 2013; **Accepted:** August 22, 2013; **Published:** October 2, 2013

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This work was supported by a grant from ViroPharma, Incorporated to CJD by the Canadian Institutes of Health Research (CIHR) to AS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. TA was supported by a scholarship from King Abdulaziz University, through the Saudi Arabian Cultural Bureau in Canada.

Competing interests: The authors received funding for the study from ViroPharma, Incorporated. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: astintzi@uottawa.ca (AS); curtisd123@yahoo.com (CJD)

Introduction

Clostridium difficile is the most important cause of healthcare-associated diarrhea in developed countries [1]. Oral vancomycin and metronidazole are the agents most often used to treat *Clostridium difficile* infection (CDI) [1]. These antibiotics are effective in suppressing *C. difficile*, but they are nonselective agents that also cause significant disruption of the indigenous microbiota of the colon [2–4]. For example, oral vancomycin achieves high concentrations in the intestinal tract, resulting in suppression of anaerobic organisms, including *Bacteroides* spp. [4,5]. Such disruption of the indigenous

microbiota may predispose patients to recurrent CDI due to regrowth of the original infecting *C. difficile* strains or acquisition of new strains [1]. Although it is known that recurrences may occur days to weeks after completion of CDI therapy, the period of vulnerability to re-establishment of colonization by *C. difficile* after therapy is not well defined. In theory, this period will extend from the time vancomycin or metronidazole levels reach subinhibitory concentrations in the colon to the time when the intestinal microbiota recovers sufficiently to inhibit *C. difficile* growth. A better understanding of the timing of the vulnerable period after CDI therapy is needed for the design of interventions to reduce recurrences.

For example, administration of biotherapeutic agents such as nontoxicogenic *C. difficile* could be targeted to coincide with the vulnerable period for re-establishment of colonization [6].

Limited information is available regarding the duration of excretion of vancomycin and metronidazole in stool after treatment. Edlund et al. [4] administered oral vancomycin 0.5 gram/day for 7 days to healthy volunteers; the mean concentration of vancomycin in stool after 7 days of treatment was 520 ± 197 µg/g, but only 1 of 10 subjects had high levels of vancomycin in stool 1 week after discontinuation of treatment (i.e., 135 µg/g). Metronidazole achieved much lower concentrations in stool of CDI patients during therapy (mean $\pm$ SD, 9.3 ± 7.5 µg/g), and levels became undetectable as diarrhea resolved during treatment [7].

Here, we examined the duration of excretion of vancomycin and metronidazole in stool after treatment of CDI and evaluated the period of vulnerability to re-establishment of colonization by *C. difficile*. Because patients with CDI are often elderly and have received multiple prior courses of broad-spectrum antibiotics, we hypothesized that they may have a prolonged period of vulnerability to *C. difficile* colonization after oral vancomycin or oral metronidazole treatment.

Materials and Methods

Ethics Statement

The Institutional Review Board of the Cleveland Veterans Affairs Medical Center approved the study protocol. *C. difficile* strains isolated from patients were collected from the Cleveland VA Medical Center. Written informed consent was obtained from all subjects.

C. difficile Strains

One epidemic North American pulsed-field gel electrophoresis type 1 (NAP1) /polymerase chain reaction ribotype 027 strain (VA 17) and one nonepidemic strain (VA 6) were tested in an in vitro assay of colonization resistance. Both strains were cultured from patients with CDI at the Cleveland VA Medical Center.

Setting and Study Design

We performed an 8-month prospective, observational study of patients treated for CDI with oral metronidazole or vancomycin at the Cleveland VA Medical Center. Patients receiving therapy for multiple recurrences or for severe, complicated CDI as defined by Cohen et al. [1] (i.e., hypotension or shock, ileus, megacolon) were excluded, but patients with severe but uncomplicated CDI were not. Severe CDI was defined as a case in which the white blood cell count was $\geq 15,000$ cells/µl and/or the creatinine was increased 1.5 or more times the baseline level [1]. During the study, diagnostic testing for CDI was performed using enzyme immunoassay (EIA) for glutamate dehydrogenase (Wampole *C. diff* Chek-60, Inverness Medical, Princeton, NJ) as an initial screen with a commercial PCR test for toxin B genes (Becton Dickinson, Cockeysville, MD) for confirmation. The laboratory rejected formed stool samples. The choice of therapy for CDI was made

by the physicians caring for the patients. Information regarding the demographic characteristics, coexisting illnesses, and antibiotic therapy was obtained through standardized chart review. The stool sample used for diagnosis of CDI was collected from the clinical microbiology laboratory and additional stool samples were collected every 3–4 days during treatment and for up to 4 weeks after treatment while the patients were hospitalized.

Microbiology and Molecular Typing

The presence of *C. difficile* in stool samples was measured by plating serially-diluted samples on *C. difficile* Brucella agar (CDBA) as previously described [8]. Plates were incubated for 48 hours at 37 °C in the anaerobic chamber. Colonies were confirmed to be *C. difficile* on the basis of typical odor and appearance of colonies and by a positive reaction using *C. difficile* latex agglutination (Microgen Bioproducts, Camberly, UK). Isolates from the highest dilution with growth were tested for in-vitro cytotoxin production using *C. difficile* Tox A/B II (Wampole Laboratories); isolates that did not produce toxin were excluded from the analysis.

Stool isolates were subjected to molecular typing to determine the prevalence of epidemic North American pulsed-field gel electrophoresis type 1 (NAP1) strains. Crude DNA was extracted from *C. difficile* isolates using QIAamp DNA Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Polymerase chain reaction (PCR)-ribotyping was used to genotype *C. difficile* isolates [9]. PCR was performed to amplify one of the genes for binary toxin (*cdtB*) using the methods of Terhes et al [10]. For each assay, a known NAP1 strain was used as a positive control and American Type Culture Collection (ATCC) *C. difficile* 9689 was used as a negative control.

Measurement of Drug Concentrations in Stool

For vancomycin-treated patients, the concentration of vancomycin in stool samples was measured using an AxSym II fluorescence polarization immunoassay, a standard assay used by clinical Pathology Laboratories for measurement of vancomycin in serum samples. The assay was modified for stool samples by including an extraction of the sample in dilute ammonia to prevent nonspecific binding of vancomycin to protein in the stool [11]. The limit of detection of the assay was 2 µg of vancomycin per mL. For metronidazole-treated patients, the concentration of metronidazole in stool was measured using a bioassay with *Clostridium perfringens* as the indicator organism [12]. For the metronidazole assay, samples from patients receiving concurrent systemic antibiotics for indications other than CDI were excluded. The limit of detection of the assay was 0.5 µg of metronidazole per gm of stool.

In vitro Assay of Colonization Resistance

To assess functional recovery of the ability of the microbiota to inhibit *C. difficile* (i.e., provide colonization resistance) after vancomycin or metronidazole treatment, we used a modification of the *in vitro* assay of colonization resistance developed by Borriello and Barclay [13]. One epidemic NAP1 strain (VA 17) and one nonepidemic strain (VA 6) were tested.

Table 1. Sequence of polymerase chain reaction (PCR) primers used for construction of the Illumina library.

Primer Set	Primer Name	Primer Sequence	Barcode sequences	V6 universal primers Adapted from [14]
Set 1	Paired-end sequencing adapter			
	GF1	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	ATAGCG	CAACGCGARGAACCTTACC-3'
	GF2	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	AGGGT	CAACGCGARGAACCTTACC-3'
	GF3	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	TTCAT	CAACGCGARGAACCTTACC-3'
	GF4	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	GATCGT	CAACGCGARGAACCTTACC-3'
	GF5	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	GCCCGT	CAACGCGARGAACCTTACC-3'
	GF6	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CTGTC	CAACGCGARGAACCTTACC-3'
	GF7	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CACGT	CAACGCGARGAACCTTACC-3'
	GF8	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CGTACG	CAACGCGARGAACCTTACC-3'
	GF9	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	GGAC	CAACGCGARGAACCTTACC-3'
	GF10	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	TAGA	CAACGCGARGAACCTTACC-3'
	GF11	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	TCAT	CAACGCGARGAACCTTACC-3'
	GF12	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT	ACTT	CAACGCGARGAACCTTACC-3'
	GR1	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	ATAGCGA	CAACACGAGTGACGAC-3'
	GR2	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	AGGGTA	CAACACGAGTGACGAC-3'
	GR3	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	TTCATA	CAACACGAGTGACGAC-3'
	GR4	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	GATCGTA	CAACACGAGTGACGAC-3'
	GR5	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	GCCCGTA	CAACACGAGTGACGAC-3'
	GR6	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	CTGTCA	CAACACGAGTGACGAC-3'
	GR7	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	CACGTA	CAACACGAGTGACGAC-3'
	GR8	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	CGTACGA	CAACACGAGTGACGAC-3'
	GR9	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	GGACA	CAACACGAGTGACGAC-3'
	GR10	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	TAGAA	CAACACGAGTGACGAC-3'
	GR11	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	TCATA	CAACACGAGTGACGAC-3'
	GR12	5'-CTCGGCATTCTGCTGAACCGCTCTTCCGATCT	ACTTA	CAACACGAGTGACGAC-3'
Set 2	Illumina flow cell adapter		Paired-end sequencing adapter	
	PCRFDW1	5'-AATGATACGGCGACCAACGAGATCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'	
	PCRRVS1	5'-CAAGCAGAAAGACGCCATACGATCGGT	CTCGGCATTCTGCTGAACCGCTCTTCCGATCT-3'	

Set 1 was used in the PCR 1st step and set 2 was used in the 2nd step.

doi: 10.1371/journal.pone.0076269.t001

Fresh stool samples were homogenized in a 1:1 dilution of pre-reduced sterile water and inoculated with 10⁴ colony-forming units (CFU) of vegetative *C. difficile* inside an anaerobic chamber (Coy Laboratories, Grass Lake, MI). Samples from patients receiving concurrent systemic antibiotics for indications other than CDI and samples with detectable levels of *C. difficile* prior to the inoculation of vegetative *C. difficile* were excluded. The samples were incubated at 37 °C for 24 hours and then serially-diluted and plated on CDBA plates to determine the concentration of *C. difficile*. Loss of colonization resistance was defined as >1 log₁₀CFU increase in the *C. difficile* concentration. In addition to the microbiota, residual metronidazole or vancomycin in stool samples could suppress growth of inoculated *C. difficile*; therefore, growth of the *C. difficile* isolates was concurrently assessed in sterile filtrates of the stool suspensions. The filtrates were produced by centrifuging the suspensions at 10,000 rpm for 10 minutes, followed by filtering the supernatant through a 0.22 µm filter.

Analysis of Stool Microbiota Using Deep 16S rRNA Sequencing

Deep 16SrRNA sequencing was performed on stool specimens from a subset of vancomycin-treated patients with samples available before, during, and after CDI therapy who did not receive concomitant non-CDI antibiotics during or after CDI therapy. DNA was extracted from fecal samples using FastDNA® Spin Kit (MP Biomedicals) utilizing two mechanical lysis cycles in FastPrep® Instrument (MP Biomedicals) at speed 6.0 for 40 seconds. Extracted DNA was then used for library construction. Two sets of primers were designed (Table 1). In set 1, the sequence of V6 universal primers was modified by adding barcodes of 4 to 6 nucleotides, to allow for multiplexing, and the Illumina paired-end sequencing adapters as previously described [14]. In total, 12 different barcodes were generated for each end. Set 2 was designed to have the Illumina paired-end sequencing adapter and the flow-cell adapters.

Illumina Library Construction

The V6 hypervariable region of 16S rRNA was PCR-amplified from extracted DNA in two steps as described by Arthur et al. [14]. Briefly, step 1 used set 1 primers allowing for

a unique combination of bar codes for each sample. The PCR reaction contained 50 ng of extracted DNA. PCR reactions were held at 94 °C for 3 min followed by 10 amplification cycles using a touchdown protocol with denaturation at 94 °C for 45 s, annealing at 61 °C for 45 s with 1 °C drop in each cycle, and an extension at 72 °C for 45 s. Amplification was continued afterward with additional 15 cycles using 51 °C as annealing temperature. Then, PCR was terminated with a final elongation at 72 °C for 2 min. In the 2nd PCR, 15 µl of the 1st PCR products were utilized with set 2 primers using the same concentrations of PCR reagents. Reaction started with one step at 94 °C for 3 min, with amplification proceeding for 15 cycles at 94 °C for 45 s, 65 °C for 45 s, and 72 °C for 45 s with a final extension step at 72 °C for 2 min. The DNA concentration of cleaned PCR amplicons from the second PCR reaction was quantified, and an equimolar combination of each reaction tube was prepared. Finally, the library was sequenced at The Centre for Applied Genomics at the Hospital for Sick Children in Toronto, Canada using one lane of a HiSeq 2000 to generate paired-end reads of 2x 100 bases. The Illumina reads were deposited in the Sequence Reads Archive at National Center for Biotechnology Information (NCBI) under accession number SRP028823.

16S rRNA Sequence Analysis and Characterization of Microbiota Composition and Diversity

Analysis of the sequencing reads was conducted as follows. First, paired-end reads were merged into extended reads using the Fast Length Adjustment of SHort reads algorithm (FLASH) and the fastq file as input [15]. More than 95% of the reads were efficiently merged with a perfect overlap between 10 and 80 nucleotides. Reads that failed to overlap were discarded. Second, reads were extracted and binned by their 5' and 3' barcodes using the NovoBarCode command (www.novocraft.com). During this process, the barcode sequences were stripped from the classified reads. Third, the reads were quality filtered using the fastq_quality_filter command from the Fastx toolkit (Version 0.0.13.1; <http://hannonlab.cshl.edu/>) with a minimum quality score of 33. Finally, the quality filtered classified reads were analyzed using Quantitative Insights Into Microbial Ecology (QIIME) software version 1.5.0 [16]. Specifically, reads were clustered into Operational Taxonomic Units (OTUs) by running a closed-reference OTU picking process using UCLUST against the Greengenes database (version 4feb2011) at 97% sequence identity. The OTUs inherited the taxonomy associated with the corresponding Greengenes sequences. After removal of singletons and doubletons, the relative abundance of taxa (from the phylum to the genus levels) was determined for each sample using the QIIME package. OTUs or taxa exhibiting significant differential representation between groups were selected by linear discriminant analysis using the LefSe package [17].

Data Analysis

Data were analyzed using STATA 9.0 (StataCorp, College Station, TX). Continuous data were analyzed using unpaired *t* tests and categorical data were assessed using Fisher's exact

test. Differences in microbiota relative abundance were analyzed using two-tailed Mann-Whitney test.

Results

Comparison of Vancomycin and Metronidazole-Treated CDI Patients

During the study period, 76 CDI patients had stool samples available for analysis (range, 1 to 8 stool samples), including 42 (55%) patients treated with oral vancomycin and 34 (45%) patients treated with oral metronidazole. Table 2 shows a comparison of the characteristics of the metronidazole and vancomycin-treated patients. There were no statistically significant differences between the two groups. More than half of *C. difficile* isolates from both treatment groups were NAP1/ribotype 027 isolates based on PCR ribotyping and amplification of the binary toxin gene *cdtB*.

Recovery of *C. difficile* From Stool of Vancomycin and Metronidazole-Treated Patients

Figure 1 shows the frequency of recovery of *C. difficile* from stool samples of vancomycin and metronidazole-treated patients. Vancomycin-treated patients had detectable *C. difficile* in stool less frequently than metronidazole-treated patients at the end of CDI therapy but the difference was not statistically significant (4 of 19, 21% versus 5 of 14, 36%; *P* = 0.44); however, vancomycin-treated patients were significantly more likely than metronidazole-treated patients to have detectable *C. difficile* at 10 to 15 days after completion of therapy (6 of 11, 55% versus 2 of 16, 13%; *P* = 0.03). In addition to acquisition of toxigenic strains, 2 vancomycin-treated patients acquired colonization with nontoxigenic *C. difficile* between 8 and 14 days after discontinuation of therapy, whereas none of the metronidazole-treated patients acquired nontoxigenic *C. difficile*.

Antibiotic Concentrations

Figure 2 shows the concentrations of vancomycin and metronidazole present in stool during and after CDI treatment. High concentrations of vancomycin were present during therapy and detectable levels persisted in stool for up to 8 days after therapy. Metronidazole was detectable in a majority of mid-treatment stool samples (12 of 13 samples tested had detectable metronidazole). However, none of the 26 samples tested at the end of treatment or after treatment had detectable metronidazole.

In vitro Assay of Colonization Resistance

Figure 3 shows results for the *in vitro* assay of colonization resistance. For vancomycin-treated patients, growth of *C. difficile* was suppressed in a majority of stool filtrates and suspensions during therapy and for 3–5 days after completion of therapy. By 6 days after discontinuation of CDI therapy, a majority of stool filtrates supported growth of *C. difficile*, consistent with the finding that inhibitory levels of vancomycin were no longer present. From 6 to 21 days after therapy, a majority of stool suspensions supported growth of *C. difficile*,

Table 2. Patient Characteristics.

Characteristic	Metronidazole N = 34	Vancomycin N = 42	P-value
Age, years, mean (range)	68 (60–90)	64 (58–89)	0.17
Male sex	32 (94)	41 (98)	0.58
Creatinine, mg/dL, mean (range)	1.3 (0.3–3.9)	1.6 (0.6–6.1)	0.33
White blood cell count, cells/mm ³ , mean, (range)	13.5 (1.2–30.9)	12.2 (1.6–33.1)	0.52
On systemic antibiotics during or after treatment	8 (24)	7 (17)	0.57
Diabetes Mellitus	9 (26)	21 (50)	0.06
Coronary Artery Disease	10 (29)	14 (33)	0.81
Chronic Obstructive Pulmonary Disease	6 (18)	7 (17)	1.0
End-Stage Renal Disease	3 (9)	3 (7)	1.0
Severe <i>Clostridium difficile</i> Infection	11 (32)	16 (38)	0.64
Infection with epidemic NAP1 strain*	8 (67)	13 (72)	1.0
Nursing home resident	5 (15)	10 (24)	0.39

Notes: Data are No. (%) of patients, unless otherwise indicated.

Abbreviations: NAP1, North American pulsed-field gel electrophoresis type 1.

*. Molecular typing was performed on 12 isolates from metronidazole-treated patients and 18 isolates from vancomycin-treated patients.

doi: 10.1371/journal.pone.0076269.t002

whereas after 21 days most samples were inhibitory, suggesting that the components of the microbiota that provide colonization resistance had recovered sufficiently to inhibit growth.

For metronidazole-treated patients, growth of *C. difficile* was suppressed in stool filtrates and suspensions of samples with detectable metronidazole during treatment, but not in filtrates of samples with no detectable metronidazole. By 15 to 20 days after metronidazole therapy, a majority of stool suspensions inhibited *in vitro* growth of inoculated *C. difficile*.

Microbiota signature of *C. difficile* colonization resistance

Microbiota relative abundance at different phylogenetic levels was studied in stool specimens that did or did not support growth of *C. difficile* based upon the *C. difficile* *in vitro* colonization assay; for this analysis, specimens collected during or within 5 days after CDI treatment were excluded because suppression of *C. difficile* may have been due to the CDI therapy. At the phylum level, *Actinobacteria* levels were higher ($P=0.07$) in specimens with intact colonization resistance (i.e., inhibition of *C. difficile* growth) (Figure 4). Bacterial biomarkers associated with intact or disrupted colonization resistance to *C. difficile* were detected using LEFSe approach [17]. Based on LEFSe analysis, three bacterial genera were associated with absence of colonization resistance to *C. difficile* (Figure 5). The two genera with the highest effect size (*Salmonella* and *Hahella*) belong to *Gammaproteobacteria* class of the phylum *Proteobacteria*, while the third genus (*Collinsella*) is related to the phylum *Actinobacteria*. In general, bacteria belonging to three phyla (*Actinobacteria*, *Firmicutes*, and *Tenericutes*) were associated with intact colonization resistance. The highest effect size was observed in the class *Actinobacteria* and the genus *Lactobacillus*. Other bacteria, arranged from highest to lowest

effect size are unclassified genus of the *Catabacteriaceae* family, the order *Actinomycetales*, the genus *Ruminococcus*, the genus *Finnegoldia*, an unclassified genus of the *Coriobacteriaceae* family, and an unidentified bacterium genus (RFN2) of *Tenericutes*. Also, the relative abundance of the genus *Bifidobacterium* was increased in resistant compared to sensitive samples ($P = 0.055$).

Analysis of Stool Microbiota of Vancomycin-Treated Patients

Figure 6 shows data for a single vancomycin-treated patient that illustrates the typical changes seen over time in the stool microbiota before, during, and after treatment. A total of ~ 14 million high quality reads were generated from all samples with an average reads count of ~ 0.5 million reads/sample. Before treatment, the microbiota was composed of 8 phyla, with *Proteobacteria* (67.1%), *Firmicutes* (27.7%), and *Actinobacteria* (2.2%) being most abundant. The remaining 3.1% was distributed among *Bacteroidetes*, *Cyanobacteria*, *Fusobacteria*, *Tenericutes*, and *Verrucomicrobia* (data not shown). During vancomycin treatment, there was a general decrease in different families of the *Firmicutes* phylum including *Streptococcaceae*, *Lachnospiraceae*, *Leuconostocaceae*, and *Ruminococcaceae*. In addition to *Firmicutes* phylum, a marked decrease was also observed in the *Proteobacteria*, *Actinobacteria* and *Tenericutes* phyla represented by *Enterobacteriaceae*, *Bifidobacteriaceae*, and *Erysipelotrichaceae* families; respectively. This decrease in relative abundance reached its lowest level at the end of treatment period.

Following treatment, two main trends were observed in families with reduced abundance during treatment. First, there was a sudden increase in bacterial relative abundance 15–20 days following treatment. Second, the bacteria relative abundance dropped to a level lower than the pre-treatment

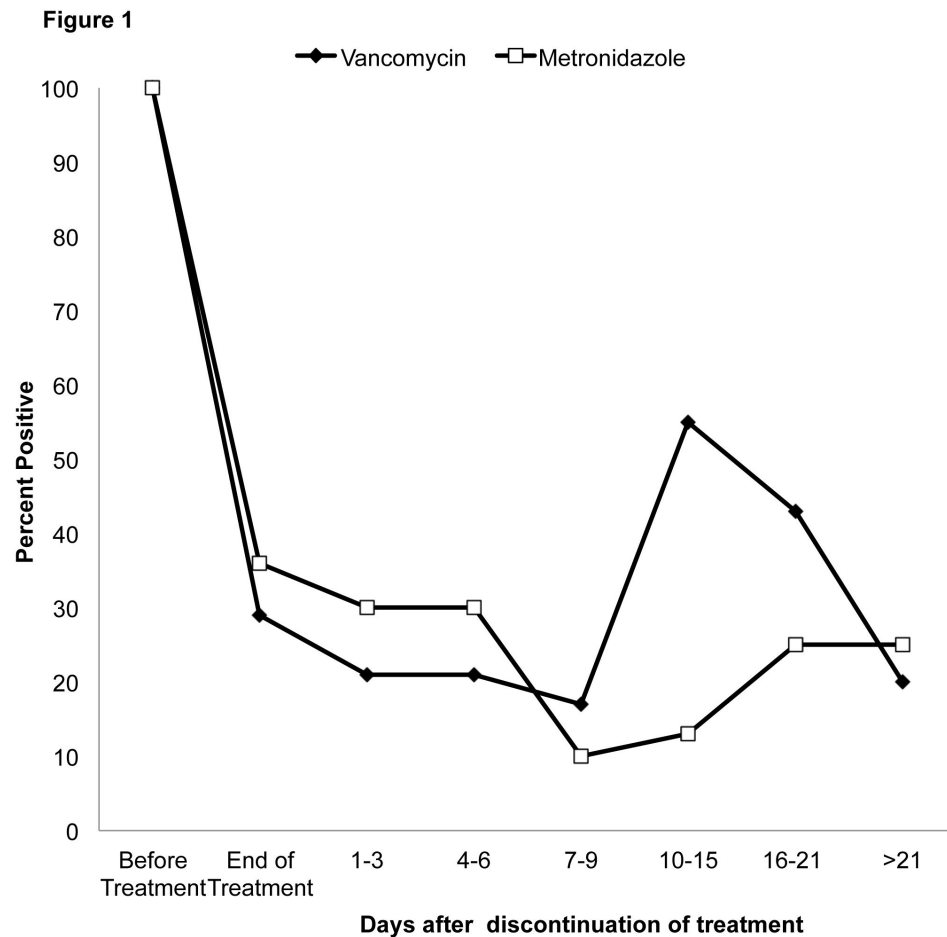


Figure 1. Frequency of recovery of toxigenic *Clostridium difficile* from stool samples of vancomycin and metronidazole-treated patients. In addition to acquisition of toxigenic strains, 2 vancomycin-treated patients acquired colonization with nontoxigenic *C. difficile* between 8 and 14 days after discontinuation of therapy.

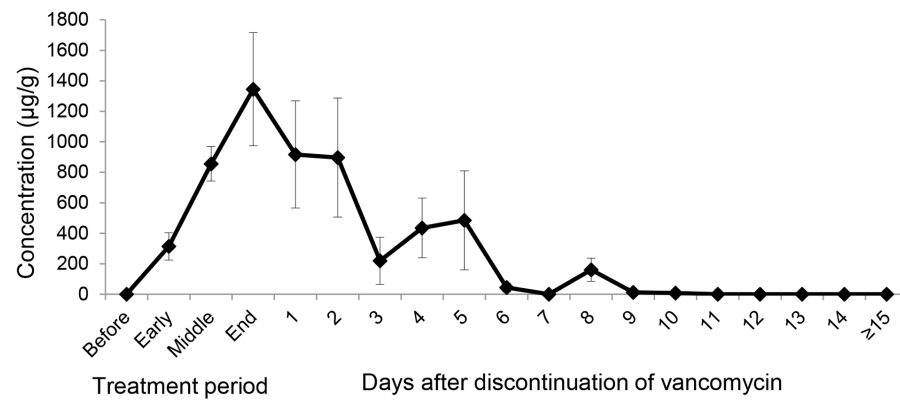
doi: 10.1371/journal.pone.0076269.g001

state. This was seen with *Streptococcaceae*, *Ruminococcaceae*, and *Enterobacteriaceae*. The second trend was characterized by an increase in relative abundance following treatment over a 30 day period, in which it exceeded the original pre-treatment state. This was observed with *Bifidobacteriaceae*, *Leuconostocaceae*, *Lachnospiraceae*, and *Erysipelotrichaceae*.

Discussion

In an observational study of CDI patients, we found that vancomycin-treated patients had high levels of vancomycin in stool during therapy that typically decreased to undetectable levels by 6 to 10 days after discontinuation of therapy. In contrast, metronidazole-treated patients had low levels of drug present during treatment that decreased to undetectable levels

2A



2B

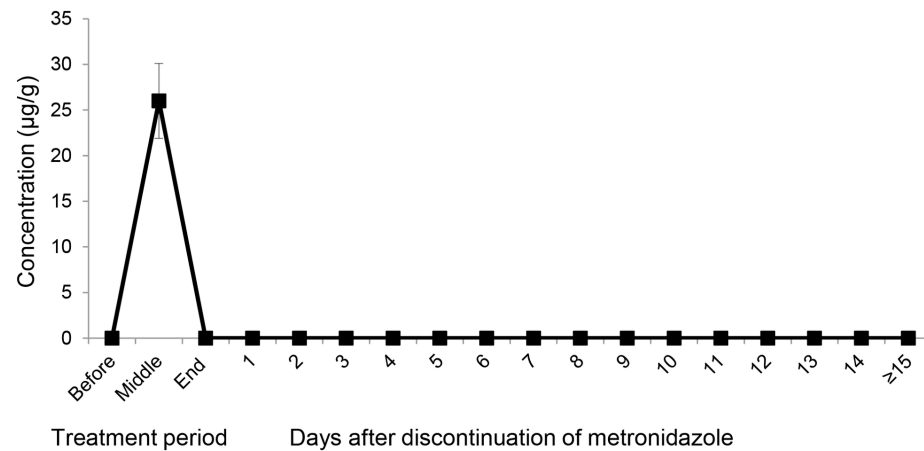


Figure 2. Concentrations of vancomycin and metronidazole in stool during and after *Clostridium difficile* infection (CDI) treatment. The concentration of vancomycin was measured using an AxSym II fluorescence polarization immunoassay that was modified for stool samples by including an extraction of the sample in dilute ammonia to prevent nonspecific binding of vancomycin to protein in the stool. The limit of detection of the assay was 2 µg of vancomycin per ml. The concentration of metronidazole in stool was measured using a bioassay with *Clostridium perfringens* as the indicator organism. Samples from patients receiving concurrent systemic antibiotics for indications other than CDI were excluded.

doi: 10.1371/journal.pone.0076269.g002

by the end of CDI therapy. For both drugs, a significant proportion of fresh stool suspensions from the period shortly

after completion of therapy (~6 to 21 days after completion of vancomycin and from end of therapy to 14 days post-treatment

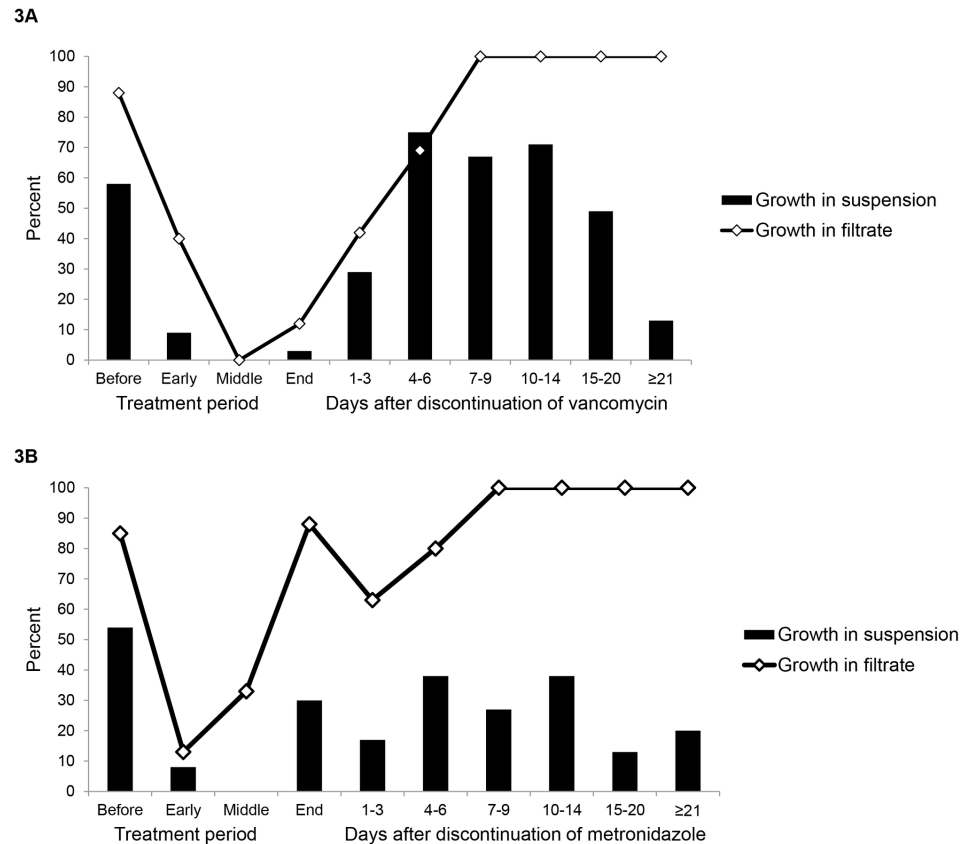


Figure 3. In vitro assay of colonization resistance during and after oral vancomycin (A) or metronidazole (B) treatment. Growth of *Clostridium difficile* was assessed in homogenized fresh stool specimens inoculated with 10^4 colony-forming units (CFU) of vegetative *C. difficile*. Loss of colonization resistance was defined as $>1 \log_{10}$ CFU increase in the *C. difficile* concentration. To assess whether residual metronidazole or vancomycin in stool samples could account for suppression of growth of inoculated *C. difficile*, growth of the *C. difficile* isolates was concurrently assessed in sterile filtrates of the stool suspensions.

doi: 10.1371/journal.pone.0076269.g003

for metronidazole) did not inhibit growth of inoculated *C. difficile*, suggesting that the microbiota that provide colonization resistance against *C. difficile* had not yet recovered. Moreover, vancomycin-treated patients frequently developed recurrent colonization with *C. difficile* 10 to 14 days after completion of therapy, whereas metronidazole-treated patients did not. These results suggest that there is a vulnerable period for re-establishment of *C. difficile* colonization after oral vancomycin or metronidazole therapy that extends from the time when drug levels are reduced to subinhibitory concentrations to the time

when the microbiota are sufficiently recovered to inhibit growth of *C. difficile*.

Previous studies have demonstrated that oral vancomycin may cause marked and prolonged suppression of the indigenous microbiota of the colon, including *Bacteroides* spp. [4,5]. Our data confirm those observations. Moreover, our findings demonstrate that oral metronidazole therapy may also be associated with prolonged disruption of the microbiota. Although previous studies have demonstrated that disturbances of the microbiota may be prolonged, they have

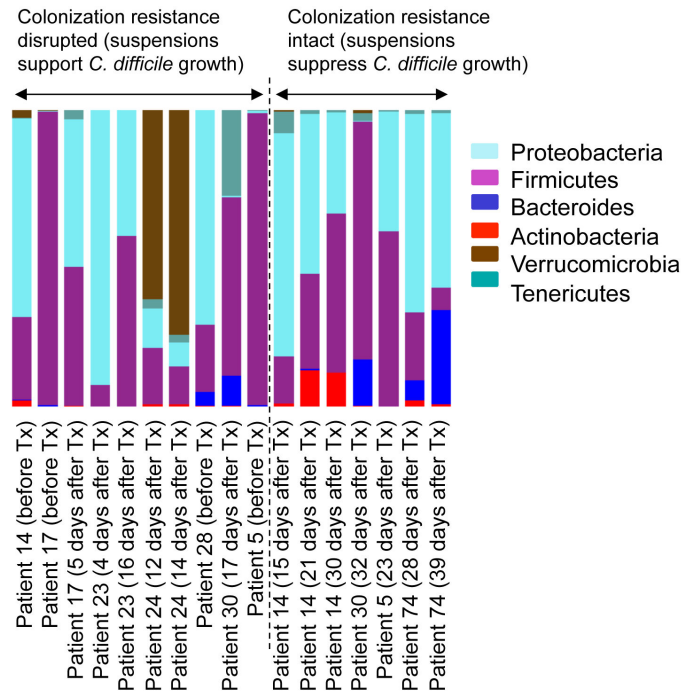
Figure 4

Figure 4. Relative abundance of bacterial phyla in stool of 8 patients treated with oral vancomycin. Stool specimens collected before or after completion of vancomycin therapy for Clostridium difficile infection (CDI), stratified by those with intact (i.e., *C. difficile* growth suppressed) or disrupted (i.e., *C. difficile* growth supported) colonization resistance based on the *in vitro* colonization assay. Microbiota of each stool specimen was evaluated by deep 16S rRNA sequencing. Each column represents total phyla distribution of one sample. Columns are labelled with sample ID and timing of stool collection in comparison to vancomycin treatment.

doi: 10.1371/journal.pone.0076269.g004

not included a functional assessment of the ability of the microbiota to inhibit colonization by *C. difficile*. Our findings from the *in vitro* assay of colonization resistance suggest that functional recovery of colonization resistance may precede full recovery of the diversity of the microbiota. Our *in vitro* data are supported by the finding that *C. difficile* was frequently acquired at 10 to 14 days but not later after discontinuation of vancomycin therapy. Similarly, Johnson et al. [18] found that vancomycin, but not metronidazole, treatment suppressed *C. difficile* in stool of asymptomatic carriers, but that 8 of 9 carriers again began shedding spores 20 $\pm$ 8 days after completing treatment.

Based upon deep 16SrRNA sequencing analysis, the presence of bacteria belonging to three phyla (*Actinobacteria*, *Firmicutes*, and *Tenericutes*) was associated with intact colonization resistance. The class *Actinobacteria* and the genus *Lactobacillus* had the strongest association with intact colonization resistance to *C. difficile*. *Actinobacteria* is a very large class that encompasses about 300 different bacterial genera, including *Bifidobacteriaceae*. We found that the family *Bifidobacteriaceae* regained colonization in the gut following vancomycin treatment at a level higher than the pretreatment state. Of particular interest, the relative abundance of the genus *Bifidobacterium* was increased in samples with intact

Figure 5

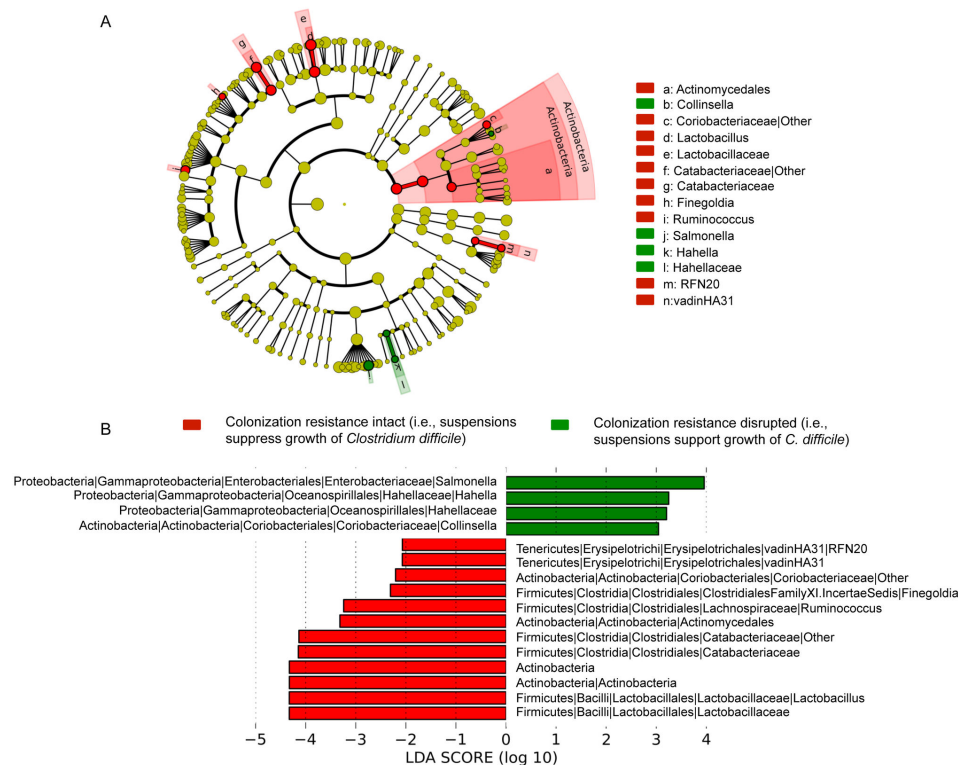


Figure 5. Microbiota signature of bacteria associated with intact versus disrupted colonization resistance to *Clostridium difficile*. (A) Phylogenetic tree display of the bacteria identified from stool samples of 8 vancomycin-treated patients with intact (i.e., *C. difficile* growth suppressed) or disrupted (i.e., *C. difficile* growth supported) colonization resistance based on the *in vitro* colonization assay. Clades of bacteria that significantly support or suppress *C. difficile* growth are highlighted in green or red, respectively. Circle diameter at each phylogenetic level is proportional to the corresponding taxon's abundance. (B) Histogram of the Linear Discriminant Analysis (LDA) score of key bacteria abundance in suppressive and supportive groups arranged according to their effect size. Positive (green bars) and negative (red bars) LDA scores represent supportive and suppressive bacteria, respectively.

doi: 10.1371/journal.pone.0076269.g005

colonization resistance. Although the efficacy of probiotics for prevention of CDI is controversial, some studies suggest that *Lactobacillus* and *Bifidobacterium*, when used as probiotics, may be associated with reduced CDI [19,20].

Ruminococcus spp. was also associated with intact colonization resistance in this study. *Ruminococcus* strains are known to produce the lantibiotic Ruminococcin A (RumA) in the

presence of trypsin, which inhibit the *in vitro* growth of various pathogenic bacteria including *C. difficile* and *C. perfringens* [21]. LEFSe has highlighted the family of *Catabacteriaceae* within the group of colonization resistant bacteria. *Catabacteriaceae* is a recently identified family which comprises one descendant, *Catabacter hongkongensis* [22,23]. The role of *C. hongkongensis* in the human gut and its

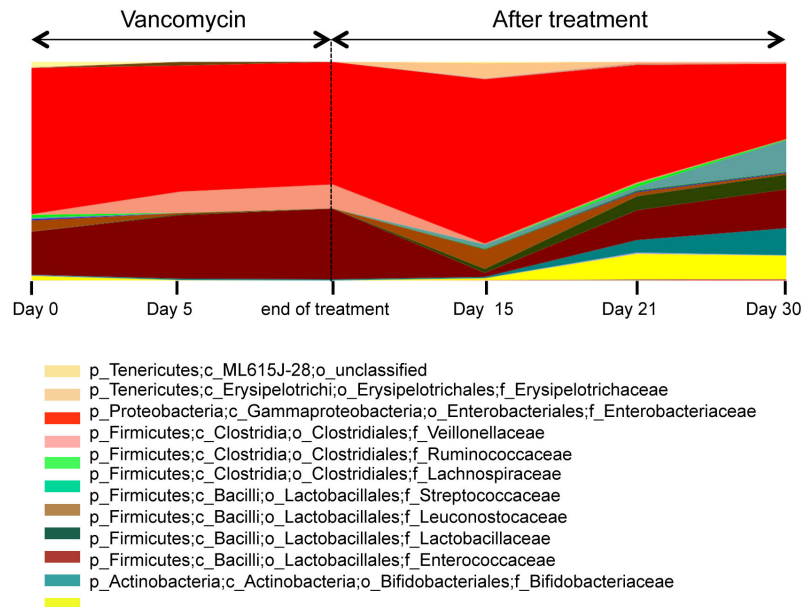
Figure 6

Figure 6. Surveillance of bacterial family relative abundance before, during, and after vancomycin treatment in one patient. Stool samples were collected before (Day 0), during, and after vancomycin treatment at the indicated time points. Changes in the abundance (abundance >0.05%) of major bacterial families are shown at each time point.

doi: 10.1371/journal.pone.0076269.g006

microbial community is unknown. Similarly, the role of *Finnegoldia* and *Coriobacteriaceae* in *C. difficile* colonization resistance requires further investigation. Finally, although the unidentified bacterium genus (RFN2) of *Tenericutes* was identified among the bacteria associated with colonization resistance, their increased abundance following vancomycin treatment may also be attributed to their natural resistance to vancomycin [24,25]. *Salmonella*, *Hahella* and *Collinsella* were associated with disrupted colonization resistance.

Our findings have important implications for development of strategies to prevent recurrences of CDI. For example, nontoxigenic *C. difficile* strains are in Phase 2 trials for secondary prevention of CDI [6,26]. Knowledge of the vulnerable period for re-establishment of *C. difficile* colonization may be useful for optimization of the timing of dosing of this agent. In addition, this knowledge could be used to optimize the timing of other biotherapeutic agents being trialed for prevention of recurrences. Other infection control measures (e.g., environmental disinfection or bathing to reduce the

burden of spores on skin) may also be most beneficial if targeted to the vulnerable period.

Our study has several limitations. The study population included mostly elderly men with multiple medical problems. Additional assessments of the impact of CDI therapies in other populations are needed. Because the study was observational, it is possible that the longer period of disruption of colonization resistance in the vancomycin versus metronidazole-treated patients was attributable to the fact that vancomycin-treated subjects were more likely to have severe CDI, rather than to differences in the 2 drugs. However, the longer period of disruption of colonization resistance by vancomycin is consistent with the higher and more prolonged levels of drug achieved in the intestinal tract. The in vitro assay of colonization resistance has been validated with in vivo findings in hamsters [13] and mice (authors' unpublished data), but further validation of this method is required in humans.

In summary, we demonstrated that there is a vulnerable period for re-establishment of *C. difficile* colonization after CDI

treatment that begins when levels of vancomycin or metronidazole reach subinhibitory concentrations and extends for about 3 weeks in most patients. These findings have important implications for development of effective interventions to prevent recurrences of CDI.

References

- Cohen SH, Gerding DN, Johnson S, Kelly CP, Loo VG et al. (2010) Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). *Infect Control Hosp Epidemiol* 31: 431-455. doi:10.1086/651706. PubMed: 20307191.
- Al-Nassir VN, Sethi AK, Li Y, Pultz MJ, Riggs MM et al. (2008) Both oral metronidazole and oral vancomycin promote persistent overgrowth of vancomycin-resistant enterococci during treatment of *Clostridium difficile*-associated disease. *Antimicrob Agents Chemother* 52: 2403-2406. doi:10.1128/AAC.00090-08. PubMed: 18443120.
- Sethi AK, Al-Nassir VN, Nerandzic MM, Donskey CJ (2009) Skin and environmental contamination with vancomycin-resistant enterococci in patients receiving oral metronidazole or oral vancomycin treatment for *Clostridium difficile*-associated disease. *Infect Control Hosp Epidemiol* 30: 13-17. doi:10.1086/592710. PubMed: 19046063.
- Edlund C, Barkholt L, Olsson-Liljequist B, Nord CE (1997) Effect of vancomycin on intestinal flora of patients who previously received antimicrobial therapy. *Clin Infect Dis* 25: 729-732. doi:10.1086/513755. PubMed: 9314469.
- Louie TJ, Emery J, Krulicki W, Byrne B, Mah M (2009) OPT-80 eliminates *Clostridium difficile* and is sparing of *Bacteroides* species during treatment of *C. difficile* infection. *Antimicrob Agents Chemother* 53: 261-263. doi:10.1128/AAC.01443-07. PubMed: 18955523.
- Sambol SP, Merrigan MM, Tang JK, Johnson S, Gerding DN (2002) Colonization for the prevention of *Clostridium difficile* disease in hamsters. *J Infect Dis* 186: 1781-1789. doi:10.1086/345676. PubMed: 12447764.
- Bolton RP, Culshaw MA (1986) Faecal metronidazole concentrations during oral and intravenous therapy for antibiotic associated colitis due to *Clostridium difficile*. *Gut* 27: 1169-1172. doi:10.1136/gut.27.10.1169. PubMed: 3781329.
- Nerandzic MM, Donskey CJ (2009) Effective and reduced-cost modified selective medium for isolation of *Clostridium difficile*. *J Clin Microbiol* 47: 397-400. doi:10.1128/JCM.01591-08. PubMed: 19073869.
- Bidet P, Lalande V, Salauze B, Burghoffer B, Avesani V et al. (2000) Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. *J Clin Microbiol* 38: 2484-2487. PubMed: 10878030.
- Terhes G, Urbán E, Söki J, Hamid KA, Nagy E (2004) Community-acquired *Clostridium difficile* diarrhea caused by binary toxin, toxin A, and toxin B gene-positive isolates in Hungary. *J Clin Microbiol* 42: 4316-4318. doi:10.1128/JCM.42.9.4316-4318.2004. PubMed: 15365032.
- Currie BP, Lemos-Filho L (2004) Evidence for biliary excretion of vancomycin into stool during intravenous therapy: potential implications for rectal colonization with vancomycin-resistant enterococci. *Antimicrob Agents Chemother* 48: 4427-4429. doi:10.1128/AAC.48.11.4427-4429.2004. PubMed: 15504873.
- Rolfe RD, Finegold SM (1983) Intestinal β -lactamase activity in ampicillin-induced, *Clostridium difficile*-associate ileocolitis. *J Infect Dis* 147: 227-235. doi:10.1093/infdis/147.2.227. PubMed: 6600772.
- Borriello SP, Barclay FE (1986) An in vitro model of colonization resistance to *Clostridium difficile* infection. *J Med Microbiol* 21: 299-309. doi:10.1099/00222615-21-4-299. PubMed: 3723582.
- Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM et al. (2012) Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 338: 120-123. doi:10.1126/science.1224820. PubMed: 22903521.
- Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27: 2957-2963. doi:10.1093/bioinformatics/btr507. PubMed: 21903629.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 108 Suppl 1: 4516-4522. doi:10.1073/pnas.1000080107. PubMed: 20534432.
- Segata N, Izard J, Waldron L, Gevers D, Miropolsky L et al. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* 12: R60. doi:10.1186/gb-2011-12-6-r60. PubMed: 21702898.
- Johnson S, Homann SR, Bettin KM, Quick JN, Clabots CR et al. (1992) Treatment of asymptomatic *Clostridium difficile* carriers (fecal excretors) with vancomycin or metronidazole. A randomized, placebo-controlled trial. *Ann Intern Med* 117: 297-302. doi:10.7326/0003-4819-117-4-297. PubMed: 1322075.
- Hickson M, D'Souza AL, Muthu N, Rogers TR, Want S et al. (2007) Use of probiotic *Lactobacillus* preparation to prevent diarrhoea associated with antibiotics: randomised double blind placebo controlled trial. *BMJ* 335: 80. doi:10.1136/bmj.39231.599815.55. PubMed: 17604300.
- Malaguarnera G, Leggio F, Vacante M, Motta M, Giordano M et al. (2012) Probiotics in the gastrointestinal diseases of the elderly. *J Nutr Health Aging* 16: 402-410. doi:10.1007/s12603-011-0357-1. PubMed: 22499466.
- Dabard J, Bridonneau C, Phillippe C, Anglade P, Molle D et al. (2001) Ruminococcin, new lantibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces. *Appl Environ Microbiol* 67: 4111-4118. doi:10.1128/AEM.67.9.4111-4118.2001. PubMed: 11526013.
- Lau SK, McNabb A, Woo GK, Hoang L, Fung AM et al. (2007) *Catabacter hongkongensis* gen. nov., sp. nov., isolated from blood cultures of patients from Hong Kong and Canada. *J Clin Microbiol* 45: 395-401. doi:10.1128/JCM.01831-06. PubMed: 17122022.
- Lau SK, Fan RY, Lo HW, Ng RH, Wong SS et al. (2012) High mortality associated with *Catabacter hongkongensis* bacteremia. *J Clin Microbiol* 50: 2239-2243. doi:10.1128/JCM.00128-12. PubMed: 22518872.
- Robinson CJ, Young VB (2010) Antibiotic administration alters the community structure of the gastrointestinal microbiota. *Gut Microbes* 1: 279-284. doi:10.4161/gmic.1.4.12614. PubMed: 20953272.
- Sommer MO, Dantas G (2011) Antibiotics and the resistant microbiome. *Curr Opin Microbiol* 14: 556-563. doi:10.1016/j.mib.2011.07.005. PubMed: 21802347.
- Villano SA, Seiberling M, Tatarowicz W, Monnot-Chase E, Gerding DN (2012) Evaluation of an oral suspension of VP20621, spores of nontoxicogenic *Clostridium difficile* strain M3, in healthy subjects. *Antimicrob Agents Chemother* 56: 5224-5229. doi:10.1128/AAC.00913-12. PubMed: 22850511.

Author Contributions

Conceived and designed the experiments: AS CJD. Performed the experiments: TA JLC LAJ VCKS SK. Analyzed the data: VCKS SK RLJ AS CJD. Wrote the manuscript: TA JLC AS CJD.