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, Sohib 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
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. Eubacterium rectale and Coprococcus are depleted and Faecalibacterium prausnitzii 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 ................................................................................................ 95

GUT PLASMIDOME IN EARLY ONSET PEDIATRIC INFLAMMATORY BOWEL DISEASE ................................................................. 95

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
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. *Eubacterium rectale* and *Faecalibacterium prausnitzii* were quantified using BCoAT and 16S rRNA primers ..............84

Figure 2.3.10: Quantitative PCR analysis of *Faecalibacterium prausnitzii* 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.

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.

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.

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.

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.

Figure 4.1: Proposed gut butyrate producers and plasmidome interactions in pediatric IBD.
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>The summary of gut microbiota characterization studies on IBD</td>
<td>20</td>
</tr>
<tr>
<td>Table 2.2.1</td>
<td>Primers used for 16S rRNA library construction</td>
<td>47</td>
</tr>
<tr>
<td>Table 2.2.2</td>
<td>Primers used for BCoAT library construction</td>
<td>52</td>
</tr>
<tr>
<td>Table 2.2.3</td>
<td>Primers used for qPCR</td>
<td>56</td>
</tr>
<tr>
<td>Table 2.3.1</td>
<td>Cohort characteristics</td>
<td>60</td>
</tr>
<tr>
<td>Table 2.3.2</td>
<td>Summary of BCoAT pyrosequencing data</td>
<td>73</td>
</tr>
<tr>
<td>Table 3.2.1</td>
<td>Oligonucleotide sequences used for plasmidome library construction</td>
<td>106</td>
</tr>
<tr>
<td>Table 3.3.1</td>
<td>Participating patients and disease characteristics</td>
<td>113</td>
</tr>
<tr>
<td>Table 3.3.2</td>
<td>Number of plasmidome reads generated from each sample</td>
<td>117</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>ACLAME</td>
<td>A CLAssification of Mobile genetic Elements database</td>
<td></td>
</tr>
<tr>
<td>APS</td>
<td>Antigen-presenting cells</td>
<td></td>
</tr>
<tr>
<td>BCD</td>
<td>Butyryl-CoA dehydrogenase</td>
<td></td>
</tr>
<tr>
<td>BCoAT</td>
<td>Butyryl CoA:acetate CoA-transferase</td>
<td></td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
<td></td>
</tr>
<tr>
<td>BHBD</td>
<td>β-hydroxybutyryl-CoA dehydrogenase</td>
<td></td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
<td></td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
<td></td>
</tr>
<tr>
<td>BUK</td>
<td>Butyrate-kinase</td>
<td></td>
</tr>
<tr>
<td>CARD15</td>
<td>Caspase Recruitment Domain-Containing Protein 15</td>
<td></td>
</tr>
<tr>
<td>CCD</td>
<td>Colonic CD</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>Crohn’s Disease</td>
<td></td>
</tr>
<tr>
<td>CHEO</td>
<td>Children’s Hospital of Eastern Ontario</td>
<td></td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
<td></td>
</tr>
<tr>
<td>COGs</td>
<td>Clusters of Orthologous Groups</td>
<td></td>
</tr>
<tr>
<td>COX2</td>
<td>Cyclooxygenase 2</td>
<td></td>
</tr>
<tr>
<td>CRO</td>
<td>Crotonase</td>
<td></td>
</tr>
<tr>
<td>Ct</td>
<td>Threshold Cycle</td>
<td></td>
</tr>
<tr>
<td>DCs</td>
<td>Dendritic Cells</td>
<td></td>
</tr>
<tr>
<td>DGGE</td>
<td>Denaturing gradient gel electrophoresis</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>DSS</td>
<td>Dextran Sodium Sulfate</td>
<td></td>
</tr>
<tr>
<td>EAEC</td>
<td>Enteroaggregative <em>Escherichia coli</em></td>
<td></td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
<td></td>
</tr>
<tr>
<td>FLASh</td>
<td>Fast Length Adjustment of Short Reads</td>
<td></td>
</tr>
<tr>
<td>GI</td>
<td>Gastrointestinal</td>
<td></td>
</tr>
<tr>
<td>GIs</td>
<td>Genbank identifiers</td>
<td></td>
</tr>
<tr>
<td>GIT</td>
<td>Gastrointestinal tract</td>
<td></td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-Wide Association Study</td>
<td></td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone Deacetylase</td>
<td></td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
<td></td>
</tr>
<tr>
<td>HMP</td>
<td>Human Microbiome Project</td>
<td></td>
</tr>
<tr>
<td>HMP-GIT</td>
<td>Human Microbiome Project Gastrointestinal_tract reference sequence</td>
<td></td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
<td></td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory Bowel Disease</td>
<td></td>
</tr>
<tr>
<td>IBD-U</td>
<td>IBD-unclassified</td>
<td></td>
</tr>
<tr>
<td>IBS</td>
<td>Irritable Bowel Syndrome</td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>Indeterminate colitis</td>
<td></td>
</tr>
<tr>
<td>ICD</td>
<td>Ileal CD</td>
<td></td>
</tr>
<tr>
<td>IgA</td>
<td>Immunoglobulin A</td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
<td></td>
</tr>
</tbody>
</table>
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
NFKB  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
Treg  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\textsubscript{reg}) cells [27, 28]. The T\textsubscript{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-κB 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-κB 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<sup>−/−</sup>* 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 show 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
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.
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.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Number of participants</th>
<th>Sample type</th>
<th>Technique</th>
<th>Observed dysbiosis</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HC</td>
<td>CD</td>
<td>UC</td>
<td>Firmicutes</td>
<td>Bacteroidetes</td>
</tr>
<tr>
<td>Adult</td>
<td>31</td>
<td>26</td>
<td>31</td>
<td>Biopsy</td>
<td>SSCP 16S rRNA sequencing qPCR</td>
</tr>
<tr>
<td></td>
<td>Adult</td>
<td>34</td>
<td>35</td>
<td>Tissue</td>
<td>16S rRNA sequencing qPCR</td>
</tr>
<tr>
<td>Adult</td>
<td>7</td>
<td>13</td>
<td>0</td>
<td>Biopsy</td>
<td>16S rRNA sequencing qPCR FISH</td>
</tr>
<tr>
<td>Adult</td>
<td>139</td>
<td>68</td>
<td>0</td>
<td>Stool</td>
<td>DGGE</td>
</tr>
<tr>
<td>Adult</td>
<td>28</td>
<td>0</td>
<td>8</td>
<td>Biopsy</td>
<td>16S rRNA sequencing</td>
</tr>
<tr>
<td>Adult</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>Stool</td>
<td>16S rRNA Microarray 16S rRNA sequencing FISH</td>
</tr>
<tr>
<td>Adult</td>
<td>35</td>
<td>29</td>
<td>16</td>
<td>Stool Biopsy</td>
<td>16S rRNA NGS</td>
</tr>
<tr>
<td>-------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Adult</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>Biopsy</td>
<td>16S rRNA sequencing</td>
</tr>
<tr>
<td>Adult</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>Stool</td>
<td>16S rRNA Microarray qPCR</td>
</tr>
<tr>
<td>Adult</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>Biopsy</td>
<td>16S rRNA sequencing qPCR</td>
</tr>
<tr>
<td>Adult</td>
<td>27</td>
<td>121</td>
<td>75</td>
<td>Biopsy Stool</td>
<td>16S rRNA NGS Metagenomic sequencing</td>
</tr>
<tr>
<td>Adult</td>
<td>21</td>
<td>21</td>
<td>34</td>
<td>Stool</td>
<td>qPCR</td>
</tr>
<tr>
<td>Pediatric</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>Biopsy</td>
<td>16S rRNA NGS qPCR</td>
</tr>
<tr>
<td>Pediatric</td>
<td>221</td>
<td>447</td>
<td>0</td>
<td>Biopsy Stool</td>
<td>16S rRNA NGS Metagenomic sequencing</td>
</tr>
<tr>
<td>Pediatric</td>
<td>7</td>
<td>12</td>
<td>7</td>
<td>Biopsy</td>
<td>Conventional culturing qPCR</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>40</td>
<td>29</td>
<td>Stool</td>
<td>qPCR</td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td><strong>Pediatric</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 *Il10*<sup>−/−</sup> 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 *Il10*<sup>−/−</sup> 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 though 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 Biomedical, Santa Ana, California) utilizing two mechanical lysis cycles in a FastPrep® Instrument (MP Biomedical) 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.

- 16S rRNA V6 Hypervariable region
- Purify Amplicon
- Prepare a pool of 50 ng from each sample
- Gel purify samples pool
- Illumina HiSeq 2000 platform

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

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF1</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF2</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF3</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF4</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF5</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF6</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF7</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF8</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF9</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF10</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF11</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MF12</td>
<td>AACTCTTTTTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR1</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR2</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR3</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR4</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR5</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR6</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR7</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR8</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR9</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR10</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR11</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>MR12</td>
<td>CTCGGCATCTCTGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>PCR-FWD1</td>
<td>AATGATACGGCGACCACGAGATCTACAGTGTGACGATCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
<tr>
<td>PCR-RVS1</td>
<td>CAAGCAGAAAGACGCACTACGAGATCAGTGCCTGATTACCGCTGACTCAGCAGCCTCTTCCTGCAATAGCGAAACTCAAAGTAGGACG</td>
</tr>
</tbody>
</table>
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 doubletones 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.
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 µl PCR reaction containing 1.5 mM MgCl₂, 2.5 µM 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.
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 ¼ 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.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence 5'-3'</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCoATscrF</td>
<td>GCIGAICATTTCACITGGAAYWSITGGCAYATG</td>
<td>[264]</td>
</tr>
<tr>
<td>BCoATscrR</td>
<td>CCTGCCCRTTTGCAATRRCRAANGC</td>
<td>[264]</td>
</tr>
<tr>
<td>BCoAT-F1</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGACGACTGTGCGTGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F2</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGACGACTGTGCGTGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F3</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F4</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F5</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F6</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F7</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F8</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F10</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F11</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F13</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-F14</td>
<td>CCATCTCATCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
<tr>
<td>BCoAT-R</td>
<td>CCTATCCCCTGCGTGTCTCCGACCTCAGATACAGACACGTGGCIGAICATTTTCACITGGAAYWSITGGCAYATG</td>
<td>This study</td>
</tr>
</tbody>
</table>
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 doubletones 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 (ΔCt) 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 ΔCt 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 µM of each primer. ΔCt calculation and comparison was done as described for the BCoAT and BUK genes.
Table 2.2.3: Primers used for qPCR.

<table>
<thead>
<tr>
<th>Target</th>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal 16S rRNA</td>
<td>UniF</td>
<td>GTGSTGCAYGGYYGTCGTCA</td>
<td>[290]</td>
</tr>
<tr>
<td></td>
<td>UniR</td>
<td>ACGTCTGCCCMSTCTCCCTCTC</td>
<td></td>
</tr>
<tr>
<td>BCoAT gene</td>
<td>BCoATscrF</td>
<td>GCIGAICATTTCACITGGGAAYWSIT</td>
<td>[264]</td>
</tr>
<tr>
<td></td>
<td>BCoATscrR</td>
<td>GCAACIGCYYYYGTTTATATGATCAGGAGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGCMCTTTGCAATRCAANGCTGACATGTG</td>
<td></td>
</tr>
<tr>
<td>BUK gene</td>
<td>G_buk_F</td>
<td>TGCTGTWGTGWWAGAGAGAGA</td>
<td>[263]</td>
</tr>
<tr>
<td></td>
<td>G_buk_R</td>
<td>GCAACIGCYYYYGTTTATATGATCAGGAGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGCMCTTTGCAATRCAANGCTGACATGTG</td>
<td></td>
</tr>
<tr>
<td>E. rectale/ Clostridium coccoides group 16S rRNA gene</td>
<td>UniF338</td>
<td>ACTCCTACGGAGGAGGAGG</td>
<td>[269]</td>
</tr>
<tr>
<td></td>
<td>C.cocR491</td>
<td>GCTTCTTAGTCAGGCTACGGATGATG</td>
<td></td>
</tr>
<tr>
<td>F. prausnitzii 16S rRNA gene</td>
<td>Fprau 07</td>
<td>CCATGAATTGGCTTCAAAACTGTT</td>
<td>[291]</td>
</tr>
<tr>
<td></td>
<td>Fprau 02</td>
<td>GAGCCTAGGCCTACGGATG</td>
<td></td>
</tr>
<tr>
<td>E. rectale BCoAT gene</td>
<td>RosEub_F</td>
<td>TCAAATCMGGIGACTGGGTG</td>
<td>[263]</td>
</tr>
<tr>
<td></td>
<td>Eub_R</td>
<td>TGTTCTTAGTCAGGCTACGGATG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GACAGGGGCCGTAGGTCTA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGACAGGGCAGATRAAGCTCITGT</td>
<td></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th></th>
<th>Crohn’s disease</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Patients</td>
<td>68</td>
<td>51</td>
</tr>
<tr>
<td>Age at Colonoscopy (years)</td>
<td>13±3.81</td>
<td>13.1±2.64</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>45</td>
<td>24</td>
</tr>
<tr>
<td>Female</td>
<td>23</td>
<td>27</td>
</tr>
<tr>
<td>Disease Severity*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Severe</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Macroscopic Mucosal Appearance</td>
<td>30:38</td>
<td>51:0</td>
</tr>
<tr>
<td>(Normal: Inflamed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site of Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD: Upper Disease</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>CD: Ileal</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>CD: Ileocolonic</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>CD: Colonic only</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Number of Biopsies</td>
<td>12:10</td>
<td>10:0</td>
</tr>
<tr>
<td>(Non-affected: affected)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Fresh Stool Samples</td>
<td>6:4</td>
<td>5:0</td>
</tr>
<tr>
<td>(Normal right colon: Inflamed right colon)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 $\Delta$Ct value compared to control subjects (Figure 2.3.1 F). The $\Delta$Ct 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 ΔCt 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 ΔCt 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 hypervariable 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.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Group</th>
<th>Colonoscopy visual appearance of right colon</th>
<th>Number of high quality reads</th>
<th>OTUs at 0.05</th>
<th>Chao-1 at 0.05</th>
<th>Number of Classified Taxa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>Control</td>
<td>Normal</td>
<td>12,444</td>
<td>299.9</td>
<td>428.3</td>
<td>7</td>
</tr>
<tr>
<td>Subject 2</td>
<td>Control</td>
<td>Normal</td>
<td>15,606</td>
<td>248.2</td>
<td>388.5</td>
<td>9</td>
</tr>
<tr>
<td>Subject 3</td>
<td>Control</td>
<td>Normal</td>
<td>5,953</td>
<td>21.3</td>
<td>22.3</td>
<td>3</td>
</tr>
<tr>
<td>Subject 4</td>
<td>Control</td>
<td>Normal</td>
<td>7,949</td>
<td>139.4</td>
<td>216.3</td>
<td>7</td>
</tr>
<tr>
<td>Subject 5</td>
<td>Control</td>
<td>Normal</td>
<td>14,269</td>
<td>146.3</td>
<td>192.6</td>
<td>6</td>
</tr>
<tr>
<td>Subject 6</td>
<td>Control</td>
<td>Normal</td>
<td>44,092</td>
<td>132.9</td>
<td>173.6</td>
<td>7</td>
</tr>
<tr>
<td>Subject 7</td>
<td>Control</td>
<td>Normal</td>
<td>15,566</td>
<td>146.9</td>
<td>202.3</td>
<td>5</td>
</tr>
<tr>
<td>Subject 8</td>
<td>Control</td>
<td>Normal</td>
<td>14,769</td>
<td>125.1</td>
<td>156.5</td>
<td>5</td>
</tr>
<tr>
<td>Subject 9</td>
<td>Control</td>
<td>Normal</td>
<td>19,747</td>
<td>216.2</td>
<td>288.4</td>
<td>7</td>
</tr>
<tr>
<td>Subject 10</td>
<td>Control</td>
<td>Normal</td>
<td>30,680</td>
<td>114</td>
<td>162.0</td>
<td>5</td>
</tr>
<tr>
<td>Subject 11</td>
<td>Control</td>
<td>Normal</td>
<td>2,919</td>
<td>64.1</td>
<td>142.2</td>
<td>6</td>
</tr>
<tr>
<td>Subject 12</td>
<td>Control</td>
<td>Normal</td>
<td>6,993</td>
<td>148.8</td>
<td>248.1</td>
<td>10</td>
</tr>
<tr>
<td>Subject 13</td>
<td>Control</td>
<td>Normal</td>
<td>28,212</td>
<td>285.2</td>
<td>425.9</td>
<td>11</td>
</tr>
<tr>
<td>Subject 14</td>
<td>CD</td>
<td>Normal</td>
<td>22,687</td>
<td>126.6</td>
<td>163.7</td>
<td>7</td>
</tr>
<tr>
<td>Subject 15</td>
<td>CD</td>
<td>Inflamed</td>
<td>5,713</td>
<td>54.4</td>
<td>71.5</td>
<td>6</td>
</tr>
<tr>
<td>Subject 16</td>
<td>CD</td>
<td>Normal</td>
<td>16,912</td>
<td>141.1</td>
<td>202.5</td>
<td>6</td>
</tr>
<tr>
<td>Subject 17</td>
<td>CD</td>
<td>Inflamed</td>
<td>3,816</td>
<td>54.7</td>
<td>87.5</td>
<td>6</td>
</tr>
<tr>
<td>Subject 18</td>
<td>CD</td>
<td>Inflamed</td>
<td>15,512</td>
<td>137</td>
<td>180.0</td>
<td>5</td>
</tr>
<tr>
<td>Subject 19</td>
<td>CD</td>
<td>Normal</td>
<td>21,122</td>
<td>144.8</td>
<td>205.7</td>
<td>7</td>
</tr>
<tr>
<td>Subject 20</td>
<td>CD</td>
<td>Normal</td>
<td>19,264</td>
<td>162.2</td>
<td>233.1</td>
<td>8</td>
</tr>
<tr>
<td>Subject 21</td>
<td>CD</td>
<td>Normal</td>
<td>5,092</td>
<td>55.3</td>
<td>84.2</td>
<td>6</td>
</tr>
<tr>
<td>Subject 22</td>
<td>CD</td>
<td>Inflamed</td>
<td>16,557</td>
<td>250.2</td>
<td>313.1</td>
<td>8</td>
</tr>
<tr>
<td>Subject 23</td>
<td>CD</td>
<td>Normal</td>
<td>14,178</td>
<td>82.7</td>
<td>97.9</td>
<td>8</td>
</tr>
<tr>
<td>Subject 24</td>
<td>CD</td>
<td>Inflamed</td>
<td>14,992</td>
<td>139</td>
<td>214.6</td>
<td>9</td>
</tr>
<tr>
<td>Subject 25</td>
<td>CD</td>
<td>Inflamed</td>
<td>4,742</td>
<td>90.6</td>
<td>144.7</td>
<td>8</td>
</tr>
<tr>
<td>Subject 26</td>
<td>CD</td>
<td>Normal</td>
<td>13,194</td>
<td>68.3</td>
<td>78.9</td>
<td>9</td>
</tr>
<tr>
<td>Subject 27</td>
<td>CD</td>
<td>Inflamed</td>
<td>10,260</td>
<td>148</td>
<td>213.3</td>
<td>7</td>
</tr>
<tr>
<td>Subject 28</td>
<td>CD</td>
<td>Normal</td>
<td>16,539</td>
<td>198.5</td>
<td>308.2</td>
<td>7</td>
</tr>
<tr>
<td>Subject 29</td>
<td>CD</td>
<td>Normal</td>
<td>20,768</td>
<td>157.5</td>
<td>229.7</td>
<td>8</td>
</tr>
<tr>
<td>Subject 30</td>
<td>CD</td>
<td>Normal</td>
<td>29,275</td>
<td>97.6</td>
<td>137.5</td>
<td>6</td>
</tr>
<tr>
<td>Subject 31</td>
<td>CD</td>
<td>Normal</td>
<td>28,242</td>
<td>68.4</td>
<td>72.9</td>
<td>8</td>
</tr>
<tr>
<td>Subject 32</td>
<td>CD</td>
<td>Normal</td>
<td>5,252</td>
<td>63.4</td>
<td>92.0</td>
<td>5</td>
</tr>
<tr>
<td>Subject 33</td>
<td>CD</td>
<td>Normal</td>
<td>19,054</td>
<td>102.5</td>
<td>148.7</td>
<td>7</td>
</tr>
</tbody>
</table>
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 (2.7x10^{-6}±9.8x10^{-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 ΔCt of targeted butyrate producers relative to total bacteria 16S rRNA. Each bar represents the mean ΔCt 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 ΔCt 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 NFκB [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\textsubscript{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. 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 *Coprococcus* 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 *Coprococcus*, 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 *Coprococcus* 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 Gever et al. if the inflammation in their cohort is only restricted to the terminal ileum and rectum or if 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/phophohydrolases, 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 Φ29 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.

1. Digest linear DNA with Plasmid-Safe™ ATP-Dependent DNase
2. In vitro insertion of oriV/KAN-2 transposon
3. Use captured plasmids to transform E. coli cells
4. Plate on LB agar containing kanamycin (50 μg/ml)
5. Inoculate each clone in 96-deep well sterile plates containing LB broth with 50 μg/ml kanamycin
   - Incubate at 37 °C for 8 hours with vigorous shaking
6. Dilute all clones from each sample 1:100 in 500 ml LB broth containing 50 μg/ml kanamycin
   - Incubate overnight at 37 °C with vigorous shaking
7. Purify plasmid using QIAGEN Plasmid Maxi Kit
8. Store remaining individual clones in 30% glycerol at -80 °C
9. Purify plasmid using QIAGEN Plasmid Maxi Kit
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 digestes 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.

<table>
<thead>
<tr>
<th>Name</th>
<th>Oligonucleotide Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1BCF1</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTATAGT</td>
</tr>
<tr>
<td>A1BCF2</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTAGGT</td>
</tr>
<tr>
<td>A1BCF3</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTTCAAT</td>
</tr>
<tr>
<td>A1BCF4</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTGTAT</td>
</tr>
<tr>
<td>A1BCF5</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCCTGCTT</td>
</tr>
<tr>
<td>A1BCF6</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTGTGT</td>
</tr>
<tr>
<td>A1BCF7</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTGCAGT</td>
</tr>
<tr>
<td>A1BCF8</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTGCTAT</td>
</tr>
<tr>
<td>A1BCF9</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTGTGT</td>
</tr>
<tr>
<td>A1BCF10</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTGTGT</td>
</tr>
<tr>
<td>A1BCF11</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTGTGT</td>
</tr>
<tr>
<td>A1BCF12</td>
<td>AACTCTTTCTCTTACACGACGCTCTTCCAGATCTGTGT</td>
</tr>
<tr>
<td>A1BCR1</td>
<td>CTATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR2</td>
<td>CCCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR3</td>
<td>TGAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR4</td>
<td>GATCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR5</td>
<td>GCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR6</td>
<td>ACAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR7</td>
<td>CTGCAATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR8</td>
<td>TACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR9</td>
<td>GTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR10</td>
<td>TCCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR11</td>
<td>ATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A1BCR12</td>
<td>AAGTATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF1</td>
<td>GATAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF2</td>
<td>GAGGAATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF3</td>
<td>ACTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF4</td>
<td>CTAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF5</td>
<td>CCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF6</td>
<td>TGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF7</td>
<td>GCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF8</td>
<td>ATGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF9</td>
<td>CAGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF10</td>
<td>AGATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF11</td>
<td>TACTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCF12</td>
<td>TTCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGGT</td>
</tr>
<tr>
<td>A2BCR1</td>
<td>CTGCGATCTCTCTGCTTACACGCGGTCTTCCAGATCTATT</td>
</tr>
<tr>
<td>A2BCR2</td>
<td>CTGCGATCTCTCTGCTTACACGCGGTCTTCCAGATCTATT</td>
</tr>
<tr>
<td>A2BCR3</td>
<td>CTGCGATCTCTCTGCTTACACGCGGTCTTCCAGATCTATT</td>
</tr>
<tr>
<td>A2BCR4</td>
<td>CTGCGATCTCTCTGCTTACACGCGGTCTTCCAGATCTATT</td>
</tr>
<tr>
<td>A2BCR5</td>
<td>CTGCGATCTCTCTGCTTACACGCGGTCTTCCAGATCTATT</td>
</tr>
<tr>
<td>A2BCR6</td>
<td>CTGCGATCTCTCTGCTTACACGCGGTCTTCCAGATCTATT</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>A2BCR7</td>
<td>CTCGGCATTCCTGCTGAACCGCTCTTTCCGATCTGTGCTT</td>
</tr>
<tr>
<td>A2BCR8</td>
<td>CTCGGCATTCCTGCTGAACCGCTCTTTCCGATCTGCACTT</td>
</tr>
<tr>
<td>A2BCR9</td>
<td>CTCGGCATTCCTGCTGAACCGCTCTTTCCGATCTCTGTT</td>
</tr>
<tr>
<td>A2BCR10</td>
<td>CTCGGCATTCCTGCTGAACCGCTCTTTCCGATCTATCTT</td>
</tr>
<tr>
<td>A2BCR11</td>
<td>CTCGGCATTCCTGCTGAACCGCTCTTTCCGATCTAGATT</td>
</tr>
<tr>
<td>A2BCR12</td>
<td>CTCGGCATTCCTGCTGAACCGCTCTTTCCGATCTGAAAT</td>
</tr>
<tr>
<td>PCRFWD1</td>
<td>AATGATACGGCGACCACCGAGATCTACACTCTCTCCCTACACGACGCTCTTTCCGATCTT</td>
</tr>
<tr>
<td>PCRRVS1</td>
<td>CAAGCAGAAGACGCCATACGAGATCGGTCTCGGCAATTCCCTGCTGAACCGCTCTTTCCGATCTT</td>
</tr>
</tbody>
</table>
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 \(1 \times 10^{-5}\) and Bit-score of 35 using the skimmer.pl script (http://alrlab.research.pdx.edu). Next, only the best hit for each read was selected using the blast_best.pl script (http://alrlab.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://alrlab.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 I). 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.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>UC</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong></td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><strong>Age at Colonoscopy (years)</strong></td>
<td>11.75±4.2</td>
<td>14.75±2.2</td>
<td>13.25±1.3</td>
</tr>
<tr>
<td><strong>Gender (M : F)</strong></td>
<td>2:2</td>
<td>4:0</td>
<td>4:0</td>
</tr>
<tr>
<td><strong>Disease Severity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Moderate : Severe)</td>
<td>NA</td>
<td>3:1</td>
<td>0:4</td>
</tr>
<tr>
<td><strong>Macroscopic Mucosal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appearance of Right Colon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Normal : Inflamed)</td>
<td>4:0</td>
<td>4:0</td>
<td>3:1</td>
</tr>
<tr>
<td><strong>Site of Disease</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ileal : Ileocolonic CD</td>
<td>NA</td>
<td>NA</td>
<td>2:2</td>
</tr>
<tr>
<td>Colon Extensive : Pancolitis</td>
<td>NA</td>
<td>3:1</td>
<td>NA</td>
</tr>
</tbody>
</table>
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% ± 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.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of raw reads</th>
<th>Number of reads post <em>E. coli</em>/transposon filtration step</th>
<th>Remaining reads after quality filtering</th>
<th>% Reads filtered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control 1</td>
<td>28,747,012</td>
<td>11,675,344</td>
<td>8,902,621</td>
<td>69.03</td>
</tr>
<tr>
<td>Control 2</td>
<td>30,952,670</td>
<td>13,276,520</td>
<td>10,575,089</td>
<td>65.83</td>
</tr>
<tr>
<td>Control 3</td>
<td>32,413,858</td>
<td>12,409,791</td>
<td>9,941,283</td>
<td>69.33</td>
</tr>
<tr>
<td>Control 4</td>
<td>30,448,836</td>
<td>13,595,683</td>
<td>11,277,616</td>
<td>62.96</td>
</tr>
<tr>
<td>UC 1</td>
<td>23,591,962</td>
<td>10,679,412</td>
<td>8,925,847</td>
<td>62.17</td>
</tr>
<tr>
<td>UC 2</td>
<td>24,085,724</td>
<td>11,031,841</td>
<td>9,259,280</td>
<td>61.56</td>
</tr>
<tr>
<td>UC 3</td>
<td>25,208,470</td>
<td>11,552,456</td>
<td>9,717,539</td>
<td>61.45</td>
</tr>
<tr>
<td>UC 4</td>
<td>25,580,554</td>
<td>10,092,409</td>
<td>8,170,620</td>
<td>68.06</td>
</tr>
<tr>
<td>CD 1</td>
<td>28,814,416</td>
<td>12,206,395</td>
<td>10,461,650</td>
<td>63.69</td>
</tr>
<tr>
<td>CD 2</td>
<td>31,407,352</td>
<td>13,286,368</td>
<td>10,243,112</td>
<td>67.39</td>
</tr>
<tr>
<td>CD 3</td>
<td>26,706,410</td>
<td>11,508,233</td>
<td>9,472,077</td>
<td>64.53</td>
</tr>
<tr>
<td>CD 4</td>
<td>28,792,622</td>
<td>12,521,162</td>
<td>9,906,293</td>
<td>65.59</td>
</tr>
</tbody>
</table>
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% ± 2.35) with a minor representation of Eukaryota (5.3% ± 1.1), Archea (0.0006% ± 0.0002), Viruses (0.32% ± 0.16), unclassified sequences (0.12% ± 0.04) and other sequences (22.15% ± 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% ± 3.7), Firmicutes (20.34% ± 4.8), Bacteroidetes (12.05% ± 8.9), Spirochaetes (1.26% ± 0.21) and Actinobacteria (0.96% ± 0.27) in addition to environmental bacteria (14.13% ± 3.9) and unclassified bacteria (1.33% ± 0.39) (Figure 3.3.2 B). At the genus level, the plasmidomes mainly originated from *Escherichia* (25.3% ± 3.4), *Bacteroides* (12.1% ± 9.5), *Klebsiella* (6.5% ± 1.1), *Tyzzerella* (5.7% ± 2.3) and *Salmonella* (5.5% ± 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% ± 19.7),
Firmicutes (37.2% ± 15.6) and, to a lesser extent, by Proteobacteria (9.27% ± 13.3)
and Actinobacteria (2.47% ± 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% ± 11.4), *Faecalibacterium* (5.04% ± 6.9), *Ruminococcus* (4.8% ±
3.1), *Parabacteroides* (4.3% ± 5.6), and *Dorea* (3.4% ± 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 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% ± 0.8 of the total assigned functions. These categories were lipid metabolism (43.04% ± 6.1), amino acid metabolism and transport (25.9% ± 8.2), translation (19.8% ± 3.5) and nucleotide metabolism and transport (10.24% ± 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% ± 0.21), carbohydrate transport and metabolism (9.43% ± 1.04), transcription (8.6% ± 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 rarefication 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 $1 \times 10^{-5}$ 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 ($P<0.05$). Differential 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.

<table>
<thead>
<tr>
<th>ID</th>
<th>Annotation</th>
<th>(Number of genes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3-hydroxyacyl-CoA dehydrogenase</td>
<td>(1)</td>
</tr>
<tr>
<td>B</td>
<td>Acetyl-CoA acetyltransferase</td>
<td>(4)</td>
</tr>
<tr>
<td>C</td>
<td>Arabinose operon regulatory protein</td>
<td>(3)</td>
</tr>
<tr>
<td>D</td>
<td>DNA-damage repair protein</td>
<td>(3)</td>
</tr>
<tr>
<td>E</td>
<td>Hypothetical protein</td>
<td>(9)</td>
</tr>
<tr>
<td>F</td>
<td>Phage proteins</td>
<td>(4)</td>
</tr>
<tr>
<td>A</td>
<td>4-hydroxyphenylacetate 3-monooxygenase</td>
<td>(1)</td>
</tr>
<tr>
<td>B</td>
<td>Acetyl-CoA acetyltransferase</td>
<td>(1)</td>
</tr>
<tr>
<td>C</td>
<td>DNA-damage repair protein</td>
<td>(1)</td>
</tr>
<tr>
<td>D</td>
<td>Hypothetical protein</td>
<td>(3)</td>
</tr>
<tr>
<td>E</td>
<td>Phage proteins</td>
<td>(2)</td>
</tr>
</tbody>
</table>
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\times10^{-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 ($P<0.05$). Differential 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% ± 2.35). Three bacterial phyla represented up to 80% of the bacterial plasmidome hosts: Proteobacteria (49.84% ± 3.7), Firmicutes (20.34% ± 4.8) and Bacteroidetes (12.05% ± 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% ± 19.7) followed by Firmicutes (37.2% ± 15.6) and then Proteobacteria (9.27% ± 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 at 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 ($H_2S$) 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$_2$S 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$_2$S by sulphate-reducing bacteria (SRB), which could result in H$_2$S 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 produces, 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 inintestinal 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


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


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.

<table>
<thead>
<tr>
<th>Sample</th>
<th>HMC ID</th>
<th>Gender</th>
<th>Age</th>
<th>Group</th>
<th>Colonoscopy visual appearance of right colon</th>
<th>Paris Classification</th>
<th>Category</th>
<th>PCDAI/PUCAI</th>
<th>Severity</th>
<th>BCoAT Sequencing</th>
<th>16S rRNA sequencing</th>
<th>Mucosal wash Used for qPCR</th>
<th>Biopsies used for qPCR</th>
<th>Stool used for qPCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>18</td>
<td>M</td>
<td>13</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject 2</td>
<td>28</td>
<td>M</td>
<td>13</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject 3</td>
<td>43</td>
<td>F</td>
<td>16</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject 4</td>
<td>55</td>
<td>M</td>
<td>6</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject 5</td>
<td>59</td>
<td>M</td>
<td>14</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject 6</td>
<td>67</td>
<td>F</td>
<td>17</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject 7</td>
<td>87</td>
<td>F</td>
<td>18</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject 8</td>
<td>98</td>
<td>F</td>
<td>16</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject 9</td>
<td>208</td>
<td>F</td>
<td>16</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject 10</td>
<td>221</td>
<td>F</td>
<td>9</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Subject 11</td>
<td>222</td>
<td>M</td>
<td>10</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Subject 12</td>
<td>232</td>
<td>F</td>
<td>15</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Subject 13</td>
<td>235</td>
<td>M</td>
<td>11</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Subject 14</td>
<td>14</td>
<td>F</td>
<td>14</td>
<td>CD</td>
<td>Normal</td>
<td>A1b, L3, L4a, B1, G1</td>
<td>Reocolonic</td>
<td>37.5</td>
<td>Moderate</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subject</td>
<td>Age</td>
<td>Gender</td>
<td>Diagnosis</td>
<td>CD Extent</td>
<td>Inflammation</td>
<td>Healing</td>
<td>Surgery</td>
<td>Remission</td>
<td>Duodenal Biopsy</td>
<td>Ascending Biopsy</td>
<td>Terminal Biopsy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>--------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
<td>-----------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>17</td>
<td>M</td>
<td>Inflamed</td>
<td>Colonic</td>
<td>57.5</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>22</td>
<td>M</td>
<td>Normal</td>
<td>Ileal</td>
<td>37.5</td>
<td>Moderate</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>29</td>
<td>M</td>
<td>Inflamed</td>
<td>Ileocolonic</td>
<td>32.5</td>
<td>Moderate</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>38</td>
<td>F</td>
<td>Normal</td>
<td>Colonic</td>
<td>20</td>
<td>Mild</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>39</td>
<td>F</td>
<td>Normal</td>
<td>Ileal</td>
<td>52.5</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>41</td>
<td>F</td>
<td>Inflamed</td>
<td>Ileocolonic</td>
<td>60</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>44</td>
<td>F</td>
<td>Normal</td>
<td>Ileal</td>
<td>57.5</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>61</td>
<td>M</td>
<td>Inflamed</td>
<td>Colonic</td>
<td>32.5</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>63</td>
<td>M</td>
<td>Normal</td>
<td>Ileal</td>
<td>65</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>72</td>
<td>M</td>
<td>Inflamed</td>
<td>Ileocolonic</td>
<td>55</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>75</td>
<td>M</td>
<td>Inflamed</td>
<td>Colonic</td>
<td>50</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>78</td>
<td>F</td>
<td>Normal</td>
<td>Ileal</td>
<td>45</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>79</td>
<td>M</td>
<td>Inflamed</td>
<td>Ileocolonic</td>
<td>45</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>85</td>
<td>M</td>
<td>Normal</td>
<td>Ileal</td>
<td>62.5</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>86</td>
<td>M</td>
<td>Normal</td>
<td>Ileal</td>
<td>22.5</td>
<td>Mild</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>94</td>
<td>M</td>
<td>Normal</td>
<td>Ileal</td>
<td>45</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>97</td>
<td>M</td>
<td>Normal</td>
<td>Ileal</td>
<td>40</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>105</td>
<td>M</td>
<td>Normal</td>
<td>Ileal</td>
<td>20</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>106</td>
<td>F</td>
<td>Normal</td>
<td>Ileal</td>
<td>45</td>
<td>Severe</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>109</td>
<td>F</td>
<td>Control</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>110</td>
<td>F</td>
<td>Control</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>112</td>
<td>M</td>
<td>Control</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Age</td>
<td>Gender</td>
<td>Group</td>
<td>Condition</td>
<td>Test 1</td>
<td>Test 2</td>
<td>Test 3</td>
<td>Test 4</td>
<td>Test 5</td>
<td>Test 6</td>
<td>Test 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>--------</td>
<td>---------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>17</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>14</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>16</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>13</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>15</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>14</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>15</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>17</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>15</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>13</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>16</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>17</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>16</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>5</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>17</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>17</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>16</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>15</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>16</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>10</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>16</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>15</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Age</td>
<td>Sex</td>
<td>Diagnosis</td>
<td>Localization</td>
<td>VA</td>
<td>RA</td>
<td>IA</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>-----------</td>
<td>--------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>225</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>241</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>321</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>324</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>368</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>371</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>376</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>381</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>387</td>
<td>M</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>403</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>397</td>
<td>F</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>217</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>A1b, L1, L4b, B1, G0</td>
<td>Ileal</td>
<td>55</td>
<td>Severe</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>231</td>
<td>F</td>
<td>CD</td>
<td>Normal</td>
<td>A1b, L1, G0</td>
<td>Ileal</td>
<td>37.5</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>49</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>A1b, L1, B1, G1</td>
<td>Ileal</td>
<td>40</td>
<td>Severe</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>266</td>
<td>F</td>
<td>CD</td>
<td>Normal</td>
<td>A1a, L2, B1, G1</td>
<td>Colonic</td>
<td>27.5</td>
<td>Inactive</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>269</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>A1a, L4a, B1, G0</td>
<td>Upper disease</td>
<td>2.5</td>
<td>Inactive</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>278</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>A1a, L1, B1, G0</td>
<td>Ileal</td>
<td>10</td>
<td>Mild</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>298</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>A1b, L1, B1, G0</td>
<td>Ileal</td>
<td>52.5</td>
<td>Severe</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>267</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>A1a, L3, B1, G0</td>
<td>Ileocolonic</td>
<td>10</td>
<td>Mild</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>256</td>
<td>F</td>
<td>CD</td>
<td>Normal</td>
<td>A1b, L2, B1, G0</td>
<td>Colonic</td>
<td>30</td>
<td>Moderate</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>209</td>
<td>F</td>
<td>CD</td>
<td>Normal</td>
<td>A1b, L1, L4a, B1, G0</td>
<td>Ileal</td>
<td>35</td>
<td>Moderate</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>240</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>A1b, L1, B1, G0</td>
<td>Ileal</td>
<td>65</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Age</td>
<td>Gender</td>
<td>Disease</td>
<td>Location</td>
<td>Stage</td>
<td>Pain</td>
<td>Anorexia</td>
<td>diarrhea</td>
<td>weight loss</td>
<td>答题</td>
<td>Fever</td>
<td>Anemia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
<td>-------</td>
<td>------</td>
<td>----------</td>
<td>-----------</td>
<td>-------------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>319</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>ileal</td>
<td>45</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>268</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>ileal</td>
<td>50</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>65</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>ileal</td>
<td>50</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>261</td>
<td>F</td>
<td>CD</td>
<td>Normal</td>
<td>ileal</td>
<td>32.5</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>262</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>ileal</td>
<td>27.5</td>
<td>Mild</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>374</td>
<td>M</td>
<td>CD</td>
<td>Normal</td>
<td>ileal</td>
<td>37.5</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>201</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>45</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>203</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>45</td>
<td>Severe</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>206</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>45</td>
<td>Severe</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>219</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>37.5</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>223</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>45</td>
<td>Severe</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>248</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>42.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>47</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>65</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>62</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>57.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>90</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>52.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>311</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>32.5</td>
<td>Severe</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>253</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>40</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>239</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>27.5</td>
<td>Mild</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>295</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>25</td>
<td>Mild</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>316</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>27.5</td>
<td>Mild</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>251</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>47.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>280</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>ileal</td>
<td>45</td>
<td>Severe</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Age</td>
<td>Sex</td>
<td>Diagnosis</td>
<td>Disease Location</td>
<td>Course</td>
<td>Steroid</td>
<td>Chemotherapy</td>
<td>Anti-TNF</td>
<td>Immune Mod</td>
<td>Antibiotics</td>
<td>Trasjaco</td>
<td>Cortico</td>
<td>Other Therapies</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>-----------</td>
<td>------------------</td>
<td>--------</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>------------</td>
<td>-------------</td>
<td>----------</td>
<td>---------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>13</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, B1, G0</td>
<td>Ileocolonic</td>
<td>52.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>14</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, L4a, B1, G0</td>
<td>Ileocolonic</td>
<td>60</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>15</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, B1P, G0</td>
<td>Ileocolonic</td>
<td>67.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>12</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, B1P, G1</td>
<td>Ileocolonic</td>
<td>65</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>12</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, L4a, B1, G0</td>
<td>Ileocolonic</td>
<td>77.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>15</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L1, B2, G0</td>
<td>Ileal</td>
<td>27.5</td>
<td>Mild</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>10</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L2, L4a, B1, G0</td>
<td>Colonic</td>
<td>57.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>12</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L2, B1P, G0</td>
<td>Colonic</td>
<td>47.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>16</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, B1, G0</td>
<td>Ileocolonic</td>
<td>55</td>
<td>Severe</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>14</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, B1, G0</td>
<td>Ileocolonic</td>
<td>47.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>12</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, B1p, G0</td>
<td>Ileocolonic</td>
<td>52.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>16</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, B1, G0</td>
<td>Ileocolonic</td>
<td>57.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>17</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, L4a, B1, G0</td>
<td>Ileocolonic</td>
<td>57.5</td>
<td>Severe</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>10</td>
<td>F</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L1, B1p, G0</td>
<td>Ileal</td>
<td>27.5</td>
<td>Mild</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>16</td>
<td>M</td>
<td>CD</td>
<td>Inflamed</td>
<td>A1b, L3, B1, G0</td>
<td>Ileocolonic</td>
<td>32.5</td>
<td>Moderate</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of isolated clones</th>
<th>Sample</th>
<th>Number of isolated clones</th>
<th>Sample</th>
<th>Number of isolated clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control 1</td>
<td>362</td>
<td>UC 1</td>
<td>278</td>
<td>CD 1</td>
<td>196</td>
</tr>
<tr>
<td>Control 2</td>
<td>489</td>
<td>UC 2</td>
<td>250</td>
<td>CD 2</td>
<td>1,644</td>
</tr>
<tr>
<td>Control 3</td>
<td>353</td>
<td>UC 3</td>
<td>1,595</td>
<td>CD 3</td>
<td>374</td>
</tr>
<tr>
<td>Control 4</td>
<td>1,296</td>
<td>UC 4</td>
<td>377</td>
<td>CD 4</td>
<td>286</td>
</tr>
<tr>
<td>Total Clones</td>
<td>2,500</td>
<td>Total Clones</td>
<td>2,500</td>
<td>Total Clones</td>
<td>2,500</td>
</tr>
</tbody>
</table>
Appendix IV: Clinical data of participating subject in the study. Table also illustrates which subjects were used for 16S rRNA sequencing.

<table>
<thead>
<tr>
<th>Sample</th>
<th>HMC ID</th>
<th>Gender</th>
<th>Age</th>
<th>Group</th>
<th>Colonoscopy visual appearance of right colon</th>
<th>Paris Classification</th>
<th>Category</th>
<th>PCDAI/PUCAI</th>
<th>Severity</th>
<th>16S rRNA sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control 1</td>
<td>20</td>
<td>F</td>
<td>12</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td>Control 2</td>
<td>28</td>
<td>M</td>
<td>13</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>Control 3</td>
<td>55</td>
<td>M</td>
<td>6</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>Control 4</td>
<td>43</td>
<td>F</td>
<td>16</td>
<td>Control</td>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>UC 1</td>
<td>77</td>
<td>M</td>
<td>17</td>
<td>UC</td>
<td>Normal</td>
<td>E3S0</td>
<td>Extensive (hepatic flexure distally); never severe</td>
<td>50</td>
<td>Moderate</td>
<td>Yes</td>
</tr>
<tr>
<td>UC 2</td>
<td>92</td>
<td>M</td>
<td>12</td>
<td>UC</td>
<td>Normal</td>
<td>E3S1</td>
<td>Extensive (hepatic flexure distally); ever severe</td>
<td>75</td>
<td>Severe</td>
<td>Yes</td>
</tr>
<tr>
<td>UC 3</td>
<td>23</td>
<td>M</td>
<td>14</td>
<td>UC</td>
<td>Normal</td>
<td>E4S1</td>
<td>Pancolitis (proximal to hepatic flexure); ever severe</td>
<td>50</td>
<td>Moderate</td>
<td>Yes</td>
</tr>
<tr>
<td>UC 4</td>
<td>24</td>
<td>M</td>
<td>16</td>
<td>UC</td>
<td>Normal</td>
<td>E3S1</td>
<td>Extensive (hepatic flexure distally); ever severe</td>
<td>40</td>
<td>Moderate</td>
<td>Yes</td>
</tr>
<tr>
<td>CD 1</td>
<td>90</td>
<td>M</td>
<td>15</td>
<td>CD</td>
<td>Inflamed</td>
<td>Ab1 L3L4a B1 G0</td>
<td>ileocolonic</td>
<td>52.5</td>
<td>Severe</td>
<td>Yes</td>
</tr>
<tr>
<td>CD 2</td>
<td>97</td>
<td>M</td>
<td>13</td>
<td>CD</td>
<td>Normal</td>
<td>A1b L3L4a B1 G0</td>
<td>ileocolonic</td>
<td>40</td>
<td>Severe</td>
<td>No</td>
</tr>
<tr>
<td>CD 3</td>
<td>49</td>
<td>M</td>
<td>12</td>
<td>CD</td>
<td>Normal</td>
<td>A1b L1 B1 G1</td>
<td>ileal</td>
<td>40</td>
<td>Severe</td>
<td>Yes</td>
</tr>
<tr>
<td>CD 4</td>
<td>63</td>
<td>M</td>
<td>13</td>
<td>CD</td>
<td>Normal</td>
<td>A1b L1L4a B1 G1</td>
<td>ileal with upper disease</td>
<td>65</td>
<td>Severe</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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 $1 \times 10^{-5}$, 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.

<p>| 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|AAA26475.1| Control | 0.0481 | 1 | 0.096 | 0.062 |
| gi|169655589|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|AAH79196.1|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|AAC32255.1| Control | 0.0481 | 1 | 0.281 | 0.014 |
| gi|220683460|gb|ALC80551.1| 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 |</p>
<table>
<thead>
<tr>
<th>gi</th>
<th>Description</th>
<th>Control 1</th>
<th>Control 2</th>
<th>Control 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>291471672</td>
<td>gb</td>
<td>EFF14155.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>292809761</td>
<td>gb</td>
<td>EFF68966.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>295115075</td>
<td>emb</td>
<td>CBL35922.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>309400374</td>
<td>gb</td>
<td>AD079610.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>315632435</td>
<td>gb</td>
<td>ADU53763.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>303752427</td>
<td>emb</td>
<td>CBL87378.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>333972755</td>
<td>gb</td>
<td>AEG39560.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>375058382</td>
<td>dbj</td>
<td>BAL60567.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>37693755</td>
<td>gb</td>
<td>AAQ98890.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>378939967</td>
<td>gb</td>
<td>AFC75523.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>3808121</td>
<td>emb</td>
<td>CAA09780.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>383842778</td>
<td>gb</td>
<td>AFH53881.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>3953663</td>
<td>dbj</td>
<td>BAA34762.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>402950453</td>
<td>gb</td>
<td>EJX68456.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>403017426</td>
<td>gb</td>
<td>EJY30174.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>4210487</td>
<td>gb</td>
<td>AAD11974.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>43739</td>
<td>emb</td>
<td>CAAl23889.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>446100002</td>
<td>ref</td>
<td>WP_000177857.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>446100092</td>
<td>ref</td>
<td>WP_000177947.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>446100096</td>
<td>ref</td>
<td>WP_000177951.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>446100119</td>
<td>ref</td>
<td>WP_000177974.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>446832557</td>
<td>ref</td>
<td>WP_000909813.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>447057677</td>
<td>ref</td>
<td>WP_001134933.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>451928584</td>
<td>pdb</td>
<td>3ZDJ</td>
<td>A</td>
</tr>
<tr>
<td>gi</td>
<td>46277646</td>
<td>gb</td>
<td>AAS87025.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>46850195</td>
<td>gb</td>
<td>AAT02530.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>476116409</td>
<td>gb</td>
<td>EMV39178.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>477137054</td>
<td>gb</td>
<td>ENE15873.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>481041386</td>
<td>ref</td>
<td>WP_001295762.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>485746862</td>
<td>ref</td>
<td>WP_001373242.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>485793337</td>
<td>ref</td>
<td>WP_001415026.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>485806691</td>
<td>ref</td>
<td>WP_001426975.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>485860877</td>
<td>ref</td>
<td>WP_001460269.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>485871440</td>
<td>ref</td>
<td>WP_001468384.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>487285635</td>
<td>ref</td>
<td>WP_001632456.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>487405847</td>
<td>ref</td>
<td>WP_001674007.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>487564528</td>
<td>ref</td>
<td>WP_001717547.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>487661010</td>
<td>ref</td>
<td>WP_001751343.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>487853220</td>
<td>ref</td>
<td>WP_001926686.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>487907108</td>
<td>ref</td>
<td>WP_001980574.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>487910581</td>
<td>ref</td>
<td>WP_001984047.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>488180921</td>
<td>ref</td>
<td>WP_002252129.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>488182696</td>
<td>ref</td>
<td>WP_002253904.1</td>
<td>Control</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>p-value</td>
<td>FDR</td>
<td>Control p-value</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>---------</td>
<td>-----</td>
<td>----------------</td>
</tr>
<tr>
<td>gi</td>
<td>488279268</td>
<td>ref</td>
<td>WP_002350476.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>489810083</td>
<td>ref</td>
<td>WP_003713939.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>490540999</td>
<td>ref</td>
<td>WP_004406128.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>490714494</td>
<td>ref</td>
<td>WP_004577211.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>490987317</td>
<td>ref</td>
<td>WP_004849050.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>490987753</td>
<td>ref</td>
<td>WP_004849486.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>491160894</td>
<td>ref</td>
<td>WP_005019281.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>491162098</td>
<td>ref</td>
<td>WP_005020484.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>491266388</td>
<td>ref</td>
<td>WP_005124526.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>492292168</td>
<td>ref</td>
<td>WP_005800676.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>492677637</td>
<td>ref</td>
<td>WP_005920083.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>493329074</td>
<td>ref</td>
<td>WP_006286292.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>494817674</td>
<td>ref</td>
<td>WP_007553082.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>495674769</td>
<td>ref</td>
<td>WP_008399348.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>495674836</td>
<td>ref</td>
<td>WP_008399415.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>498572590</td>
<td>ref</td>
<td>WP_010835044.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>499964299</td>
<td>ref</td>
<td>WP_011645017.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>500821251</td>
<td>ref</td>
<td>WP_012008451.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>501083106</td>
<td>ref</td>
<td>WP_012133644.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>502533835</td>
<td>ref</td>
<td>WP_012814673.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>503044338</td>
<td>ref</td>
<td>WP_013279314.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>503320404</td>
<td>ref</td>
<td>WP_013555065.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>505348778</td>
<td>ref</td>
<td>WP_015535880.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>505381927</td>
<td>ref</td>
<td>WP_015569029.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>507590704</td>
<td>gb</td>
<td>EOR83978.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>510927756</td>
<td>ref</td>
<td>WP_016248400.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>514663242</td>
<td>gb</td>
<td>EPJ19297.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>519784041</td>
<td>gb</td>
<td>EPP30712.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>523614619</td>
<td>ref</td>
<td>WP_020762354.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>531494529</td>
<td>gb</td>
<td>EQH80598.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>531556724</td>
<td>gb</td>
<td>EQI39009.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>531626711</td>
<td>gb</td>
<td>EQJ05633.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>534787977</td>
<td>gb</td>
<td>EQP40445.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>536846312</td>
<td>gb</td>
<td>ERC29327.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>536879613</td>
<td>gb</td>
<td>ERC61643.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>53756912</td>
<td>gb</td>
<td>AAU91203.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>544967051</td>
<td>ref</td>
<td>WP_021370747.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>544980124</td>
<td>ref</td>
<td>WP_021380190.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>545048784</td>
<td>ref</td>
<td>WP_021423038.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>545296843</td>
<td>ref</td>
<td>WP_021577058.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>546734476</td>
<td>ref</td>
<td>WP_021888857.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>547220975</td>
<td>ref</td>
<td>WP_021960263.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>547313703</td>
<td>ref</td>
<td>WP_022045363.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td></td>
<td>ref</td>
<td>Control</td>
<td>0.0481</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>gi</td>
<td>547811649</td>
<td>ref</td>
<td>WP_022220766.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>547941520</td>
<td>ref</td>
<td>WP_022342824.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>548203089</td>
<td>ref</td>
<td>WP_022423198.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>548226340</td>
<td>ref</td>
<td>WP_022445215.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>548329222</td>
<td>ref</td>
<td>WP_022514978.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>548374420</td>
<td>ref</td>
<td>WP_022518728.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>552996</td>
<td>gb</td>
<td>AAA32208.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>553928417</td>
<td>ref</td>
<td>WP_023140126.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>555421419</td>
<td>ref</td>
<td>WP_023263970.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>564587202</td>
<td>gb</td>
<td>AHB86974.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>564909311</td>
<td>gb</td>
<td>ETE09584.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>568816383</td>
<td>gb</td>
<td>AHE38449.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>571577779</td>
<td>dbj</td>
<td>BAD83862.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>57157820</td>
<td>dbj</td>
<td>BAD83864.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>572727913</td>
<td>gb</td>
<td>ETS26555.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>575535143</td>
<td>gb</td>
<td>ETX35607.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>585339574</td>
<td>ref</td>
<td>WP_024221205.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>585348282</td>
<td>ref</td>
<td>WP_024229899.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>596222225</td>
<td>gb</td>
<td>EYA67038.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>596233172</td>
<td>gb</td>
<td>EYA76433.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>599342835</td>
<td>gb</td>
<td>AHN62075.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>603578</td>
<td>emb</td>
<td>CAA58446.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>606748092</td>
<td>gb</td>
<td>EYX90533.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>607741474</td>
<td>gb</td>
<td>EZE23258.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>635649404</td>
<td>ref</td>
<td>WP_024296277.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>642975646</td>
<td>ref</td>
<td>WP_025188709.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>643562994</td>
<td>ref</td>
<td>WP_025232309.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>6457346</td>
<td>gb</td>
<td>AAF09485.1</td>
<td>AF191496_1</td>
</tr>
<tr>
<td>gi</td>
<td>654789027</td>
<td>ref</td>
<td>WP_028423020.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>660355405</td>
<td>gb</td>
<td>KEO42207.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>667745630</td>
<td>gb</td>
<td>KFA83936.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>671561717</td>
<td>ref</td>
<td>WP_031542377.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>671564761</td>
<td>ref</td>
<td>WP_031545132.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>690410436</td>
<td>ref</td>
<td>WP_031942887.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>691347822</td>
<td>ref</td>
<td>WP_032072697.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>692193070</td>
<td>ref</td>
<td>WP_032106118.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>692983954</td>
<td>ref</td>
<td>WP_032173612.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>693003576</td>
<td>ref</td>
<td>WP_032189073.1</td>
<td>CD</td>
</tr>
<tr>
<td>gi</td>
<td>693025390</td>
<td>ref</td>
<td>WP_032202178.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>693113791</td>
<td>ref</td>
<td>WP_032264873.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>693120738</td>
<td>ref</td>
<td>WP_032269776.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>693123444</td>
<td>ref</td>
<td>WP_032271733.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td>693199697</td>
<td>ref</td>
<td>WP_032320913.1</td>
<td>Control</td>
</tr>
<tr>
<td>gi</td>
<td></td>
<td>ref</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
<td>-----</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>693210867</td>
<td>ref</td>
<td>WP_032329776.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>694087296</td>
<td>ref</td>
<td>WP_032432063.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>695263285</td>
<td>ref</td>
<td>WP_032490155.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>695269199</td>
<td>ref</td>
<td>WP_032491376.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>695317687</td>
<td>ref</td>
<td>WP_032516898.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>695319432</td>
<td>ref</td>
<td>WP_032518622.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>695323975</td>
<td>ref</td>
<td>WP_032522284.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>695348632</td>
<td>ref</td>
<td>WP_032545056.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>695401614</td>
<td>ref</td>
<td>WP_032556970.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>695539704</td>
<td>ref</td>
<td>WP_032596024.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>71842885</td>
<td>gb</td>
<td>AAZ49080.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>71842945</td>
<td>gb</td>
<td>AAZ49135.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>727108235</td>
<td>ref</td>
<td>WP_033623324.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>727367979</td>
<td>ref</td>
<td>WP_033801914.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>727372691</td>
<td>ref</td>
<td>WP_033802872.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>727386467</td>
<td>ref</td>
<td>WP_033805340.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>729796164</td>
<td>gb</td>
<td>KHG91472.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>729951767</td>
<td>gb</td>
<td>KHI42278.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>737108003</td>
<td>ref</td>
<td>WP_035095871.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>738288061</td>
<td>ref</td>
<td>WP_036241831.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>739010638</td>
<td>ref</td>
<td>WP_036892165.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>739495614</td>
<td>ref</td>
<td>WP_037355341.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>746822997</td>
<td>gb</td>
<td>AJD80684.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>749549041</td>
<td>ref</td>
<td>WP_040182562.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>749939258</td>
<td>ref</td>
<td>WP_040324867.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>971143</td>
<td>dbj</td>
<td>BAA09901.1</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>992929</td>
<td>dbj</td>
<td>BAA06208.1</td>
</tr>
</tbody>
</table>
Appendix VI: 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 $1 \times 10^{-5}$, 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.

<table>
<thead>
<tr>
<th>HMP-GIT ID</th>
<th>Control</th>
<th>IBD</th>
<th>Control</th>
<th>UC</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANASTE_01800</td>
<td>0.0481</td>
<td>1</td>
<td>0.021</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>BACPEC_02943</td>
<td>0.0481</td>
<td>1</td>
<td>0.021</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>BUTYVIB_00219</td>
<td>0.0481</td>
<td>1</td>
<td>0.021</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>CATMIT_00867</td>
<td>0.0481</td>
<td>1</td>
<td>0.023</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>CLOM621_09087</td>
<td>0.0481</td>
<td>1</td>
<td>0.075</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td>CLOSTASPAR_00404</td>
<td>0.0481</td>
<td>1</td>
<td>0.08</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td>COLAER_01671</td>
<td>0.0481</td>
<td>1</td>
<td>0.281</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>COPEUT_01255</td>
<td>0.0481</td>
<td>1</td>
<td>0.141</td>
<td>0.039</td>
<td></td>
</tr>
<tr>
<td>ERE_22580</td>
<td>0.0481</td>
<td>1</td>
<td>0.114</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td>EUBIFOR_01208</td>
<td>0.0041</td>
<td>1</td>
<td>0.016</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>FAEPRAM212_00170</td>
<td>0.0081</td>
<td>1</td>
<td>0.039</td>
<td>0.019</td>
<td></td>
</tr>
<tr>
<td>HMPREF0101_00539</td>
<td>0.0081</td>
<td>1</td>
<td>0.018</td>
<td>0.038</td>
<td></td>
</tr>
<tr>
<td>HMPREF0105_4636</td>
<td>0.5701</td>
<td>0.062</td>
<td>0.011</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HMPREF0866_03206</td>
<td>0.0481</td>
<td>1</td>
<td>0.023</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>HMPREF1416_00108</td>
<td>0.0481</td>
<td>1</td>
<td>0.327</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>HMPREF1431_00347</td>
<td>0.6831</td>
<td>0.096</td>
<td>0.014</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HMPREF1431_00348</td>
<td>0.0081</td>
<td>1</td>
<td>0.17</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>HMPREF1444_00159</td>
<td>0.0481</td>
<td>1</td>
<td>0.116</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>HMPREF1452_00433</td>
<td>0.0281</td>
<td>1</td>
<td>0.044</td>
<td>0.095</td>
<td></td>
</tr>
<tr>
<td>HMPREF9345_01139</td>
<td>0.0481</td>
<td>1</td>
<td>0.078</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>HMPREF9356_03350</td>
<td>0.1091</td>
<td>0.004</td>
<td>0.016</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HMPREF9541_03330</td>
<td>0.0481</td>
<td>1</td>
<td>0.023</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>HMPREF9547_02737</td>
<td>0.0481</td>
<td>1</td>
<td>0.078</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>HMPREF9552_04932</td>
<td>0.0481</td>
<td>1</td>
<td>0.023</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>RO1_13080</td>
<td>0.0481</td>
<td>1</td>
<td>0.021</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>ROSINTL182_05038</td>
<td>0.0481</td>
<td>1</td>
<td>0.021</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>RUMLAC_00163</td>
<td>0.0081</td>
<td>1</td>
<td>0.042</td>
<td>0.011</td>
<td></td>
</tr>
</tbody>
</table>
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 $1 \times 10^{-5}$, 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                  | $P|IBD$ | $P|Control$ | $P|UC$ | $P|CD$ |
|---------------------------|--------|------------|-------|-------|
| protein:plasmid:19382     | 0.035  | 1          | 0.624 | 0.014 |
| protein:vir:100911        | 0.024  | 1          | 0.768 | 0.012 |
| protein:vir:101234        | 0.036  | 1          | 0.326 | 0.011 |
| protein:vir:101241        | 0.025  | 1          | 1     | 0.019 |
| protein:vir:102267        | 0.022  | 1          | 0.096 | 0.006 |
| protein:vir:102270        | 0.01   | 0.036      | 0.003 | 1     |
| protein:vir:102286        | 0.023  | 0.031      | 0.011 | 1     |
| protein:vir:102302        | 0.038  | 1          | 0.011 | 0.134 |
| protein:vir:106677        | 0.028  | 1          | 0.021 | 0.021 |
| protein:vir:2149          | 0.022  | 1          | 0.623 | 0.009 |
| protein:vir:97987         | 0.018  | 1          | 0.433 | 0.006 |
| protein:vir:98404         | 0.007  | 1          | 0.117 | 0.002 |
| protein:vir:98574         | 0.017  | 0.155      | 0.004 | 1     |
| protein:vir:99488         | 0.049  | 1          | 0.281 | 0.014 |
| protein:vir:99706         | 0.04   | 1          | 0.105 | 0.012 |
| protein:vir:99718         | 0.028  | 1          | 0.021 | 0.021 |
| protein:vir:99892         | 0.01   | 1          | 0.17  | 0.002 |
| protein:vir:99893         | 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.

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. Cadnum1, Lucy A. Jury2, Venkata C. K. Sunkesula3, Sirisha Kundrapu1, Robin L. Jump1, Alain C. Stintzi1, Curtis J. Donskey1,4

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.


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 biobehavioral agents such as non-toxic, C. difficile could be targeted to colonize 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 g/day for 7 days to healthy volunteers; the mean concentration of vancomycin in stool after 7 days of treatment was $250 \pm 197$ 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 ± 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 non-epidemic 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 ≥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. difficile A/B, Wampole Laboratories, Wallingford, CT) and for C. difficile toxins A and B (Cambridge Biotech, Cambridge, MA) 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, comorbidities, 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, Cambridge, UK). Isolates from the highest dilution with growth were tested for in-vitro cytokin 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 Ternes et al [10]. For each assay, a known NAP1 strain was used as a positive control and American Type Culture Collection (ATCC) C. difficile 6359 as 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 g of stool.

In vitro Assay of Colonization Resistance

To assess functional recovery of the ability of the microbe 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 Bornfie1 and Barclay [13]. One epidemic NAP1 strain (VA 17) and one non-epidemic strain (VA 6) were tested.
Table 1. Sequence of polymerase chain reaction (PCR) primers used for construction of the Illumina library.

<table>
<thead>
<tr>
<th>Primer Set</th>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>Barcode sequences</th>
<th>Vs universal primers Adapted from [14]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCTCCTAGATCT</td>
<td>ATACG</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF4</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>AGCTCTAGT</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF1</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF2</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF3</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF4</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF5</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF6</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF7</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF8</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF9</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF10</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF11</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GF12</td>
<td>5'-ACACTCTTTCCCTACACGACTGCTATCT</td>
<td>CATCTGTC</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR1</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR2</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR3</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR4</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR5</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR6</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR7</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR8</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR9</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR10</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR11</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>GR12</td>
<td>5'-CTGCGGATGTTCCTGCGTGAACGCTTGAGTCACTAC</td>
<td>ATAGCGA</td>
<td>CAACGCGGAGAAGCTATGAC</td>
<td></td>
</tr>
</tbody>
</table>

Fresh stool samples were homogenized in a 1:10 dilution of pre-reduced sterile water and inoculated with 10⁷ 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 CDCA 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 16S rRNA 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 30. 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 doublertons, 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 6 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.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Metronidazole N = 54</th>
<th>Vancomycin N = 42</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, mean (range)</td>
<td>68 (60–90)</td>
<td>64 (56–89)</td>
<td>0.17</td>
</tr>
<tr>
<td>Male sex</td>
<td>33 (98)</td>
<td>41 (98)</td>
<td>0.58</td>
</tr>
<tr>
<td>Creatinine, mg/dL, mean (range)</td>
<td>1.3 (0.3–3.9)</td>
<td>1.6 (0.4–6.1)</td>
<td>0.33</td>
</tr>
<tr>
<td>White blood cell count, cells/mm$^3$, mean (range)</td>
<td>13.5 (12.2–30.9)</td>
<td>12.2 (1.6–33.1)</td>
<td>0.52</td>
</tr>
<tr>
<td>On systemic antibiotics during or after treatment</td>
<td>9 (24)</td>
<td>7 (17)</td>
<td>0.57</td>
</tr>
<tr>
<td>Diabetes Mellitus</td>
<td>9 (26)</td>
<td>21 (50)</td>
<td>0.06</td>
</tr>
<tr>
<td>Coronary Artery Disease</td>
<td>10 (29)</td>
<td>14 (33)</td>
<td>0.81</td>
</tr>
<tr>
<td>Chronic Obstructive Pulmonary Disease</td>
<td>6 (18)</td>
<td>7 (17)</td>
<td>1.0</td>
</tr>
<tr>
<td>End-Stage Renal Disease</td>
<td>3 (9)</td>
<td>3 (7)</td>
<td>1.0</td>
</tr>
<tr>
<td>Severe Clostridium difficile infection</td>
<td>9 (26)</td>
<td>16 (38)</td>
<td>0.54</td>
</tr>
<tr>
<td>Infection with epidemic HAP1 strain$^*$</td>
<td>8 (47)</td>
<td>13 (72)</td>
<td>1.0</td>
</tr>
<tr>
<td>Nursing home resident</td>
<td>5 (15)</td>
<td>10 (24)</td>
<td>0.39</td>
</tr>
</tbody>
</table>

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

Abbreviation: HAP1, 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.

Vulnerable Period for C. difficile Colonization

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 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 Haemophilus) belong to Gamma proteobacteria class of the phylum Proteobacteria, while the third genus (Clostridia) 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 Caetabacteriaceae family, the order Actinomycetales, the genus Ruminococcus, the genus Finegoldia, 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 = 0.005).

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 (57.1%), Firmicutes (27.7%), and Actinobacteria (22.2%) 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

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.
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
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. 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.0037626.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 to 36 days after completing treatment.

Based upon deep 16S rRNA 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
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]. LELSe 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 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.0376269.g005
microbial community is unknown. Similarly, the role of *Finegaldia* and *Coriobacteriaceae* in *C. difficile* colonization resistance requires further investigation. Finally, although the unidentified bacterium genus (RTFN2) 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, Hafnia* 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 methicillin 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


