Australian National University

Unravelling the role of microbiome and *Escherichia coli* as a trigger for Inflammatory Bowel Disease with a particular focus on Crohn's disease

Mukta Das Gupta

March 2022

A thesis submitted for the Degree of Doctor of Philosophy of
The Australian National University

© Copyright by Mukta Das Gupta 2022 All Rights Reserved

Declaration

I declare that this thesis titled ‘Unravelling the role of microbiome and *Escherichia coli* as a trigger for Inflammatory Bowel Disease with a particular focus on Crohn's disease is the product of my original work. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge, it contains no material previously published or written by another person, except where due reference is made in the text.

Mukta Das Gupta

March 2022

Acknowledgements

I express my sincere gratitude to my primary supervisor Dr Claire O'Brien, for the inspiration, encouragement, and guidance throughout my PhD journey. The knowledge and skills I learned from you are immense. Your understanding during difficult times was also very much appreciated.

I am incredibly grateful to my supervisory panel Prof. Dr Paul Pavli and Prof. David Gordon, for their guidance. I have been fortunate to be part of such a great team. Thank you, Paul, for the constructive feedback on my work and for guiding me to improve my writing skill. David, the bioinformatics skills that I learned from you will be an outstanding achievement of my scientific career. Thank you for all your support throughout my PhD research.

Thank you, all past and present lab members, especially Phil, Rasel, Natalia, Sam, Angel, Buddhie, Judith, Mah, Truc, Veronica, and Ojas, for helping me whenever needed and for the beautiful moments we enjoyed together.

I also like to thank all my friends at Graduate House, ANU and family friends in and around Canberra. You all made my life more beautiful in this beautiful Australia.

Thank you, mom and dad, for all your love and mental support during my PhD. It was the most challenging decision for me to pursue a PhD leaving my family in Bangladesh if you were not by my side.

Thank you, my dear husband, Ashutosh and my lovely son Atri for your unconditional love and sacrifice. You are my strength and a huge source of inspiration.

Finally, thanks to The Australian National University and Endeavour Post Graduate scholarship programme for giving me this opportunity to fulfil my dream.

Thank you, everyone!

Abstract

Inflammatory bowel diseases (IBD) are chronic inflammatory conditions of the digestive tract consisting of two primary forms: Crohn's disease (CD) and ulcerative colitis (UC). The pathogenesis of IBD is still not precisely known. However, an imbalance in the intestinal microbiome (dysbiosis) and putative pathogens have been implicated. Several lines of evidence implicate bacteria in the pathogenesis of IBD, and *Escherichia coli* (*E. coli*) is one of the leading candidate triggers. These studies aimed to investigate the role of gut bacteria with a particular focus on *E. coli*, to determine if genes or molecular properties of *E. coli* are associated with tissue type, including lymph nodes and the initial CD lesion (aphthous ulcer), or adult and paediatric isolates. We also aimed to determine the association of genes and molecular properties of *E. coli* with NLRP3 inflammasome activation, which may be a mechanism that causes CD.

The microbiome analysis was based on 16 non-BD controls and 72 CD tissue samples. CD tissue samples consisted of tissues from active and inactive stages of disease; 'early-active-stage' tissues, including aphthous ulcers; 'mid-stage tissues', which included biopsies obtained during colonoscopy; and 'late-stage' tissues, obtained from surgical resections. We found that CD at different disease stages was associated with a microbial signature characterised by reduced alpha diversity and greater dysbiosis compared to controls. An imbalance in bacterial genera was common in all patient groups, and the microbial diversity of CD samples was lower than non-IBD controls. Several genera were significantly increased in CD (overall) and different disease stages of CD compared to control, including *Escherichia_Shigella*, *Flavonifractor*, *Klebsiella*, and *Ralstonia*. On the other hand, *Faecalibacterium* was reduced significantly in CD and different disease stages compared to the normal biopsies in controls.

E. coli have been implicated in CD, but the molecular properties or genes that may contribute to their potential virulence are undetermined. The comparative genomics analysis involved the whole-genome comparisons of 179 *E. coli* strains, isolated from 62 patients with CD (n = 23 paediatric CD), 18 with UC, and 20 healthy controls. The source of the isolates included lymph nodes (n = 13), aphthous ulcers (n = 7), intestinal mucosa (n = 133), and other sources (n = 26). Bottromycin bacteriocin, ABC capsule, and *E. coli* pUTI89 plasmid were significantly associated with paediatric CD. In

contrast, Group IV e capsule and Bottromycin bacteriocin were significantly associated with lymph node CD isolates compared to other isolates controls. CD isolates were more likely to have genes associated with adhesion, the toxin-antitoxin system, plasmid partitioning, and conjugation than healthy controls. Lymph node CD isolates were associated with genes involved in adhesion, phosphate binding and catalase-peroxidase activity compared to controls. Catalase-peroxidase activity may allow *E. coli* to survive within lymph nodes, consequently providing an ongoing stimulus to the immune system. Mechanisms that allow *E. coli* to remain viable in lymph nodes of patients with CD should be explored by in vitro studies and with a larger sample size.

It has been hypothesised that damage to the intestinal mucosa occurs in patients with CD due to a dysregulated innate immune response triggered by microbial antigens in genetically susceptible individuals. Previous studies have shown that the NLRP3 inflammasome was associated with the secretion and release of mature IL-1β and IL-18, enhancing intestinal inflammation. We explored whether *E. coli* activate the NLRP3 inflammasome in mouse bone marrow-derived macrophages, focusing on different molecular properties and genes and their ability to activate the NLRP3 inflammasome. Although activation of NLRP3 was not associated with the clearance of *E. coli* from macrophages, we found the association of some genes of *E. coli* strains with inflammasome activation. Several features, such as small toxic polypeptides that are beneficial for *E. coli* to survive inside the macrophages, may be a mechanism that causes CD.

Abbreviations

AC	Ascending colon
AIEC	Adherent-invasive *E. coli*
ANOVA	Analysis of variance
ANU	Australian National University
APCs	Antigen-presenting cells
ATG16L1	Autophagy-related protein 16-1
AU	Aphthous ulcer
BLAST	Basic local alignment tool
BMDM	Bone marrow-derived macrophages
CCR2	Chemokine receptor type 2
CD	Crohn's disease
CDI	*Clostridioides difficile* infection
CEACAM6	Carcinoembryonic antigen-related cell adhesion molecule 6
CFU	Colony forming units
COGs	Clusters of orthologous genes
DAMPs	Damage-associated molecular patterns
DCs	Dendritic cells
DNA	Deoxyribonucleic acid
EMP	Erythro-myeloid precursor cells
ERIC	Enterobacterial repetitive intergenic consensus
FAE	Follicle associated epithelium
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GWAS	Genome-wide association studies
HSC	Hematopoietic stem cells
HTS	High throughput sequencing
IBD	Inflammatory Bowel Disease
IC	Indeterminate colitis
IFN-γ	Interferon-gamma
IL23R	Interleukin 23 receptor
ILCs	Innate lymphoid cells

Abbreviations continued

IM	Involved mucosa
IRGM	Immunity-related GTPase family M protein
LB	Luria-Bertani
LB	Large bowel
LEfSe	Linear discriminant analysis effect size
LN	Lymph node
LPS	Lipopolysaccharide
MAP	*Mycobacterium avium* subspecies *paratuberculosis*
MHC	Major histocompatibility complex
NLRP3	NOD-, LRR- and pyrin domain-containing protein 3
NLRs	NOD-like receptors
NM	Uninvolved mucosa
NOD2	Nucleotide-binding oligomerization domain-containing protein 2
NSAID	Non-steroidal anti-inflammatory drug
OTU	Operational taxonomic unit
PAMPs	Pathogen-associated molecular patterns
PBS	Phosphate buffered saline
PCoA	Principal component analysis
PCR	Polymerase chain reaction
PERMANOVA	Permutational multivariate analysis of variance
POCER	Postoperative Crohn's endoscopic recurrence
PP	Peyer's patches
PRRs	Pattern recognition receptors
QIIME	Quantitative Insights into Microbial Ecology
REP	Repetitive extragenic palindromic
Rep-PCR	Repetitive element PCR
RLRs	RIG-1-like receptors
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rRNA	Ribosomal RNA

Abbreviations continued

SB	Small bowel
SNP	Single nucleotide polymorphism
TBE	Tris-borate-EDTA buffer
TE	Tris-EDTA buffer
TI	Terminal Ileum
TLRs	Toll-like receptors
TNF-α	Tumour necrosis factor alpha
UC	Ulcerative colitis
WT	Wildtype

Table of contents

Declaration **2**

Acknowledgements **3**

Abstract **4**

Abbreviations **6**

Table of contents **9**

Statement of contribution to projects **14**

Chapter 1: General Introduction **15**

1. General Introduction **16**

1.1 Epidemiology of CD *17*

1.2 Causes and risk factors of CD *20*

1.3 Clinical features of CD *30*

1.4 Clinical features of ulcerative colitis (UC) *34*

1.5 Overview of paediatric CD *34*

1.6 Diagnosis of CD and/or IBD *36*

1.7 History of CD *37*

1.8 Therapeutic management of CD *38*

1.9 Surgical management of CD *40*

1.10 Role of gut microbiota in CD *41*

1.11 Individual bacterial species implicated in CD *45*

1.12 Bacterial translocation in CD *54*

1.13 Inflammation and role of the monocyte-macrophage cell lineage in CD *57*

1.14 The NLRP3 inflammasome and CD *59*

1.15 Aims of the PhD project *62*

1.16 References *64*

Chapter 2: Microbiome analysis of different disease stages and tissue types from Crohn's disease patients comparing with healthy controls **94**

2.1 Abstract **95**

2.2 Introduction **97**

2.3 Methods **101**

2.3.1 Selection of patients *101*

2.3.2 16S amplification and sequencing *104*

2.3.3 Pre-processing of 16S rRNA gene sequencing data *105*

2.3.4 Statistical analysis *107*

2.4 Results **108**

2.4.1 Alpha rarefaction curves based on observed OTUs in disease status, disease activity and tissue types *109*

2.4.2 Alpha diversity analysis to compare microbiota richness and diversity between Crohn's disease and controls *110*

2.4.3 Alpha diversity analyses based on disease activity and different tissue types in Crohn's disease and controls *111*

2.4.4 Differences in microbial community composition (beta diversity) between all Crohn's disease and controls *116*

2.4.5 Comparative beta diversity in Crohn's disease and controls based on disease stages, disease activity and tissue types *117*

2.4.6 The gut microbial abundance at taxonomic levels phylum and genus in Crohn's disease patients and controls *123*

2.4.7 Comparison of the core intestinal microbiome in disease status and different disease stages in Crohn's disease and controls *128*

2.4.8 Microbial biomarkers in Crohn's disease compared with controls *129*

2.4.9 Co-abundance of different bacterial genera in the gut *132*

2.5 Discussion **134**

2.6 Conclusion **139**

2.7 References **140**

2.8 Supplemental Materials **146**

Chapter 3: Comparative Genomics of *E. coli* Isolated from Patients with Inflammatory Bowel Disease and Controls **166**

3.1 Abstract **167**

3.2 Introduction **169**

3.3 Materials and Methods **174**

3.3.1 Sample information *174*

3.3.2 Isolation and identification of E. coli isolates from the biopsy tissues and bowel resected samples *176*

3.3.3 Characterization of E. coli by Rep-PCR and Clermont PCR *176*

3.3.4 Whole genome sequencing using high throughput sequencing (HTS) technology *177*

3.3.5 E. coli genome contig assembly *177*

3.3.6 WGS data analysis for molecular characterization of the E. coli isolates *178*

3.3.7 Plasmid analysis *178*

3.3.8 Phylogenetic Analysis *179*

3.3.9 Genome annotation and pan-genome association studies *180*

3.3.10 Principal component analysis (PCA) and statistical methods *180*

3.3.11 Bacterial genome-wide association study (GWAS) using Pyseer *181*

3.3.12 Genome alignment and visualization *181*

3.4 Results **182**

3.4.1 Phylogroups of isolates with unique sequence types (STs) *182*

3.4.2 Serotype diversity among the E. coli isolates *185*

3.4.3 Virulence properties *186*

3.4.4 Plasmid analysis of the E. coli isolates *188*

3.4.5 Presence of conjugative elements in different groups of isolates *196*

3.4.6 Antibiotic resistance in E. coli isolates *198*

3.4.7 Capsule typing *200*

3.4.8 Bacteriocin production *201*

3.4.9 Prophage analysis *203*

3.4.10 CRISPR analysis *203*

3.4.11 Phylogenetic Analysis *204*

3.4.12 Principal component analysis *206*

3.4.13 Pan-genome analysis *207*

3.4.14 Phenotype and genotype association study using Scoary *208*

3.4.15 Genome-wide association study (GWAS) by Pyseer *219*

3.4.16 Comparison of SNPs and genes of E. coli strains in bowel resected samples.....229

3.5 Discussion.....231

3.6 Conclusion 240

3.7 References 241

3.8 Supplemental Materials.....250

Chapter 4: Review on Macrophage- *E. coli* interactions in Crohn's disease.....277

4.1 Introduction.....278

4.2 Methodology for literature search for this review 278

4.3 Objective of this review.....279

4.4 The complex biology of macrophage: origin, function, and activation.....280

4.5 Lineage tracking of tissue-resident macrophages.....282

4.6 Role of macrophages in innate immunity 283

4.7 Role of macrophages in inflammation and tissue repair.....286

4.8 Macrophages in the intestine 286

4.9 Role of macrophages in intestinal injury and repair 287

4.10 Macrophages in CD mucosa.....290

4.11 Genetic variants associated with CD and their effect on macrophage functions 290

4.12 Host macrophage and bacterial interactions in inflammatory bowel disease.....292

4.13 Immune evasion of *E. coli* in IBD 293

4.14 The interplay between macrophages and *E. coli* in CD 295

4.15 Persistence of *E. coli* in macrophages of patients with CD 300

4.16 Handling of AIEC by macrophages and the role of virulence factors 300

4.17 Macrophages and secreted cytokines in response to AIEC.....303

4.18 Conclusion 303

4.19 References 304

Chapter 5: NLRP3-mediated inflammasome activation by *E. coli* isolates from patients with inflammatory bowel disease and controls 314

5.1 Abstract 315

5.2 Introduction 316

5.3 Methods 320

5.3.1 Mouse model 320

5.3.2 Bone marrow-derived macrophages (BMDMs) 320

5.3.3 Preparation of E. coli culture 321

5.3.4 Bacterial stimulation and activation of inflammasome in BMDMs 321

5.3.5 Western blot analysis 321

5.3.6 E. coli genome analysis 323

5.4 Results 326

5.4.1 Inflammasome study results 326

5.4.2 Association of sequence types (STs) and phylogroup of E. coli with inflammasome activation 328

5.4.3 Association of molecular properties and genes of E. coli with inflammasome activation 328

5.4.4 Association of E. coli genes with inflammasome activation 329

5.4.5 Association of k-mers in the E. coli with inflammasome activation based on k-mers analysis 331

5.4.6 Phylogenetic analysis 332

5.5 Discussion 333

5.6 Conclusion 337

5.7 References 337

5.8 Supplemental materials 342

Chapter 6: General Discussion 347

6.1 General discussion 348

6.1.1 Characterisation of microbial communities in different disease stages and tissues from Crohn's disease patients and non-IBD controls 348

6.1.2 The plausible association of Escherichia coli in CD 354

6.1.3 NLRP3 mediated inflammasome activation by IBD and non-IBD associated E. coli isolates 356

6.2 References 358

Statement of contribution to projects

Project one: Microbiome analysis of different disease stages and tissue types from Crohn's disease patients comparing with healthy controls

We planned a microbiome study of different disease stages and tissue types from CD patients and healthy controls with my supervisors. I performed all the microbiome data processing, analysis, comparison, and interpretation of the results.

Project two: Comparative genomics of *E. coli* isolated from patients with Inflammatory Bowel Disease and controls

I assisted with the library preparation of the newly sequenced *E. coli* isolates for this study together with the lab technician. I determined which *E. coli* isolates from samples were unique, using Rep-PCR and determined their phylogroup using the Clermont et al. multiplex PCR. After sequencing the new isolates, I completed the contig assembly for all the *E. coli* genomes in this study. I conducted the whole genome study, genome-wide association study and all the data analysis for molecular characterization of all the *E. coli* isolates.

Project three: NLRP3-mediated inflammasome activation by *E. coli* isolates from patients with inflammatory bowel disease and controls

I learnt the NLRP3-mediated inflammasome activation assay in the lab of Professor Si Ming Man, and performed some of the first experiments. I conducted all of the data analysis related to identifying molecular characteristics of *E. coli* isolates, pangenome analysis and genome-wide association studies, to identify the inflammasome activation associated genes in the *E. coli* isolates used for the inflammasome assay.

Chapter 1: General Introduction

1. General Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are the two major types of inflammatory bowel disease (IBD). Indeterminate colitis (IC) is the term used when pathologic findings are not sufficient to distinguish between CD and UC. The IBDs are chronic conditions, which significantly affect the quality of life of patients, and place significant demands on the healthcare system. There is no cure for the IBDs, and its incidence is increasing globally (Shivashankar et al., 2017). While significant improvement has been achieved in the treatment of these diseases, not all patients respond to therapy. Treatment is solely aimed at controlling the clinical manifestations of disease. While the aetiology of IBD is unknown, it is thought to arise from an atypical immune response to intestinal luminal microbiota, in genetically susceptible individuals.

CD and UC share many clinical features but are distinguished based on their microscopic and macroscopic appearances (Sartor, 2006). Inflammatory lesions in UC are limited to the superficial mucosa and the colon, and inflammation starts at the rectum and moves proximally. CD lesions are transmural and can occur anywhere along the gastrointestinal tract (Baumgart and Carding, 2007). CD most frequently affects the terminal ileum and colon (Stange et al., 2006). IBD and CD are the general and specific foci, respectively, of this PhD research, and CD will be the primary focus of this literature review. In particular, how bacteria in the gastrointestinal tract are involved in the development of CD.

So far, the following theories of the microbial pathogenesis of CD have been proposed:

1) Microbial dysbiosis is resulting in an increase in 'harmful' bacteria and a decrease in 'beneficial' bacteria in CD (Van Kruiningen et al., 2005; Png et al.,

2010; Fava and Danese, 2011; Mondot et al., 2011; Fujimoto et al., 2013; Cao et al., 2014; Vester-Andersen et al., 2019);

2) Many putative bacterial triggers have been suggested, including a novel class of *Escherichia coli*, called Adherent invasive *E. coli* (AIEC) (Barnich et al., 2007; Carvalho et al., 2009; Roberts et al., 2010; Chassaing et al., 2011; Strober, 2011; Agus et al., 2014; O'Brien et al., 2014) and

3) The presence of an unidentified pathogen in CD (De Hertogh et al., 2008).

Although various culture-based and molecular studies supported a plausible role of *Escherichia coli (E. coli)* in CD; however, none could confirm any gene or molecular properties of this bacteria that could trigger the pathogenesis in CD. Our research interest was to investigate the genes and molecular properties of *E. coli* linked with the translocation of this bacteria across the gut mucosa. In this chapter we begin with a brief introduction and provide background knowledge of the gut microbiome and *E. coli* in the context of CD.

1.1 Epidemiology of CD

IBD has appeared as a global disease in the 21st century; however, in the 20th century, IBD was recognised as a condition that affected people living in industrialised countries. Plenty of studies have stated the worldwide incidence and prevalence of IBD (Ng et al., 2017). According to a report published in 2007 by the Australian Crohn's and Colitis Association (ACCA) on the economic costs of CD and UC, in 2005, approximately 61,000 Australians had IBD, and around 28,000 was living with CD. According to Crohn’s and Colitis Australia (CCA), the number of Australians living with IBD is supposed to increase to approximately 100, 000 by 2022. According to a 2010

study based in Australia, the rough annual incidence rates for IBD and Crohn's disease were 29.3 per 100,000 and 17.4 per 100,000, respectively (Wilson et al., 2010). According to this study, the incidence of IBD in Australia is higher compared with other countries. Another report in 2013 by Crohn's and Colitis, Australia, stated that approximately 1 in 250 people in Australia are affected by IBD. Between 2011-2012, a population-based study shows that IBD incidence and prevalence was increased within Asia-pacific regions (Ng et al., 2013b). In this study, they observed that the incidence rates per 100,000 individuals were 23.67 for IBD and 14.00 for CD in Australia: much greater than in Asia (1.37 for IBD and 0.054 for CD). Recently, one population-based study investigated the prevalence and incidence of IBD in the Tasmanian region from 1 June 2013 to 31 May 2014 (Bhatia et al., 2019). This study shows a prevalence of 1719 cases in Tasmania, with age-standardised rates (ASR) of IBD, CD and UC of 303.9, 165.5, and 131.4 per 100 000, respectively. The incidence of IBD, CD, and UC were 29.5, 15.4, and 12.4 in 100 000, respectively. Here, prevalent CD cases had more endoscopy and bowel resections; and incident CD cases were more common than UC (Bhatia et al., 2019). These data indicate the growing number of patients with IBD, particularly CD in Australia, and a thorough study to explore its cause is essential for managing CD and overcoming the challenges associated with this disease.

A recent study stated that around 6·8 million cases of IBD had been found worldwide in 2017 (Alatab et al., 2020). According to this study, the USA has the greatest age-standardised prevalence of IBD (464·5 in 100 000), and the UK had the second-highest prevalence of IBD (449·6 per 100 000). However, in another study, the highest prevalence was observed in Europe (UC: 500/ 100 000 in Norway; CD: 322 / 100 000 in Germany) and North America (UC: 286 / 100 000 in the USA; CD: 319 / 100 000 in

Canada) (Ng et al., 2018). It has been estimated that around 1.3% of the US population has been diagnosed with IBD (Dahlhamer et al., 2016). In Europe, incidence does not differ significantly between sex, race or social class; however, in the USA, the CD is more common amongst Jewish ethnicity than cohorts of other people in the same area (Cho and Brant, 2011). In Asia, the highest incidence rate was found in China (3.44/100,000). CD associated complications such as strictures and penetration were more common in Asia than in Australia ($P = 0.001$), but the family correlation of IBD was not as strong in Asia ($P < 0.001$) (Ng et al., 2013b). Global data indicates that CD is becoming more prevalent. The therapeutic management is getting more complex due to the heterogeneous lesions, overlapping manifestation with other inflammatory bowel diseases, and unidentified etiology, which shows the significance of in-depth studies to identify the actual cause of this disease (Gajendran et al., 2018).

CD onset usually occurs between 15-50 years of age and has its peak in early life (between age 20 and 30 years), with a second smaller peak among the elderly (around age 50 years). It is commonly observed in developed countries and among Jews. Whether the condition is linked to occupation, sex, stress, infection, oral contraceptives, diet, and smoking is less evident. According to different epidemiological studies, first-degree relative (parent, child, or sibling) of patients with IBD (either CD or UC) have almost a 3 to 20 times higher likelihood of getting CD compared to normal individuals (Kevans et al., 2016).

Cigarette smoking plays a protective role in UC but is a significant risk factor for CD and intensifies the course of this disease (Berkowitz et al., 2018). In patients with CD, termination of smoking helps improve the condition of the disease. It has been reported that following cessation of smoking, the activity of the disease, rates of relapse, the necessity of immunosuppressant medication, and need for and surgery is

reduced compared to CD patients who smoke (Cosnes et al., 2001; Lewis et al., 2007; Cosnes, 2010; van der Heide et al., 2011). The role of nonsteroidal anti-inflammatory drugs (NSAID) in CD has been reported (Long et al., 2016). Their finding suggested that regular use of NSAID could increase the disease activity and may be a risk factor for relapse in patients with CD.

1.2 Causes and risk factors of CD

CD is supposed to result from a dysregulated immune response to gut microbiota in genetically prone individuals. Diet, psychological stress, smoking, antibiotics, NSAIDs, as well as epigenetics (Ananthakrishnan, 2013; Loddo and Romano, 2015; Dutta and Chacko, 2016) may also contribute to CD pathogenesis (Molodecky and Kaplan, 2010). The major risk factors of CD are described below.

1.2.1 Genetic factors

Genetic factors are the essential components in the pathogenesis of CD. Epidemiological investigation for genetic association revealed that 2–14% of patients with CD have a family background of IBD (Halme et al., 2006). Several concordance studies in twins also indicated a genetic contribution to CD. (Bengtson et al., 2009b; Baumgart and Sandborn, 2012). Tysk *et al*. (1988) reported the concordance for CD was as high as 58% in monozygotic twins. An increased risk for first-degree relatives (5%-35%), and a less evident, but still increased risk for more distant relatives have been observed in CD (Russell and Satsangi, 2004). A nationwide study in Germany also indicated that 35% of monozygotic twins, and 3% of dizygotic twins, were concordant for CD (Spehlmann et al., 2008). All these figures reflect a pronounced genetic predisposition in CD.

An *in-silico* genome-wide association study (GWAS) has identified 241 IBD susceptibility loci (de Lange et al., 2017) and 71 of those were associated with CD (Lee and Parkes, 2011). A model selection analysis by Jostins *et al.* (2012) revealed that out of 163 IBD susceptibility loci, 110 are linked with both CD and UC; of the remaining loci, 30 are considered CD-specific and 23 UC-specific. Before this study, Barrett *et al.* (2008) identified more than 30 susceptibility loci associated with CD. Identifying susceptibility loci has contributed to understanding the fundamental biology of how disturbed immunological pathways of the intestinal immune system could trigger CD pathogenesis. Although not all the susceptibility loci recognised in GWAS mapped to known genes, those that were mapped have been implicated in different pathophysiologic processes in CD. Several mutations in CD-associated loci are found to be strongly associated with CD pathogenesis. Those are involved in bacterial recognition (NOD2 mutations) (Hugot et al., 2001), autophagy (ATG16L1 and IRGM) (Hampe et al., 2007; Glas et al., 2013b), and cytokine signalling (IL23R) (Xu et al., 2015). The polymorphisms in the susceptibility loci indicate that patients with CD may have an abnormal and defective ongoing innate immune response to microbes that may be implicated in CD pathogenesis.

Polymorphisms in NOD2 (nucleotide-binding oligomerization domain 2) gene on chromosome 16 was the first polymorphism known to be associated with CD (Hugot et al., 2001; Ogura et al., 2001; Glas et al., 2010). NOD2 is specialized NOD-like receptor (NLR), a family of intracellular pattern recognition receptors associated with detecting damage and microbial-associated molecular patterns. NOD2 gene can recognize and respond to pathogenic microorganisms that can invade and multiply intracellularly (Ogura et al., 2001). Three main NOD2 mutations have been found to

be overrepresented in CD patients so far. Some studies also identified further NOD2 variants being associated with IBD pathogenesis. These NOD2 variants play a role in defective bacterial recognition in the innate immune system in CD (Maeda et al., 2005; Glas et al., 2010). Several studies revealed that NOD2 plays an important role in regulating stability and homeostasis between microbiome and host immune responses (Rehman et al., 2011; Al Nabhani et al., 2017; Balasubramanian and Gao, 2017). It also plays a key role in xenophagy; the engulfment and degradation of intracellular microorganisms in autolysosomes (Levine, 2005; Cho and Brant, 2011).

Genetic studies have revealed that the ATG16L1 gene encodes a protein that is part of a multimeric protein complex essential for autophagy and is associated with CD susceptibility (Hampe et al., 2007). Autophagy is a degradation process of cell's self-components such as protein aggregates and damaged organelles to be degraded and recycled (Pavel and Rubinsztein, 2017). The Atg16L1 T300A variant, a significant risk polymorphism in CD, disrupts autophagy which can lead to intestinal inflammation and predispose to CD, but the process remains unclear (Palomino-Morales et al., 2009; Tschurtschenthaler et al., 2017). The T300A variant is associated with a 1.65-fold risk of CD overall and a 2.2-fold risk of ileal disease, suggesting its association with ileal CD (Prescott et al., 2007). Furthermore, some studies show an interaction between NOD2 and ATG16L1 (Cooney et al., 2010; Travassos et al., 2010; Plantinga et al., 2011). NOD2 can recruit ATG16L1 to the bacterial site of entry. NOD2 variants disrupt the recruitment process of ATG16L1, leading to impaired autophagy of invading bacteria in dendritic cells (Travassos et al., 2010; Sidiq et al., 2016). Therefore, both ATGL16 and NOD2 play additive roles in controlling microbial invasion (Cho and Brant, 2011). Moreover, ATG16L1 T300A polymorphism-expressing mice exhibit abnormalities in Paneth cells, causing decreased protein stability, altered antibacterial

autophagy and reduced inflammatory cytokine production that triggers induction and progression of colitis (Lassen et al., 2014; Cohen et al., 2019). The immunity related GTPase family M (*IRGM*) protein is an enzyme that induces autophagy to remove intracellular pathogens (Xavier et al., 2008; Fritz et al., 2011) such as *Mycobacterium tuberculosis* (Singh et al., 2006). Three common polymorphisms in the *IRGM* gene are related to autophagy and intracellular bacterial handling and are associated with CD susceptibility (Parkes et al., 2007; Franke et al., 2008; Van Limbergen et al., 2009; Prescott et al., 2010). This gene encodes a GTP-binding protein involved in the regulation of autophagy to eliminate intracellular pathogens. IFN- γ can also induce autophagy in macrophages (Gutierrez et al., 2004). IRGM has been found to restabilise autophagy in the absence of IFN-γ stimulation (Singh et al., 2006). A study has confirmed the significant association of three polymorphisms (rs13361189, rs4958847 and rs10065172) in IRGM genes with the susceptibility of CD in European populations (Lu et al., 2013). A recent study has identified that risk alleles at rs13361189 and rs10065172 loci decreased the expression of IRGM in whole blood and the terminal ileum, leading to the dysregulation of other IBD-associated genes that play a role in autophagy and inflammation (Ajayi et al., 2019).

Genome-wide association studies (GWAS) and murine model studies revealed a strong role protective or susceptibility role for the interleukin 23 receptor (IL23R) in CD (Duerr et al., 2006; Rioux et al., 2007a; Amre et al., 2008; Sivanesan et al., 2016). Nevertheless, most previous study results were inconsistent and inconclusive whether IL23R plays a protective or susceptibility role.

The presence of several variants of the IL23R, particularly G149R, V362I, and R381Q variants, play a protective role against the development of CD (Sivanesan et al., 2016). A recent meta-analysis has also found that the rs11209026 polymorphism in IL23R

may play a protective factor against developing CD (Zhu et al., 2019). Around 14% of healthy humans of European ancestry harbour a variant of the IL23R gene (p. R381Q) that reduces CD risk by approximately 3-fold (Cho and Brant, 2011). IL-23 also plays a critical regulatory role in intestinal inflammation through innate immune cells and T cells (Izcue et al., 2008).

In addition to that, studies show that the IL-23/IL-17 pathway is implicated in the pathogenesis of CD (McGovern et al., 2009; Kim et al., 2011). In IL-23/IL-17 pathway, IL-23 upregulates the secretion of the pro-inflammatory cytokine IL-17 (Hoeve et al., 2006), further implicating its role in the development of CD. It appears that a IL23R polymorphism is protective against the development of CD (Palmieri et al., 2017).

Our understanding of CD genetics is essential to know the relative contributions of the genetic factors in CD pathogenesis. In adult-onset CD, genome-wide association studies (GWAS) and meta-analysis have confirmed above 30 distinct susceptibility loci that could explain around 20% of the genetic variability in CD (Barrett et al., 2008). However, recently, another GWAS has identified 163 genetic loci that may explain only 13.6% of CD disease variance, and most of these loci can contribute very little towards explaining the IBD heritability (Jostins et al., 2012). Adult and paediatric onset CD are phenotypically heterogeneous; however, the causal mechanisms of these differences are still unknown. Genetic susceptibility may play a significant role in the pathogenesis of early-onset CD, and gene mutations are expected to be present in higher frequency in paediatric-onset CD. A previous study suggested that disease course and manifestation in very early-onset (0 to <10 yr) CD are likely to be associated with common genetic background (Bianco et al., 2012). It has been found that

polymorphisms 3020insC in NOD2 and SNP rs3792876 in SLC22A4/5 were significantly higher in patients with paediatric-onset CD compared to patients with adult-onset CD (de Ridder et al., 2007). However, not all studies have found that 100% of CD patients have mutations in CD-associated genes, and some mutations are also present in healthy subjects (Jess et al., 2005).

This indicates that other factors, such as intestinal microbiota, are probably playing an important role in the progression of CD. In this regard Jostins et al. (2012) suggests that an interaction between host's mucosal immune system and microbes in the intestinal mucosal layer and gut lumen could be a potential trigger for this disease. In particular, susceptibility genes possibly play a role to turn the symbiotic relationship of commensal microbiota to a pathogenic with the host. Consequently, CD research mostly focuses on the interaction between genetic and immunological factors, and the microbial community in the host's gut.

1.2.2 Immunological factors

There is a notable role of innate immunity in CD pathogenesis. The mucus layer overlying the intestinal epithelium acts as the primary physical barrier between intestinal bacterial pathogens and the mucosal immune system (Peterson and Artis, 2014; Yue et al., 2019). It is suggested that a defect or alteration of this mucosal barrier function by infection or other environmental factors may lead to increased intestinal permeability in patients with CD. An early study reported that, children with active small bowel CD had six-fold increase in intestinal permeability (Pearson et al., 1982). Since then, many studies have reported an altered intestinal barrier leading to increased intestinal permeability in patients with CD (Jäger et al., 2013; Antoni et al., 2014; Michielan and D'Incà, 2015). Some studies have also indicated that inflammatory

cytokines such as tumour necrosis factor (TNF)-α may cause defects in the intestinal barrier function, as these were responsible for the dysfunction of the intestinal barrier monolayer cultures, and anti-TNF therapy stabilises intestinal permeability in patients with CD (Suenaert et al., 2002; Wang et al., 2005; Noth et al., 2012). A deficiency in the protective function of innate immunity that keeps microbes from breaching the epithelial barrier could be responsible for CD pathogenesis. However, whether the defect in intestinal barrier function is the primary cause or an outcome is still under debate (Michielan and D'Incà, 2015).

Numerous innate immune cells have been implicated in CD pathogenesis. Neutrophils intensify intestinal inflammation by impairing intestinal barrier function and secretion of multiple inflammatory cytokines (de Souza and Fiocchi, 2016). Macrophages and dendritic cells (DCs) regulate gut immune homeostasis through sensing microbial antigens in the gut. This microbial sensing is maintained by pattern-recognition receptors (PRRs), consisting of toll-like receptors (TLRs) and NOD-like receptors (NLRs). These are stimulated by pathogen-associated molecular patterns (PAMPs), which is crucial and well maintained to prevent inappropriate immune responses to commensal microbes (Geremia et al., 2014; Ahluwalia et al., 2017). It is assumed that mutations in genes encoding TLRs and NLRs, such as polymorphisms in NOD2 may be associated with a dysregulated innate immune response, and are therefore implicated in CD pathogenesis (Hugot et al., 2001; Bonen et al., 2003; Noguchi et al., 2009; Saruta et al., 2009).

Macrophages are a key source of various cytokines and inflammatory mediators, which have been linked to CD pathogenesis (Caprilli and Frieri, 2009; Gren and Grip, 2016). A study reported an increased number of infiltrating macrophages induced the production of pro-inflammatory cytokines and their contribution to inflammation using

the experimental colitis murine model (2,4,6-trinitrobenzene sulfonic acid-treated C57BL/6 mice) (Kanai et al., 2001). Highly expressed genes in intestinal macrophages could aggravate intestinal inflammation in mice. Study reported that Adherent-Invasive *Escherichia coli* (AIEC)-infected, monocyte-derived macrophage (MDM) from patients with CD secreted significantly higher IL-6 and TNF-alpha than those from healthy controls (Vazeille et al., 2015). This observation implicates the ability of macrophages from CD patients to support the survival of pathogenic bacteria (Vazeille et al., 2015).

Current effective treatments for CD include immunosuppressants and TNF-α antagonists (Podolsky 2002; Melsheimer et al., 2019), supporting a connection with adaptive immune responses in CD pathogenesis. To initiate the adaptive immune response, antigen-presenting cells (APCs) play a critical role. APCs consists of DCs, macrophages, and B cells. Antigens are recognised by APCs through PRR and play a crucial role to initiate and modulate T-cell-mediated responses. The key role of intestinal DCs is to recruit antigens from gut lumen then transfer them to the mesenteric lymph nodes to initiate the T-cell responses (Voedisch et al., 2009). Intestinal macrophages in the lamina propria present antigens to T cells (Platt and Mowat, 2008) and are generally hyporesponsive to PRR, leading to a weaker immune response to TLR antigens (Smith et al., 2011). Intestinal macrophages under normal conditions do not produce pro-inflammatory cytokines. Lamina propria macrophages also produce anti-inflammatory IL-10 cytokines (Denning et al., 2007). Consequently, in the normal intestinal mucosa, resident macrophages express anti-inflammatory responses and retain the scavenger and host defence function to gut microbiota (Smith et al., 2011). Macrophages also can stimulate inflammation. Increased numbers of recently recruited CD14+ macrophages that produce greater amounts of proinflammatory cytokines, for example, IL-23, TNF-α, and IL-6, relative to CD14-

negative macrophages, were found in the intestines of CD patients (Kamada et al., 2008). It has also been observed that activated macrophages accumulate at sites of inflammation with increased TLR and inflammatory receptor expression (Hausmann et al., 2002; Smith et al., 2011). This observation illustrates a strong association between gut microbiota and macrophages in initiating inflammation in CD.

The initial steps in the adaptive immune response involve the activation of the Th lymphocytes (Iwasaki and Medzhitov, 2010). Th-1 cells are stimulated by IL-12 and induce the secretion of IFN-gamma, TNF, and IL-12 (Romagnani, 1994). Dysregulated Th-1 responses due to the increased amount of IL-12 and IL-18, are assumed to be linked with inflammation in CD (Monteleone et al., 1997; Monteleone et al., 1999). Furthermore, T cells are functionally more active and secrete an increased amount of Th1 cytokines (IL-2 and IFN-gamma) in CD than in controls (Breese et al., 1993). Kugathasan et al. reported that mucosal T-cell immunoregulation varies between the early and late stages of the disease, where Th-1 responses are lost in late CD (Kugathasan et al., 2007). Studies also found that Th-17 cells, IL-23 and IL-17A are upregulated in the lamina propria in CD compared with controls (Kobayashi et al., 2008; Rovedatti et al., 2009; Olsen et al., 2011). Taken together, this evidence supports the hypothesis that innate and adaptive immune responses are linked with CD pathogenesis. However, the specific mechanism behind this is still not fully elucidated.

1.2.3 Environmental factors

Numerous epidemiologic studies have revealed an increased incidence of CD and geographic variation in response to westernized lifestyle and industrialization over the past decades, indicating an environmental influence on the pathogenesis of IBD

(Molodecky et al., 2012; Ng et al., 2013a; Ng et al., 2013b). It is suggested that environmental factors can influence intestinal permeability (Lewis and Abreu, 2017; Celiberto et al., 2018), dysregulate the intestinal mucosal immune system (Xu et al., 2014), and modulate the intestinal microbiota (Hasan and Yang, 2019), thus playing an important role in IBD (O'Toole and Korzenik, 2014; Rogler et al., 2016; Altajar and Moss, 2020). These environmental risk factors include smoking, hygiene practice during childhood, lifestyle, diet, geographical variation, microbiota, influences of infections and medications etc. In childhood, factors such as the hygiene hypothesis is supposed to have a significant influence on the pathogenesis of CD. It has been hypothesized that due to industrialization, childhood hygiene practices have changed, leading to decreased exposure to microbial agents from the environment and causing dysbiosis or an imbalance between commensal and pathogenic intestinal microbiota (Bernstein, 2010). The alteration in the intestinal microbiota could be associated with the initiation and perpetuation of CD.

Lifestyle factors, such as smoking have also been identified as risk factors for CD (Higuchi et al., 2012), and a greater number of CD patients smoke than the general population (Cosnes, 2010). Stress, depression, and anxiety have also been associated with CD (Mawdsley and Rampton, 2005; Walker et al., 2008; Graff et al., 2009; Goodhand et al., 2012). However, these could be both the cause and consequences of CD. It has been found that acute experimental stress increased mucosal inflammation, intestinal permeability and altered the bacterial-host relationship both in patients with IBD and in animal models of colitis (Mawdsley and Rampton, 2005). Study suggested that stress stimulated the dysfunction of the hypothalamic-pituitary-adrenal axis and altered interactions of gut mucosa and luminal bacteria resulted in the gastrointestinal inflammation (Mawdsley and Rampton, 2005).

Food is also implicated in the pathogenesis of CD as it can influence mucosal inflammation by acting as an antigen and by altering the intestinal microbiota. During the two last years, several papers have investigated the role of industrialized food in CD (Cui et al., 2021; Narula et al., 2021; Chassaing et al.,2022), linking industrialized countries and prevalence of CD.

The Western diet rich in fat and refined carbohydrates and low in fibre has been hypothesized to play a role in CD pathogenesis (Loftus, 2004; Thia et al., 2008). A study found that increased dietary intake of fibre was associated with a decreased risk of CD, and fibre from vegetables or fruits had the most significant influence in reducing the risk compared to fibre from cereals or whole grains (Ananthakrishnan et al., 2013). Another important part of the lifestyle factor associated with CD is exercise (Narula and Fedorak, 2008; Perez Pérez, 2009). Studies have reported a lower prevalence of IBD among German people with occupations that needed more outdoor physical activity (Sonnenberg, 1990).

However, environmental risk factors have not fully illuminated the actual and specific cause of CD and may only partially contribute to disease pathogenesis. Therefore, future studies are required to understand better the effects of environmental factors on the onset and progression of CD. In 70% of discordant monozygotic twins, the first-born child got IBD (Spehlmann et al., 2008). This highlights the critical role of environmental factors alongside genetic predisposition (Spencer et al., 2020).

1.3 Clinical features of CD

CD can affect any part of the gastrointestinal tract from mouth to the anus, although the most common inflammation site is the terminal ileum, caecum, perianal area, and colon (Khor et al., 2011). In 1991, to categorise CD, a classification was proposed

based on anatomical sites, history of surgery, and clinical behaviour (inflammation, fistula, stenosis and other complications of CD). However, this classification was felt unsuitable for clinical use, and in 1998 the World Congress of Gastroenterology in Vienna proposed a review and revision of this classification (Gasche et al., 2000). The Vienna classification of CD has taken the age of onset (A), anatomic location of disease (L), and behaviour of disease (B) into consideration as the major phenotypic features. In 2003, the Montreal Working Party established a revised classification system for CD (Satsangi et al., 2006). The Montreal revision of the Vienna classification has not changed the three primary parameters (age of onset, site of inflammation, and behaviour), but revisions were made within each of these groups. The Montreal classification permits categorising the early onset of disease as A1 group for those patients diagnosed at 16 years old or younger, whereas A2 and A3 are those who were diagnosed at 17–40 years and 40 years or older, respectively (Table 1.1).

Table 1.1. Vienna and Montreal classification for Crohn's disease

	Vienna	Montreal
Age at diagnosis	A1 below 40 y	A1 below 16 y
	A2 above 40 y	A2 between 17 and 40 y
		A3 above 40 y
Location	L1 ileal	L1 ileal
	L2 colonic	L2 colonic
	L3 ileocolonic	L3 ileocolonic
	L4 upper	L4 isolated upper disease*
Behaviour	B1 non-stricturing, non-penetrating	B1 non-stricturing, non-penetrating
	B2 stricturing	B2 stricturing
	B3 penetrating	B3 penetrating
		p perianal disease modifier†

*L4 is a modifier added to L1–L3 when concurrent upper gastrointestinal disease is present. †“p” is added to L1-L3 when concomitant perianal disease is present. Adapted from Satsangi et al. (2006).

This adjustment allows several studies to represent those specific phenotypes that are more commonly found in early-onset Crohn's disease (Uhlig et al., 2014; Snapper, 2015; Ashton et al., 2017; Shim, 2019).

The most common gastrointestinal symptoms in CD are diarrhoea, abdominal pain, unexplained fever and unexplained weight loss. Extra-intestinal manifestations include anaemia, skin disorders, osteoporosis, arthritis, ocular inflammation, erythema nodosum, pyoderma gangrenosum, and occasionally gallbladder or liver disease. CD is characterised by transmural inflammation (affecting all layers from mucosa to serosa), deep ulcers and fissures. In some cases, affected areas have a ‘cobblestone appearance’ (see Figure 1.1 and Figure 1.2). The inflammation in CD occurs in a discontinuous fashion, leaving damaged segments of mucosa interspersed with normal sections referred to as ‘skip’ lesions. The microscopic features of CD include thickened submucosa, and non-caseating granulomas are found in approximately 40% of cases. Granulomas are aggregates of immune cells known as epithelioid macrophages forming tiny nodules at the infection or inflammation site and surrounded by lymphocytes. Other complications commonly associated with CD are strictures, abscesses, fistulas (abnormal tunnelling of tissue between any two visceral organs or the skin) and stenosis (narrowing of the GI tract). Treatment in CD often involves surgical intervention.

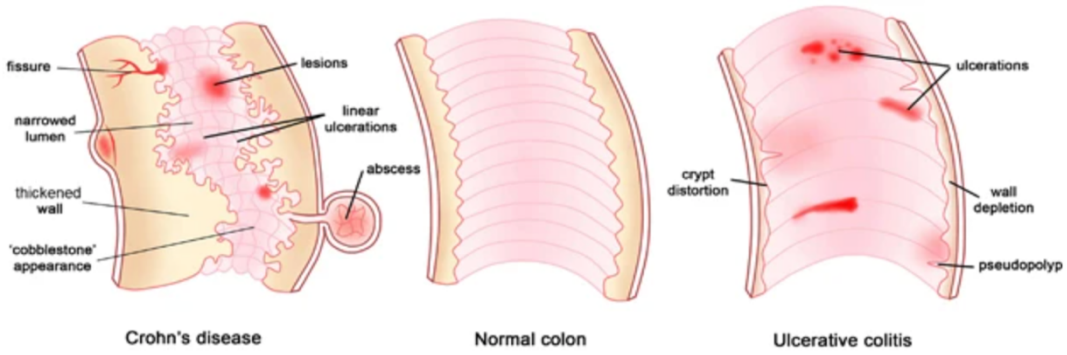
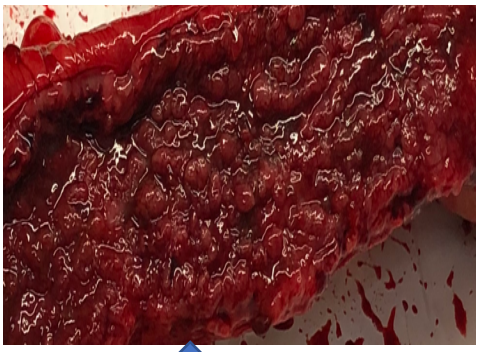


Figure 1.1. Comparative pathological Features of CD and UC. Source: *In vivo* reprogramming in inflammatory bowel disease (Wagnerova and Gardlik, 2013)

A

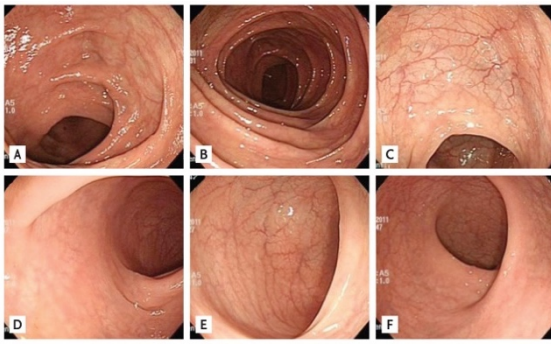
.



Involved mucosa/ cobblestone appearance

B

.



Normal mucosa throughout the gut (A, B, C, D, and E) the colon and (F) sigmoid colon (Hyun et al., 2013).

Figure 1.2 A. Bowel resection from a patient with CD showing a 'cobblestone' appearance. B. normal mucosa of the gut.

1.4 Clinical features of ulcerative colitis (UC)

The most consistent feature of UC is inflammation that is restricted to the colon, start from the rectum and spreads proximally in a continuous pattern to involve other parts of the colon. Histologically unlike CD, UC causes superficial inflammation typically restricted to the mucosa and submucosa with cryptitis and crypt abscesses. Symptoms may include abdominal cramping, malnutrition, tenesmus and diarrhoea with blood and mucus. In contrast to CD, other complications like fistulae, stenosis and granulomas are rare in UC. However, very few people with UC can develop inflammation in other body parts, like the liver, skin, joints, and eyes. Indeterminant colitis (IC) is used to describe any condition that cannot be distinguished clinically from UC or CD based on diagnostic tests.

1.5 Overview of paediatric CD

A new classification approach by Levine et al. (2011) (the Paris classification) was implemented to determine dynamic features of paediatric IBD phenotypes based on disease location and disease activity, and growth failure. It has improved the classification approach in children and established a uniform standard for defining IBD phenotypes. According to this classification, age at diagnosis is used to categorise paediatric IBD as very early-onset (VEO, 0 to < 10 years) and early onset (EO, 10 to < 17 years). VEO phenotypes have a greater tendency to produce disease lesions in the colon, extensive in nature, and a change in the location of disease manifestation over time. However, the clinical features of EO phenotypes are similar to adult-onset IBD. And, like adult CD, the incidence of CD in childhood is rising worldwide (Benchimol et al., 2017; Sýkora et al., 2018).

The paediatric CD is considered as a more complex disorder with distinct heterogeneity in onset, disease course, presentation and progression, and extensive intestinal involvement (Paul et al., 2006; Van Limbergen et al., 2008; Pigneur et al., 2009; Haberman et al., 2014; Ledder et al., 2014; Rigoli and Caruso, 2014; Herzog et al., 2017). Previous reports have identified distinct differences in paediatric CD phenotypes compared to adult CD (Van Limbergen et al., 2008; Sagiv-Friedgut et al., 2010; Lovasz et al., 2013; Ostrowski et al., 2016). Numerous studies suggest that this phenotypic variation between early-onset and adult-onset IBD is due to the differences in genetics, microbiome composition and possibly environmental factors (Thompson et al., 1996; Orholm et al., 2000). There is an increased risk for paediatric IBD in patients with some genetic defects that causes intestinal inflammation (Okou and Kugathasan, 2014), and many studies have reported a higher rate of family history with very early onset CD (Paul et al., 2006). Earlier disease onset in the progenies of parents with CD has been established in previous studies (Bengtson et al., 2009a; Moller et al., 2015). Several gene variants were associated with early onset CD; variants in chromosome 16 IBD1 susceptibility locus have been associated with CD pedigrees at the early onset of disease (Brant et al., 2000). A particular variant in NOD2 was linked with early-onset CD compared with patients with adult CD (Posovszky et al., 2013; Okou and Kugathasan, 2014; Shaw et al., 2019; Horowitz et al., 2021). Recently, it has also been found that a defect in the NADPH oxidase gene (NCF4) was associated with a very early onset CD (Wright et al., 2019). Moreover, very early onset IBD is more likely to be associated with a spectrum of rare Mendelian disorders, and environmental factors are less critical. In contrast, adult-onset IBD possesses a strong association with the environmental components (Snapper, 2015). GWAS and meta-analysis based on adult-onset CD cases have identified more than

30 risk alleles describing approximately 20% of the genetic variance in CD (Rioux et al., 2007b; Barrett et al., 2008). A study has determined the genotypes and phenotypes of patients with paediatric-onset CD and compared them with adult-onset CD and controls (de Ridder et al., 2007; Ostrowski et al., 2016). Phenotypic variability between adult and paediatric CD is well reported, but the specific mechanisms behind these differences are still unclear. This heterogeneity is supposed to be largely driven by heredity (Colombel et al., 1996; Satsangi et al., 1996) and genotype-phenotype correlations (Leshinsky-Silver et al., 2005; de Ridder et al., 2007; Levine et al., 2007). Several CD susceptibility alleles are present in patients with adult and paediatric CD (Biank et al., 2007; Dubinsky et al., 2007; Peterson et al., 2008). However, no one factor is associated with a particular age of onset.

1.6 Diagnosis of CD and/or IBD

A diagnosis of CD may take months or years from the onset of symptoms, as it often takes time to exclude other diseases such as bowel infections or irritable bowel syndrome (IBS). For appropriate diagnosis, colonoscopy is used as a reliable tool to evaluate symptoms of a flare of the disease. Intestinal mucosal biopsies are taken during colonoscopy to identify pathological changes, including inflammation, ulceration, and granulomas. In addition to that, blood tests (tests for anaemia or infection), endoscopy, sigmoidoscopy, gastroscopy, X-rays, CT scan, magnetic resonance imaging (MRI) and small bowel series are used in the diagnosis and monitoring of CD. No single test can reliably diagnose all the cases of CD, and most individuals require several other tests such as estimation of faecal calprotectin and inflammatory markers, such as C-reactive protein (CRP). Faecal calprotectin is elevated during disease activity as it correlates with intestinal inflammation. Results of a meta-analysis showed 89-98% sensitivity and 81-91% specificity for measurement

of calprotectin in faeces (von Roon et al., 2007). In addition to that, an increased level of faecal lactoferrin is also indicative of IBD (Gisbert et al., 2009). Therefore, doctors usually use a combination of tests to diagnose and confirm Crohn's disease.

1.7 History of CD

The IBD was described by Giovanni Battista Morgagni and by Scottish surgeon T Kennedy Dalziel in 1913 (Kirsner, 1988). Crohn's disease was named after the renowned gastroenterologist, Dr Burrill Crohn, and it became an official medical entity in 1932 (Naser et al., 2012a). Crohn and colleagues reported fourteen cases in 1932 and published the article to the American Medical Association under the title of "Terminal ileitis: A new clinical entity". However, the first description of Crohn's was given by Giovanni Battista Morgagni in 1769, who diagnosed a patient with typical regional ileitis and suffering from a debilitating condition with chronic diarrhoea. After the death of that patient, Morgagni performed his autopsy and described characteristic CD lesions in the intestine, including ulcerations, erosions, transmural inflammation, perforations and lymphadenopathy (Mulder et al., 2014). John Berg described reports of similar cases in 1898 and Antoni Lesniowski in 1904. Crohn et al. (1932) gave the first detailed report of 'regional ileitis'. However, they could not identify the aetiology of the disease and eliminated several infectious and other potential causes of CD, including recognized granulomatous and other known inflammatory conditions (Crohn, et al., 1932).

1.8 Therapeutic management of CD

The treatment of CD depends on activity and severity of disease, location of inflammation, and subtype of infection (i.e., inflammatory, stricturing, or penetrating), side effects of the disease, and patient's response to treatment. Over the last decades, several treatments have been developed with the aim of reducing inflammation or maintaining remission. They include immunosuppressive agents such as azathioprine (AZA), 6-mercaptopurine (MP) or methotrexate (MTX), and monoclonal antibodies, such as anti-tumour necrosis factor (anti-TNF) infliximab or adalimumab. Among several pro-inflammatory cytokines, TNF has been the prime focus. This cytokine is supposed to trigger the initiation and progression of chronic inflammation in CD by upregulating the transcription of other pro-inflammatory cytokines and stimulating the phagocytic activity of macrophages (Neurath, 2014). The first use of anti-TNF therapy, such as neutralising antibody against TNF (infliximab) in patients with CD resulted in significant improvement and healing of the mucosal inflammation (Van Dullemen et al., 1995), considered as a major success in CD therapy. In that study, a total of eight patients had complete healing of the ulceration and normalization of CD Activity Index scores within four weeks of anti-TNF therapy. After that treatment, one patient developed a perforation and entirely recovered after surgery. Unfortunately, one elderly CD patient exhibited a poor response to the therapy (Van Dullemen et al., 1995).

Some antibiotics such as amoxicillin, metronidazole, ciprofloxacin, and azithromycin are also used to treat CD, but studies supporting their effectiveness are limited (Arnold et al., 2002; Su et al., 2015; Gomollón et al., 2017). A recent study based on 13 randomised controlled trials (RCTs) could not draw a firm conclusion regarding the role of antibiotics in maintaining remission in CD (Townsend et al., 2019).

Some anti-mycobacterial antibiotic therapy (AMAT) trials targeting atypical mycobacteria have demonstrated efficacy in CD, with higher clinical remission rates (Borody et al., 2002; Agrawal et al., 2020). However, another study did not observe the sustained benefit of using combined antibiotic therapy against *Mycobacterium avium* subspecies *paratuberculosis* and does not support a significant role of this bacteria in the pathogenesis of CD (Selby et al., 2007). The critical function of antibiotics is to treat the abscess, fistula and other perianal complications of CD (Lal and Steinhart, 2006; Rahimi et al., 2006; Aguilera-Castro et al., 2017). Antibiotics are used to treat infections in perianal CD as well as in the management of other infectious complications, like abscesses. However, no steady and sustained recovery has been reported with antibiotic treatment alone (Vavricka and Rogler, 2010). Ciprofloxacin and metronidazole are the most studied antibiotics used to treat perianal CD and fistula, and appear to have short-term benefits either alone or in combination (Wolf, 1990; Solomon et al., 1993).

The newly approved drugs are monoclonal antibodies targeting specific integrins or interleukins (IL-12/IL-23). For example, Ustekinumab is a monoclonal antibody that blocks the p40 subunit of IL-12 and IL-23 to stop interaction between cell surface receptors and cytokine activation. Ustekinumab has been tested and proven effective to induce and maintain clinical remission in CD (Sandborn et al., 2008; Sandborn et al., 2012; Feagan et al., 2016). In addition, Vedolizumab, a monoclonal antibody against α4β7 integrin that inhibits lymphocyte trafficking, has been effective in treating UC (Ha and Kornbluth, 2014; Mühl et al., 2021). However, a phase III trial showed that at week 6, patients with refractory CD on vedolizumab were no better off than individuals who received the placebo (Sands et al., 2014). Treating patients with refractory CD using haematopoietic stem cell transplant (HSCT) can be an option for

exceptional beneficial outcome (Burt et al., 2010; Snowden et al., 2014). Some recent trials have shown exciting results for HSCT therapy as an alternative therapy for patients with inflammatory bowel diseases (Hawkey and Hommes, 2017).

1.9 Surgical management of CD

Over the last few decades, the therapeutic management of patients with CD has significantly improved (Hazlewood et al., 2015; Singh et al., 2015; Kaplan and Ng, 2017). However, surgery continues to play a major role in the management of complications in CD, specifically for patients with the aggressive and penetrating disease, complete bowel obstruction, or who have failed to respond to medical treatment (Fichera and Michelassi, 2007). Some risk factors associated with aggressive CD include early age of diagnosis (<30 years), extensive involvement of intestine, deep ulceration, previous history of surgery, stenosis, stricture and penetrating disease (Dassopoulos et al., 2013). The most significant challenge in the management of CD is that following 20 years of disease activity, 80% of the CD patients will have had surgery, and around 30% will need surgery within five years of diagnosis (Bernstein et al., 2012; Cheifetz, 2013). A randomised postoperative Crohn's endoscopic recurrence (POCER) trial was conducted between 2009 and 2011 based on patients with CD undergoing intestinal resection at 17 hospitals in Australia and New Zealand (De Cruz et al., 2015). In that study, patients were randomly selected and divided into two groups: colonoscopy at six months of surgery (active care) or no colonoscopy after surgery (standard care). They enrolled 174 patients (122 active care, 52 standard care) who have taken at least one dose of the study drug. After the primary endpoint, at 18 months, endoscopic recurrence was observed in 49% of the patients in the active care group and 67% of the patients in the standard care group (p=0.03) (De Cruz et al., 2015). The POCER study outcomes reveal that treatment

based on risk of recurrence, treatment step-up for recurrence and early colonoscopy is more efficient compared to other available drug therapies alone to effectively prevent postoperative CD recurrence (De Cruz et al., 2015).

Recently a study identified that microbial factors are likely to be associated with recurrence of CD after bowel resection (Hamilton et al., 2020). Surgery is not curative for CD; in some cases, patients will need several surgeries throughout their life (Bernstein et al., 2012). The most common causes for surgical resection in CD are perforation, recurrent obstruction, stenosis, abscess, intractable haemorrhage, and dysplasia. In addition, endoscopic recurrence (Rutgeerts' score ≥i1) may reappear at the neo-terminal ileum in 30%-93% of patients within 12 months of surgery (Rutgeerts et al., 1984; Rutgeerts et al., 1990; Vantrappen and Rutgeerts, 1990; Achkar, 2004; Buisson et al., 2012; Gajendran et al., 2018).

1.10 Role of gut microbiota in CD

The gastrointestinal tract harbours abundant, diverse, and complex microbial communities that play a critical role in human health. It has been estimated that the microorganisms inhabiting the GI tract exceed 100 trillion (Thursby and Juge, 2017), 90% of which are bacteria. The human gut microbiome is composed of 1000 to 1150 common bacterial species, and each individual harbours at least 160 species (Qin et al., 2010). Most of these microbes have a significant influence on physiology and nutrition and are critical for human life (Bäckhed et al., 2005; Clemente et al., 2012); therefore, also known as "forgotten organ" (O'Hara and Shanahan, 2006). The microbiota is linked with several functions such as food fermentation and absorption of nutrients (Gill et al., 2006). Some bacteria in the gut can ferment fibres and yield

short-chain fatty acids like acetate, propionate and butyrate. They play a vital role in controlling intestinal homeostasis by strengthening the intestinal barrier function (Parada Venegas et al., 2019). Moreover, the gut microbiota stimulates the immune system, promotes the maturation of immune cells and develops immunological functions (Chow et al., 2010; Gensollen et al., 2016; Zheng et al., 2020). In healthy individuals, the microbiome and host have a symbiotic relationship and play a unique metabolic function (i.e., digestion of dietary compounds and fermentation of indigestible compounds). In addition, gut microbiota can significantly influence the metabolism and modulation of some specific drugs and control the efficacy of medication prescribed for some diseases (Crouwel et al., 2021). Throughout the life of an individual, the gut microbiota is influenced by disease status, age, diet, geographical location, intake of food and drugs, consumption of antibiotics; and other environmental factors (Yatsunenko et al., 2012; Greenhalgh et al., 2016). It is important to note that changes in the physiological condition along the gut like acidity, pH and transit time influences the distribution, diversity, and concentration of different bacterial phyla.

In a healthy gut, the major gut microbial phyla are the Firmicutes and Bacteroidetes, followed by Proteobacteria and Actinobacteria (Arumugam et al., 2011). The Firmicutes and Bacteroidetes, on average, compose 90% of gut microbiota, but their relative abundance varies widely across healthy individuals (Huttenhower et al., 2012). Several studies have tried to define a healthy gut microbiota, with some previous research on the microbiome identifying “core” microbial taxa universally present in healthy persons. It is assumed that the loss of these “core” microbes could initiate dysbiosis (Turnbaugh et al., 2007). However, other studies show that no taxa are universally present among all healthy individuals. Only a few studies described the

healthy hallmarks of the intestinal microbiota with a predominance of members of the Firmicutes (approximately 50%), Bacteroidetes (about 40%), and Proteobacteria (around 10%) (Hong et al., 2011). According to Jandhyala and colleagues, Firmicutes, Bacteroidetes, Actinobacteria, and Verrucomicrobia phyla are the most common components of the healthy intestinal microbiota (Jandhyala et al., 2015). Some of the Firmicutes, especially *Faecalibacterium prausnitzii,* a dominant species of healthy gut benefits their host by producing anti-inflammatory metabolites (Maioli et al., 2021).

Apart from host-derived factors, several lines of evidence implicate bacteria in the pathogenesis of CD. The terminal ileum and colon, the predominant site for CD, harbour the highest concentrations and diversity of bacteria in the gut (Flint et al., 2012); diversion of the faecal stream reduces pathology, whilst reinfusion reactivates inflammation (Rutgeerts et al., 1991; D'Haens et al., 1998); oral antibiotics improve the clinical manifestations in some CD patients (Feller et al., 2010; Su et al., 2015; Nitzan et al., 2016); antibiotics affect recurrence and remission rates (Doherty et al., 2010); gene variants most significantly associated with CD are in genes involved in bacterial recognition [NOD2 mutation (Kennedy et al., 2018)]; autophagy [ATG16L1 (Salem et al., 2015; Wildenberg et al., 2017)],; and signalling in response to bacteria [IRGM] (Naser et al., 2012b; Glas et al., 2013a)]; granulomas contain bacteria (Ryan et al., 2004) and bacteria are isolated significantly more frequently from CD mesenteric lymph nodes (33% - 48%) than control nodes (5% - 15%) (Ambrose et al., 1984; O'Boyle et al., 1998; Takesue et al., 2002a; O'Brien et al., 2014; Kiely et al., 2018b). Elevated antibody levels to microbial antigens have also been associated with a higher chance of minor bowel surgery in patients with CD (Mow et al., 2004). Results from animal studies also provide evidence that commensal enteric bacteria deliver a continuing stimulus for spontaneous or induced chronic, immune-mediated colitis and

ileitis (Strober et al., 2007; Mizoguchi and Mizoguchi, 2008; Sartor, 2008; Bloom et al., 2011; Loh and Blaut, 2012; Nakanishi et al., 2015). The role of commensals in triggering intestinal inflammation is also evident in other studies where germ-free susceptible mice with defective immune system typically cannot develop colitis until non-pathogenic commensals are introduced (Veltkamp et al., 2001; Scharl and Rogler, 2010). Germ-free mice have less developed immune systems with fewer Peyer's patches, mesenteric lymph nodes, spleen, and impaired mucosal immune function. The existence of luminal, mucosa-associated and/or translocating bacteria could trigger immune responses to induce and perpetuate intestinal inflammation in genetically susceptible hosts. Microbial dysbiosis (an alteration of the gut microbial community) has been suggested to be a crucial factor involved in the pathogenesis of CD (Abraham and Medzhitov, 2011; Joossens et al., 2011; Saleh and Elson, 2011; Nagalingam and Lynch, 2012; Gevers et al., 2014; Chamaillard and Radulovic, 2016; Pascal et al., 2017; Caparrós et al., 2021)

Numerous studies have demonstrated an alteration in the gut microbiome of patients with CD, including a reduction in bacterial richness and diversity. Alpha diversity is the measurement of the number of distinguishable taxa in a community. Beta diversity is the variation of microbial communities between samples, with a primary focus on the difference in taxonomic abundance profiles. Patients with CD also displayed a reduced alpha diversity in the faecal microbiome compared with healthy controls (Manichanh et al., 2006; McIlroy et al., 2018; Galazzo et al., 2019; Kowalska-Duplaga et al., 2019; Clooney et al., 2020; Machiels et al., 2020) as well as decreased diversity in inflamed mucosa compared to non-inflamed mucosa in same patient with a reduced bacterial load in the inflamed tissue in patients with CD (Sepehri et al., 2007).

A consistent finding of the dysbiosis in CD consists of the alterations in the proportion of Bacteroides and Firmicutes and an increase in the abundance of members of the Proteobacteria (particularly *Escherichia coli*) compared to controls (Kostic et al., 2014; McIlroy et al., 2018). A decrease in the abundance of Firmicutes, more particularly the Clostridium cluster IV, is common in CD (Manichanh et al., 2006; Frank et al., 2007; Sokol et al., 2008). Metagenomic studies of gut microbiota in CD patients have also shown a marked shift in microbial abundances compared to healthy controls, characterized by enrichment in pathobionts from Bacteroidetes and enterobacteria with a simultaneous reduction of Firmicutes, Bifidobacteria, and Clostridia (Gevers et al., 2017). An expansion in the abundance of Proteobacteria has also been exhibited constantly in studies (Gophna et al., 2006; Baumgart et al., 2007; Lupp et al., 2007; Martinez-Medina et al., 2009; Willing et al., 2009; Kiely et al., 2018b; Machiels et al., 2020). Reports on the abundance of Bacteroides in CD have shown conflicting results; some show increases (Dicksved et al., 2008), and others show decreases in CD (Takahashi et al., 2016; Alam et al., 2020). In previous studies, *Faecalibacterium prausnitzii* was significantly and consistently decreased in abundance in CD patients than healthy individuals (Baumgart et al., 2007; Frank et al., 2007; Sokol et al., 2009; Willing et al., 2009; Mondot et al., 2011; Hansen et al., 2012; Fujimoto et al., 2013; Lopez-Siles et al., 2014; Takahashi et al., 2016; Kowalska-Duplaga et al., 2019). However, the importance of alterations in the microbiota is still under investigation. It is not clear yet, how changes in the gastrointestinal microbiota could be involved in the pathogenesis of CD.

1.11 Individual bacterial species implicated in CD

Although many studies have been conducted to look for a specific microbial cause of CD, no infectious agent has been confirmed (Eckburg and Relman, 2007; Knights et

al., 2014). Metagenomic studies comparing healthy controls and CD patients have not always revealed consistent outcomes due to different types of samples used in these studies. It is clear that there are significant differences in microbial communities isolated from intestinal mucosal biopsies and faecal samples (Zoetendal et al., 2002; Tang et al., 2020). Applications of different molecular techniques have clarified the massive complexity of the gut microbiota. Still, the individual diversity of the microbiome complicates the search for a specific microbe that causes CD. Previous studies were undertaken on several microorganisms to know the role of these microorganisms in the pathogenesis of CD, but none could provide a strong association. The most investigated bacterial species are described below.

1.11.1 *Mycobacterium avium* subspecies paratuberculosis (MAP)

Many studies support and refute the role of MAP as a causative agent for CD (Chiodini et al., 2012; Lamont et al., 2012; Liverani et al., 2014; McNees et al., 2015; Oken et al., 2017; Zarei-Kordshouli et al., 2019). A meta-analysis revealed that the prevalence of MAP DNA was higher in patients with CD than controls in 16 of 18 studies (Feller et al., 2007). However, many studies had conflicting results and failed to provide convincing evidence of a causal role of MAP in CD (Parrish et al., 2009; Sasikala et al., 2009; Ricanek et al., 2010). An investigation to find the associations between MAP and single nucleotide polymorphisms [rs2241880 (ATG16L1) and rs10045431 (IL12B) in host linked with CD was carried out on 84 MAP-positive CD patients; however, no evidence of association was identified (Dalton et al., 2014). Mycobacterium was also not detected in bowel resected samples (Lymph nodes, involved and uninvolved mucosal biopsies) from 58 patients with CD (O'Brien et al., 2014). Moreover, no association was confirmed in patient with CD between NOD2 gene (a gene significantly associated with CD) polymorphism and MAP serology (Bernstein et al.,

2007). Mycobacterial infections are a well-known consequence of anti-TNF use, while they are very effective in CD, long-term treatment with antibiotics against MAP did not improve the risk of relapse. Therefore, the role of MAP in the pathogenesis of CD remains a controversial subject.

1.11.2 *Helicobacter* spp.

A role of *Helicobacter spp*. in IBD has also been suggested (Papamichael et al., 2014). In paediatric CD patients, *Helicobacter spp*. were more likely to be detected (41.5%) than controls (22.5%) (Kaakoush et al., 2010). Several studies have reported a lower *H. pylori* infection rate in CD and/or UC compared to controls (Sonnenberg and Genta, 2012; Xiang et al., 2013; Wu et al., 2015; Bartels et al., 2016; Castaño-Rodríguez et al., 2017; Shah et al., 2017). *H. pylori* infection might also protect against CD development (Bartels *et al.*, 2016), and a recent epidemiological study suggests an inverse association between *H. pylori* and CD (Shah et al., 2017). However, many studies did not find evidence of association of *Helicobacter Pylori* with CD (LAHARIE et al., 2009; Kaakoush et al., 2010; Lahat et al., 2017). In addition, higher prevalence of *H. pylori* negative gastritis among paediatric CD patients (Sharif et al., 2002; Genta and Sonnenberg, 2012) other bacteria may play a causative role in the development of gastritis in those patients.

1.11.3 *Campylobacter* spp.

Campylobacter spp. have recently been found to be linked with CD (Guslandi, 2009). A significantly higher prevalence of *Campylobacter concisus* in intestinal biopsies of adult and paediatric CD patients were identified than controls in two different studies

(Zhang et al., 2009; Mahendran et al., 2011). Another study also reported a higher prevalence of *C. concisus* in stool samples in children with CD (Man et al., 2010). Although the confirmed clinical association of *C. concisus* remains undecided, a review reported that C. concisus could be an emerging pathogen with putative roles in intestinal infectious diseases (Newell, 2005). *Campylobacter jejuni* has also been implicated in CD (Chiba et al., 1990) and flare-ups in IBD (Newman and Lambert, 1980). However, another study did not etiologically implicate *C. jejuni* in CD (Blaser et al., 1984). Future studies are necessary for uncovering the association to determine the role of *Campylobacter spp*. in the pathogenesis of CD.

1.11.4 *Clostridioides difficile*

Clostridioides difficile (*C. difficile*) is a spore-forming gram-positive bacteria that produces two toxins: toxin A (enterotoxin) and toxin B (cytotoxin). It is an important cause of antibiotic-associated diarrhoea, gastrointestinal illness, colitis, and even death. Recently this bacterium is also reported in having a higher rate of infection in IBD (Musa et al., 2010; Bien et al., 2013; Rodríguez et al., 2020). Several reports stated that *C. difficile* infection (CDI) incidence and severity is increased in CD patients. *C. difficile*-associated small bowel enteritis was also found in a CD patient after total proctocolectomy (Kurtz et al., 2010). This CDI may develop as *C. difficile* is highly resistant to most antibiotics. Under normal conditions, *C. difficile* is suppressed by the normal gut bacteria. However, when normal gut bacteria are reduced due to the antibiotic intake, *C. difficile* starts to proliferate and leads to colitis development (Fordtran, 2006). The toxins of *C. difficile* may cause damage to the intestinal mucosa. In patients with CD, the proliferation of this bacteria may result from dysbiosis that reduces the protective and healthy microbiota and produce a favourable environment for *C. difficile* infection. Some predisposing factors like antibiotic overuse (Diggs and

Surawicz, 2009) and corticosteroid therapy seemed to contribute to *C. difficile* colitis in patients with IBD (Schneeweiss et al., 2009). However, no certain evidence has explored the role of *C. difficile* in the onset of CD.

1.11.5 *Listeria* and *Yersinia* spp.

Listeria spp. are intracellular Gram-positive intracellular bacilli belonging to the Firmicutes phylum, Listeriaceae family, and has been found in a higher proportion in CD patients than controls in some reports (Kang et al., 2010; Magin et al., 2013). *Listeria monocytogenes* has been suggested as a potential cause of CD. However, there are several data elements that militate against the aetiologic significance of *L. monocytogenes* in CD (Chiba et al., 1998; Chen et al., 2000). Another question regarding the association of *L. monocytogenes* and CD is that, is it primary or secondary? Anti (TNF)-α therapy may increase the chance of *L. monocytogenes* infection in patients with CD (Ramos et al., 2010). Therefore, it is presumed that since patients with CD are treated with immunosuppressive drugs, which may lead to the deprivation of immunological protective barrier; thereby, the patient could be more vulnerable to the colonisation of *L. monocytogenes* (Miranda-Bautista et al., 2014). In general, to disclose the direct or indirect association between *L. monocytogenes* and CD, more investigations are needed.

Yersinia is a Gram-negative bacillus belonging to the Proteobacteria phylum and Enterobacteriaceae family and has been detected in CD patients more often than control in some studies (Saebo et al., 2005; Leu et al., 2013). It has also been suggested that *Yersinia* species may play a role in the initiation or progression of gut inflammation in CD patients (Hugot et al., 2003). However, a recent study denied any

association of *Yersinia* with CD (Le Baut et al., 2018). Other genera of the family Enterobacteriaceae have a stable connection with CD (O'Brien et al., 2014).

1.11.6 Adherent-invasive *Escherichia coli* (AIEC)

Escherichia coli belongs to the Enterobacteriaceae family, a predominant group of the human gastrointestinal microbiota that plays a critical role in maintaining intestinal function. It comprises ~1% of a gut microbiome. *E. coli* can be assigned to one of eight phylogroups: seven (A, B1, B2, C, D, E, F) belong to *E. coli* sensu stricto, and the eighth is the Escherichia cryptic clade I (Clermont et al., 2013). This bacteria is usually a gut commensal; however, some strains of *E. coli* may acquire virulence properties, become pathogenic, and play a role in the pathogenesis of CD. Generally, most pathogenic and human-associated *E. coli* belong to the B2 and D phylogroup. No particular phylogroup of *E. coli* is definitively associated with CD, although phylogroups B2 and D are predominantly found in CD (Kotlowski et al., 2007a). Decades ago, in 1978, the first report regarding the association of *E. coli* with the pathogenesis of IBD was published by Tabaqchali *et al.*, where they have found a higher antibody titre against O-antigens of *E. coli* in patients with IBD (Tabaqchali et al., 1978) and by Keighley *et al*., exhibiting a significant increase in *E. coli* in the ileum and colon of patients with CD compared with controls (Keighley et al., 1978). Since then, many studies have shown an expansion of mucosa-associated *E. coli* population with CD (Darfeuille-Michaud et al., 1998; Darfeuille-Michaud et al., 2004; Martin et al., 2004; Kotlowski et al., 2007b; Willing et al., 2009; de Souza et al., 2012; Lopez-Siles et al., 2014; Proctor et al., 2019). The increased abundance of *E. coli* adherent to

inflamed mucosa of patients with CD has driven great interest in these bacteria over the last couple of decades.

A specific pathovar of *E. coli*, known as adherent-invasive *E. coli* (AIEC), has been implicated in CD. AIEC strains are characterized by two phenotypic criteria: (i) their ability to adhere to and invade intestinal epithelial cells, and (ii) survive and replicate intracellularly in macrophages (Boudeau et al., 2001; Darfeuille-Michaud et al., 2004). The higher prevalence of AIEC in CD patients compared to healthy subjects suggested the plausible association of *E. coli* invasiveness with the pathogenesis of CD (Boudeau et al., 2001; Darfeuille-Michaud et al., 2004; Martin et al., 2004; Baumgart et al., 2007; Martinez-Medina et al., 2009; Agus et al., 2014; Martinez-Medina and Garcie-Gil, 2014). However, the comparisons between several CD-associated strains and strains isolated from patients without IBD failed to detect any virulence genes unique to the CD-associated strains (Darfeuille-Michaud et al., 1998; O'Brien et al., 2016). In addition, the pathway used by AIEC to colonize the intestinal mucosa is not clearly known. The initial lesion in CD, the aphthous ulcer, can be observed macroscopically during colonoscopy. They overlie the follicle-associated epithelium of Peyer's patches in the small intestine and lymphoid follicles in the large intestine (Hizawa et al., 1994; Fujimura et al., 1996; Kinjo et al., 2004). M cells (membranous/microfold cells) are specialised epithelial cells present in the follicle-associated epithelium that transcytose antigens and microorganisms across the epithelial barrier, delivering them to underlying immune cells for processing (Kraehenbuhl and Neutra, 2000). Chassaing and colleagues demonstrated the ability of AIEC bacteria to interact with Peyer's patches and the virulence factors associated with this phenotype. They found that the prevalence of AIEC strains possessing the *lpf* operon was significantly higher in isolates from patients with CD than controls and

showed that long polar fimbriae plays a critical role in the attachment of AIEC to M cells (Chassaing et al., 2011).

A recent study found that ileal CD had a higher proportion of *Escherichia coli* in their lymph nodes, indicating their plausible association in initiation and perpetuation of inflammation in ileal CD (O'Brien et al., 2014). In AIEC, Type one pili can bind to host carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) (Barnich et al., 2007; Carvalho et al., 2009). CEACAM6 has been reported to be overexpressed in ileal CD tissue than healthy controls (Glasser et al., 2001; Barnich et al., 2007). In CD patients, upregulated expression of CEACAM6 on ileal enterocytes could be stimulated by the colonisation of AIEC in the ileal mucosa and through the stimulation of proinflammatory cytokines IFN-γ and TNF-α (Barnich et al., 2007). Since AIEC within macrophages induce the secretion of large amounts of TNF-α (Glasser et al., 2001), this bacterium can stimulate their self-colonisation in patients indirectly through the induction of TNF-α secretion from infected macrophages (Figure 1.3).

Another proposed mechanism by which AIEC could worsen colitis in patients with CD is through the increased expression of flagella. Flagellin can be detected by TLR5 and IPAF receptors of the host's immune system that trigger the innate immune response (Miao et al., 2006; Miao et al., 2007). Upregulated expression of flagella in virulent AIEC LF82 strain exacerbated an intestinal innate immune response in DSS-treated mice compared to nonpathogenic *E. coli* via increased expression of TLR5 (toll-like receptor 5) and IPAF flagellin receptors (Carvalho et al., 2008). Another study suggests that AIEC strains induce the secretion of pro-inflammatory cytokine IL-1β through an NLRP3-dependent mechanism that may contribute to CD pathogenesis (De la Fuente et al., 2014).

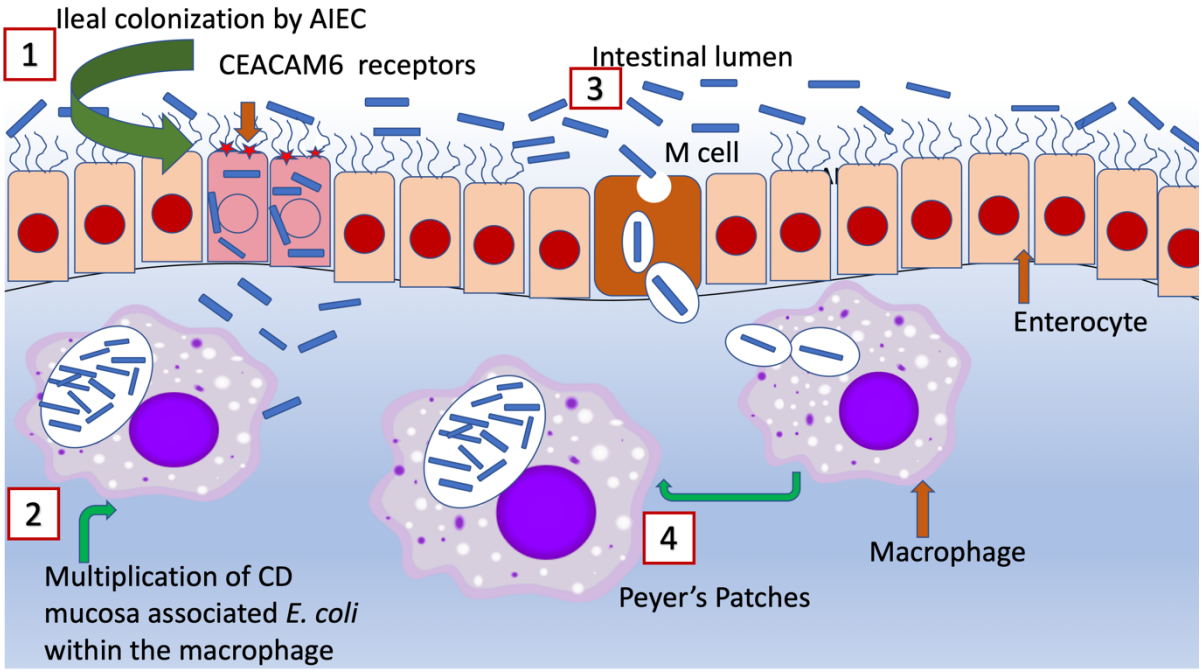


Figure 1.3. Invasion of enterocytes by mucosa associated AIEC in CD. (1) The colonisation of AIEC in the ileal mucosa is initiated through an interaction between AIEC and intestinal epithelial cells. AIEC attaches to the CEACAM6 receptor, which is increased in number in the ileum. (2) AIEC enters, survives, and replicates within the cell cytoplasm. (3) Using the invasive ability, AIEC crosses the intestinal barrier via intestinal epithelial cells or M cells, invades host cells, and translocates into Peyer's patches in the small intestine, and lymphoid follicle in the large intestine (4) AIEC is able to survive within resident macrophages and induces the secretion of large amounts of inflammatory cytokines such as TNF-α which leads to inflammation.

There are no unique molecular features that characterise AIEC; they are classified by their phenotypic properties: adherence and invasion of epithelial cells and their ability to proliferate in macrophages (O'Brien et al., 2016). Moreover, no particular virulence factor is exclusively present in AIEC that may enable them to identify by molecular techniques. Not all but most of the AIEC strains possess a type 1 pilus variant that allows them to adhere to epithelial cells; without it, they show much less adherence (Boudeau et al., 2001). Till now, six genes or their variants (fimH, pduC, lpfA, gipA, chuA, ompC) have been proposed as putative AIEC biomarkers (Rolhion et al., 2007; Chassaing et al., 2011; Iebba et al., 2012; Dreux et al., 2013; Dogan et al., 2014; Desilets et al., 2016; Vazeille et al., 2016; Céspedes et al., 2017; Camprubí-Font et

al., 2019). Acquired nonsynonymous substitutions in fimH, leading to a pathoadaptive point mutant of AIEC, enabled AIEC to colonise in the intestinal mucosa leading to the initiation and progression of inflammation in a genetically susceptible host (Dreux et al., 2013). Therefore, fimH SNP analysis could be used as an indicator of virulence in CD-associated *E. coli* strains for diagnosis. The study demonstrated that chuA (associated with iron uptake) significantly associated with AIEC strains than non-AIEC strains (Céspedes et al., 2017). Furthermore, OmpC, which regulates the expression of many virulence factors via the SigmaE pathway is supposed to play a vital role in AIEC for adhesion and invasion (Rolhion et al., 2007). The oxidoreductase, DsbA, and HtrA stress protein are also assumed to be involved in AIEC's ability to survive and replicate within macrophages (Bringer et al., 2005; Barnich et al., 2007). Future research is needed to identify distinguishing factors of AIEC that enable them to invade enterocytes as well as replicate within the macrophage of CD patients.

1.12 Bacterial translocation in CD

From the intestine, viable luminal endogenous bacteria can cross the mucosal barrier and migrate to the mesenteric lymph nodes and other internal organs, which is defined as bacterial translocation (Berg, 1999; Ignacio et al., 2019). Luminal enteric bacteria are thought to be the most significant inflammation-causing environmental factor in CD (Sartor, 1997; Swidsinski et al., 2002). The intestinal epithelial barrier is essential in controlling the uptake of luminal antigens and regulating the immune responses to luminal bacteria in the gut. Some clear evidence suggested that in CD, the gut epithelial barrier function is disrupted, allowing luminal bacteria to cross the barrier, stimulating the mucosal immune system in the submucosa. Polymorphisms in the bacterial recognition and autophagy associated genes lead to the abnormal immune

response to the invading bacteria. Consequently, these bacteria are not eliminated and trigger an ongoing stimulus leading to the chronic inflammation typical of CD.

Translocation is likely to occur at this location because mucin (the primary barrier for the luminal bacteria) is relatively thin. Musin layer is thinner in CD patients than in healthy subjects (Gullberg and Söderholm, 2006). However, the essential feature of these sites is the presence of M cells that overlie the Peyer's patches and lymphoid follicles. These cells act at the portal of entry to transport microbes across the epithelial cell layer to underlying lymphocytes and Dendritic cells (Neutra, 1999), thereby initiating an immune response. Studies have revealed the role of M cells in the mucosal immune system and suggested that microbial antigen uptake by M cells also triggers the systemic immune responses (Kobayashi et al., 2019). In addition to that, M cells are also increased in number in NOD2 knockout mice, implying increased bacterial translocation through the Peyer's patches in patients with this mutation (Chassaing et al., 2011).

Since the primary lesions, aphthous ulcers, in CD appear at Peyer's patches and lymphoid follicles, it seems likely that CD mucosa-associated microbes are translocated via the M-cells and consequently delivered to the underlying macrophages where they can resist killing within the phagolysosome and replicate (Bringer et al., 2005). Macrophages are able to mediate bacterial translocation to draining lymph nodes (Hooper et al., 2012; O'Brien et al., 2014), and it is evidenced that bacteria can be cultured from these lymph nodes in patients with CD (Ambrose et al., 1984; Laffineur et al., 1992; Takesue et al., 2002b; O'Brien et al., 2016).

Furthermore, damage to the intestinal barrier function can play a major role in bacterial translocation leading to intestinal inflammation in CD. The incidence of pathogenic bacteria in mesenteric lymph nodes was increased in patients. Results from some studies clearly demonstrated a higher rate of bacterial translocation in CD patients (Takesue et al., 2002a; O'Brien et al., 2016). Recent reports have revealed that surgical recurrence in ileocolic CD patients is markedly decreased when the Mesenteric lymph nodes and mesentery are removed during bowel resection (Coffey et al., 2016; Coffey et al., 2018), suggesting the probable role of the mesentery in CD. In addition to that, a study also found an increased incidence of bacterial translocation in fistulizing CD (Kroesen et al., 2000). Recently, Gutierrez et al. (2014) assessed the influence of bacterial DNA translocation on relapse in CD during remission. They found that the presence of bacterial DNA in the blood of CD patients increases the risk for relapse at six months.

To our knowledge, only one study has explored the microbial composition of mesenteric lymph nodes in CD using culture-independent methods using high throughput sequencing technique (O'Brien et al., 2014). They also detected Bacterial RNA in the lymph nodes of IBD patients (Kiely et al., 2018a). Culture-dependent reports also demonstrated bacterial translocation to mesenteric lymph nodes and mesenteric fat in CD (Laffineur et al., 1992; Peyrin-Biroulet et al., 2012). Four independent studies have demonstrated bacterial translocation in 28%- 75% of draining mesenteric lymph nodes from CD patients (Ambrose et al., 1984; Laffineur et al., 1992; O'Brien et al., 2014; Kiernan et al., 2019). Irrespective of culture-dependent or culture-independent techniques, constantly, the most common bacteria identified in the mesenteric lymph nodes in CD was *Escherichia coli* (Ambrose et al., 1984; O'Brien et al., 2014; Kiernan et al., 2019). Indeed, follicle-associated epithelial biopsies from

ileal CD patients showed an increased number of adherent non-pathogenic *E. coli* compared to controls (Keita et al., 2008). However, the mechanisms of invasion and translocation of *Escherichia coli* through the intestinal mucosa of CD patients remains unknown. It is suggested that some bacteria have mechanisms that enable them to escape the physical barrier and immunological defence mechanisms of the host. Moreover, it has been shown that AIEC possesses long polar fimbriae (LPF) that allow them to translocate across an in vitro model of follicle associated epithelium (FAE) and interact with murine and human Peyer's patches (Chassaing et al., 2011). These may cause an enhanced antigen presentation in Peyer's patches and trigger pro-inflammatory immune responses in the host (Salim et al., 2009; Powell et al., 2015). However, bacterial translocation in CD could be passive, getting entrance through the host cells by the primary lesions like aphthous ulcers that are part of the disease progression. Or via the active way, using one of the bacterial properties (pili, fimbriae) or through a novel strategy to gain access to host cells. However, clarifying whether bacteria are passively or actively translocating the gut barrier will be hard to prove. Investigators will be needed to determine this using in vitro systems or animal models.

1.13 Inflammation and role of the monocyte-macrophage cell lineage in CD

Antigen-presenting cells (APC) are cells that present antigens complexed with major histocompatibility complex (MHC) on their surfaces. APCs engulf and destroy foreign antigens like bacteria and are responsible for the immune responses against intracellular and extracellular pathogens. In this project, we are focusing mainly on macrophages rather than the other antigen-presenting cells, because the granuloma (aggregates of macrophages) is a feature in CD. Furthermore, the first susceptibility gene described for CD, NOD2, is expressed in monocytes, the precursors of macrophages. Monocytes are leucocytes or white blood cells produced in the bone

marrow, which circulate through the bloodstream for about one to three days to reach the tissues throughout the body where they differentiate into macrophages. They have many functions, including phagocytosis, antigen presentation, cytokine production and resolution of inflammation (Murray and Wynn, 2011; Zhang et al., 2021). Phagocytosis is the process of engulfing microbes and foreign particles, followed by digestion and destruction of the ingested material. During phagocytosis by macrophages, bacteria are engulfed by the pseudopodia (arm-like projections), which surround and trap the bacteria in a vessel called a phagosome. The phagosome fuses with a lysosome, forming a phagolysosome, leading to degradation. This phagosome contains a variety of antimicrobials, including lysozyme, proteases, and antimicrobial peptides. Other enzymes include hyaluronidase, lipase, collagenase, elastase, ribonuclease, and deoxyribonuclease. These play an essential role to prevent the spread of infection and degradation of microbes leading to cell death. Phagosomes can also produce reactive oxygen species (ROS), hydrogen peroxide, superoxides, hypochlorous acid, and nitric oxide that leads to the destruction of bacteria. Once the bacteria are killed, the bacterial particles are expelled from the macrophage by a process called exocytosis.

In a recent study, Baillie *et al.* examined the hypothesis that dysregulation of intestinal macrophage responses to bacterial lipopolysaccharide (LPS) is a crucial component of susceptibility to IBD. To prove this, they demonstrated the transcriptional changes and gene expression in human monocyte-derived macrophages grown in CSF1. In this study, monocyte-derived macrophages were treated with bacterial lipopolysaccharide (LPS) and then harvested at time points from 15 minutes to 48 hours. Expression value for each CTSS (clusters of transcription start sites) was compared to the expression values of time points immediately before or after LPS treatment across the LPS time course. They have found that IBD loci are highly

expressed for monocyte-specific genes. Moreover, based on the previously published GWAS, they have identified a 134 more candidate genes linked with IBD susceptibility. Finally, they have proposed that dysregulated adaptation of monocyte to the gut mucosa might be associated with the pathogenesis of IBD (Baillie et al., 2017). In addition to that, live bacteria in the macrophage after phagocytosis may provide an ongoing stimulus to the immune system leading to chronic inflammation in CD. Loss of downregulation of pro-inflammatory cytokines like interferon-γ (IFN-γ) and tumour necrosis factor-α (TNF-α), and increased stimulation of the pro-inflammatory response, plays a critical role in altered intestinal permeability. This stimulus may lead to the expansion of AIEC within macrophages and thereby chronic inflammation in patients with CD (Hugot et al., 2001).

1.14 The NLRP3 inflammasome and CD

Inflammasomes were formerly considered as a fundamental part of the innate immune system, but their dysregulated activation may lead to several diseases. Inflammasomes are intracellular multiprotein complexes that sense different stresses: exogenous microbes and endogenous danger signals; and are responsible for the initiation of inflammatory responses. Among different subtypes of inflammasomes, NLRP3 inflammasome plays a significant role in maintaining immune responses and controlling the integrity of intestinal homeostasis in various inflammatory diseases (Elinav et al., 2013; Pellegrini et al., 2017). NLRP3 multiprotein complex consists of sensors - NLRP3, the adaptor-apoptosis-associated speck-like protein containing a caspase recruitment domain (ASC) and the effector protein - caspase-1 (Agostini et al., 2004; Hauenstein et al., 2015). The NLRP3 inflammasome is formed by bone marrow-derived macrophages (BMDM) and activated in response to pathogen-associated patterns (PAMPs) and damage-associated molecular patterns (DAMPs)

such as bacterial toxins, exogenous particulate substances, and lipopolysaccharide (LPS) (Martinon et al., 2006; Chen and Núñez, 2011). The mechanism of NLRP3 activation remains obscure. It is proposed that the activation of NLRP3 inflammasome in macrophages requires two steps: priming (signal 1) and activation (signal 2) (Figure 1.4). The priming step or signal one is induced by inflammatory stimuli, which induce NF-κB-mediated NLRP3 and pro-IL-1β expression. The activation step or signal two is initiated by PAMPs and DAMPs, thus promoting NLRP3 inflammasome activation and caspase-1-induced IL-1β and IL-18 secretion (Bauernfeind et al., 2009) and a soluble cytosolic protein gasdermin D that mediate inflammatory form of cell death termed pyroptosis (He et al., 2015; Kayagaki et al., 2015; Shi et al., 2015). Ion fluxes involving K+ efflux, Ca2+ mobilisation, and chloride efflux have been recognized as significant events in NLRP3 inflammasome activation (Murakami et al., 2012; Muñoz-Planillo et al., 2013; Green et al., 2018). Some other molecular and cellular mechanisms suggested to be involved in triggering the inflammasome activation includes mitochondrial dysfunction (Nakahira et al., 2011; Zhou et al., 2011) and production of reactive oxygen species (ROS) (He et al., 2016). In addition, several particulate substances like alum, silica, asbestos, amyloid-β, cholesterol crystals, and calcium crystals associated with lysosomal damage after phagocytosis also trigger NLRP3 inflammasome activation (Kelley et al., 2019). Particulate matter also triggers the NLRP3 inflammasome by endocytosis and lysosomal damage, causing the release of cathepsin B. Cathepsin B is also thought to play a role (Amaral et al., 2018; Chevriaux et al., 2020). NLRP3 initiates the formation of the inflammasome by binding with ASC, which recruits and activates procaspase-1 to produce active caspase-1. This activated caspase-1 then promotes the transformation and maturation of pro-inflammatory cytokines precursors pro-IL-1β and pro-IL-18 into mature and biologically

active molecules. After activation, IL-1β and IL-18 initiate a sequence of inflammatory responses and pyroptotic cell death (Latz et al., 2013; Lu et al., 2014; Liston and Masters, 2017; Próchnicki and Latz, 2017; Mangan et al., 2018), which plays a critical role in the pathogenesis of several autoinflammatory diseases (LaRock et al., 2016).

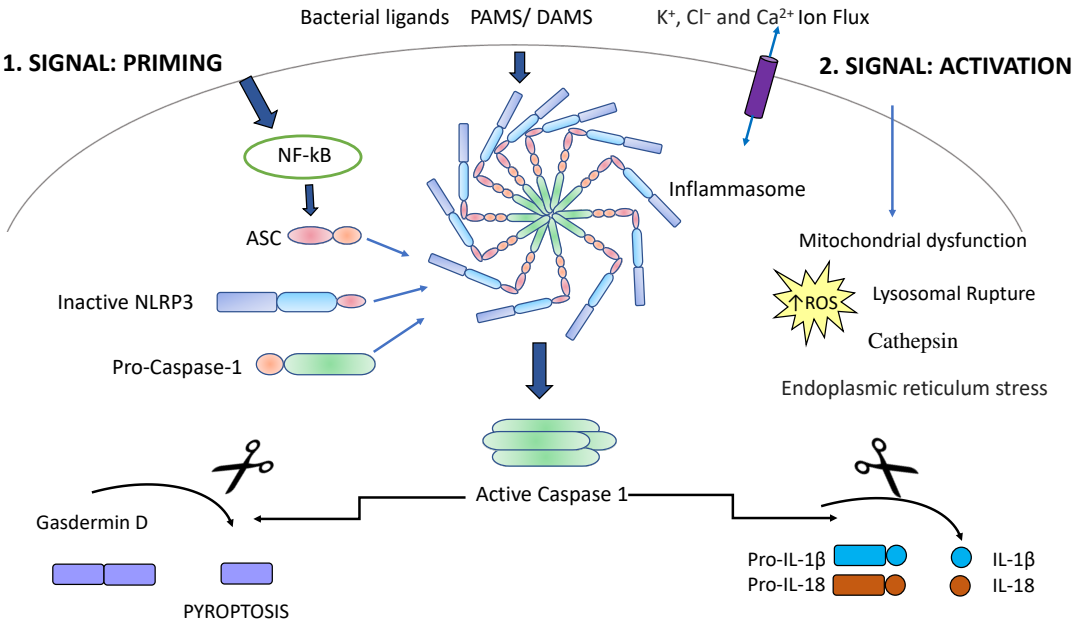


Figure 1.4. The mechanism of NLRP3 activation. Pathogen-associated molecular pattern (PAMP)- or danger-associated molecular pattern (DAMP)-associated ligands such as LPS or DNA (often through TLRs) induce the priming signals to trigger the expression of NLRP3. Signals by PAMPs and DAMPs result in ion flux and physiological cellular changes resulting in activation of NLRP3 inflammasome. As a result, activated caspase-1 cleaves gasdermin D, which leads to pore formation in the cell membrane and a form of cell death known as pyroptosis. Similarly, active caspase 1 leads to cleavage of pro-IL-1b and pro-IL-18, and cytokine secretion and activates the inflammasome further.

The NLRP3 inflammasome has been associated with IBD. NLRP3 in IBD is reported to play both pathogenic and protective roles. Studies have also shown that genetic variants and mutations in the NLRP3 gene are associated with IBD pathogenesis. The NLRP3 inflammasome was overexpressed in the colonic mucosa of patients (Liu et

al., 2016). Similarly, NLRP3 inflammasome activation is observed in 60% peripheral blood mononuclear cells when induced with LPS compared with only 28.6% of controls (Lazaridis et al., 2017). In the following steps, for inflammation resolution, it is essential to clear out apoptotic cells to prevent secondary necrosis and perpetuating inflammation. Macrophages are particularly equipped to sense and engulf apoptotic cells, following which macrophages undertake a shift from a pro-inflammatory to an anti-inflammatory phenotype (Gordon and Plüddemann, 2018). As apoptosis occurs in immune cells at a fast rate in the gut mucosa, phagocytosis of apoptotic cells by intestinal macrophages is essential to regulate homeostasis and prevent intestinal inflammation (Cummings et al., 2016; A-Gonzalez et al., 2017). However, intestinal macrophages bearing the apoptotic intestinal epithelial cells expressed several genes overlapped with IBD susceptibility genes (Cummings et al., 2016). In addition to that, colonic macrophages in CD patients exhibit abnormal maturation, persistent intracellular bacterial survival, and alterations in the expression of surface markers (Dige et al., 2016). Targeting NLRP3 inflammasome may provide new insights into the development of future therapeutic policies in CD.

1.15 Aims of the PhD project

Theories of the role of microbes in the aetiology of CD include:

1) Dysbiosis in gut microbiota causing an increase in ‘harmful’ bacteria and a decrease in the beneficial bacteria.

2) Bacterial translocation and dysregulated immune response towards them; and

3) An expansion of *E. coli* in gut mucosa.

We obtained mucosal biopsy samples during colonoscopy from patients with IBD and non-IBD controls; and lymph node, normal mucosa, involved mucosa samples from surgically resected specimens from patients with CD. We investigated the alteration of gut microbiota in patients with CD (Chapter 2). The microbial composition and existence of putative pathogens were compared by sequencing data obtained from 16s rRNA analysis. These results were analysed using the patient's disease status, disease stages and tissue types (example, Terminal ileum, colon, rectum and active or inactive, etc.).

We aimed to identify unique/signature genes of *E. coli* isolated in patients with CD compared with UC and healthy control, adults vs paediatric, and lymph node and aphthous ulcer isolates.

Our aim was to determine if particular *E. coli* genes or molecular properties are associated with tissue type, including lymph nodes and initial CD lesion (aphthous ulcer), or adult and paediatric isolates (Chapter 3).

CD is thought to be a consequence of a dysregulated immune response to intestinal microbiota in genetically predisposed individuals, and *E. coli* was one of the putative pathogens associated with the pathogenesis of CD. So, we attempted a small pilot study to investigate whether any gene in the *E. coli* obtained from CD patients is associated with NLRP3 inflammasome activation in bone marrow derives macrophage (BMDM) from wild type and NLRP3 knockout mice (Chapter 5).

1.16 References

A-Gonzalez, N., Quintana, J.A., García-Silva, S., Mazariegos, M., González de la Aleja, A., Nicolás-Ávila, J.A., et al. (2017). Phagocytosis imprints heterogeneity in tissue-resident macrophages. *Journal of Experimental Medicine* 214(5), 1281-1296.

Abraham, C., and Medzhitov, R. (2011). Interactions between the host innate immune system and microbes in inflammatory bowel disease. *Gastroenterology* 140(6), 1729-1737.

Achkar, J.-P. (2004). Endoscopic evaluation of postoperative recurrence in the patient with Crohn's disease. *Techniques in Gastrointestinal Endoscopy* 6(4), 165-168.

Agostini, L., Martinon, F., Burns, K., McDermott, M.F., Hawkins, P.N., and Tschopp, J. (2004). NALP3 forms an IL-1β-processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. *Immunity* 20(3), 319-325.

Agrawal, G., Clancy, A., Sharma, R., Huynh, R., Ramrakha, S., and Borody, T. (2020). Targeted Combination Antibiotic Therapy Induces Remission in Treatment-Naïve Crohn's Disease: A Case Series. *Microorganisms* 8(3), 371. doi: 10.3390/microorganisms8030371.

Aguilera-Castro, L., Ferre-Aracil, C., Garcia-Garcia-de-Paredes, A., Rodriguez-de-Santiago, E., and Lopez-Sanroman, A. (2017). Management of complex perianal Crohn's disease. *Annals of gastroenterology* 30(Goodhand et al.), 33-44. doi: 10.20524/aog.2016.0099.

Agus, A., Massier, S., Darfeuille-Michaud, A., Billard, E., and Barnich, N. (2014). Understanding host-adherent-invasive Escherichia coli interaction in Crohn's disease: opening up new therapeutic strategies. *BioMed research international* 2014.

Ahluwalia, B., Magnusson, M.K., and Öhman, L. (2017). Mucosal immune system of the gastrointestinal tract: maintaining balance between the good and the bad. *Scandinavian Journal of Gastroenterology* 52(11), 1185-1193.

Ajayi, T.A., Innes, C.L., Grimm, S.A., Rai, P., Finethy, R., Coers, J., et al. (2019). Crohn's disease IRGM risk alleles are associated with altered gene expression in human tissues. *American journal of physiology. Gastrointestinal and liver physiology* 316(Goodhand et al.), G95-G105. doi: 10.1152/ajpgi.00196.2018.

Al Nabhani, Z., Dietrich, G., Hugot, J.-P., and Barreau, F. (2017). Nod2: The intestinal gate keeper. *PLoS pathogens* 13(3), e1006177.

Alam, M.T., Amos, G.C.A., Murphy, A.R.J., Murch, S., Wellington, E.M.H., and Arasaradnam, R.P. (2020). Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels. *Gut Pathogens* 12(Goodhand et al.), 1. doi: 10.1186/s13099-019-0341-6.

Alatab, S., Sepanlou, S.G., Ikuta, K., Vahedi, H., Bisignano, C., Safiri, S., et al. (2020). The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology & Hepatology* 5(Goodhand et al.), 17-30. doi: 10.1016/S2468-1253(19)30333-4.

Altajar, S., and Moss, A. (2020). Inflammatory Bowel Disease Environmental Risk Factors: Diet and Gut Microbiota. *Current Gastroenterology Reports* 22(12), 57. doi: 10.1007/s11894-020-00794-y.

Amaral, E.P., Riteau, N., Moayeri, M., Maier, N., Mayer-Barber, K.D., Pereira, R.M., et al. (2018). Lysosomal Cathepsin Release Is Required for NLRP3-Inflammasome Activation by Mycobacterium tuberculosis in Infected Macrophages. *Frontiers in immunology* 9, 1427-1427. doi: 10.3389/fimmu.2018.01427.

Ambrose, N.S., Johnson, M., Burdon, D.W., and Keighley, M.R. (1984). Incidence of pathogenic bacteria from mesenteric lymph nodes and ileal serosa during Crohn's disease surgery. *Br J Surg* 71(8)**,** 623-625.

Amre, D.K., Mack, D., Israel, D., Morgan, K., Lambrette, P., Law, L., et al. (2008). Association between genetic variants in the IL-23R gene and early-onset Crohn's disease: results from a case-control and family-based study among Canadian children. *American Journal of Gastroenterology* 103(3)**,** 615-620.

Ananthakrishnan, A.N. (2013). Environmental Risk Factors for Inflammatory Bowel Disease. *Gastroenterology & Hepatology* 9(6)**,** 367-374.

Ananthakrishnan, A.N., Khalili, H., Konijeti, G.G., Higuchi, L.M., de Silva, P., Korzenik, J.R., et al. (2013). A prospective study of long-term intake of dietary fiber and risk of Crohn's disease and ulcerative colitis. *Gastroenterology* 145(5)**,** 970-977. doi: 10.1053/j.gastro.2013.07.050.

Antoni, L., Nuding, S., Wehkamp, J., and Stange, E.F. (2014). Intestinal barrier in inflammatory bowel disease. *World journal of gastroenterology: WJG* 20(5)**,** 1165.

Arnold, G.L., Beaves, M.R., Pryjdun, V.O., and Mook, W.J. (2002). Preliminary study of ciprofloxacin in active Crohn's disease. *Inflamm Bowel Dis* 8(Goodhand et al.)**,** 10-15.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473(7346)**,** 174-180. doi: 10.1038/nature09944.

Ashton, J.J., Ennis, S., and Beattie, R.M. (2017). Early-onset paediatric inflammatory bowel disease. *The Lancet Child & Adolescent Health* 1(Polito et al.)**,** 147-158. doi: 10.1016/S2352-4642(17)30017-2.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. *science* 307(5717)**,** 1915-1920.

Baillie, J.K., Arner, E., Daub, C., De Hoon, M., Itoh, M., Kawaji, H., et al. (2017). Analysis of the human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease. *PLoS Genet* 13(3)**,** e1006641. doi: 10.1371/journal.pgen.1006641.

Balasubramanian, I., and Gao, N. (2017). From sensing to shaping microbiota: insights into the role of NOD2 in intestinal homeostasis and progression of Crohn’s disease. *American Journal of Physiology-Gastrointestinal and Liver Physiology* 313(Goodhand et al.)**,** G7-G13.

Barnich, N., Carvalho, F.A., Glasser, A.-L., Darcha, C., Jantscheff, P., Allez, M., et al. (2007). CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease. *Journal of Clinical Investigation* 117(6)**,** 1566-1574. doi: 10.1172/JCI30504.

Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., et al. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 40(8)**,** 955-962. doi: 10.1038/ng.175.

Bartels, L.E., Jepsen, P., Christensen, L.A., Gerdes, L.U., Vilstrup, H., and Dahlerup, J.F. (2016). Diagnosis of Helicobacter Pylori Infection is Associated with Lower Prevalence and Subsequent Incidence of Crohn's Disease. *Journal of Crohn's & colitis* 10(4)**,** 443-448. doi: 10.1093/ecco-jcc/jjv229.

Bauernfeind, F.G., Horvath, G., Stutz, A., Alnemri, E.S., MacDonald, K., Speert, D., et al. (2009). Cutting edge: NF-κB activating pattern recognition and cytokine receptors license NLRP3 inflammasome activation by regulating NLRP3 expression. *The Journal of Immunology* 183(Polito et al.)**,** 787-791.

Baumgart, D.C., and Carding, S.R. (2007). Inflammatory bowel disease: cause and immunobiology. *The Lancet* 369(9573), 1627-1640. doi: https://doi.org/10.1016/S0140-6736(07)60750-8.

Baumgart, D.C., and Sandborn, W.J. (2012). Crohn's disease. *Lancet* 380(9853), 1590-1605. doi: 10.1016/s0140-6736(12)60026-9.

Baumgart, M., Dogan, B., Rishniw, M., Weitzman, G., Bosworth, B., and Yantiss, R. (2007). Culture independent analysis of ilealmucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *The ISME Journal* 1. doi: 10.1038/ismej.2007.52.

Benchimol, E.I., Bernstein, C.N., Bitton, A., Carroll, M.W., Singh, H., Otley, A.R., et al. (2017). Trends in Epidemiology of Pediatric Inflammatory Bowel Disease in Canada: Distributed Network Analysis of Multiple Population-Based Provincial Health Administrative Databases. *Am J Gastroenterol* 112(7), 1120-1134. doi: 10.1038/ajg.2017.97.

Bengtson, M.B., Solberg, C., Aamodt, G., Jahnsen, J., Moum, B., Sauar, J., et al. (2009a). Clustering in time of familial IBD separates ulcerative colitis from Crohn's disease. *Inflamm Bowel Dis* 15(12), 1867-1874. doi: 10.1002/ibd.20978.

Bengtson, M.B., Solberg, C., Aamodt, G., Sauar, J., Jahnsen, J., Moum, B., et al. (2009b). Familial aggregation in Crohn's disease and ulcerative colitis in a Norwegian population-based cohort followed for ten years. *J Crohns Colitis* 3(Polito et al.), 92-99. doi: 10.1016/j.crohns.2008.11.002.

Berg, R.D. (1999). Bacterial translocation from the gastrointestinal tract. *Adv Exp Med Biol* 473, 11-30. doi: 10.1007/978-1-4615-4143-1_2.

Berkowitz, L., Schultz, B.M., Salazar, G.A., Pardo-Roa, C., Sebastián, V.P., Álvarez-Lobos, M.M., et al. (2018). Impact of Cigarette Smoking on the Gastrointestinal Tract Inflammation: Opposing Effects in Crohn's Disease and Ulcerative Colitis. *Frontiers in Immunology* 9(74). doi: 10.3389/fimmu.2018.00074.

Bernstein, C.N. (2010). Epidemiologic clues to inflammatory bowel disease. *Current gastroenterology reports* 12(6), 495-501.

Bernstein, C.N., Loftus, E.V., Ng, S.C., Lakatos, P.L., and Moum, B. (2012). Hospitalisations and surgery in Crohn's disease. *Gut* 61(4), 622-629.

Bernstein, C.N., Wang, M.-H., Sargent, M., Brant, S.R., and Collins, M.T. (2007). Testing the interaction between NOD-2 status and serological response to Mycobacterium paratuberculosis in cases of inflammatory bowel disease. *Journal of clinical microbiology* 45(3), 968-971. doi: 10.1128/JCM.02062-06.

Bhatia, R., Yeoh, S.W., Vaz, K., Studd, C., Wilson, J., Bell, S., et al. (2019). Inflammatory bowel disease incidence, prevalence and 12-month initial disease course in Tasmania, Australia. *Internal Medicine Journal* 49(5), 622-630. doi: 10.1111/imj.14111.

Bianco, A.M., Zanin, V., Girardelli, M., Magnolato, A., Martellossi, S., Tommasini, A., et al. (2012). A common genetic background could explain early-onset Crohn's disease. *Medical Hypotheses* 78(4), 520-522. doi: https://doi.org/10.1016/j.mehy.2012.01.023.

Biank, V., Broeckel, U., and Kugathasan, S. (2007). Pediatric inflammatory bowel disease: clinical and molecular genetics. *Inflammatory bowel diseases* 13(11), 1430-1438.

Bien, J., Palagani, V., and Bozko, P. (2013). The intestinal microbiota dysbiosis and Clostridium difficile infection: is there a relationship with inflammatory bowel disease? *Therap Adv Gastroenterol* 6(Goodhand et al.), 53-68. doi: 10.1177/1756283x12454590.

Blaser, M.J., Hoverson, D., Ely, I.G., Duncan, D.J., Wang, W.-L.L., and Brown, W.R. (1984). Studies of Campylobacter jejuni in patients with inflammatory bowel

disease. *Gastroenterology* 86(Goodhand et al.), 33-38. doi: 10.1016/0016-5085(84)90587-0.

Bloom, Seth M., Bijanki, Vinieth N., Nava, Gerardo M., Sun, L., Malvin, Nicole P., Donermeyer, David L., et al. (2011). Commensal Bacteroides Species Induce Colitis in Host-Genotype-Specific Fashion in a Mouse Model of Inflammatory Bowel Disease. *Cell Host & Microbe* 9(5), 390-403. doi: https://doi.org/10.1016/j.chom.2011.04.009.

Bonen, D.K., Ogura, Y., Nicolae, D.L., Inohara, N., Saab, L., Tanabe, T., et al. (2003). Crohn's disease-associated NOD2 variants share a signaling defect in response to lipopolysaccharide and peptidoglycan. *Gastroenterology* 124(Goodhand et al.), 140-146.

Borody, T., Leis, S., Warren, E., and Surace, R. (2002). Treatment of severe Crohn's disease using antimycobacterial triple therapy—approaching a cure? *Digestive and Liver Disease* 34(Goodhand et al.), 29-38.

Boudeau, J., Barnich, N., and Darfeuille-Michaud, A. (2001). Type 1 pili-mediated adherence of Escherichia coli strain LF82 isolated from Crohn's disease is involved in bacterial invasion of intestinal epithelial cells. *Molecular microbiology* 39(5), 1272-1284.

Brant, S.R., Panhuysen, C.I., Bailey–Wilson, J.E., Rohal, P.M., Lee, S., Mann, J., et al. (2000). Linkage heterogeneity for the IBD1 locus in Crohn's disease pedigrees by disease onset and severity. *Gastroenterology* 119(6), 1483-1490.

Breese, E., Braegger, C.P., Corrigan, C.J., Walker-Smith, J.A., and MacDonald, T.T. (1993). Interleukin-2- and interferon-gamma-secreting T cells in normal and diseased human intestinal mucosa. *Immunology* 78(Goodhand et al.), 127-131.

Bringer, M.A., Barnich, N., Glasser, A.L., Bardot, O., and Darfeuille-Michaud, A. (2005). HtrA stress protein is involved in intramacrophagic replication of adherent and invasive Escherichia coli strain LF82 isolated from a patient with Crohn's disease. *Infect Immun* 73(Polito et al.), 712-721. doi: 10.1128/iai.73.2.712-721.2005.

Buisson, A., Chevaux, J.-B., Allen, P.B., Bommelaer, G., and Peyrin-Biroulet, L. (2012). Review article: the natural history of postoperative Crohn's disease recurrence. *Alimentary Pharmacology & Therapeutics* 35(6), 625-633. doi: 10.1111/j.1365-2036.2012.05002.x.

Burt, R.K., Craig, R.M., Milanetti, F., Quigley, K., Gozdziak, P., Bucha, J., et al. (2010). Autologous nonmyeloablative hematopoietic stem cell transplantation in patients with severe anti-TNF refractory Crohn disease: long-term follow-up. Blood, The Journal of the American Society of Hematology 116(26), 6123-6132.

Cao, Y., Shen, J., and Ran, Z.H. (2014). Association between Faecalibacterium prausnitzii Reduction and Inflammatory Bowel Disease: A Meta-Analysis and Systematic Review of the Literature. *Gastroenterol Res Pract* 2014, 872725. doi: 10.1155/2014/872725.

Camprubí-Font, C., Ewers, C., Lopez-Siles, M., and Martinez-Medina, M. (2019). Genetic and Phenotypic Features to Screen for Putative Adherent-Invasive Escherichia coli. *Front Microbiol* 10, 108. doi: 10.3389/fmicb.2019.00108.

Caparrós, E., Wiest, R., Scharl, M., Rogler, G., Gutiérrez Casbas, A., Yilmaz, B., et al. (2021). Dysbiotic microbiota interactions in Crohn's disease. *Gut Microbes* 13(Goodhand et al.), 1949096. doi: 10.1080/19490976.2021.1949096.

Caprilli, R., and Frieri, G. (2009). The dyspeptic macrophage 30 years later: an update in the pathogenesis of Crohn's disease. *Dig Liver Dis* 41(Polito et al.), 166-168. doi: 10.1016/j.dld.2008.09.012.

Carvalho, F.A., Barnich, N., Sauvanet, P., Darcha, C., Gelot, A., and Darfeuille-Michaud, A. (2008). Crohn's disease-associated Escherichia coli LF82 aggravates colitis in injured mouse colon via signaling by flagellin. *Inflammatory bowel diseases* 14(8), 1051-1060.

Carvalho, F.A., Barnich, N., Sivignon, A., Darcha, C., Chan, C.H., and Stanners, C.P. (2009). Crohn's disease adherent-invasive Escherichia coli colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM. *J Exp Med* 206. doi: 10.1084/jem.20090741.

Castaño-Rodríguez, N., Kaakoush, N.O., Lee, W.S., and Mitchell, H.M. (2017). Dual role of Helicobacter and Campylobacter species in IBD: a systematic review and meta-analysis. *Gut* 66(Polito et al.), 235-249. doi: 10.1136/gutjnl-2015-310545.

Cătană, C.-S., Berindan Neagoe, I., Cozma, V., Magdaş, C., Tăbăran, F., and Dumitraşcu, D.L. (2015). Contribution of the IL-17/IL-23 axis to the pathogenesis of inflammatory bowel disease. *World journal of gastroenterology* 21(19), 5823-5830. doi: 10.3748/wjg.v21.i19.5823.

Celiberto, L.S., Graef, F.A., Healey, G.R., Bosman, E.S., Jacobson, K., Sly, L.M., et al. (2018). Inflammatory bowel disease and immunonutrition: novel therapeutic approaches through modulation of diet and the gut microbiome. *Immunology* 155(Goodhand et al.), 36-52.

Céspedes, S., Saitz, W., Del Canto, F., De la Fuente, M., Quera, R., Hermoso, M., et al. (2017). Genetic Diversity and Virulence Determinants of Escherichia coli Strains Isolated from Patients with Crohn's Disease in Spain and Chile. *Frontiers in Microbiology* 8, 639. doi: 10.3389/fmicb.2017.00639.

Chamaillard, M., and Radulovic, K. (2016). Defining dysbiosis threatens Koch's postulates and current dogma on the role of Paneth cells in Crohn's disease. *Gut* 65(Polito et al.), 190-191. doi: 10.1136/gutjnl-2015-309714.

Chassaing, B., Compher, C., Bonhomme, B., Liu, Q., Tin, Y., Walters, W., et al. (2022). Randomized controlled-feeding study of dietary emulsifier carboxymethylcellulose reveals detrimental impacts on the gut microbiota and metabolome. Gastroenterology 162(3), 743-756. doi: 10.1053/j.gastro.2021.11.006.

Chassaing, B., Rolhion, N., de Vallée, A., Salim, S.a.Y., Prorok-Hamon, M., Neut, C., et al. (2011). Crohn disease–associated adherent-invasive E. coli bacteria target mouse and human Peyer's patches via long polar fimbriae. *The Journal of Clinical Investigation* 121(3), 966-975. doi: 10.1172/JCI44632.

Cheifetz, A.S. (2013). Management of active Crohn disease. *Jama* 309(20), 2150-2158.

Chen, G.Y., and Núñez, G. (2011). Inflammasomes in intestinal inflammation and cancer. *Gastroenterology* 141(6), 1986-1999.

Chen, W., Li, D., Paulus, B., Wilson, I., and Chadwick, V.S. (2000). Detection of Listeria monocytogenes by polymerase chain reaction in intestinal mucosal biopsies from patients with inflammatory bowel disease and controls. *Journal of Gastroenterology and Hepatology* 15(10), 1145-1150. doi: 10.1046/j.1440-1746.2000.02331.x.

Chevriaux, A., Pilot, T., Derangère, V., Simonin, H., Martine, P., Chalmin, F., et al. (2020). Cathepsin B Is Required for NLRP3 Inflammasome Activation in Macrophages, Through NLRP3 Interaction. *Frontiers in Cell and Developmental Biology* 8(167). doi: 10.3389/fcell.2020.00167.

Chiba, M., Fukushima, T., Inoue, S., Horie, Y., Iizuka, M., and Masamune, O. (1998). Listeria monocytogenes in Crohn's disease. *Scand J Gastroenterol* 33(4), 430-434. doi: 10.1080/00365529850171071.

Chiba, M., Goto, T., Funaoka, M., Masamune, O., and Koizumi, R. (1990). A Case of Crohn's Disease-Significance of Campylobacter Jejuni Detected from the Stool during the Relapse. *Nippon Daicho Komonbyo Gakkai Zasshi* 43(Polito et al.), 192-196. doi: 10.3862/jcoloproctology.43.192.

Chiodini, R.J., Chamberlin, W.M., Sarosiek, J., and McCallum, R.W. (2012). Crohn's disease and the mycobacterioses: a quarter century later. Causation or simple association? *Critical reviews in microbiology* 38(Goodhand et al.)**,** 52-93.

Cho, J.H., and Brant, S.R. (2011). Recent insights into the genetics of inflammatory bowel disease. *Gastroenterology* 140(6)**,** 1704-1712. doi: 10.1053/j.gastro.2011.02.046.

Chow, J., Lee, S.M., Shen, Y., Khosravi, A., and Mazmanian, S.K. (2010). "Host–bacterial symbiosis in health and disease," in *Advances in immunology*. Elsevier), 243-274.

Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. *Cell* 148(6)**,** 1258-1270. doi: 10.1016/j.cell.2012.01.035.

Clermont, O., Christenson, J.K., Denamur, E., and Gordon, D.M. (2013). The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5(Goodhand et al.)**,** 58-65. doi: 10.1111/1758-2229.12019.

Clooney, A.G., Eckenberger, J., Laserna-Mendieta, E., Sexton, K.A., Bernstein, M.T., Vagianos, K., et al. (2020). Ranking microbiome variance in inflammatory bowel disease: a large longitudinal intercontinental study. *Gut***,** gutjnl-2020-321106. doi: 10.1136/gutjnl-2020-321106.

Coffey, C.J., Kiernan, M.G., Sahebally, S.M., Jarrar, A., Burke, J.P., Kiely, P.A., et al. (2018). Inclusion of the mesentery in ileocolic resection for Crohn's disease is associated with reduced surgical recurrence. *Journal of Crohn's and Colitis* 12(10)**,** 1139-1150.

Coffey, J.C., O'Leary, D.P., Kiernan, M.G., and Faul, P. (2016). The mesentery in Crohn's disease: friend or foe? *Current opinion in gastroenterology* 32(4)**,** 267-273.

Cohen, L.J., Cho, J.H., Gevers, D., and Chu, H. (2019). Genetic factors and the intestinal microbiome guide development of microbe-based therapies for inflammatory bowel diseases. *Gastroenterology* 156(8)**,** 2174-2189.

Colombel, J.F., Grandbastien, B., Gower-Rousseau, C., Plegat, S., Evrard, J.P., Dupas, J.L., et al. (1996). Clinical characteristics of Crohn's disease in 72 families. *Gastroenterology* 111(3)**,** 604-607.

Cooney, R., Baker, J., Brain, O., Danis, B., Pichulik, T., Allan, P., et al. (2010). NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nat Med* 16(Goodhand et al.)**,** 90-97. doi: 10.1038/nm.2069.

Cosnes, J. (2010). Smoking, physical activity, nutrition and lifestyle: environmental factors and their impact on IBD. *Digestive diseases* 28(3)**,** 411-417.

Cosnes, J., Beaugerie, L., Carbonnel, F., and Gendre, J.P. (2001). Smoking cessation and the course of Crohn's disease: an intervention study. *Gastroenterology* 120(5)**,** 1093-1099.

Crohn, B.B., Ginzburg, L., and Oppenheimer, G.D. (1932). Regional ileitis: A pathologic and clinical entity. *Journal of the American Medical Association* 99(16)**,** 1323-1329. doi: 10.1001/jama.1932.02740680019005.

Crouwel, F., Buiter, H.J.C., and de Boer, N.K. (2021). Gut Microbiota-driven Drug Metabolism in Inflammatory Bowel Disease. *Journal of Crohn's and Colitis* 15(Polito et al.)**,** 307-315. doi: 10.1093/ecco-jcc/jjaa143.

Cui, G., Liu, H., Xu, G., Laugsand, J.-B., and Pang, Z. (2021). Exploring Links Between Industrialization, Urbanization, and Chinese Inflammatory Bowel Disease. *Frontiers in Medicine* 8. doi: 10.3389/fmed.2021.757025.

Cummings, R.J., Barbet, G., Bongers, G., Hartmann, B.M., Gettler, K., Muniz, L., et al. (2016). Different tissue phagocytes sample apoptotic cells to direct distinct homeostasis programs. *Nature* 539(7630)**,** 565-569.

D'Haens, G.R., Geboes, K., Peeters, M., Baert, F., Penninckx, F., and Rutgeerts, P. (1998). Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded ileum. *Gastroenterology* 114(Polito et al.), 262-267.

Dahlhamer, J.M., Zammitti, E.P., Ward, B.W., Wheaton, A.G., and Croft, J.B. (2016). Prevalence of inflammatory bowel disease among adults aged≥ 18 years—United States, 2015. *Morbidity and Mortality Weekly Report* 65(42), 1166-1169.

Dalton, J.P., Desmond, A., Shanahan, F., and Hill, C. (2014). Detection of Mycobacterium avium subspecies paratuberculosis in patients with Crohn's disease is unrelated to the presence of single nucleotide polymorphisms rs2241880 (ATG16L1) and rs10045431 (IL12B). *Med Microbiol Immunol* 203(3), 195-205. doi: 10.1007/s00430-014-0332-7.

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

Darfeuille-Michaud, A., Neut, C., Barnich, N., Lederman, E., Di Martino, P., Desreumaux, P., et al. (1998). Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease. *Gastroenterology* 115(6), 1405-1413.

Dassopoulos, T., Sultan, S., Falck–Ytter, Y.T., Inadomi, J.M., and Hanauer, S.B. (2013). American Gastroenterological Association Institute technical review on the use of thiopurines, methotrexate, and anti–TNF-α biologic drugs for the induction and maintenance of remission in inflammatory Crohn's disease. *Gastroenterology* 145(6), 1464-1478. e1465.

De Cruz, P., Kamm, M.A., Hamilton, A.L., Ritchie, K.J., Krejany, E.O., Gorelik, A., et al. (2015). Crohn's disease management after intestinal resection: a randomised trial. *Lancet* 385(9976), 1406-1417. doi: 10.1016/s0140-6736(14)61908-5.

De Hertogh, G., Aerssens, J., Geboes, K.P., and Geboes, K. (2008). Evidence for the involvement of infectious agents in the pathogenesis of Crohn's disease. *World journal of gastroenterology: WJG* 14(6), 845.

De la Fuente, M., Franchi, L., Araya, D., Díaz-Jiménez, D., Olivares, M., Álvarez-Lobos, M., et al. (2014). Escherichia coli isolates from inflammatory bowel diseases patients survive in macrophages and activate NLRP3 inflammasome. *International Journal of Medical Microbiology* 304(3), 384-392. doi: https://doi.org/10.1016/j.ijmm.2014.01.002.

de Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature genetics* 49(Polito et al.), 256.

de Ridder, L., Weersma, R.K., Dijkstra, G., van der Steege, G., Benninga, M.A., Nolte, I.M., et al. (2007). Genetic susceptibility has a more important role in pediatric-onset Crohn's disease than in adult-onset Crohn's disease. *Inflammatory bowel diseases* 13(9), 1083-1092.

de Souza, H.L., de Carvalho, V.R., Romeiro, F.G., Sassaki, L.Y., Keller, R., and Rodrigues, J. (2012). Mucosa-associated but not luminal Escherichia coli is augmented in Crohn's disease and ulcerative colitis. *Gut Pathogens* 4, 21-21. doi: 10.1186/1757-4749-4-21.

de Souza, H.S., and Fiocchi, C. (2016). Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol* 13(Goodhand et al.), 13-27. doi: 10.1038/nrgastro.2015.186.

Denning, T.L., Wang, Y.-c., Patel, S.R., Williams, I.R., and Pulendran, B. (2007). Lamina propria macrophages and dendritic cells differentially induce regulatory and interleukin 17–producing T cell responses. *Nature immunology* 8(10), 1086-1094.

Desilets, M., Deng, X., Rao, C., Ensminger, A.W., Krause, D.O., Sherman, P.M., et al. (2016). Genome-based Definition of an Inflammatory Bowel Disease-associated Adherent-Invasive Escherichia coli Pathovar. *Inflamm Bowel Dis* 22(Goodhand et al.), 1-12. doi: 10.1097/mib.0000000000000574.

Dicksved, J., Halfvarson, J., Rosenquist, M., Järnerot, G., Tysk, C., Apajalahti, J., et al. (2008). Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *The ISME journal* 2(7), 716-727.

Dige, A., Magnusson, M.K., Öhman, L., Hvas, C.L., Kelsen, J., Wick, M.J., et al. (2016). Reduced numbers of mucosal DRint macrophages and increased numbers of CD103+ dendritic cells during anti-TNF-α treatment in patients with Crohn's disease. *Scandinavian journal of gastroenterology* 51(6), 692-699.

Diggs, N.G., and Surawicz, C.M. (2009). Evolving concepts in Clostridium difficile colitis. *Current gastroenterology reports* 11(5), 400-405.

Dogan, B., Suzuki, H., Herlekar, D., Sartor, R.B., Campbell, B.J., and Roberts, C.L. (2014). Inflammation-associated adherent-invasive Escherichia coli are enriched in pathways for use of propanediol and iron and M-cell translocation. *Inflamm Bowel Dis* 20. doi: 10.1097/mib.0000000000000183.

Doherty, G.A., Bennett, G.C., Cheifetz, A.S., and Moss, A.C. (2010). Meta-analysis: targeting the intestinal microbiota in prophylaxis for post-operative Crohn's disease. *Aliment Pharmacol Ther* 31(8), 802-809. doi: 10.1111/j.1365-2036.2010.04231.x.

Dreux, N., Denizot, J., Martinez-Medina, M., Mellmann, A., Billig, M., Kisiela, D., et al. (2013). Point mutations in FimH adhesin of Crohn's disease-associated adherent-invasive Escherichia coli enhance intestinal inflammatory response. *PLoS pathogens* 9(Goodhand et al.), e1003141.

Dubinsky, M.C., Wang, D., Picornell, Y., Wrobel, I., Katzir, L., Quiros, A., et al. (2007). IL-23 receptor (IL-23R) gene protects against pediatric Crohn's disease. *Inflammatory bowel diseases* 13(5), 511-515.

Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., et al. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *science* 314(5804), 1461-1463.

Dutta, A.K., and Chacko, A. (2016). Influence of environmental factors on the onset and course of inflammatory bowel disease. *World Journal of Gastroenterology* 22(3), 1088-1100. doi: 10.3748/wjg.v22.i3.1088.

Eckburg, P.B., and Relman, D.A. (2007). The Role of Microbes in Crohn's Disease. *Clinical Infectious Diseases* 44(Polito et al.), 256-262. doi: 10.1086/510385.

Elinav, E., Henao-Mejia, J., and Flavell, R. (2013). Integrative inflammasome activity in the regulation of intestinal mucosal immune responses. *Mucosal immunology* 6(Goodhand et al.), 4-13.

Fava, F., and Danese, S. (2011). Intestinal microbiota in inflammatory bowel disease: friend of foe? *World journal of gastroenterology: WJG* 17(5), 557.

Feagan, B.G., Sandborn, W.J., Gasink, C., Jacobstein, D., Lang, Y., Friedman, J.R., et al. (2016). Ustekinumab as induction and maintenance therapy for Crohn's disease. *New England Journal of Medicine* 375(20), 1946-1960.

Feller, M., Huwiler, K., Schoepfer, A., Shang, A., Furrer, H., and Egger, M. (2010). Long-Term Antibiotic Treatment for Crohn's Disease: Systematic Review and Meta-Analysis of Placebo-Controlled Trials. *Clinical Infectious Diseases* 50(4), 473-480. doi: 10.1086/649923.

Feller, M., Huwiler, K., Stephan, R., Altpeter, E., Shang, A., Furrer, H., et al. (2007). Mycobacterium avium subspecies paratuberculosis and Crohn's disease: a systematic

review and meta-analysis. *The Lancet Infectious Diseases* 7(9), 607-613. doi: https://doi.org/10.1016/S1473-3099(07)70211-6.

Fichera, A., and Michelassi, F. (2007). Surgical treatment of Crohn's disease. *Journal of Gastrointestinal Surgery* 11(6), 791-803.

Flint, H.J., Scott, K.P., Louis, P., and Duncan, S.H. (2012). The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* 9(10), 577-589. doi: 10.1038/nrgastro.2012.156.

Fordtran, J.S. (2006). Colitis due to Clostridium difficile toxins: underdiagnosed, highly virulent, and nosocomial. *Proceedings (Baylor University. Medical Center)* 19(Goodhand et al.), 3-12. doi: 10.1080/08998280.2006.11928114.

Frank, D.N., St. Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences* 104(34), 13780-13785. doi: 10.1073/pnas.0706625104.

Franke, A., Balschun, T., Karlsen, T.H., Hedderich, J., May, S., Lu, T., et al. (2008). Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. *Nature genetics* 40(6), 713-715.

Fritz, T., Niederreiter, L., Adolph, T., Blumberg, R.S., and Kaser, A. (2011). Crohn's disease: NOD2, autophagy and ER stress converge. *Gut* 60(11), 1580-1588.

Fujimoto, T., Imaeda, H., Takahashi, K., Kasumi, E., Bamba, S., Fujiyama, Y., et al. (2013). Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease. *J Gastroenterol Hepatol* 28(4), 613-619. doi: 10.1111/jgh.12073.

Fujimura, Y., Kamoi, R., and Iida, M. (1996). Pathogenesis of aphthoid ulcers in Crohn's disease: correlative findings by magnifying colonoscopy, electron microscopy, and immunohistochemistry. *Gut* 38(5), 724-732.

Gajendran, M., Loganathan, P., Catinella, A.P., and Hashash, J.G. (2018). A comprehensive review and update on Crohn's disease. *Dis Mon* 64(Polito et al.), 20-57. doi: 10.1016/j.disamonth.2017.07.001.

Galazzo, G., Tedjo, D.I., Wintjens, D.S.J., Savelkoul, P.H.M., Masclee, A.A.M., Bodelier, A.G.L., et al. (2019). Faecal Microbiota Dynamics and their Relation to Disease Course in Crohn's Disease. *J Crohns Colitis* 13(10), 1273-1282. doi: 10.1093/ecco-jcc/jjz049.

Gasche, C., Scholmerich, J., Brynskov, J., D'Haens, G., Hanauer, S.B., Irvine, E.J., et al. (2000). A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 6(Goodhand et al.), 8-15. doi: 10.1097/00054725-200002000-00002.

Gensollen, T., Iyer, S.S., Kasper, D.L., and Blumberg, R.S. (2016). How colonization by microbiota in early life shapes the immune system. *Science* 352(6285), 539-544.

Genta, R.M., and Sonnenberg, A. (2012). Non-Helicobacter pylori gastritis is common among paediatric patients with inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics* 35(11), 1310-1316. doi: https://doi.org/10.1111/j.1365-2036.2012.05090.x.

Geremia, A., Biancheri, P., Allan, P., Corazza, G.R., and Di Sabatino, A. (2014). Innate and adaptive immunity in inflammatory bowel disease. *Autoimmun Rev* 13(Goodhand et al.), 3-10. doi: 10.1016/j.autrev.2013.06.004.

Gevers, D., Kugathasan, S., Denson, L.A., Vazquez-Baeza, Y., Treuren, W., and Ren, B. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15. doi: 10.1016/j.chom.2014.02.005.

Gevers, D., Kugathasan, S., Knights, D., Kostic, A.D., Knight, R., and Xavier, R.J. (2017). A microbiome foundation for the study of Crohn's disease. *Cell host & microbe* 21(3), 301-304.

Gill, S.R., Pop, M., DeBoy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., et al. (2006). Metagenomic analysis of the human distal gut microbiome. *science* 312(5778), 1355-1359.

Gisbert, J.P., Bermejo, F., Pérez-Calle, J.-L., Taxonera, C., Vera, I., McNicholl, A.G., et al. (2009). Fecal calprotectin and lactoferrin for the prediction of inflammatory bowel disease relapse. *Inflammatory bowel diseases* 15(8), 1190-1198.

Glas, J., Seiderer, J., Bues, S., Stallhofer, J., Fries, C., Olszak, T., et al. (2013a). IRGM variants and susceptibility to inflammatory bowel disease in the German population. *PLoS One* 8(Goodhand et al.), e54338. doi: 10.1371/journal.pone.0054338.

Glas, J., Seiderer, J., Bues, S., Stallhofer, J., Fries, C., Olszak, T., et al. (2013b). IRGM Variants and Susceptibility to Inflammatory Bowel Disease in the German Population. *PLOS ONE* 8(Goodhand et al.), e54338. doi: 10.1371/journal.pone.0054338.

Glas, J., Seiderer, J., Tillack, C., Pfennig, S., Beigel, F., Jürgens, M., et al. (2010). The NOD2 single nucleotide polymorphisms rs2066843 and rs2076756 are novel and common Crohn's disease susceptibility gene variants. *PLoS One* 5(12), e14466. doi: 10.1371/journal.pone.0014466.

Glasser, A.-L., Boudeau, J., Barnich, N., Perruchot, M.-H., Colombel, J.-F., and Darfeuille-Michaud, A. (2001). Adherent invasive Escherichia coli strains from patients with Crohn's disease survive and replicate within macrophages without inducing host cell death. *Infection and immunity* 69(9), 5529-5537.

Gomollón, F., Dignass, A., Annese, V., Tilg, H., Van Assche, G., Lindsay, J.O., et al. (2017). 3rd European evidence-based consensus on the diagnosis and management of Crohn's disease 2016: part 1: diagnosis and medical management. *Journal of Crohn's and Colitis* 11(Goodhand et al.), 3-25.

Goodhand, J., Wahed, M., Mawdsley, J., Farmer, A., Aziz, Q., and Rampton, D. (2012). Mood disorders in inflammatory bowel disease: relation to diagnosis, disease activity, perceived stress, and other factors. *Inflammatory bowel diseases* 18(12), 2301-2309.

Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W.F., and Veldhuyzen van Zanten, S.J.O. (2006). Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *Journal of clinical microbiology* 44(11), 4136-4141. doi: 10.1128/jcm.01004-06.

Gordon, S., and Plüddemann, A. (2018). Macrophage clearance of apoptotic cells: a critical assessment. *Frontiers in immunology* 9, 127.

Graff, L.A., Walker, J.R., and Bernstein, C.N. (2009). Depression and anxiety in inflammatory bowel disease: a review of comorbidity and management. *Inflamm Bowel Dis* 15(7), 1105-1118. doi: 10.1002/ibd.20873.

Green, J.P., Yu, S., Martín-Sánchez, F., Pelegrin, P., Lopez-Castejon, G., Lawrence, C.B., et al. (2018). Chloride regulates dynamic NLRP3-dependent ASC oligomerization and inflammasome priming. *Proceedings of the National Academy of Sciences* 115(40), E9371-E9380.

Greenhalgh, K., Meyer, K.M., Aagaard, K.M., and Wilmes, P. (2016). The human gut microbiome in health: establishment and resilience of microbiota over a lifetime. *Environmental microbiology* 18(7), 2103-2116. doi: 10.1111/1462-2920.13318.

Gren, S.T., and Grip, O. (2016). Role of Monocytes and Intestinal Macrophages in Crohn's Disease and Ulcerative Colitis. *Inflamm Bowel Dis* 22(8), 1992-1998. doi: 10.1097/mib.0000000000000824.

Gullberg, E., and Söderholm, J.D. (2006). Peyer's patches and M cells as potential sites of the inflammatory onset in Crohn's disease. *Annals of the New York Academy of Sciences* 1072(Goodhand et al.), 218-232.

Guslandi, M. (2009). Campylobacter concisus: a New Character in the Crohn's Disease Story? *Journal of Clinical Microbiology* 47(5), 1614. doi: 10.1128/JCM.00310-09.

Gutiérrez, A., Scharl, M., Sempere, L., Holler, E., Zapater, P., Almenta, I., et al. (2014). Genetic susceptibility to increased bacterial translocation influences the response to biological therapy in patients with Crohn's disease. *Gut* 63(Polito et al.), 272-280.

Gutiérrez, A., Zapater, P., Juanola, O., Sempere, L., García, M., Laveda, R., et al. (2016). Gut Bacterial DNA Translocation is an Independent Risk Factor of Flare at Short Term in Patients With Crohn's Disease. *Am J Gastroenterol* 111(4), 529-540. doi: 10.1038/ajg.2016.8.

Gutierrez, M.G., Master, S.S., Singh, S.B., Taylor, G.A., Colombo, M.I., and Deretic, V. (2004). Autophagy Is a Defense Mechanism Inhibiting BCG and Mycobacterium tuberculosis Survival in Infected Macrophages. *Cell* 119(6), 753-766. doi: 10.1016/j.cell.2004.11.038.

Ha, C., and Kornbluth, A. (2014). Vedolizumab as a Treatment for Crohn's Disease and Ulcerative Colitis. *Gastroenterol Hepatol (N Y)* 10(12), 793-800.

Haberman, Y., Tickle, T.L., Dexheimer, P.J., Kim, M.-O., Tang, D., Karns, R., et al. (2014). Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *The Journal of Clinical Investigation* 124(8), 3617-3633. doi: 10.1172/JCI75436.

Halme, L., Paavola-Sakki, P., Turunen, U., Lappalainen, M., Farkkila, M., and Kontula, K. (2006). Family and twin studies in inflammatory bowel disease. *World J Gastroenterol* 12(23), 3668-3672.

Hamilton, A.L., Kamm, M.A., De Cruz, P., Wright, E.K., Feng, H., Wagner, J., et al. (2020). Luminal microbiota related to Crohn's disease recurrence after surgery. *Gut Microbes* 11(6), 1713-1728. doi: 10.1080/19490976.2020.1778262.

Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., et al. (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 39(Polito et al.), 207-211. doi: 10.1038/ng1954.

Hansen, R., Russell, R.K., Reiff, C., Louis, P., McIntosh, F., Berry, S.H., et al. (2012). Microbiota ofDe-NovoPediatric IBD: IncreasedFaecalibacterium Prausnitziiand Reduced Bacterial Diversity in Crohn's But Not in Ulcerative Colitis. *American Journal of Gastroenterology* 107(12), 1913-1922.

Hasan, N., and Yang, H. (2019). Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* 7, e7502-e7502. doi: 10.7717/peerj.7502.

Hauenstein, A.V., Zhang, L., and Wu, H. (2015). The hierarchical structural architecture of inflammasomes, supramolecular inflammatory machines. *Current opinion in structural biology* 31, 75-83.

Hausmann, M., Kiessling, S., Mestermann, S., Webb, G., Spöttl, T., Andus, T., et al. (2002). Toll-like receptors 2 and 4 are up-regulated during intestinal inflammation. *Gastroenterology* 122(7), 1987-2000.

Hawkey, C.J., and Hommes, D.W. (2017). Is stem cell therapy ready for prime Ttme in treatment of inflammatory bowel diseases? Gastroenterology 152(2), 389-397.e382. doi: 10.1053/j.gastro.2016.11.003.

Hazlewood, G.S., Rezaie, A., Borman, M., Panaccione, R., Ghosh, S., Seow, C.H., et al. (2015). Comparative effectiveness of immunosuppressants and biologics for inducing and maintaining remission in Crohn's disease: a network meta-analysis. *Gastroenterology* 148(Polito et al.), 344-354. e345.

He, W.-t., Wan, H., Hu, L., Chen, P., Wang, X., Huang, Z., et al. (2015). Gasdermin D is an executor of pyroptosis and required for interleukin-1β secretion. *Cell research* 25(12), 1285.

He, Y., Hara, H., and Núñez, G. (2016). Mechanism and regulation of NLRP3 inflammasome activation. *Trends in biochemical sciences* 41(12), 1012-1021.

Herzog, D., Fournier, N., Buehr, P., Rueger, V., Koller, R., Heyland, K., et al. (2017). Prevalence of intestinal complications in inflammatory bowel disease: a comparison between paediatric-onset and adult-onset patients. *Eur J Gastroenterol Hepatol* 29(8), 926-931. doi: 10.1097/meg.0000000000000896.

Higuchi, L.M., Khalili, H., Chan, A.T., Richter, J.M., Bousvaros, A., and Fuchs, C.S. (2012). A prospective study of cigarette smoking and the risk of inflammatory bowel disease in women. *The American journal of gastroenterology* 107(9), 1399.

Hizawa, K., Iida, M., Kohrogi, N., Kuroki, F., Yao, T., Sakamoto, K., et al. (1994). Crohn disease: early recognition and progress of aphthous lesions. *Radiology* 190(Polito et al.), 451-454.

Hoeve, M.A., Savage, N.D., de Boer, T., Langenberg, D.M., de Waal Malefyt, R., Ottenhoff, T.H., et al. (2006). Divergent effects of IL-12 and IL-23 on the production of IL-17 by human T cells. *European journal of immunology* 36(3), 661-670.

Hong, P.-Y., Croix, J.A., Greenberg, E., Gaskins, H.R., and Mackie, R.I. (2011). Pyrosequencing-based analysis of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-heterogeneity. *PloS one* 6(9), e25042-e25042. doi: 10.1371/journal.pone.0025042.

Hooper, L.V., Littman, D.R., and Macpherson, A.J. (2012). Interactions Between the Microbiota and the Immune System. *Science* 336(6086), 1268.

Horowitz, J.E., Warner, N., Staples, J., Crowley, E., Gosalia, N., Murchie, R., et al. (2021). Mutation spectrum of NOD2 reveals recessive inheritance as a main driver of Early Onset Crohn's Disease. *Scientific Reports* 11(Goodhand et al.), 5595. doi: 10.1038/s41598-021-84938-8.

Hugot, J.-P., Alberti, C., Berrebi, D., Bingen, E., and Cézard, J.-P. (2003). Crohn's disease: the cold chain hypothesis. *The Lancet* 362(9400), 2012-2015.

Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., et al. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411(6837), 599-603. doi: 10.1038/35079107.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., et al. (2012). Structure, function and diversity of the healthy human microbiome. *nature* 486(7402), 207.

Hyun, K., Shin, H., and Kim, D. (2013). Malakoplakia in a healthy young female patient. *The Korean journal of internal medicine* 28, 475-480. doi: 10.3904/kjim.2013.28.4.475.

Iebba, V., Conte, M.P., Lepanto, M.S., Di Nardo, G., Santangelo, F., Aloi, M., et al. (2012). Microevolution in fimH gene of mucosa-associated Escherichia coli strains isolated from pediatric patients with inflammatory bowel disease. *Infection and immunity* 80(4), 1408-1417.

Ignacio, A., Terra, F.F., Watanabe, I.K.M., Basso, P.J., and Câmara, N.O.S. (2019). "Chapter 13 - Role of the Microbiome in Intestinal Barrier Function and Immune Defense," in *Microbiome and Metabolome in Diagnosis, Therapy, and other Strategic Applications,* eds. J. Faintuch & S. Faintuch. Academic Press), 127-138.

Iwasaki, A., and Medzhitov, R. (2010). Regulation of adaptive immunity by the innate immune system. *science* 327(5963), 291-295.

Izcue, A., Hue, S., Buonocore, S., Arancibia-Cárcamo, C.V., Ahern, P.P., Iwakura, Y., et al. (2008). Interleukin-23 restrains regulatory T cell activity to drive T cell-dependent colitis. *Immunity* 28(4), 559-570. doi: 10.1016/j.immuni.2008.02.019.

Jäger, S., Stange, E.F., and Wehkamp, J. (2013). Inflammatory bowel disease: an impaired barrier disease. *Langenbeck's archives of surgery* 398(Goodhand et al.), 1-12.

Jandhyala, S.M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., and Nageshwar Reddy, D. (2015). Role of the normal gut microbiota. *World J Gastroenterol* 21(29), 8787-8803. doi: 10.3748/wjg.v21.i29.8787.

Jess, T., Riis, L., Jespersgaard, C., Hougs, L., Andersen, P.S., Orholm, M.K., et al. (2005). Disease concordance, zygosity, and NOD2/CARD15 status: follow-up of a population-based cohort of Danish twins with inflammatory bowel disease. *Am J Gastroenterol* 100(11), 2486-2492. doi: 10.1111/j.1572-0241.2005.00224.x.

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., et al. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60(5), 631-637.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *nature* 491(7422), 119.

Kaakoush, N.O., Holmes, J., Octavia, S., Man, S.M., Zhang, L., Castano-Rodriguez, N., et al. (2010). Detection of Helicobacteraceae in intestinal biopsies of children with Crohn's disease. *Helicobacter* 15(6), 549-557. doi: 10.1111/j.1523-5378.2010.00792.x.

Kamada, N., Hisamatsu, T., Okamoto, S., Chinen, H., Kobayashi, T., Sato, T., et al. (2008). Unique CD14+ intestinal macrophages contribute to the pathogenesis of Crohn disease via IL-23/IFN-γ axis. *The Journal of clinical investigation* 118(6), 2269-2280.

Kanai, T., Watanabe, M., Okazawa, A., Sato, T., Yamazaki, M., Okamoto, S., et al. (2001). Macrophage-derived IL-18–mediated intestinal inflammation in the murine model of Crohn's disease. *Gastroenterology* 121(4), 875-888. doi: https://doi.org/10.1053/gast.2001.28021.

Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., et al. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis* 16(12), 2034-2042. doi: 10.1002/ibd.21319.

Kaplan, G.G., and Ng, S.C. (2017). Understanding and preventing the global increase of inflammatory bowel disease. *Gastroenterology* 152(Polito et al.), 313-321. e312.

Kayagaki, N., Stowe, I.B., Lee, B.L., O'Rourke, K., Anderson, K., Warming, S., et al. (2015). Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. *Nature* 526(7575), 666-671.

Keighley, M.R., Arabi, Y., Dimock, F., Burdon, D.W., Allan, R.N., and Alexander-Williams, J. (1978). Influence of inflammatory bowel disease on intestinal microflora. *Gut* 19(12), 1099. doi: 10.1136/gut.19.12.1099.

Keita, Å., Salim, S.Y., Jiang, T., Yang, P.C., Franzén, L., Söderkvist, P., et al. (2008). Increased uptake of non-pathogenic E. coli via the follicle-associated epithelium in longstanding ileal Crohn's disease. *The Journal of Pathology* 215(Polito et al.), 135-144. doi: 10.1002/path.2337.

Keita, Å.V., Alkaissi, L.Y., Holm, E.B., Heil, S.D.S., Chassaing, B., Darfeuille-Michaud, A., et al. (2019). Enhanced E. coli LF82 Translocation through the Follicle-associated Epithelium in Crohn's Disease is Dependent on Long Polar Fimbriae and CEACAM6 expression, and Increases Paracellular Permeability. *Journal of Crohn's and Colitis* 14(Polito et al.), 216-229. doi: 10.1093/ecco-jcc/jjz144.

Kelley, N., Jeltema, D., Duan, Y., and He, Y. (2019). The NLRP3 Inflammasome: An Overview of Mechanisms of Activation and Regulation. *International journal of molecular sciences* 20(13), 3328. doi: 10.3390/ijms20133328.

Kennedy, N.A., Lamb, C.A., Berry, S.H., Walker, A.W., Mansfield, J., Parkes, M., et al. (2018). The Impact of NOD2 Variants on Fecal Microbiota in Crohn's Disease and Controls Without Gastrointestinal Disease. *Inflammatory Bowel Diseases* 24(3), 583-592. doi: 10.1093/ibd/izx061.

Kevans, D., Silverberg, M.S., Borowski, K., Griffiths, A., Xu, W., Onay, V., et al. (2016). IBD Genetic Risk Profile in Healthy First-Degree Relatives of Crohn's Disease Patients. *Journal of Crohn's & colitis* 10(Polito et al.), 209-215. doi: 10.1093/ecco-jcc/jjv197.

Khor, B., Gardet, A., and Xavier, R.J. (2011). Genetics and pathogenesis of inflammatory bowel disease. *Nature* 474(7351), 307.

Kiely, C.J., Pavli, P., and O'Brien, C.L. (2018a). The microbiome of translocated bacterial populations in patients with and without inflammatory bowel disease. *Internal Medicine Journal* 48(11), 1346-1354. doi: https://doi.org/10.1111/imj.13998.

Kiely, C.J., Pavli, P., and O'Brien, C.L. (2018b). The microbiome of translocated bacterial populations in patients with and without inflammatory bowel disease. *Intern Med J* 48(11), 1346-1354. doi: 10.1111/imj.13998.

Kiernan, M.G., Coffey, J.C., McDermott, K., Cotter, P.D., Cabrera-Rubio, R., Kiely, P.A., et al. (2019). The Human Mesenteric Lymph Node Microbiome Differentiates Between Crohn's Disease and Ulcerative Colitis. *J Crohns Colitis* 13(Goodhand et al.), 58-66. doi: 10.1093/ecco-jcc/jjy136.

Kim, S.W., Kim, E.S., Moon, C.M., Park, J.J., Kim, T.I., Kim, W.H., et al. (2011). Genetic polymorphisms of IL-23R and IL-17A and novel insights into their associations with inflammatory bowel disease. *Gut* 60(11), 1527. doi: 10.1136/gut.2011.238477.

Kinjo, F., Tomiyama, R., Hokama, A., Matayoshi, R., Yonamine, Y., Sunagawa, T., et al. (2004). Gastrointestinal: Aphthous lesions to overt Crohn's disease. *Journal of gastroenterology and hepatology* 19(6), 707-707.

Kirsner, J.B. (1988). Historical aspects of inflammatory bowel disease. *J Clin Gastroenterol* 10(3), 286-297. doi: 10.1097/00004836-198806000-00012.

Knights, D., Silverberg, M.S., Weersma, R.K., Gevers, D., Dijkstra, G., Huang, H., et al. (2014). Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome medicine* 6(12), 107.

Kobayashi, N., Takahashi, D., Takano, S., Kimura, S., and Hase, K. (2019). The Roles of Peyer's Patches and Microfold Cells in the Gut Immune System: Relevance to Autoimmune Diseases. *Frontiers in Immunology* 10(2345). doi: 10.3389/fimmu.2019.02345.

Kobayashi, T., Okamoto, S., Hisamatsu, T., Kamada, N., Chinen, H., Saito, R., et al. (2008). IL23 differentially regulates the Th1/Th17 balance in ulcerative colitis and Crohn's disease. *Gut* 57(12), 1682-1689.

Kostic, A.D., Xavier, R.J., and Gevers, D. (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 146(6), 1489-1499. doi: 10.1053/j.gastro.2014.02.009.

Kotlowski, R., Bernstein, C.N., Sepehri, S., and Krause, D.O. (2007a). High prevalence of Escherichia coli belonging to the B2+ D phylogenetic group in inflammatory bowel disease. *Gut* 56(5), 669-675.

Kotlowski, R., Bernstein, C.N., Sepehri, S., and Krause, D.O. (2007b). High prevalence of Escherichia coli belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut* 56(5), 669-675. doi: 10.1136/gut.2006.099796.

Kowalska-Duplaga, K., Gosiewski, T., Kapusta, P., Sroka-Oleksiak, A., Wędrychowicz, A., Pieczarkowski, S., et al. (2019). Differences in the intestinal microbiome of healthy children and patients with newly diagnosed Crohn's disease. *Scientific Reports* 9(Goodhand et al.)**,** 18880. doi: 10.1038/s41598-019-55290-9.

Kraehenbuhl, J.P., and Neutra, M.R. (2000). Epithelial M cells: differentiation and function. *Annu Rev Cell Dev Biol* 16**,** 301-332. doi: 10.1146/annurev.cellbio.16.1.301.

Kroesen, A.J., Becker, B., Fromm, M., Foitzik, T., Schulzke, J.D., and Buhr, H.J. (2000). Bacterial translocation in IBD: Increased incidence only in fistulizing Crohn's disease. *Gastroenterology* 118(4, Part 2)**,** A1352. doi: https://doi.org/10.1016/S0016-5085(00)81278-0.

Kugathasan, S., Saubermann, L.J., Smith, L., Kou, D., Itoh, J., Binion, D.G., et al. (2007). Mucosal T-cell immunoregulation varies in early and late inflammatory bowel disease. *Gut* 56(12)**,** 1696-1705. doi: 10.1136/gut.2006.116467.

Kurtz, L.E., Yang, S.S., and Bank, S. (2010). Clostridium difficile-associated small bowel enteritis after total proctocolectomy in a Crohn's disease patient. *J Clin Gastroenterol* 44(Goodhand et al.)**,** 76-77. doi: 10.1097/MCG.0b013e3181a7481b.

Laffineur, G., Lescut, D., Vincent, P., Quandalle, P., Wurtz, A., and Colombel, J.F. (1992). [Bacterial translocation in Crohn disease]. *Gastroenterol Clin Biol* 16(10)**,** 777-781.

LAHARIE, D., ASENCIO, C., ASSELINEAU, J., BULOIS, P., BOURREILLE, A., MOREAU, J., et al. (2009). Association between entero-hepatic Helicobacter species and Crohn's disease: a prospective cross-sectional study. *Alimentary Pharmacology & Therapeutics* 30(3)**,** 283-293. doi: 10.1111/j.1365-2036.2009.04034.x.

Lahat, A., Kopylov, U., Neuman, S., Levhar, N., Yablecovitch, D., Avidan, B., et al. (2017). Helicobacter pylori prevalence and clinical significance in patients with quiescent Crohn's disease. *BMC Gastroenterol* 17(Goodhand et al.)**,** 27. doi: 10.1186/s12876-017-0588-7.

Lal, S., and Steinhart, A.H. (2006). Antibiotic therapy for Crohn's disease: a review. *Can J Gastroenterol* 20(10)**,** 651-655.

Lamont, E.A., Bannantine, J.P., Armién, A., Ariyakumar, D.S., and Sreevatsan, S. (2012). Identification and characterization of a spore-like morphotype in chronically starved Mycobacterium avium subsp. paratuberculosis cultures. *PLoS One* 7(Goodhand et al.).

LaRock, C.N., Todd, J., LaRock, D.L., Olson, J., O'Donoghue, A.J., Robertson, A.A.B., et al. (2016). IL-1β is an innate immune sensor of microbial proteolysis. *Science immunology* 1(Polito et al.)**,** eaah3539. doi: 10.1126/sciimmunol.aah3539.

Lassen, K.G., Kuballa, P., Conway, K.L., Patel, K.K., Becker, C.E., Peloquin, J.M., et al. (2014). Atg16L1 T300A variant decreases selective autophagy resulting in altered cytokine signaling and decreased antibacterial defense. *Proceedings of the National Academy of Sciences of the United States of America* 111(21)**,** 7741-7746. doi: 10.1073/pnas.1407001111.

Latz, E., Xiao, T.S., and Stutz, A. (2013). Activation and regulation of the inflammasomes. *Nature Reviews Immunology* 13(6)**,** 397-411.

Lazaridis, L.D., Pistiki, A., Giamarellos-Bourboulis, E.J., Georgitsi, M., Damoraki, G., Polymeros, D., et al. (2017). Activation of NLRP3 Inflammasome in Inflammatory Bowel Disease: Differences Between Crohn's Disease and Ulcerative Colitis. *Dig Dis Sci* 62(9)**,** 2348-2356. doi: 10.1007/s10620-017-4609-8.

Le Baut, G., O'Brien, C., Pavli, P., Roy, M., Seksik, P., Tréton, X., et al. (2018). Prevalence of Yersinia Species in the Ileum of Crohn's Disease Patients and Controls. *Frontiers in Cellular and Infection Microbiology* 8. doi: 10.3389/fcimb.2018.00336.

Ledder, O., Catto-Smith, A.G., Oliver, M.R., Alex, G., Cameron, D.J., and Hardikar, W. (2014). Clinical patterns and outcome of early-onset inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 59(5), 562-564. doi: 10.1097/mpg.0000000000000465.

Lee, J.C., and Parkes, M. (2011). Genome-wide association studies and Crohn's disease. *Brief Funct Genomics* 10(Polito et al.), 71-76. doi: 10.1093/bfgp/elr009.

Leshinsky-Silver, E., Karban, A., Buzhakor, E., Fridlander, M., Yakir, B., Eliakim, R., et al. (2005). Is age of onset of Crohn's disease governed by mutations in NOD2/caspase recruitment domains 15 and Toll-like receptor 4? Evaluation of a pediatric cohort. *Pediatric research* 58(3), 499-504.

Leu, S.B., Shulman, S.C., Steelman, C.K., Lamps, L.W., Bulut, O.P., Abramowsky, C.R., et al. (2013). Pathogenic Yersinia DNA in intestinal specimens of pediatric patients with Crohn's disease. *Fetal Pediatr Pathol* 32(5), 367-370. doi: 10.3109/15513815.2013.768744.

Levine, A., Griffiths, A., Markowitz, J., Wilson, D.C., Turner, D., Russell, R.K., et al. (2011). Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflamm Bowel Dis* 17(6), 1314-1321. doi: 10.1002/ibd.21493.

Levine, A., Kugathasan, S., Annese, V., Biank, V., Leshinsky-Silver, E., Davidovich, O., et al. (2007). Pediatric onset Crohn's colitis is characterized by genotype-dependent age-related susceptibility. *Inflammatory bowel diseases* 13(12), 1509-1515.

Levine, B. (2005). Eating oneself and uninvited guests: autophagy-related pathways in cellular defense. *Cell* 120(Polito et al.), 159-162. doi: 10.1016/j.cell.2005.01.005.

Lewis, C.M., Whitwell, S.C., Forbes, A., Sanderson, J., Mathew, C.G., and Marteau, T.M. (2007). Estimating risks of common complex diseases across genetic and environmental factors: the example of Crohn disease. *Journal of medical genetics* 44(11), 689-694.

Lewis, J.D., and Abreu, M.T. (2017). Diet as a trigger or therapy for inflammatory bowel diseases. *Gastroenterology* 152(Polito et al.), 398-414. e396.

Liston, A., and Masters, S.L. (2017). Homeostasis-altering molecular processes as mechanisms of inflammasome activation. *Nature Reviews Immunology* 17(3), 208.

Liu, L., Dong, Y., Ye, M., Jin, S., Yang, J., Joosse, M., et al. (2016). The Pathogenic Role of NLRP3 Inflammasome Activation in Inflammatory Bowel Diseases of Both Mice and Humans. *Journal of Crohn's & colitis* 11. doi: 10.1093/ecco-jcc/jjw219.

Liverani, E., Scaioli, E., Cardamone, C., Dal Monte, P., and Belluzzi, A. (2014). Mycobacterium avium subspecies paratuberculosis in the etiology of Crohn's disease, cause or epiphenomenon? *World journal of gastroenterology* 20(36), 13060-13070. doi: 10.3748/wjg.v20.i36.13060.

Loddo, I., and Romano, C. (2015). Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. *Frontiers in Immunology* 6, 551. doi: 10.3389/fimmu.2015.00551.

Loftus, E.V., Jr. (2004). Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 126(6), 1504-1517.

Loh, G., and Blaut, M. (2012). Role of commensal gut bacteria in inflammatory bowel diseases. *Gut microbes* 3(6), 544-555. doi: 10.4161/gmic.22156.

Long, M.D., Kappelman, M.D., Martin, C.F., Chen, W., Anton, K., and Sandler, R.S. (2016). Role of Nonsteroidal Anti-Inflammatory Drugs in Exacerbations of Inflammatory Bowel Disease. *J Clin Gastroenterol* 50(Polito et al.), 152-156. doi: 10.1097/mcg.0000000000000421.

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., Duncan, S.H., Flint, H.J., et al. (2014). Mucosa-associated Faecalibacterium prausnitzii and Escherichia coli co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int J Med Microbiol* 304(3-4), 464-475. doi: 10.1016/j.ijmm.2014.02.009.

Lovasz, B.D., Lakatos, L., Horvath, A., Szita, I., Pandur, T., Mandel, M., et al. (2013). Evolution of disease phenotype in adult and pediatric onset Crohn's disease in a population-based cohort. *World J Gastroenterol* 19(14), 2217-2226. doi: 10.3748/wjg.v19.i14.2217.

Lu, A., Magupalli, V.G., Ruan, J., Yin, Q., Atianand, M.K., Vos, M.R., et al. (2014). Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes. *Cell* 156(6), 1193-1206.

Lu, X.C., Tao, Y., Wu, C., Zhao, P.L., Li, K., Zheng, J.Y., et al. (2013). Association between Variants of the Autophagy Related Gene – IRGM and Susceptibility to Crohn's Disease and Ulcerative Colitis: A Meta-Analysis. *PLOS ONE* 8(11), e80602. doi: 10.1371/journal.pone.0080602.

Lupp, C., Robertson, M.L., Wickham, M.E., Sekirov, I., Champion, O.L., Gaynor, E.C., et al. (2007). Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell host & microbe* 2(Polito et al.), 119-129.

Machiels, K., Pozuelo del Río, M., Martinez-De la Torre, A., Xie, Z., Pascal Andreu, V., Sabino, J., et al. (2020). Early Postoperative Endoscopic Recurrence in Crohn's Disease Is Characterised by Distinct Microbiota Recolonisation. *Journal of Crohn's and Colitis*. doi: 10.1093/ecco-jcc/jjaa081.

Maeda, S., Hsu, L.C., Liu, H., Bankston, L.A., Iimura, M., Kagnoff, M.F., et al. (2005). Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. *Science* 307(5710), 734-738. doi: 10.1126/science.1103685.

Magin, W.S., Van Kruiningen, H.J., and Colombel, J.F. (2013). Immunohistochemical search for viral and bacterial antigens in Crohn's disease. *J Crohns Colitis* 7(Polito et al.), 161-166. doi: 10.1016/j.crohns.2012.03.021.

Mahendran, V., Riordan, S.M., Grimm, M.C., Tran, T.A.T., Major, J., Kaakoush, N.O., et al. (2011). Prevalence of Campylobacter species in adult Crohn's disease and the preferential colonization sites of Campylobacter species in the human intestine. *PLoS One* 6(9), e25417.

Maioli, T.U., Borras-Nogues, E., Torres, L., Barbosa, S.C., Martins, V.D., Langella, P., et al. (2021). Possible Benefits of Faecalibacterium prausnitzii for Obesity-Associated Gut Disorders. *Frontiers in Pharmacology* 12. doi: 10.3389/fphar.2021.740636.

Man, S.M., Zhang, L., Day, A.S., Leach, S.T., Lemberg, D.A., and Mitchell, H. (2010). Campylobacter concisus and other Campylobacter species in children with newly diagnosed Crohn's disease. *Inflammatory bowel diseases* 16(6), 1008-1016.

Mangan, M.S., Olhava, E.J., Roush, W.R., Seidel, H.M., Glick, G.D., and Latz, E. (2018). Targeting the NLRP3 inflammasome in inflammatory diseases. *Nature Reviews Drug Discovery* 17(8), 588.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., et al. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55(Polito et al.), 205-211. doi: 10.1136/gut.2005.073817.

Martin, H.M., Campbell, B.J., Hart, C.A., Mpofu, C., Nayar, M., and Singh, R. (2004). Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology* 127. doi: 10.1053/j.gastro.2004.03.054.

Martinez-Medina, M., and Garcie-Gil, L.G. (2014). Escherichia coli In chronic inflammatory bowel diseases: an update on adherent invasive Escherichia coli pathogenicity. *World J Gastrointest Pathophysiol* 5.

Martinez-Medina, M.X., Aldeguer, M., Lopez-Siles, F., González-Huix, C., López-Oliu, G., and Dahbi, G. (2009). molecular diversity of Escherichia coli in the human gut: new ecological evidence supporting the role of adherent-invasive E. coli (AIEC) in Crohn's disease. *Inflamm Bowel Dis* 15. doi: 10.1002/ibd.20860.

Martinon, F., Pétrilli, V., Mayor, A., Tardivel, A., and Tschopp, J. (2006). Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* 440(7081)**,** 237-241.

Mawdsley, J., and Rampton, D. (2005). Psychological stress in IBD: new insights into pathogenic and therapeutic implications. *Gut* 54(10)**,** 1481-1491.

McGovern, D.P.B., Rotter, J.I., Mei, L., Haritunians, T., Landers, C., Derkowski, C., et al. (2009). Genetic epistasis of IL23/IL17 pathway genes in Crohn's disease. *Inflammatory bowel diseases* 15(6)**,** 883-889. doi: 10.1002/ibd.20855.

McIlroy, J., Ianiro, G., Mukhopadhya, I., Hansen, R., and Hold, G. (2018). the gut microbiome in inflammatory bowel disease—avenues for microbial management. *Alimentary pharmacology & therapeutics* 47(Goodhand et al.)**,** 26-42.

McNees, A.L., Markesich, D., Zayyani, N.R., and Graham, D.Y. (2015). Mycobacterium paratuberculosis as a cause of Crohn's disease. *Expert review of gastroenterology & hepatology* 9(12)**,** 1523-1534. doi: 10.1586/17474124.2015.1093931.

Melsheimer, R., Geldhof, A., Apaolaza, I., and Schaible, T. (2019). Remicade(®) (infliximab): 20 years of contributions to science and medicine. *Biologics : targets & therapy* 13**,** 139-178. doi: 10.2147/BTT.S207246.

Miao, E.A., Alpuche-Aranda, C.M., Dors, M., Clark, A.E., Bader, M.W., Miller, S.I., et al. (2006). Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1β via Ipaf. *Nature Immunology* 7(6)**,** 569-575. doi: 10.1038/ni1344.

Miao, E.A., Andersen-Nissen, E., Warren, S.E., and Aderem, A. (2007). TLR5 and Ipaf: dual sensors of bacterial flagellin in the innate immune system. *Seminars in Immunopathology* 29(3)**,** 275-288. doi: 10.1007/s00281-007-0078-z.

Michielan, A., and D'Incà, R. (2015). Intestinal Permeability in Inflammatory Bowel Disease: Pathogenesis, Clinical Evaluation, and Therapy of Leaky Gut. *Mediators Inflamm* 2015**,** 628157. doi: 10.1155/2015/628157.

Michielan, A., and D'Incà, R. (2015). Intestinal permeability in inflammatory bowel disease: pathogenesis, clinical evaluation, and therapy of leaky gut. *Mediators of inflammation* 2015.

Miranda-Bautista, J., Padilla-Suárez, C., Bouza, E., Muñoz, P., Menchén, L., and Marín-Jiménez, I. (2014). Listeria monocytogenes infection in inflammatory bowel disease patients: case series and review of the literature. *Eur J Gastroenterol Hepatol* 26(11)**,** 1247-1252. doi: 10.1097/meg.0000000000000188.

Mizoguchi, A., and Mizoguchi, E. (2008). Inflammatory bowel disease, past, present and future: lessons from animal models. *J Gastroenterol* 43(Goodhand et al.)**,** 1-17. doi: 10.1007/s00535-007-2111-3.

Moller, F.T., Andersen, V., Wohlfahrt, J., and Jess, T. (2015). Familial risk of inflammatory bowel disease: a population-based cohort study 1977-2011. *Am J Gastroenterol* 110(4)**,** 564-571. doi: 10.1038/ajg.2015.50.

Molodecky, N.A., and Kaplan, G.G. (2010). Environmental Risk Factors for Inflammatory Bowel Disease. *Gastroenterology & Hepatology* 6(5)**,** 339-346.

Molodecky, N.A., Soon, S., Rabi, D.M., Ghali, W.A., Ferris, M., Chernoff, G., et al. (2012). Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 142(Goodhand et al.)**,** 46-54. e42.

Mondot, S., Kang, S., Furet, J.P., Aguirre de Carcer, D., McSweeney, C., Morrison, M., et al. (2011). Highlighting new phylogenetic specificities of Crohn's disease microbiota. *Inflamm Bowel Dis* 17(Goodhand et al.)**,** 185-192. doi: 10.1002/ibd.21436.

Monteleone, G., Biancone, L., Marasco, R., Morrone, G., Marasco, O., Luzza, F., et al. (1997). Interleukin 12 is expressed and actively released by Crohn's disease intestinal lamina propria mononuclear cells. *Gastroenterology* 112(4)**,** 1169-1178.

Monteleone, G., Trapasso, F., Parrello, T., Biancone, L., Stella, A., Iuliano, R., et al. (1999). Bioactive IL-18 expression is up-regulated in Crohn's disease. *J Immunol* 163(Goodhand et al.)**,** 143-147.

Mow, W.S., Landers, C.J., Steinhart, A.H., Feagan, B.G., Croitoru, K., Seidman, E., et al. (2004). High-level serum antibodies to bacterial antigens are associated with antibiotic-induced clinical remission in Crohn's disease: a pilot study. *Digestive diseases and sciences* 49(7)**,** 1280-1286.

Mühl, L., Becker, E., Müller, T.M., Atreya, R., Atreya, I., Neurath, M.F., et al. (2021). Clinical experiences and predictors of success of treatment with vedolizumab in IBD patients: a cohort study. *BMC Gastroenterology* 21(Goodhand et al.)**,** 33. doi: 10.1186/s12876-021-01604-z.

Mulder, D.J., Noble, A.J., Justinich, C.J., and Duffin, J.M. (2014). A tale of two diseases: The history of inflammatory bowel disease. *Journal of Crohn's and Colitis* 8(5)**,** 341-348. doi: 10.1016/j.crohns.2013.09.009.

Muñoz-Planillo, R., Kuffa, P., Martínez-Colón, G., Smith, B.L., Rajendiran, T.M., and Núñez, G. (2013). K+ efflux is the common trigger of NLRP3 inflammasome activation by bacterial toxins and particulate matter. *Immunity* 38(6)**,** 1142-1153.

Murakami, T., Ockinger, J., Yu, J., Byles, V., McColl, A., Hofer, A.M., et al. (2012). Critical role for calcium mobilization in activation of the NLRP3 inflammasome. *Proceedings of the National Academy of Sciences* 109(28)**,** 11282-11287.

Murray, P.J., and Wynn, T.A. (2011). Protective and pathogenic functions of macrophage subsets. *Nature Reviews Immunology* 11(11)**,** 723-737. doi: 10.1038/nri3073.

Musa, S., Thomson, S., Cowan, M., and Rahman, T. (2010). Clostridium difficile infection and inflammatory bowel disease. *Scandinavian journal of gastroenterology* 45(3)**,** 261-272.

Nagalingam, N.A., and Lynch, S.V. (2012). Role of the microbiota in inflammatory bowel diseases. *Inflammatory bowel diseases* 18(5)**,** 968-984.

Nakahira, K., Haspel, J.A., Rathinam, V.A., Lee, S.-J., Dolinay, T., Lam, H.C., et al. (2011). Autophagy proteins regulate innate immune responses by inhibiting the release of mitochondrial DNA mediated by the NALP3 inflammasome. *Nature immunology* 12(3)**,** 222.

Nakanishi, Y., Sato, T., and Ohteki, T. (2015). Commensal Gram-positive bacteria initiates colitis by inducing monocyte/macrophage mobilization. *Mucosal Immunol* 8(Goodhand et al.)**,** 152-160. doi: 10.1038/mi.2014.53.

Narula, N., Wong, E.C.L., Dehghan, M., Mente, A., Rangarajan, S., Lanas, F., et al. (2021). Association of ultra-processed food intake with risk of inflammatory bowel disease: prospective cohort study. *BMJ* 374**,** n1554. doi: 10.1136/bmj.n1554.

Narula, N., and Fedorak, R.N. (2008). Exercise and inflammatory bowel disease. *Canadian Journal of Gastroenterology* 22.

Naser, S.A., Arce, M., Khaja, A., Fernandez, M., Naser, N., Elwasila, S., et al. (2012a). Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. *World J Gastroenterol* 18(5)**,** 412-424. doi: 10.3748/wjg.v18.i5.412.

Naser, S.A., Arce, M., Khaja, A., Fernandez, M., Naser, N., Elwasila, S., et al. (2012b). Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. *World journal of gastroenterology* 18(5)**,** 412-424. doi: 10.3748/wjg.v18.i5.412.

Neurath, M.F. (2014). Cytokines in inflammatory bowel disease. *Nature Reviews Immunology* 14(5)**,** 329-342. doi: 10.1038/nri3661.

Neutra, M. (1999). "M cells in antigen sampling in mucosal tissues," in *Defense of Mucosal Surfaces: Pathogenesis, Immunity and Vaccines*. Springer), 17-32.

Newell, D.G. (2005). Campylobacter concisus: an emerging pathogen? *European journal of gastroenterology & hepatology* 17(10), 1013-1014.

Newman, A., and Lambert, J.R. (1980). Campylobacter jejuni causing flare-up in inflammatory bowel disease. *Lancet* 2(8200), 919. doi: 10.1016/s0140-6736(80)92078-4.

Ng, S.C., Bernstein, C.N., Vatn, M.H., Lakatos, P.L., Loftus, E.V., Tysk, C., et al. (2013a). Geographical variability and environmental risk factors in inflammatory bowel disease. *Gut* 62(4), 630-649.

Ng, S.C., Shi, H.Y., Hamidi, N., Underwood, F.E., Tang, W., Benchimol, E.I., et al. (2017). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *The Lancet* 390(10114), 2769-2778.

Ng, S.C., Shi, H.Y., Hamidi, N., Underwood, F.E., Tang, W., Benchimol, E.I., et al. (2018). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 390(10114), 2769-2778. doi: 10.1016/s0140-6736(17)32448-0.

Ng, S.C., Tang, W., Ching, J.Y., Wong, M., Chow, C.M., Hui, A.J., et al. (2013b). Incidence and Phenotype of Inflammatory Bowel Disease Based on Results From the Asia-Pacific Crohn's and Colitis Epidemiology Study. *Gastroenterology* 145(Goodhand et al.), 158-165.e152. doi: https://doi.org/10.1053/j.gastro.2013.04.007.

Nitzan, O., Elias, M., Peretz, A., and Saliba, W. (2016). Role of antibiotics for treatment of inflammatory bowel disease. *World Journal of Gastroenterology* 22(3), 1078-1087. doi: 10.3748/wjg.v22.i3.1078.

Noguchi, E., Homma, Y., Kang, X., Netea, M.G., and Ma, X. (2009). A Crohn's disease-associated NOD2 mutation suppresses transcription of human IL10 by inhibiting activity of the nuclear ribonucleoprotein hnRNP-A1. *Nat Immunol* 10(5), 471-479. doi: 10.1038/ni.1722.

Noth, R., Stüber, E., Häsler, R., Nikolaus, S., Kühbacher, T., Hampe, J., et al. (2012). Anti-TNF-α antibodies improve intestinal barrier function in Crohn's disease. *Journal of Crohn's and Colitis* 6(4), 464-469.

O'Brien, C.L., Bringer, M.-A., Holt, K.E., Gordon, D.M., Dubois, A.L., Barnich, N., et al. (2016). Comparative genomics of Crohn's disease-associated adherent-invasive Escherichia coli. *Gut*. doi: 10.1136/gutjnl-2015-311059.

O'Brien, C.L., Pavli, P., Gordon, D.M., and Allison, G.E. (2014). Detection of bacterial DNA in lymph nodes of Crohn's disease patients using high throughput sequencing. *Gut*. doi: 10.1136/gutjnl-2013-305320.

O'Hara, A.M., and Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO reports* 7(7), 688-693.

O'Boyle, C., MacFie, J., Mitchell, C., Johnstone, D., Sagar, P., and Sedman, P. (1998). Microbiology of bacterial translocation in humans. *Gut* 42(Goodhand et al.), 29-35.

O'Toole, A., and Korzenik, J. (2014). Environmental Triggers for IBD. *Current Gastroenterology Reports* 16(7), 396. doi: 10.1007/s11894-014-0396-y.

Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., et al. (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411(6837), 603-606. doi: 10.1038/35079114.

Oken, H.A., Saleeb, P.G., Redfield, R.R., and Schimpff, S.C. (2017). Is Mycobacterium avium paratuberculosis the Trigger in the Crohn's Disease Spectrum? *Open Forum Infectious Diseases* 4(3). doi: 10.1093/ofid/ofx104.

Okou, D.T., and Kugathasan, S. (2014). Role of genetics in pediatric inflammatory bowel disease. *Inflammatory bowel diseases* 20(10), 1878-1884. doi: 10.1097/MIB.0000000000000085.

Olsen, T., Rismo, R., Cui, G., Goll, R., Christiansen, I., and Florholmen, J. (2011). TH1 and TH17 interactions in untreated inflamed mucosa of inflammatory bowel disease, and their potential to mediate the inflammation. *Cytokine* 56(3), 633-640.

Orholm, M., Binder, V., Sorensen, T.I., Rasmussen, L.P., and Kyvik, K.O. (2000). Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand J Gastroenterol* 35(10), 1075-1081.

Ostrowski, J., Paziewska, A., Lazowska, I., Ambrozkiewicz, F., Goryca, K., Kulecka, M., et al. (2016). Genetic architecture differences between pediatric and adult-onset inflammatory bowel diseases in the Polish population. *Scientific Reports* 6, 39831. doi: 10.1038/srep39831

Palomino-Morales, R., Oliver, J., Gomez-Garcia, M., Lopez-Nevot, M., Rodrigo, L., Nieto, A., et al. (2009). Association of ATG16L1 and IRGM genes polymorphisms with inflammatory bowel disease: a meta-analysis approach. *Genes and immunity* 10(4), 356-364.

Palmieri, O., Bossa, F., Valvano, M.R., Corritore, G., Latiano, T., Martino, G., et al. (2017). Crohn's disease localization displays different predisposing genetic variants. PLoS One 12(1), e0168821. doi: 10.1371/journal.pone.0168821.

Papamichael, K., Konstantopoulos, P., and Mantzaris, G.J. (2014). Helicobacter pylori infection and inflammatory bowel disease: is there a link? *World J Gastroenterol* 20(21), 6374-6385. doi: 10.3748/wjg.v20.i21.6374.

Parada Venegas, D., De la Fuente, M.K., Landskron, G., González, M.J., Quera, R., Dijkstra, G., et al. (2019). Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. *Frontiers in Immunology* 10. doi: 10.3389/fimmu.2019.00277.

Parkes, M., Barrett, J.C., Prescott, N.J., Tremelling, M., Anderson, C.A., Fisher, S.A., et al. (2007). Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 39(7), 830-832. doi: 10.1038/ng2061.

Parrish, N.M., Radcliff, R.P., Brey, B.J., Anderson, J.L., Clark, D.L., Jr., Koziczkowski, J.J., et al. (2009). Absence of mycobacterium avium subsp. paratuberculosis in Crohn's patients. *Inflamm Bowel Dis* 15(4), 558-565. doi: 10.1002/ibd.20799.

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., et al. (2017). A microbial signature for Crohn's disease. *Gut* 66(5), 813-822.

Paul, T., Birnbaum, A., Pal, D.K., Pittman, N., Ceballos, C., LeLeiko, N.S., et al. (2006). Distinct phenotype of early childhood inflammatory bowel disease. *Journal of clinical gastroenterology* 40(7), 583-586.

Pavel, M., and Rubinsztein, D.C. (2017). Mammalian autophagy and the plasma membrane. *The FEBS journal* 284(5), 672-679.

Pearson, A.D., Eastham, E.J., Laker, M.F., Craft, A.W., and Nelson, R. (1982). Intestinal permeability in children with Crohn's disease and coeliac disease. *Br Med J (Clin Res Ed)* 285(6334), 20-21. doi: 10.1136/bmj.285.6334.20.

Pellegrini, C., Antonioli, L., Lopez-Castejon, G., Blandizzi, C., and Fornai, M. (2017). Canonical and non-canonical activation of NLRP3 inflammasome at the crossroad between immune tolerance and intestinal inflammation. *Frontiers in immunology* 8, 36.

Perez Pérez, C.A. (2009). Prescription of physical exercise in Crohn's disease. *Journal of Crohn's and Colitis* 3(4), 225-231.

Peterson, L.W., and Artis, D. (2014). Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat Rev Immunol* 14(3), 141-153. doi: 10.1038/nri3608.

Peterson, N., Guthery, S., Denson, L., Lee, J., Saeed, S., Prahalad, S., et al. (2008). Genetic variants in the autophagy pathway contribute to paediatric Crohn's disease. *Gut* 57(9), 1336-1337.

Peyrin-Biroulet, L., Gonzalez, F., Dubuquoy, L., Rousseaux, C., Dubuquoy, C., Decourcelle, C., et al. (2012). Mesenteric fat as a source of C reactive protein and as a target for bacterial translocation in Crohn's disease. *Gut* 61(Goodhand et al.), 78-85.

Pigneur, B., Seksik, P., Viola, S., Viala, J., Beaugerie, L., Girardet, J.-P., et al. (2009). Natural history of Crohn's disease: comparison between childhood-and adult-onset disease. *Inflammatory bowel diseases* 16(6), 953-961.

Plantinga, T.S., Joosten, L.A., and Netea, M.G. (2011). ATG16L1 polymorphisms are associated with NOD2-induced hyperinflammation. *Autophagy* 7(9), 1074-1075.

Platt, A.M., and Mowat, A.M. (2008). Mucosal macrophages and the regulation of immune responses in the intestine. *Immunology letters* 119(1-2), 22-31.

Png, C.W., Lindén, S.K., Gilshenan, K.S., Zoetendal, E.G., McSweeney, C.S., Sly, L.I., et al. (2010). Mucolytic Bacteria With Increased Prevalence in IBD Mucosa AugmentIn VitroUtilization of Mucin by Other Bacteria. *American Journal of Gastroenterology* 105(11).

Podolsky , D.K. (2002). Inflammatory Bowel Disease. *New England Journal of Medicine* 347(6), 417-429. doi: 10.1056/NEJMra020831.

Posovszky, C., Pfalzer, V., Lahr, G., Niess, J.H., Klaus, J., Mayer, B., et al. (2013). Age-of-onset-dependent influence of NOD2 gene variants on disease behaviour and treatment in Crohn's disease. *BMC gastroenterology* 13, 77-77. doi: 10.1186/1471-230X-13-77.

Powell, J.J., Thomas-McKay, E., Thoree, V., Robertson, J., Hewitt, R.E., Skepper, J.N., et al. (2015). An endogenous nanomineral chaperones luminal antigen and peptidoglycan to intestinal immune cells. *Nature nanotechnology* 10(4), 361.

Prescott, N.J., Dominy, K.M., Kubo, M., Lewis, C.M., Fisher, S.A., Redon, R., et al. (2010). Independent and population-specific association of risk variants at the IRGM locus with Crohn's disease. *Human molecular genetics* 19(9), 1828-1839.

Prescott, N.J., Fisher, S.A., Franke, A., Hampe, J., Onnie, C.M., Soars, D., et al. (2007). A nonsynonymous SNP in ATG16L1 predisposes to ileal Crohn's disease and is independent of CARD15 and IBD5. *Gastroenterology* 132(5), 1665-1671. doi: 10.1053/j.gastro.2007.03.034.

Próchnicki, T., and Latz, E. (2017). Inflammasomes on the crossroads of innate immune recognition and metabolic control. *Cell metabolism* 26(Goodhand et al.), 71-93.

Proctor, L.M., Creasy, H.H., Fettweis, J.M., Lloyd-Price, J., Mahurkar, A., Zhou, W., et al. (2019). The Integrative Human Microbiome Project. *Nature* 569(7758), 641-648. doi: 10.1038/s41586-019-1238-8.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., et al. (2010). A human gut microbial gene catalog established by metagenomic sequencing. *Nature* 464(7285), 59-65. doi: 10.1038/nature08821.

Rahimi, R., Nikfar, S., Rezaie, A., and Abdollahi, M. (2006). A meta-analysis of broad-spectrum antibiotic therapy in patients with active Crohn's disease. *Clinical therapeutics* 28(12), 1983-1988.

Ramos, J., García-Sepulcre, M., Masiá, M., Brotons, A., Grau, M., and Gutiérrez, F. (2010). Listeria monocytogenes infection in patients with inflammatory bowel diseases receiving anti-tumor necrosis factor therapy. *Revista espanola de enfermedades digestivas: organo oficial de la Sociedad Espanola de Patologia Digestiva* 102(10), 614-616.

Rehman, A., Sina, C., Gavrilova, O., Häsler, R., Ott, S., Baines, J.F., et al. (2011). Nod2 is essential for temporal development of intestinal microbial communities. *Gut* 60(10), 1354-1362.

Ricanek, P., Lothe, S.M., Szpinda, I., Jorde, A.T., Brackmann, S., Perminow, G., et al. (2010). Paucity of mycobacteria in mucosal bowel biopsies from adults and children with early inflammatory bowel disease. *Journal of Crohn's and Colitis* 4(5), 561-566. doi: 10.1016/j.crohns.2010.05.003.

Rigoli, L., and Caruso, R.A. (2014). Inflammatory bowel disease in pediatric and adolescent patients: a biomolecular and histopathological review. *World journal of gastroenterology* 20(30), 10262-10278. doi: 10.3748/wjg.v20.i30.10262.

Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., et al. (2007a). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature genetics* 39(5), 596-604.

Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., et al. (2007b). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 39(5), 596-604. doi: 10.1038/ng2032.

Roberts, C.L., Keita, A.V., Duncan, S.H., O'Kennedy, N., Soderholm, J.D., Rhodes, J.M., et al. (2010). Translocation of Crohn's disease Escherichia coli across M-cells: contrasting effects of soluble plant fibres and emulsifiers. *Gut* 59(10), 1331-1339. doi: 10.1136/gut.2009.195370.

Rodríguez, C., Romero, E., Garrido-Sanchez, L., Alcaín-Martínez, G., Andrade, R.J., Taminiau, B., et al. (2020). Microbiota insights in Clostridium difficile infection and inflammatory bowel disease. *Gut Microbes*, 1-25. doi: 10.1080/19490976.2020.1725220.

Rogler, G., Zeitz, J., and Biedermann, L. (2016). The Search for Causative Environmental Factors in Inflammatory Bowel Disease. *Digestive Diseases* 34(suppl 1)(Suppl. 1), 48-55. doi: 10.1159/000447283.

Rolhion, N., Carvalho, F.A., and Darfeuille-Michaud, A. (2007). OmpC and the sigma(E) regulatory pathway are involved in adhesion and invasion of the Crohn's disease-associated Escherichia coli strain LF82. *Mol Microbiol* 63. doi: 10.1111/j.1365-2958.2007.05638.x.

Romagnani, S. (1994). Lymphokine production by human T cells in disease states. *Annual review of immunology* 12(Goodhand et al.), 227-257.

Rovedatti, L., Kudo, T., Biancheri, P., Sarra, M., Knowles, C., Rampton, D.S., et al. (2009). Differential regulation of interleukin 17 and interferon γ production in inflammatory bowel disease. *Gut* 58(12), 1629-1636.

Russell, R.K., and Satsangi, J. (2004). IBD: a family affair. *Best Pract Res Clin Gastroenterol* 18(3), 525-539. doi: 10.1016/j.bpg.2003.12.006.

Rutgeerts, P., Geboes, K., Vantrappen, G., Beyls, J., Kerremans, R., and Hiele, M. (1990). Predictability of the postoperative course of Crohn's disease. *Gastroenterology* 99(4), 956-963. doi: 10.1016/0016-5085(90)90613-6.

Rutgeerts, P., Geboes, K., Vantrappen, G., Kerremans, R., Coenegrachts, J., and Coremans, G. (1984). Natural history of recurrent Crohn's disease at the ileocolonic anastomosis after curative surgery. *Gut* 25(6), 665-672.

Rutgeerts, P., Goboes, K., Peeters, M., Hiele, M., Penninckx, F., Aerts, R., et al. (1991). Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum. *Lancet* 338(8770), 771-774.

Ryan, P., Kelly, R.G., Lee, G., Collins, J.K., O'Sullivan, G.C., O'Connell, J., et al. (2004). Bacterial DNA within Granulomas of Patients with Crohn's Disease[mdash]Detection by Laser Capture Microdissection and PCR. *Am J Gastroenterol* 99(8), 1539-1543.

Saebo, A., Vik, E., Lange, O.J., and Matuszkiewicz, L. (2005). Inflammatory bowel disease associated with Yersinia enterocolitica O: 3 infection. *European journal of internal medicine* 16(3), 176-182.

Sagiv-Friedgut, K., Karban, A., Weiss, B., Shaoul, R., Shamir, R., Bujanover, Y., et al. (2010). Early-onset Crohn disease is associated with male sex and a polymorphism in the IL-6 promoter. *J Pediatr Gastroenterol Nutr* 50(Goodhand et al.), 22-26. doi: 10.1097/MPG.0b013e3181b7a6a4.

Saleh, M., and Elson, C.O. (2011). Experimental inflammatory bowel disease: insights into the host-microbiota dialog. *Immunity* 34(3), 293-302.

Salem, M., Ammitzboell, M., Nys, K., Seidelin, J.B., and Nielsen, O.H. (2015). ATG16L1: A multifunctional susceptibility factor in Crohn disease. *Autophagy* 11(4), 585-594. doi: 10.1080/15548627.2015.1017187.

Salim, S.a.Y., Silva, M.A., Keita, Å.V., Larsson, M., Andersson, P., Magnusson, K.-E., et al. (2009). CD83(+)CCR7(−) Dendritic Cells Accumulate in the Subepithelial Dome and Internalize Translocated Escherichia coli HB101 in the Peyer's Patches of Ileal Crohn's Disease. *The American Journal of Pathology* 174(Goodhand et al.), 82-90. doi: 10.2353/ajpath.2009.080273.

Sandborn, W.J., Feagan, B.G., Fedorak, R.N., Scherl, E., Fleisher, M.R., Katz, S., et al. (2008). A randomized trial of Ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with moderate-to-severe Crohn's disease. *Gastroenterology* 135(4), 1130-1141.

Sandborn, W.J., Gasink, C., Gao, L.-L., Blank, M.A., Johanns, J., Guzzo, C., et al. (2012). Ustekinumab induction and maintenance therapy in refractory Crohn's disease. *New England Journal of Medicine* 367(16), 1519-1528.

Sands, B.E., Feagan, B.G., Rutgeerts, P., Colombel, J.-F., Sandborn, W.J., Sy, R., et al. (2014). Effects of vedolizumab induction therapy for patients with Crohn's disease in whom tumor necrosis factor antagonist treatment failed. Gastroenterology 147(3), 618-627. e613.

Sartor, R.B. (1997). The influence of normal microbial flora on the development of chronic mucosal inflammation. *Research in Immunology* 148(8), 567-576. doi: https://doi.org/10.1016/S0923-2494(98)80151-X.

Sartor, R.B. (2006). Mechanisms of Disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nature Clinical Practice Gastroenterology &Amp; Hepatology* 3, 390. doi: 10.1038/ncpgasthep0528.

Sartor, R.B. (2008). Microbial Influences in Inflammatory Bowel Diseases. *Gastroenterology* 134(Polito et al.), 577-594. doi: 10.1053/j.gastro.2007.11.059.

Saruta, M., Targan, S.R., Mei, L., Ippoliti, A.F., Taylor, K.D., and Rotter, J.I. (2009). High-frequency haplotypes in the X chromosome locus TLR8 are associated with both CD and UC in females. *Inflammatory bowel diseases* 15(3), 321-327.

Sasikala, M., Reddy, D.N., Pratap, N., Sharma, S.K., Balkumar, P.R., Sekaran, A., et al. (2009). Absence of Mycobacterium avium ss paratuberculosis-specific IS900 sequence in intestinal biopsy tissues of Indian patients with Crohn's disease. *Indian journal of gastroenterology* 28(5), 169-174.

Satsangi, J., Grootscholten, C., Holt, H., and Jewell, D. (1996). Clinical patterns of familial inflammatory bowel disease. *Gut* 38(5), 738-741.

Satsangi, J., Silverberg, M.S., Vermeire, S., and Colombel, J.F. (2006). The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 55(6)**,** 749-753. doi: 10.1136/gut.2005.082909.

Scharl, M., and Rogler, G. (2010). Microbial Sensing by the Intestinal Epithelium in the Pathogenesis of Inflammatory Bowel Disease. *International Journal of Inflammation* 2010**,** 671258. doi: 10.4061/2010/671258.

Schneeweiss, S., Korzenik, J., Solomon, D., Canning, C., Lee, J., and Bressler, B. (2009). Infliximab and other immunomodulating drugs in patients with inflammatory bowel disease and the risk of serious bacterial infections. *Alimentary pharmacology & therapeutics* 30(3)**,** 253-264.

Selby, W., Pavli, P., Crotty, B., Florin, T., Radford-Smith, G., Gibson, P., et al. (2007). Two-Year Combination Antibiotic Therapy With Clarithromycin, Rifabutin, and Clofazimine for Crohn's Disease. *Gastroenterology* 132(7)**,** 2313-2319. doi: https://doi.org/10.1053/j.gastro.2007.03.031.

Sepehri, S., Kotlowski, R., Bernstein, C.N., and Krause, D.O. (2007). Microbial diversity of inflamed and noninflamed gut biopsy tissues in inflammatory bowel disease. *Inflammatory bowel diseases* 13(6)**,** 675-683.

Shah, A., Talley, N.J., Walker, M., Koloski, N., Morrison, M., Burger, D., et al. (2017). Is There a Link Between H. Pylori and the Epidemiology of Crohn's Disease? *Digestive Diseases and Sciences* 62(9)**,** 2472-2480. doi: 10.1007/s10620-017-4496-z.

Sharif, F., McDermott, M., Dillon, M., Drumm, B., Rowland, M., Imrie, C., et al. (2002). Focally enhanced gastritis in children with Crohn's disease and ulcerative colitis. *Am J Gastroenterol* 97(6)**,** 1415-1420. doi: 10.1111/j.1572-0241.2002.05785.x.

Shaw, K.A., Cutler, D.J., Okou, D., Dodd, A., Aronow, B.J., Haberman, Y., et al. (2019). Genetic variants and pathways implicated in a pediatric inflammatory bowel disease cohort. *Genes & Immunity* 20(Polito et al.)**,** 131-142. doi: 10.1038/s41435-018-0015-2.

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., et al. (2015). Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature* 526(7575)**,** 660-665.

Shim, J.O. (2019). Recent Advance in Very Early Onset Inflammatory Bowel Disease. *Pediatric gastroenterology, hepatology & nutrition* 22(Goodhand et al.)**,** 41-49. doi: 10.5223/pghn.2019.22.1.41.

Shivashankar, R., Tremaine, W.J., Harmsen, W.S., and Loftus, E.V., Jr. (2017). Incidence and Prevalence of Crohn's Disease and Ulcerative Colitis in Olmsted County, Minnesota From 1970 Through 2010. *Clin Gastroenterol Hepatol* 15(6)**,** 857-863. doi: 10.1016/j.cgh.2016.10.039.

Sidiq, T., Yoshihama, S., Downs, I., and Kobayashi, K.S. (2016). Nod2: A Critical Regulator of Ileal Microbiota and Crohn's Disease. *Frontiers in Immunology* 7(367). doi: 10.3389/fimmu.2016.00367.

Singh, S., Al-Darmaki, A., Frolkis, A.D., Seow, C.H., Leung, Y., Novak, K.L., et al. (2015). Postoperative mortality among patients with inflammatory bowel diseases: a systematic review and meta-analysis of population-based studies. *Gastroenterology* 149(4)**,** 928-937.

Singh, S.B., Davis, A.S., Taylor, G.A., and Deretic, V. (2006). Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science* 313(5792)**,** 1438-1441. doi: 10.1126/science.1129577.

Sivanesan, D., Beauchamp, C., Quinou, C., Lee, J., Lesage, S., Chemtob, S., et al. (2016). IL23R (Interleukin 23 Receptor) Variants Protective against Inflammatory Bowel Diseases (IBD) Display Loss of Function due to Impaired Protein Stability and

Intracellular Trafficking. *J Biol Chem* 291(16), 8673-8685. doi: 10.1074/jbc.M116.715870.

Smith, P., Smythies, L., Shen, R., Greenwell-Wild, T., Gliozzi, M., and Wahl, S. (2011). Intestinal macrophages and response to microbial encroachment. *Mucosal immunology* 4(Goodhand et al.), 31-42.

Snapper, S.B. (2015). Very-Early-Onset Inflammatory Bowel Disease. *Gastroenterology & hepatology* 11(8), 554-556.

Snowden, J., Ansari, A., Sachchithanantham, S., Jackson, G., Thompson, N., Lobo, A., et al. (2014). Autologous stem cell transplantation in severe treatment-resistant Crohn's disease: long-term follow-up of UK patients treated on compassionate basis. QJM: An International Journal of Medicine 107(11), 871-877.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L.G., Gratadoux, J.J., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 105(43), 16731-16736. doi: 10.1073/pnas.0804812105.

Sokol, H., Seksik, P., Furet, J.P., Firmesse, O., Nion-Larmurier, I., Beaugerie, L., et al. (2009). Low counts of Faecalibacterium prausnitzii in colitis microbiota. *Inflamm Bowel Dis* 15(8), 1183-1189. doi: 10.1002/ibd.20903.

Solomon, M., McLeod, R., O'Connor, B., Steinhart, A., Greenberg, G., and Cohen, Z. (1993). Combination of ciprofloxacin and metronidazole in severe perianal Crohn's disease. *Canadian Journal of Gastroenterology* 7.

Sonnenberg, A. (1990). Occupational distribution of inflammatory bowel disease among German employees. *Gut* 31(9), 1037-1040. doi: 10.1136/gut.31.9.1037.

Sonnenberg, A., and Genta, R. (2012). Low prevalence of Helicobacter pylori infection among patients with inflammatory bowel disease. *Alimentary pharmacology & therapeutics* 35(4), 469-476.

Spehlmann, M.E., Begun, A.Z., Burghardt, J., Lepage, P., Raedler, A., and Schreiber, S. (2008). Epidemiology of inflammatory bowel disease in a German twin cohort: results of a nationwide study. *Inflamm Bowel Dis* 14(7), 968-976. doi: 10.1002/ibd.20380.

Spencer, E.A., Helmus, D., Telesco, S., Colombel, J.F., and Dubinsky, M.C. (2020). Inflammatory Bowel Disease Clusters Within Affected Sibships in Ashkenazi Jewish Multiplex Families. *Gastroenterology* 159(Goodhand et al.), 381-382. doi: 10.1053/j.gastro.2020.03.023.

Stange, E.F., Travis, S.P.L., Vermeire, S., Beglinger, C., Kupcinskas, L., Geboes, K., et al. (2006). European evidence based consensus on the diagnosis and management of Crohn's disease: definitions and diagnosis. *Gut* 55(Suppl 1), i1-i15. doi: 10.1136/gut.2005.081950a.

Strober, W. (2011). Adherent-invasive E. coli in Crohn disease: bacterial "agent provocateur". *The Journal of Clinical Investigation* 121(3), 841-844. doi: 10.1172/JCI46333.

Strober, W., Fuss, I., and Mannon, P. (2007). The fundamental basis of inflammatory bowel disease. *J Clin Invest* 117(3), 514-521. doi: 10.1172/jci30587.

Su, J.W., Ma, J.J., and Zhang, H.J. (2015). Use of antibiotics in patients with Crohn's disease: a systematic review and meta-analysis. *J Dig Dis* 16(Polito et al.), 58-66. doi: 10.1111/1751-2980.12216.

Suenaert, P., Bulteel, V., Lemmens, L., Noman, M., Geypens, B., Van Assche, G., et al. (2002). Anti-tumor necrosis factor treatment restores the gut barrier in Crohn's disease. *The American journal of gastroenterology* 97(8), 2000-2004.

Swidsinski, A., Ladhoff, A., Pernthaler, A., Swidsinski, S., Loening–Baucke, V., Ortner, M., et al. (2002). Mucosal flora in inflammatory bowel disease. *Gastroenterology* 122(Goodhand et al.), 44-54.

Sýkora, J., Pomahačová, R., Kreslová, M., Cvalínová, D., Štych, P., and Schwarz, J. (2018). Current global trends in the incidence of pediatric-onset inflammatory bowel disease. *World journal of gastroenterology* 24(25)**,** 2741.

Tabaqchali, S., Donoghue, D.P., and Bettelheim, K.A. (1978). Escherichia coli antibodies in patients with inflammatory bowel disease. *Gut* 19(Polito et al.)**,** 108. doi: 10.1136/gut.19.2.108.

Takahashi, K., Nishida, A., Fujimoto, T., Fujii, M., Shioya, M., Imaeda, H., et al. (2016). Reduced Abundance of Butyrate-Producing Bacteria Species in the Fecal Microbial Community in Crohn's Disease. *Digestion* 93(Goodhand et al.)**,** 59-65. doi: 10.1159/000441768.

Takesue, Y., Ohge, H., Uemura, K., Imamura, Y., Murakami, Y., Yokoyama, T., et al. (2002a). Bacterial translocation in patients with Crohn's disease undergoing surgery. *Dis Colon Rectum* 45(12)**,** 1665-1671. doi: 10.1097/01.Dcr.0000034515.66401.8b.

Takesue, Y., Ohge, H., Uemura, K., Imamura, Y., Murakami, Y., Yokoyama, T., et al. (2002b). Bacterial translocation in patients with Crohn's disease undergoing surgery. *Diseases of the colon & rectum* 45(12)**,** 1665-1671.

Tang, Q., Jin, G., Wang, G., Liu, T., Liu, X., Wang, B., et al. (2020). Current Sampling Methods for Gut Microbiota: A Call for More Precise Devices. *Frontiers in Cellular and Infection Microbiology* 10(151). doi: 10.3389/fcimb.2020.00151.

Thia, K.T., Loftus Jr, E.V., Sandborn, W.J., and Yang, S.-K. (2008). An update on the epidemiology of inflammatory bowel disease in Asia. *American Journal of Gastroenterology* 103(12)**,** 3167-3182.

Thompson, N.P., Driscoll, R., Pounder, R.E., and Wakefield, A.J. (1996). Genetics versus environment in inflammatory bowel disease: results of a British twin study. *BMJ : British Medical Journal* 312(7023)**,** 95-96.

Thursby, E., and Juge, N. (2017). Introduction to the human gut microbiota. *Biochemical Journal* 474(11)**,** 1823-1836. doi: 10.1042/BCJ20160510.

Townsend, C.M., Parker, C.E., MacDonald, J.K., Nguyen, T.M., Jairath, V., Feagan, B.G., et al. (2019). Antibiotics for induction and maintenance of remission in Crohn's disease. *Cochrane Database Syst Rev* 2(Polito et al.)**,** Cd012730. doi: 10.1002/14651858.CD012730.pub2.

Travassos, L.H., Carneiro, L.A., Ramjeet, M., Hussey, S., Kim, Y.-G., Magalhães, J.G., et al. (2010). Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nature immunology* 11(Goodhand et al.)**,** 55-62.

Tschurtschenthaler, M., Adolph, T.E., Ashcroft, J.W., Niederreiter, L., Bharti, R., Saveljeva, S., et al. (2017). Defective ATG16L1-mediated removal of IRE1α drives Crohn's disease-like ileitis. *J Exp Med* 214(Polito et al.)**,** 401-422. doi: 10.1084/jem.20160791.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007). The human microbiome project. *Nature* 449(7164)**,** 804-810.

Tysk, C., Lindberg, E., Järnerot, G., and Flodérus-Myrhed, B. (1988). Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 29(7)**,** 990-996. doi: 10.1136/gut.29.7.990.

Uhlig, H.H., Schwerd, T., Koletzko, S., Shah, N., Kammermeier, J., Elkadri, A., et al. (2014). The Diagnostic Approach to Monogenic Very Early Onset Inflammatory Bowel Disease. *Gastroenterology* 147(5)**,** 990-1007.e1003. doi: https://doi.org/10.1053/j.gastro.2014.07.023.

van der Heide, F., Wassenaar, M., van der Linde, K., Spoelstra, P., Kleibeuker, J.H., and Dijkstra, G. (2011). Effects of active and passive smoking on Crohn's disease and

ulcerative colitis in a cohort from a regional hospital. *Eur J Gastroenterol Hepatol* 23(3), 255-261. doi: 10.1097/MEG.0b013e3283435233.

Van Dullemen, H.M., van Deventer, S.J., Hommes, D.W., Bijl, H.A., Jansen, J., Tytgat, G.N., et al. (1995). Treatment of Crohn's disease with anti-tumor necrosis factor chimeric monoclonal antibody (cA2). *Gastroenterology* 109(Goodhand et al.), 129-135.

Van Kruiningen, H.J., Joossens, M., Vermeire, S., Joossens, S., Debeugny, S., Gower-Rousseau, C., et al. (2005). Environmental factors in familial Crohn's disease in Belgium. *Inflamm Bowel Dis* 11(4), 360-365. doi: 10.1097/01.mib.0000158536.31557.90.

Van Limbergen, J., Russell, R.K., Drummond, H.E., Aldhous, M.C., Round, N.K., Nimmo, E.R., et al. (2008). Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease. *Gastroenterology* 135(4), 1114-1122. doi: 10.1053/j.gastro.2008.06.081.

Van Limbergen, J., Wilson, D.C., and Satsangi, J. (2009). The genetics of Crohn's disease. *Annual review of genomics and human genetics* 10, 89-116.

Vantrappen, G., and Rutgeerts, P. (1990). Recurrence of Crohn's lesions in the neoterminal ileum after ileal resection and ileocolonic anastomosis. *Verh K Acad Geneeskd Belg* 52(5), 373-382; discussion 383-375.

Vavricka, S.R., and Rogler, G. (2010). Fistula treatment: The unresolved challenge. *Dig Dis* 28(3), 556-564. doi: 10.1159/000320416.

Vazeille, E., Buisson, A., Bringer, M.A., Goutte, M., Ouchchane, L., Hugot, J.P., et al. (2015). Monocyte-derived macrophages from Crohn's disease patients are impaired in the ability to control intracellular adherent-invasive Escherichia coli and exhibit disordered cytokine secretion profile. *J Crohns Colitis* 9(5), 410-420. doi: 10.1093/ecco-jcc/jjv053.

Vazeille, E., Chassaing, B., Buisson, A., Dubois, A., De Vallee, A., Billard, E., et al. (2016). GipA factor supports colonization of Peyer's Patches by Crohn's disease-associated Escherichia coli. *Inflammatory bowel diseases* 22(Goodhand et al.), 68-81.

Veltkamp, C., Tonkonogy, S.L., De Jong, Y.P., Albright, C., Grenther, W.B., Balish, E., et al. (2001). Continuous stimulation by normal luminal bacteria is essential for the development and perpetuation of colitis in Tg(epsilon26) mice. *Gastroenterology* 120(4), 900-913.

Vester-Andersen, M., Mirsepasi-Lauridsen, H., Prosberg, M., Mortensen, C., Träger, C., Skovsen, K., et al. (2019). Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. *Scientific reports* 9(Goodhand et al.), 1-10.

Voedisch, S., Koenecke, C., David, S., Herbrand, H., Förster, R., Rhen, M., et al. (2009). Mesenteric lymph nodes confine dendritic cell-mediated dissemination of Salmonella enterica serovar Typhimurium and limit systemic disease in mice. *Infection and immunity* 77(8), 3170-3180.

von Roon, A.C., Karamountzos, L., Purkayastha, S., Reese, G.E., Darzi, A.W., Teare, J.P., et al. (2007). Diagnostic precision of fecal calprotectin for inflammatory bowel disease and colorectal malignancy. *Am J Gastroenterol* 102(4), 803-813. doi: 10.1111/j.1572-0241.2007.01126.x.

Wagnerova, A., and Gardlik, R. (2013). In vivo reprogramming in inflammatory bowel disease. *Gene Therapy* 20(12), 1111-1118. doi: 10.1038/gt.2013.43.

Walker, J.R., Ediger, J.P., Graff, L.A., Greenfeld, J.M., Clara, I., Lix, L., et al. (2008). The Manitoba IBD cohort study: a population-based study of the prevalence of lifetime and 12-month anxiety and mood disorders. *American Journal of Gastroenterology* 103(8), 1989-1997.

Wang, F., Graham, W.V., Wang, Y., Witkowski, E.D., Schwarz, B.T., and Turner, J.R. (2005). Interferon-γ and tumor necrosis factor-α synergize to induce intestinal epithelial barrier dysfunction by up-regulating myosin light chain kinase expression. *The American journal of pathology* 166(Polito et al.), 409-419.

Wildenberg, M.E., Koelink, P.J., Diederen, K., te Velde, A.A., Wolfkamp, S.C.S., Nuij, V.J., et al. (2017). The ATG16L1 risk allele associated with Crohn's disease results in a Rac1-dependent defect in dendritic cell migration that is corrected by thiopurines. *Mucosal Immunology* 10(Polito et al.), 352-360. doi: 10.1038/mi.2016.65.

Willing, B., Halfvarson, J., Dicksved, J., Rosenquist, M., Jarnerot, G., Engstrand, L., et al. (2009). Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis* 15(5), 653-660. doi: 10.1002/ibd.20783.

Wilson, J., Hair, C., Knight, R., Catto-Smith, A., Bell, S., Kamm, M., et al. (2010). High incidence of inflammatory bowel disease in Australia: a prospective population-based Australian incidence study. *Inflamm Bowel Dis* 16(9), 1550-1556. doi: 10.1002/ibd.21209.

Wolf, J. (1990). Ciprofloxacin may be useful in Crohn's disease. *Gastroenterology* 98(Polito et al.), A212.

Wright, M., Chandrakasan, S., Okou, D.T., Yin, H., Jurickova, I., Denson, L.A., et al. (2019). Early Onset Granulomatous Colitis Associated with a Mutation in NCF4 Resolved with Hematopoietic Stem Cell Transplantation. *The Journal of Pediatrics* 210, 220-225. doi: 10.1016/j.jpeds.2019.03.042.

Wu, X.-W., Ji, H.-Z., Yang, M.-F., Wu, L., and Wang, F.-Y. (2015). Helicobacter pylori infection and inflammatory bowel disease in Asians: a meta-analysis. *World journal of gastroenterology* 21(15), 4750-4756. doi: 10.3748/wjg.v21.i15.4750.

Xavier, R.J., Huett, A., and Rioux, J.D. (2008). Autophagy as an important process in gut homeostasis and Crohn's disease pathogenesis. *Gut* 57(6), 717-720.

Xiang, Z., Chen, Y.-P., Ye, Y.-F., Ma, K.-F., Chen, S.-H., Zheng, L., et al. (2013). Helicobacter pylori and Crohn's disease: a retrospective single-center study from China. *World journal of gastroenterology* 19(28), 4576-4581. doi: 10.3748/wjg.v19.i28.4576.

Xu, W.-D., Xie, Q.-B., Zhao, Y., and Liu, Y. (2015). Association of Interleukin-23 receptor gene polymorphisms with susceptibility to Crohn's disease: A meta-analysis. *Scientific reports* 5.

Xu, X.-R., Liu, C.-Q., Feng, B.-S., and Liu, Z.-J. (2014). Dysregulation of mucosal immune response in pathogenesis of inflammatory bowel disease. *World journal of gastroenterology* 20(12), 3255-3264. doi: 10.3748/wjg.v20.i12.3255.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486(7402), 222-227. doi: 10.1038/nature11053.

Yue, B., Luo, X., Yu, Z., Mani, S., Wang, Z., and Dou, W. (2019). Inflammatory Bowel Disease: A Potential Result from the Collusion between Gut Microbiota and Mucosal Immune System. *Microorganisms* 7(10), 440. doi: 10.3390/microorganisms7100440.

Zarei-Kordshouli, F., Geramizadeh, B., and Khodakaram-Tafti, A. (2019). Prevalence of Mycobacterium avium subspecies paratuberculosis IS 900 DNA in biopsy tissues from patients with Crohn's disease: histopathological and molecular comparison with Johne's disease in Fars province of Iran. *BMC Infectious Diseases* 19(Goodhand et al.), 23. doi: 10.1186/s12879-018-3619-2.

Zhang, C., Yang, M., and Ericsson, A.C. (2021). Function of Macrophages in Disease: Current Understanding on Molecular Mechanisms. *Frontiers in Immunology* 12(635). doi: 10.3389/fimmu.2021.620510.

Zhang, L., Man, S.M., Day, A.S., Leach, S.T., Lemberg, D.A., Dutt, S., et al. (2009). Detection and isolation of Campylobacter species other than C. jejuni from children with Crohn's disease. *Journal of clinical microbiology* 47(Polito et al.)**,** 453-455.

Zheng, D., Liwinski, T., and Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. *Cell Research* 30(6)**,** 492-506. doi: 10.1038/s41422-020-0332-7.

Zhou, R., Yazdi, A.S., Menu, P., and Tschopp, J. (2011). A role for mitochondria in NLRP3 inflammasome activation. *Nature* 469(7329)**,** 221-225.

Zhu, Y., Jiang, H., Chen, Z., Lu, B., Li, J., and Shen, X. (2019). Genetic association between IL23R rs11209026 and rs10889677 polymorphisms and risk of Crohn's disease and ulcerative colitis: evidence from 41 studies. *Inflamm Res*. doi: 10.1007/s00011-019-01296-y.

Zoetendal, E.G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A.D., and de Vos, W.M. (2002). Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl. Environ. Microbiol.* 68(7)**,** 3401-3407.

Chapter 2: Microbiome analysis of different disease stages and tissue types from Crohn's disease patients comparing with healthy controls

2.1 Abstract

The pathogenesis of Crohn's disease (CD) is still not precisely known. However, an imbalance in the intestinal microbiome (dysbiosis), i.e., reduced alpha diversity and changes in the relative abundance of taxa has been implicated in CD pathogenesis. It has been consistently reported that CD patients have a relatively increased number of Enterobacteriaceae, particularly *Escherichia coli (E. coli)* species and a decreased abundance of Firmicutes and the Bacteroidetes phyla compared with healthy controls. A more distinct difference was detected for mucosa-associated microbiota than faecal samples. This study aimed to characterise and compare microbial communities of samples taken at different stages of CD, during both active and inactive disease, and control samples (N). Samples include aphthous ulcers, the initial CD lesion, mucosa from different gut locations, including small and large bowel, obtained during colonoscopy (mucosal biopsies) or surgery. Various analyses of the different stages of disease and tissue types were carried out.

A total of 72 CD tissue samples were obtained at different stages of disease; 'early-stage' tissues, which included aphthous ulcers; 'mid-stage tissues', which included biopsies obtained during colonoscopy; and 'late-stage' tissues, obtained from surgical resections. CD tissue samples consisted of aphthous ulcers (n=12), biopsy tissues from inactive small bowel (n=18), active small bowel (n=13), inactive colon (n=22), and active colon (n=7). Biopsy samples from non-IBD controls (n=16) were collected from Peyer's patches (n=5) and normal mucosal biopsies (n=11). Biopsy tissues from different anatomical locations of gut (terminal ileum, caecum, ascending colon and rectum) were obtained from active and inactive stages from CD patients. From late stage (bowel resected samples) of CD patients, involved mucosa and unaffected mucosal samples were collected. We used universal bacterial primers targeting the V3-V4 region of the 16S rRNA gene to amplify all samples. Samples at different stages of disease and tissue type were compared between CD and controls.

Comparing CD with non-IBD controls, there were significant reductions in OTU richness and species diversity in CD patients compared with that of controls (observed OTUs, $p = 9.9e\text{-}05$; Fisher diversity index, $p = 0.0001$ and Shannon diversity index, $p = 0.004$). Species richness (alpha diversity), calculated using the Shannon index, was significantly reduced in CD mucosal samples, including aphthous ulcers ($p = 0.003$), inactive small bowel ($p = 0.007$), active small bowel ($p = 0.0003$), inactive colon ($P = 0.003$), and active colon ($p = 0.002$) compared with that of control mucosal biopsies. Analyses restricted to specific gut locations revealed a significant reduction in alpha diversity in CD mucosa obtained from small bowel ($p = 0.00003$) and colon ($p = 0.0006$) compared to mucosa from the same location in control subjects. Based on Shannon diversity index, alpha diversity was significantly reduced in active small bowel CD (p

= 0.01) and active colon CD (p = 0.03) compared with controls. Various measures of beta diversity revealed that CD and normal control samples varied significantly at different stages of disease and for different gut locations. Overall, patients with CD had a significant increase in Proteobacteria (p = 0.01) and decrease in Bacteroidetes (p = 0.03) compared to controls. Several genera were significantly increased in CD compared to control mucosa, including *Escherichia_Shigella* (p = 0.01), *Flavonifractor* (p = 0.01), *Klebsiella* (p = 0.01), and *Ralstonia* (p = 0.01). A significant reduction in several genera was observed in CD tissues when compared to control tissues, including: *Faecalibacterium* (p = 0.02), *Ruminococcus_torques_group* (p = 0.006). The abundance of *Escherichia-shigella* was increased significantly in active colonic (p = 0.005), small bowel (p = 0.009) and inactive colonic (p = 0.02) CD mucosa compared to the normal mucosa from the same region in controls. *Faecalibacterium* was reduced significantly in active small bowel (P = 0.02), active colonic (P = 0.03), and inactive colonic mucosa (p = 0.01) in CD compared to the normal biopsies in controls. In summary, we demonstrate that multiple phyla and genera differ between CD and normal mucosa from control subjects, depending on stage of disease/tissue type, and gut location. These results suggest that dysbiosis of the gut in CD patients related to current disease activity and different stages of the disease.

Consistent with previous studies, we found that CD is associated with a microbial signature characterised by reduced alpha diversity and greater dysbiosis compared to controls. Further studies are required to determine if the microbiome impacts inflammatory development in CD.

2.2 Introduction

The exact cause of Crohn's disease is still unknown. However, it is a complex disease involving genetic, immunological, environmental, and microbial factors. Genome-wide association studies (GWASs) have identified more than 200 IBD-associated loci (Liu et al., 2015). However, common variants in IBD could only explain a small fraction of disease occurrence 13.1% and 8.2% in CD and UC, respectively (Jostins et al., 2012).

GWAS analyses have highlighted most of the loci in several vital pathways underlying IBD susceptibility, many associated with innate immunity (TLR4, NOD1) (Toiyama et al., 2006; Zhernakova et al., 2008; Ellinghaus et al., 2016), bacterial recognition (NOD2)(Boukercha et al., 2015; Sidiq et al., 2016; Kennedy et al., 2018), autophagy (ATG16L1 and IRGM) (Iida et al., 2017; Lassen and Xavier, 2017; Simon et al., 2018; Ajayi et al., 2019; Kim et al., 2019; Nedelsky et al., 2019), epithelial barrier function (PTPN2) (Barrett et al., 2008; Spalinger et al., 2016; Yilmaz et al., 2018), lymphocyte activation (NKX2-3)(Meggyesi et al., 2010; John et al., 2011; Jostins et al., 2012) and cytokine signalling (IL23R) (Duerr et al., 2006; Abraham and Cho, 2009; Xu et al., 2015; Sivanesan et al., 2016). Several variants in CD-associated loci are found to be strongly associated with CD pathogenesis. Those are involved in bacterial recognition (Xu et al., 2015). These polymorphisms in susceptibility loci indicate that patients with CD may have an abnormal and defective ongoing innate immune response to microbes that may implicate them in CD pathogenesis.

In addition to genetic factors, gut microbiome also plays a significant role in CD. In healthy individuals, the gastrointestinal microbiota is predominated by four bacterial phyla: Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria (Eckburg et al., 2005; Human Microbiome Project, 2012). Taxonomic alteration or imbalance in a microbial community (defined as dysbiosis) has been implicated in the development

of numerous diseases, including CD (D. Gevers, S. Kugathasan, L. A. Denson, Y. Vazquez-Baeza, et al., 2014; Morgan et al., 2012). Studies have demonstrated that microbial diversity and richness (α-diversity) was significantly reduced in patients with CD (Ma et al., 2018; Galazzo et al., 2019; Vester-Andersen et al., 2019b). In addition, approximately one-third of dominant species in the representative CD microbiome belonged to undefined phylogenetic groups not usually prevailing in healthy individuals (Seksik et al., 2003). Although gut microbiome consists of viruses, bacteria, fungi, archaea and protozoa, bacteria contribute the largest impact (Tlaskalová-Hogenová et al., 2011; Hoffmann et al., 2013). The most distinct changes in gut microbiota in CD include decreased representation of several taxa in Firmicutes with an increased abundance of Proteobacteria (Loh and Blaut, 2012; Matsuoka and Kanai, 2015; Vester-Andersen et al., 2019a). Among the Firmicutes, a reduced abundance of *Faecalibacterium prausnitzii* in CD patients is commonly reported (Willing et al., 2009; Fujimoto et al., 2013; Lopez-Siles et al., 2014a; Wang et al., 2014; Prosberg et al., 2016; Zhao et al., 2021) and is associated with a higher risk of postoperative recurrence in patients with CD (Wright et al., 2017). It is assumed that there is a correlation between the reduced abundance of *F. prausnitzii* and the altered immune tolerance to the gut microbiota (Sokol et al., 2008b). Along with the butyrate production that reduces inflammation in the gut mucosa, other anti-inflammatory properties have been associated with *F. prausnitzii* (Lopez-Siles et al., 2014a). In addition to the reduced abundance of beneficial microbiota, a greater abundance of putative pathogens has also been demonstrated in CD (Carrière et al., 2014) (Table 2.1).

Table 2.1. List of bacteria that have been shown to be decreased or increased in CD

Bacteria	Comment	References
Firmicutes (*F. prausnitzii*)	Decrease	Sokol et al., 2009; Joossens et al., 2011a; Fujimoto et al., 2013; Wang et al., 2014
Bifidobacterium	Decrease	Seksik et al., 2003; Joossens et al., 2011a; Sha et al., 2013; Hall et al., 2014
Clostridium clusters (IV)	Decrease	Fava and Danese, 2011; Hedin et al., 2014
Roseburia	Decrease	Willing et al., 2010
Enterobacteriaceae (*E. coli*)	Increase	Seksik et al., 2003; Martinez-Medina et al., 2009; Willing et al., 2010; Alam et al., 2020
Ruminococcus gnavus	Increase	Darfeuille-Michaud et al., 2004; Willing et al., 2010; Joossens et al., 2011b
Pseudomonas	Increase	Scaldaferri et al., 2013
Bacteroidetes	Increase	Walker et al., 2011

Significant taxonomic shifts have been shown in CD (Gevers et al., 2014); however, explanation of these changes in different reports is challenging because of the diversity of study populations, differences in sample types (e.g., stool versus mucosa), variances in methodology and study design. Many studies investigated the gut microbiota of CD using faecal samples since those can be easily obtained (He et al., 2017; Pascal et al., 2017; Wang et al., 2018). However, the microbial community and composition of faecal microbiota are significantly different from the mucosal microbiota. It has been hypothesised that microbiota associated with intestinal epithelium is more relevant to the pathogenesis of CD (Atarashi et al., 2015; Ringel et al., 2015).

In a previous study done by Gevers and colleagues, various limitations of previous studies were addressed for a better understanding of the contribution of the microbiota in CD pathogenesis (Gevers et al., 2014). In that study, they collected samples from multiple mucosal tissues before treatment started and included stool samples. They

then added more cohorts with the new-onset treatment-naive paediatric cohort to obtain a total of 1,742 samples from either new-onset paediatric patients or adult patients with established disease. Interestingly, in that study based on the new-onset paediatric CD patients, dysbiosis was only observed in mucosal samples and not in faecal samples, which indicates that dysbiotic features in the mucosa-associated microbiota play a more significant role in the CD pathogenesis than dysbiosis in the luminal microbiota (Gevers et al., 2014). This study also suggests that analysing mucosal biopsies from different disease stages and tissue types rather than faecal samples from CD patients could be a better option for exploring the association of microbiota in CD.

The initial lesion in CD is the aphthous ulcer (AU) that overlies the Peyer's patches (PP) in the small bowel, or lymphoid follicles in the large bowel. They are the first macroscopic sign of inflammation during a disease flare and progress from this 'early' lesion to deeper transverse linear ulcers, which may require surgical resection ('late-stage') if there is a risk of perforation, fistulas, adhesions, stricturing, or other complications.

The aim of this study was to investigate the change in composition and diversity of microbial groups in CD and controls in different disease stages (active and inactive) and tissue type (aphthous ulcers, small bowel mucosa, large bowel mucosa and unaffected mucosa). We wanted to demonstrate which microbial groups differ during disease stages and in different tissue types in the CD when compared with controls.

2.3 Methods

2.3.1 Selection of patients

Human ethics protocols (ACT Health, ETH.5.07.464; Australian National University: 20212/596). CD patients were diagnosed by endoscopic and pathological findings and recruited for this study in Canberra Hospital at Canberra, Australia. All participants in this study were divided into different groups according to their disease status (CD and HC), disease stages (early stage - aphthous ulcers, active and inactive) and tissue types (initial lesion of CD - aphthous ulcers (AU), small bowel mucosa and colon mucosa from patients with CD, and Peyer's patches (PP) and normal biopsies from controls). Biopsy tissues were collected from different sites of the GI tract such as terminal ileum, caecum, ascending colon and rectum in CD patients. All biopsy samples were immediately put in RNAlater® and stored at - 80°C before processing. The study was approved by the Ethics Committee of the Canberra Hospital, and all participants provided written informed consent before enrolment (O'Brien et al., 2014).

A total of 88 samples (either biopsy or mucosa from bowel resected samples) were obtained; 72 from patients with CD, and 16 from non-IBD controls. These participants were grouped based on disease stages (active CD and inactive CD) and different types of tissues (i.e., initial lesion of CD - aphthous ulcers, small bowel mucosa and colon mucosa from patients with CD, and Peyer's patches (PP) and normal biopsies from controls as shown in Table 2.2). In some patients, two or more samples were taken at different sites of the intestine based on disease activity. For example, three samples were taken prom patient 69; an active lesion in the terminal ileum, and two inactive sites (colon and rectum). The controls correspond to non-IBD patients who came to Canberra hospital for colonoscopy.

Table 2.2. Patient details: disease status, disease stage, and tissue type.

Serial	Sample ID	Tissue-types	Disease stages	Disease status
1	HDT-1	AU TI-CD	AU-early-active-CD	CD
2	LSM-2	AU TI-CD	AU-early-active-CD	CD
3	SK-AU-TI	AU TI-CD	AU- early-active-CD	CD
4	DT-AU-TI	AU TI-CD	AU- early-active-CD	CD
5	JM-2	AU TI-CD	AU- early-active-CD	CD
6	SMP-1	AU TI-CD	AU- early-active-CD	CD
7	AH-2	AU TI-CD	AU- early-active-CD	CD
8	DRT-AU-TI	AU TI-CD	AU- early-active-CD	CD
9	JF-2	AU TI-CD	AU- early-active-CD	CD
10	AW-AU-LB	AU LB-CD	AU- early-active-CD	CD
11	TB-AU-LB	AU LB-CD	AU- early-active-CD	CD
12	JT-AU-LB	AU LB-CD	AU- early-active-CD	CD
13	108-1	inactive TI_CD	Inactive-SB-CD	CD
14	219-1	inactive TI_CD	Inactive-SB-CD	CD
15	103-1	inactive TI_CD	Inactive-SB-CD	CD
16	109-1	inactive TI_CD	Inactive-SB-CD	CD
17	66-1	inactive TI_CD	Inactive-SB-CD	CD
18	110-1	inactive TI_CD	Inactive-SB-CD	CD
19	90-1	inactive TI_CD	Inactive-SB-CD	CD
20	111-1	inactive TI_CD	Inactive-SB-CD	CD
21	76-1	inactive TI-CD	Inactive-SB-CD	CD
22	86-1	inactive TI-CD	Inactive-SB-CD	CD
23	77-1	inactive TI-CD	Inactive-SB-CD	CD
24	95-1	inactive TI-CD	Inactive-SB-CD	CD
25	79-1	inactive TI-CD	Inactive-SB-CD	CD
26	67-1	inactive TI-CD	Inactive-SB-CD	CD
27	134-1	Normal SB_CD	Inactive-SB-CD	CD
28	AW-NM-SB	NM SB/TI-CD	Inactive-SB-CD	CD
29	72-1	normal TI-CD	Inactive-SB-CD	CD
30	84-1	inactive TI-CD	Inactive-SB-CD	CD
31	65-1	active TI-CD	Active-SB-CD	CD
32	69-1	active TI-CD	Active-SB-CD	CD
33	85-1	active TI-CD	Active-SB-CD	CD
34	93-1	active TI-CD	Active-SB-CD	CD
35	107-1	active TI-CD	Active-SB-CD	CD
36	129-1	active TI_CD	Active-SB-CD	CD

37	106-1	active TI-CD	Active-SB-CD	CD
38	SK-NM-LB	NM LB-CD	Inactive-colon-CD	CD
39	GON-NM	NM LB-CD	Inactive-colon-CD	CD
40	121-2	Inactive caecum-CD	Inactive-colon-CD	CD
41	77-2	Inactive caecum-CD	Inactive-colon-CD	CD
42	86-2	Inactive caecum-CD	Inactive-colon-CD	CD
43	93-2	Inactive caecum-CD	Inactive-colon-CD	CD
44	76-2	Inactive caecum-CD	Inactive-colon-CD	CD
45	85-2	Inactive caecum-CD	Inactive-colon-CD	CD
46	84-2	Inactive caecum-CD	Inactive-colon-CD	CD
47	67-2	Inactive AC-CD	Inactive-colon-CD	CD
48	69-2	Inactive AC-CD	Inactive-colon-CD	CD
49	79-2	Inactive caecum-CD	Inactive-colon-CD	CD
50	67-3	Inactive rectum-CD	Inactive-colon-CD	CD
51	69-3	Normal rectum-CD	Inactive-colon-CD	CD
52	79-3	Inactive rectum-CD	Inactive-colon-CD	CD
53	86-3	Inactive rectum-CD	Inactive-colon-CD	CD
54	76-3	Inactive rectum-CD	Inactive-colon-CD	CD
55	85-3	Inactive rectum-CD	Inactive-colon-CD	CD
56	84-3	Inactive rectum-CD	Inactive-colon-CD	CD
57	95-2	Active caecum-CD	Active-colon-CD	CD
58	72-2	Active caecum-CD	Active-colon-CD	CD
59	95-3	Active caecum-CD	Active-colon-CD	CD
60	121-3	Active caecum-CD	Active-colon-CD	CD
61	77-3	Active caecum-CD	Active-colon-CD	CD
62	93-3	Active caecum-CD	Active-colon-CD	CD
63	72-3	Active caecum-CD	Active-colon-CD	CD
64	CD44IM	Late-involved mucosa-CD	Active-SB-CD	CD
65	CD29IM	Late-involved mucosa-CD	Active-SB-CD	CD
66	CD50IM	Late-involved mucosa-CD	Active-SB-CD	CD
67	CD62IM	Late-involved mucosa-CD	Active-SB-CD	CD
68	CD1IM	Late-involved mucosa-CD	Active-SB-CD	CD
69	CD53IM	Late-involved mucosa-CD	Active-SB-CD	CD
70	CD50NM	Late-unaffected mucosa-CD	Inactive-colon-CD	CD
71	CD62NM	Late-unaffected mucosa-CD	Inactive-colon-CD	CD
72	CD1NM	Late-unaffected mucosa-CD	Inactive-colon-CD	CD
73	JST1	PP	Normal biopsies	Controls
74	ME-1	PP	Normal biopsies	Controls
75	LJ-1	PP	Normal biopsies	Controls

76	RDV-1	PP	Normal biopsies	Controls
77	TLP-1	PP	Normal biopsies	Controls
78	LJ-2	NM SB/TI-N	Normal biopsies	Controls
79	NB11-2	NM SB/TI-N	Normal biopsies	Controls
80	NB17-2	NM SB/TI-N	Normal biopsies	Controls
81	RDV-2	NM SB/TI-N	Normal biopsies	Controls
82	JST-2	NM SB/TI-N	Normal biopsies	Controls
83	NB12-2	NM SB/TI-N	Normal biopsies	Controls
84	NB18-2	NM SB/TI-N	Normal biopsies	Controls
85	ME-2	NM SB/TI-N	Normal biopsies	Controls
86	NB15-2	NM SB/TI-N	Normal biopsies	Controls
87	NB23-2	NM SB/TI-N	Normal biopsies	Controls
88	NB13-2	NM LB-N	Normal biopsies	Controls

CD- Crohn's disease, TI- Terminal Ileum, NM- Normal mucosa, SB-Small bowel, PP- Peyer's patches, AU- aphthous ulcers, LB- Large bowel. AC – ascending colon

2.3.2 16S amplification and sequencing

Total DNA was extracted from tissue samples using Qiagen DNeasy kits (Qiagen, cat #: 69504), which included a lysis buffer and bead-beating step. DNA concentration was quantified using a Qubit fluorometer and diluted based on the total DNA concentration to 5 ng/uL.

DNA was amplified using primers targeting the V3-V4 regions of the 16S rRNA gene (Klindworth et al., 2013):

(F)

5′TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG 3′

(R)

5′GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAA TCC 3′.

A second PCR was performed to attach adapters and barcodes to the target amplicon using an Illumina indexing kit (V2 set C- Index, Illumina San Diego, California, United

States). PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter). PCR products were diluted and quantified on a Bioanalyzer 1000 DNA chip to verify the amplicons were predominantly ~630 bp. Samples were adjusted to a concentration of 4nM, and 5 ul of each sample was used combined in one tube to make a final sequencing library.

Samples were denatured with NaOH, diluted with hybridisation buffer, and heat denatured. The final sequencing library was sequenced using an Illumina MiSeq platform, using MiSeq v3 reagents in 300 bp paired-end format.

2.3.3 Pre-processing of 16S rRNA gene sequencing data

We used QIIME2-2019.10 (Bolyen et al., 2019) to perform microbiome bioinformatics. Paired-end sequenced data in fastQ.gz file format was imported into QIIME2 to combine the forward and reverse reads. Primers and adapters were removed, samples demultiplexed, and quality filtering performed using the q2-demux plugin. Sequences were denoised and filtered using the q2-DADA2 plugin (Callahan et al., 2016). We aligned all amplicon sequence variants (ASVs) using mafft (Katoh et al., 2002) (via q2-alignment) and used fasttree2 (Price et al., 2010) (via q2-phylogeny) to create a phylogeny. Taxonomy was assigned to ASVs using the q2-feature-classifier Silva_132 (Dec 13, 2017 version) (Quast et al., 2012). We extracted the V3/V4 region from the reference database to produce QIIME2 artefact 99_otus_16S. Taxonomy bar plots were used to compare the relative abundance of different bacterial phyla, genera and OTUs in different samples. For instance, the SILVA database uses the following taxonomic levels: Level 1 = Kingdom, Level 2 = Phylum, Level 3 = Class, Level 4 = Order, Level 5 = Family, Level 6 = Genus, Level 7 = Species/OTU.

Samples were normalised according to best practice using rarefaction, a normalisation procedure that uses sub-sampling without replacement, until all samples contain the same number of sequences. Rarefaction occurred in two steps: first, samples that were below the rarefaction depth were filtered out of the feature table. Then, all remaining samples were subsampled without replacement to get to the specified sequencing depth. The OTU table was rarefied to normalise for sampling effort (Weiss et al., 2015; Weiss et al., 2017) at 38739 sequences per sample (Weiss et al., 2015; Weiss et al., 2017). This cut off was chosen to get the best adjustment between sequences and samples per disease status. Finally, the OTU table and a phylogenetic tree were generated for microbial diversity analysis. To estimate the diversity of the microbial community of the sample, we calculated the within-sample (alpha) diversity by Wilcoxon rank test or t-test for two groups and Kruskal-Wallis test for multiple group comparisons. We determined the alpha-diversity using different metrics (observed OTUs, Chao1, ACE, Shannon, Simpson, and Fisher). Beta diversity was estimated using various measures, such as Bray-Curtis dissimilarity metrics, Jaccard distance metrics, Jensen-Shannon distance metrics, weighted Unifrac (Lozupone et al., 2007) and unweighted Unifrac (Lozupone and Knight, 2005). Distance matrices were visualised using principal coordinate analysis (PCoA) plots. Statistical differences ($p = < 0.05$) between two groups, according to the relative abundance of bacterial phyla present in the samples, was calculated using the Kruskal-Wallis test. Before beta diversity analyses were performed, samples were rarefied to normalise for sampling effort at 38739 sequences per sample (Weiss et al., 2015; Weiss et al., 2017). For visualisation and clustering of multivariate data, MicrobiomeAnalystR was applied (https://github.com/xia-lab/MicrobiomeAnalystR) (Chong et al., 2020). After processing the data, network plots were generated to investigate co-abundance

patterns of different bacterial genera in CD and controls. This helps us to identify the importance of the microbiome in both health and disease.

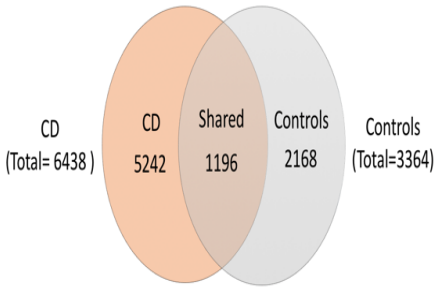
2.3.4 Statistical analysis

Statistical analyses were carried out in QIIME2 and R. To normalise the data, we considered an equal number of sequences from all groups based on the smallest library size among all samples. Normally distributed data were compared by t-test for paired or unpaired data. Alpha diversity analyses were performed for different stages of disease, between CD and controls, using the Kruskal-Wallis test for multiple comparisons. We performed analyses with the pairwise PERMANOVA (pseudo-F) to test for differences in microbial community composition. Using Pearson's correlation in the R language, correlations between overlapping genera were identified. We also used phyloseq in R for comparative analysis of different bacterial phyla and genera in different groups (CD vs controls, between different disease stages and tissue types). The Kruskal-Wallis test was used to compare the abundance of a bacterial genus among different groups (CD vs controls, and disease stages), that is, the difference in the abundance of bacteria in different groups of patients (Kruskal & Wallis, 1952). After Kruskal-Wallis analysis, a posthoc Dunn's test (Dunn, 1964) was performed with Bonferroni adjustment using the Dunn test code for R. A p-value of <0.05 was considered statistically significant. To identify bacterial genera that may act as a biomarker for CD (or controls), we used the classical univariate analysis based on the Kruskal-Wallis test and the linear discriminant analysis effect size (LEfSe) method. Both methods provided false discovery rate (FDR)-corrected p-values, where an FDR<0.05 is considered significant for both tests.

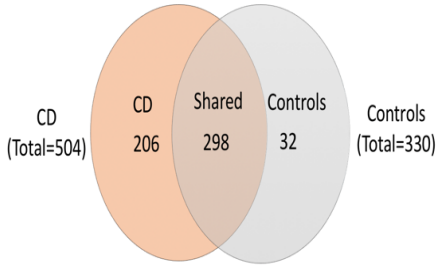
2.4 Results

We obtained 445039 reads per sample with a minimum of 38739 reads and a maximum of 1807708 reads. After filtering for low-quality reads, phiX and chimeric sequences, we detected a total of 14497 OTUs at the 99% sequence similarity level. On average, more than 70% of the input reads passed the quality filter. The minimum number of OTUs with ≥two was 3171, calculated from the filtered table (Supplemental Table 2.1). The total number of OTUs detected across all samples was 8862 (at taxonomic level 7 or L7, i.e., at species/ OTU level. After filtering out chloroplast and mitochondrial sequences, 8606 OTUs remained. In CD, total OTUs were 6438, and in controls, the total OTUs were 3364. CD and controls shared a total of 1196 OTUs (Figure 2.1A). At the genus level, the total number of genera were 536 (504 CD, 330 controls), of which 298 genera were shared. (Figure 2.1B).

A



B



A) OTUs (L7) overlaps between different groups (CD and controls). Taxonomy was assigned with a 99% similarity with the Silva 131 database.

B) Genera (L6) overlaps between different groups (CD and controls). Taxonomy was assigned with a 99% similarity with the Silva 131 database.

Figure 2.1. Venn diagram representing the total number of OTUs, and genera detected in A) CD and B) controls

2.4.1 Alpha rarefaction curves based on observed OTUs in disease status, disease activity and tissue types

Alpha rarefaction is a technique that allows the calculation of species richness for a given number of individual samples through the construction of rarefaction curves. We generated alpha rarefaction curves based on observed OTUs for CD and control tissues (Figure 2.2A), and different disease stages and tissue types compared with controls (Figure 2.2B). These rarefaction curves show that OTUs richness is lower in CD patients than controls (Figure 2.2A), and that OTU richness is much lower in active disease tissues in the small and large bowel of CD than the inactive stages of CD and the corresponding samples from controls (Figure 2.2B). OTU richness also differs in different sequencing depths.

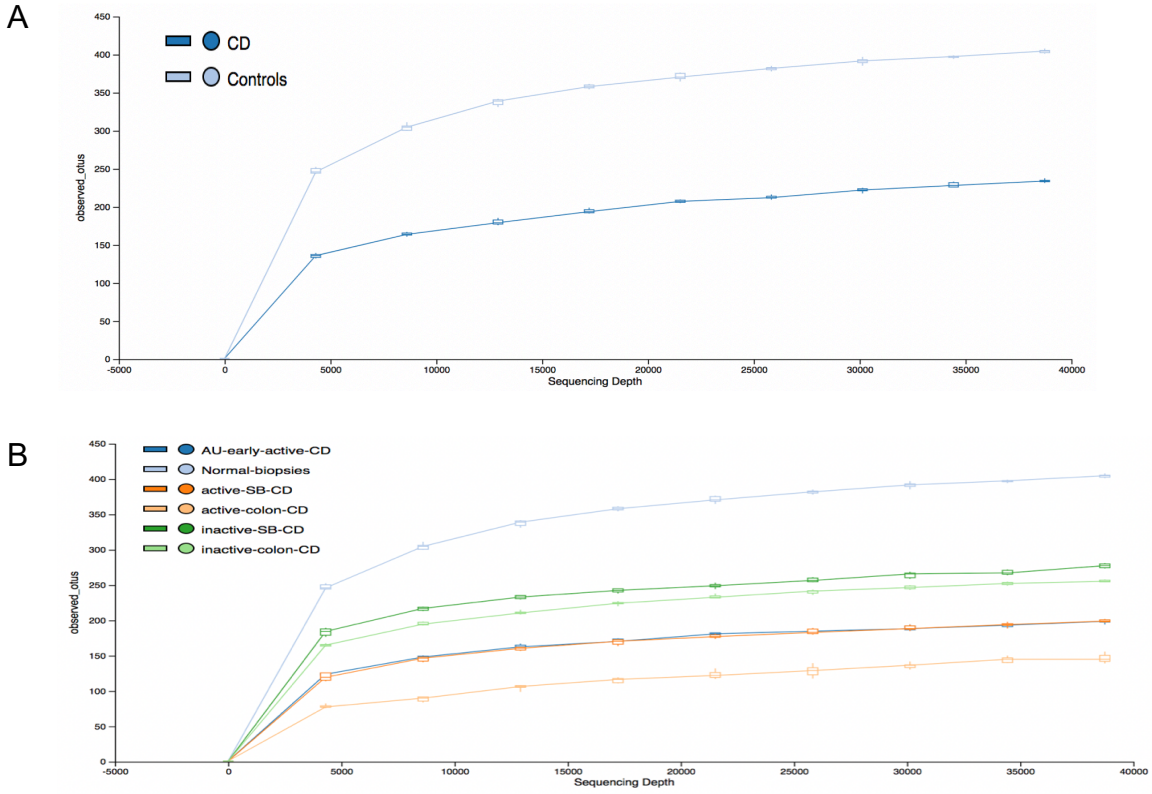


Figure 2.2. A) Alpha rarefaction curves showing the difference of richness in OTUs in gut mucosal samples between CD and controls**. B)** Alpha rarefaction curves showing the difference of richness in OTUs for active and inactive CD, and for different tissue types (AU, small bowel CD, colon CD) between CD and controls. *SB, small bowel; CD, Crohn's disease; and AU, aphthous ulcer.

2.4.2 Alpha diversity analysis to compare microbiota richness and diversity between Crohn's disease and controls

We used different metrics to characterise the alpha diversity (richness and diversity) of the microbiota of all CD and all control tissues (Figure 2.3). Microbial alpha diversity by richness was significantly lower in CD than controls for the observed species (OTUs) (observed, p = 9.9e-05), Chao1 richness estimator (Chao1, p = 0.0003) and ACE richness estimator (ACE, p = 0.0003) based on t-test. Additionally, there were significant differences in alpha diversity based on different diversity indices. Based on the Shannon Diversity Index, which relates both OTU richness and evenness, CD patients had a significantly lower bacterial richness and diversity in comparison to controls (Shannon, p = 0.004). Based on the Fisher diversity index that measures both species number and evenness of those species in a community, we observed a significant reduction in alpha diversity in CD compared with that of controls (Fisher, p = 0.0001). This indicates that species richness and diversity of microbiota was significantly reduced in CD. The summary of the alpha diversity analyses based on OTU richness (observed OTUs, Chao and Ace), and diversity index (Shannon, Fisher) for CD and controls is shown in Figure 2.3.

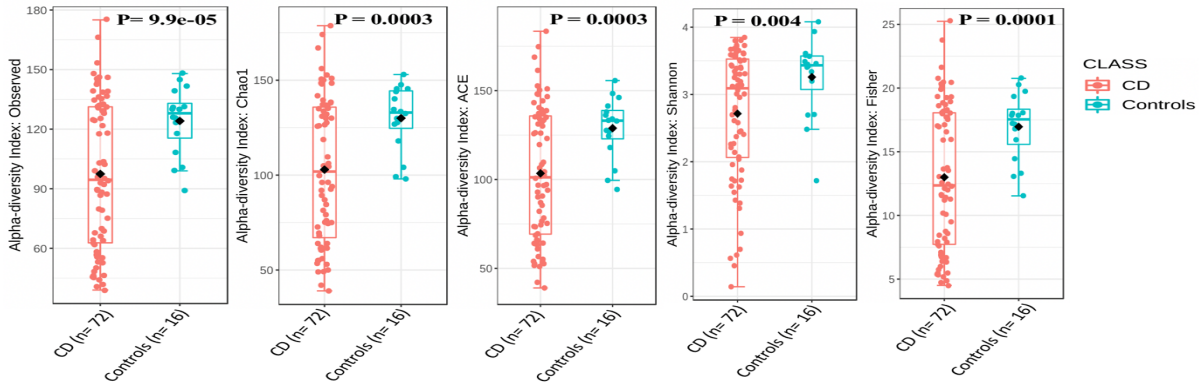


Figure 2.3 Alpha diversity boxplot representing the OTU richness (observed, Chao1 and Ace) and diversity index (Shannon, and fisher) for CD (N=72) and controls (n=16) (normal biopsies and Peyer's patches) 16S rRNA samples and the measurement of significant difference in richness and diversity between CD and controls based on t-test. CD patients had a significantly decreased bacterial richness and diversity compared to the control group, $p < 0.05$*

The alpha diversity analysis between CD and controls based on the Kruskal Wallis test is shown in the Supplemental Table 2.2-2.3, and Supplemental Figure 2.2-2.3).

2.4.3 Alpha diversity analyses based on disease activity and different tissue types in Crohn's disease and controls

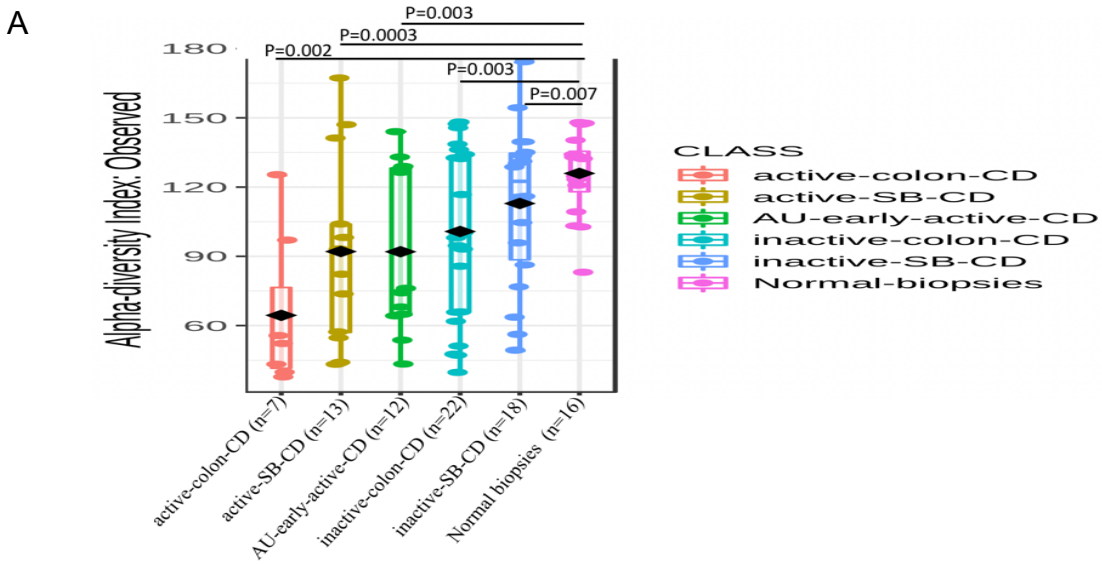
2.4.3.1 Between aphthous ulcers and healthy controls

The difference of microbiome patterns and diversity between aphthous ulcers in CD and controls is important to understand how the microbiome may play a role in the pathogenesis of CD. The healthy control biopsy samples included five Peyer's patches and 11 normal mucosal biopsies. We could not see any difference in the microbiome richness and diversity between Peyer's patches and normal mucosal biopsies from healthy controls, so we combined them for further analysis. Alpha diversity richness analysis based on observed OTUs between aphthous ulcers in the early stage of CD and normal biopsies in controls shows that OTU richness was significantly reduced ($p = 0.003$) in aphthous ulcers compared to normal biopsies (Figure 2.4 and Supplemental Table 2.4 and Supplemental Figure 2.4). However, based on the Shannon diversity index, there was no significant difference ($p = 0.06$) in the alpha diversity between aphthous ulcers (N=12) and normal biopsies (N=16) (Figure 2.4B). The Shannon diversity index gave different results from the observed OTUs because it considers the OTU richness and their relative abundance (evenness). The detailed analysis results based on the Shannon diversity index are listed in Supplemental Table 2.5 and Supplemental Figure 2.5). The difference was significantly different (p-value < 0.05).

2.4.3.2 Between small bowel Crohn's disease/SB-CD (active and inactive) vs controls

Observed OTU richness was significantly reduced in both active small bowel CD (p = 0.0003) and inactive small bowel CD (p = 0.003) microbiome compared with normal biopsies in healthy subjects (Figure 2.4A and Supplemental Table 2.4 and Supplemental Figure 2.4). Similarly, the Shannon diversity index revealed the significant reduction of microbiome diversity for active small bowel CD mucosa in (p =0.01) (Figure 2.4B and Supplemental Table 2.5 and Supplemental Figure 2.5). We did not observe any significant difference in the alpha diversity indices between the active small bowel CD and inactive small bowel CD. When the microbiome of the site is considered (both active and inactive small bowel) it was found that observed OTU richness was significantly reduced in small bowel CD (p = 0.0003) compared with that of the normal biopsies in controls (Figure 2.5B and Supplemental Table 2.6 and Supplemental Figure 2.6).

A



B

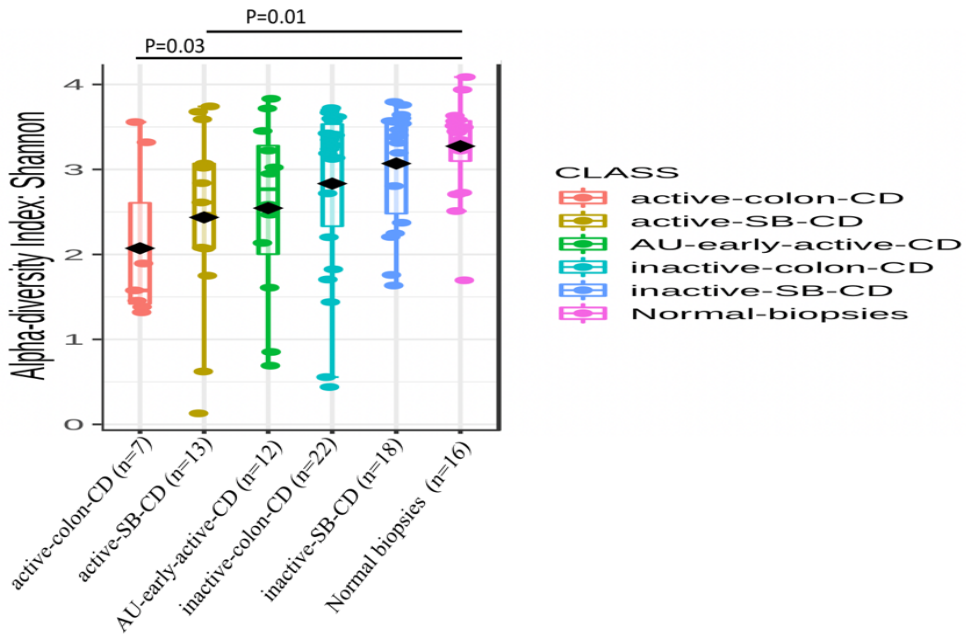


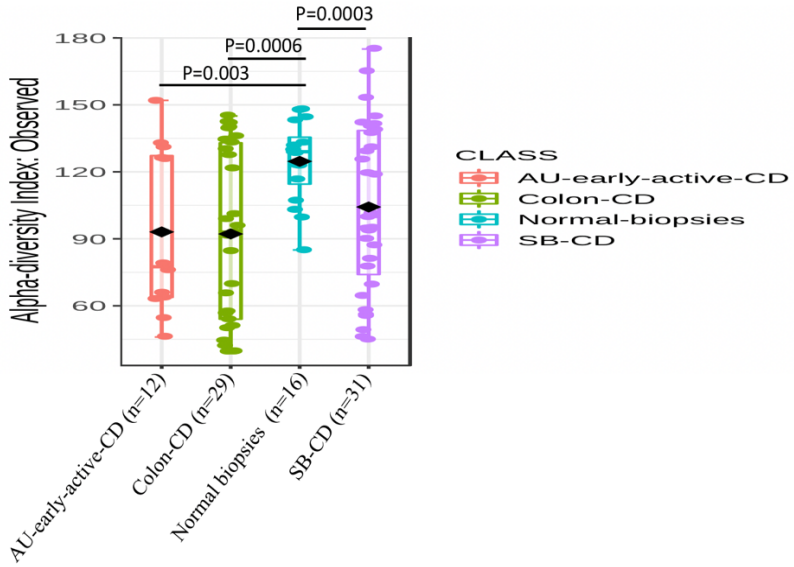
Figure 2.4. Changes in alpha-diversity indices in samples collected from active colonic mucosa (active-colon-CD), active small bowel (active-SB-CD), aphthous ulcers (AU-early-active-CD), inactive colonic mucosa (inactive-colon-CD), inactive small bowel (inactive-SB-CD) in CD patients and healthy controls (normal biopsies). Panels depict *A)* observed OTUs, and *B)* Shannon index. Significance was tested using Kruskal-Wallis (pairwise) test. The P-value above the box plots indicates significant differences in alpha diversity between different tissues.

2.4.3.3 Between patients with colonic Crohn's disease (active and inactive) and controls

The richness of active colonic CD tissues based on observed OTUs revealed significantly decreased richness when compared to control biopsies (p = 0.002). The OTU richness based on Observed OTUs was also reduced in inactive colonic CD tissues compared with the microbiome of normal biopsies in healthy subjects (p = 0.003) (Figure 2.4A and Supplemental Table 2.4 and Supplemental Figure 2.4). Similarly, based on the Shannon diversity index, the microbiome of the active colonic CD also showed reduced diversity compared with the controls (p = 0.03) (Figure 2.4B, Supplemental Table 2.5 and Supplemental Figure 2.5). When the microbiome of the colonic site was considered (combining both active and inactive colonic CD) it was

found that observed OTU richness was significantly reduced in colonic CD (p = 0.0006) compared to controls (Figure 2.5A, Supplemental Table 2.6 and Supplemental Figure 2.6). The summary of the Shannon diversity index results between colonic CD and normal biopsies is represented in the Figure 2.5B, Supplemental Table 2.7 and Supplemental Figure 2.7.

A.



B.

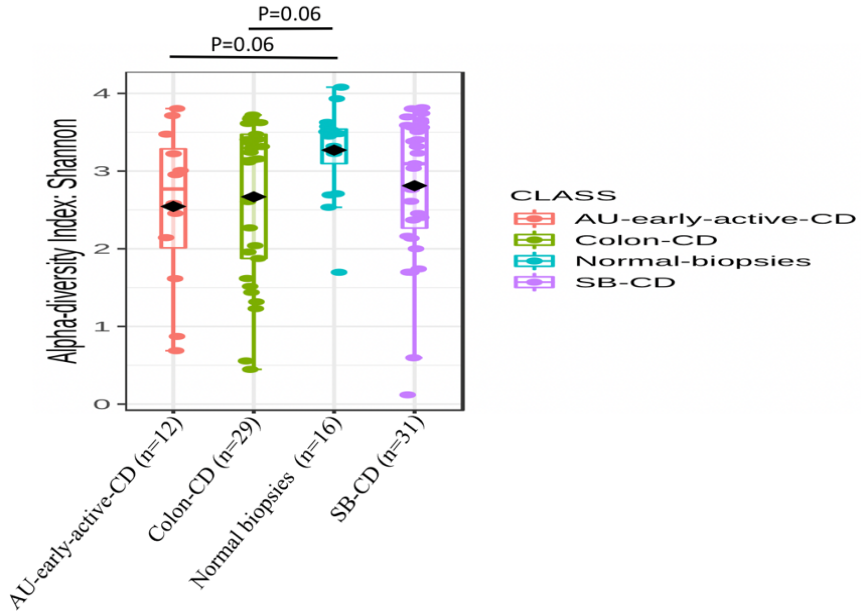


Figure 2.5. Changes in alpha-diversity indices in samples collected from aphthous ulcers (AU-early-active-CD), colonic mucosa (colon-CD) and small bowel (SB-CD) in CD patients and healthy controls (normal biopsies). Panels depict **A)** observed OTUs, and **B)** Shannon index. Observed OTU richness was significantly reduced in (AU-early-active-CD, colon- CD, and SB-CD compared to normal biopsies. Significance was tested using Kruskal-Wallis (pairwise)Test.

2.4.3.4 Between CD patients with active terminal ileum (TI) and inactive rectum

We also analysed the alpha diversity of the microbiome in active, inflamed TI and the inactive, uninflamed rectum from CD patients compared with normal controls. We observed a significant reduction ($p = 0.004$) in microbial alpha diversity (based on observed OTU) in the active terminal ileum but not in inactive rectum compared with the normal biopsies in controls (Supplemental Figure 2.8). These results showed that decreased microbiota diversity is found only in the inflamed site but not in the uninflamed site in patients with CD, which indicates that reduced alpha diversity in CD is not diffuse but focal.

2.4.3.5 Between CD patients with aphthous ulcer and inactive-small bowel

We analysed the alpha diversity of the microbiome in active aphthous ulcers compared to inactive uninflamed small bowel from CD patients. We could no observed a significant reduction ($p = 0.28$) in microbial alpha diversity (based on observed OTU) in the active aphthous ulcers when compared with inactive small bowel (Supplemental Figure 2.9).

2.4.3.6 Between CD patients with active small bowel and inactive-small bowel

We analysed the alpha diversity of the microbiome in active small bowel compared to inactive uninflamed small bowel from CD patients. We observed a significant reduction ($p = 0.04$) in microbial alpha diversity (based on observed OTU) in the active small bowel CD when compared with inactive small bowel (Supplemental Figure 2.10).

2.4.3.7 Comparison between active and inactive colon from CD patients

We also analysed the alpha diversity of the microbiome in active colon compared to inactive uninflamed colon from CD patients. We observed a significant reduction ($p = 0.03$) in microbial alpha diversity (based on observed OTU) in the active colon CD when compared with inactive colon (Supplemental Figure 2.11).

2.4.4 Differences in microbial community composition (beta diversity) between all Crohn's disease and controls

We carried out PERMANOVA analysis, which is a multivariate analysis with permutations, capable of considering multiple variables such as a large number of OTUs. To evaluate the gut microbiota beta diversity between CD and controls, we carried out principal coordinate analysis (PCoA) for all samples using analysis through application of Bray–Curtis dissimilarity (Figure 2.6A), Jensen–Shannon divergence (Figure 2.6B), Jaccard similarity index (Figure 2.6C), unweighted UniFrac (Figure 2.6D), and weighted UniFrac (Figure 2.6E). Patients with CD showed a distinct shift on PCoA compared to controls. Statistically significant dissimilarities of microbiome composition between samples were seen in all distance metrics: Bray–Curtis distance ($p = 0.001$), Jensen-Shannon-Divergence ($p = 0.001$), Jaccard distance ($p = 0.001$), unweighted UniFrac distance ($p = 0.003$), and weighted UniFrac distance ($p = 0.01$). Box plots showing statistically significant dissimilarities between CD and controls (beta diversity) based on Bray–Curtis distance metrics (Figure 2.6). Other results of beta diversity through QIIME analysis (Bray–Curtis distance, Jaccard distances, weighted UniFrac distances, unweighted UniFrac distances) are presented in the Supplemental Table 2.8 to Table 2.11)

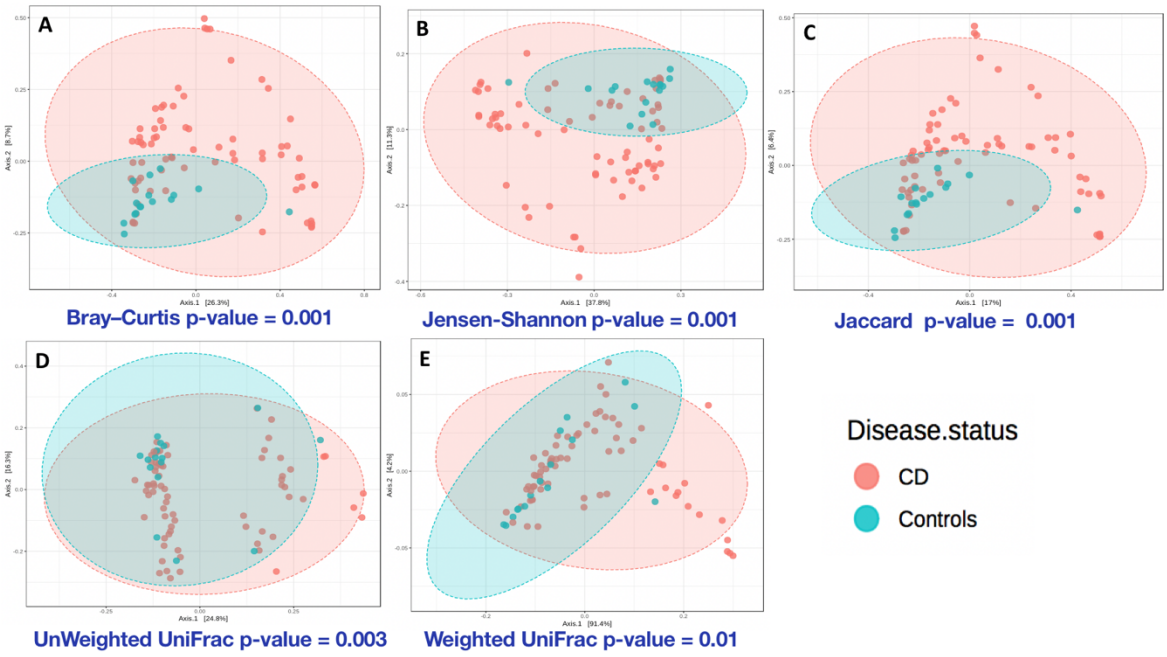


Figure 2.6. PCoA 2D plots of beta diversity analysis of CD patients (red dots) and control (blue dots). Between-sample dissimilarities were measured by **A**) Bray-Curtis distance, **B)** Jensen-Shannon divergence, **C**) Jaccard distances, **D**) Unweighted UniFrac, and **E**) Weighted UniFrac. Permutational multivariate analysis of variance (PERMANOVA) was performed to analyse statistical significance. P-value ($p<0.05$) is indicated at the bottom of each plot.

2.4.5 Comparative beta diversity in Crohn's disease and controls based on disease stages, disease activity and tissue types

2.4.5.1 Early stage of CD (aphthous ulcers) vs controls

Next, we visualized the PCoA to compare the microbial structure at the early active stage of disease stages in patients with CD and controls. We evaluated the gut microbiota diversity (beta diversity) based on permutational ANOVA (PERMANOVA) analysis carried out using QIIME-2 and microbiome analyst between early stage of stages of CD (aphthous ulcers) and controls. We observed statistically significant dissimilarities in beta diversity between aphthous ulcers samples and the normal biopsies in controls: Bray–Curtis distance ($p = 0.006$), Jensen-Shannon divergence ($p = 0.006$) Jaccard distance ($p = 0.006$), unweighted UniFrac ($p = 0.006$) and weighted

UniFrac (p = 0.02) (Figure 2.7). The detailed results of beta diversity analysis (Bray–Curtis distance, Jaccard distances, weighted UniFrac distances, unweighted UniFrac distances) are presented in the Supplemental Table 2.12 to 2.15.

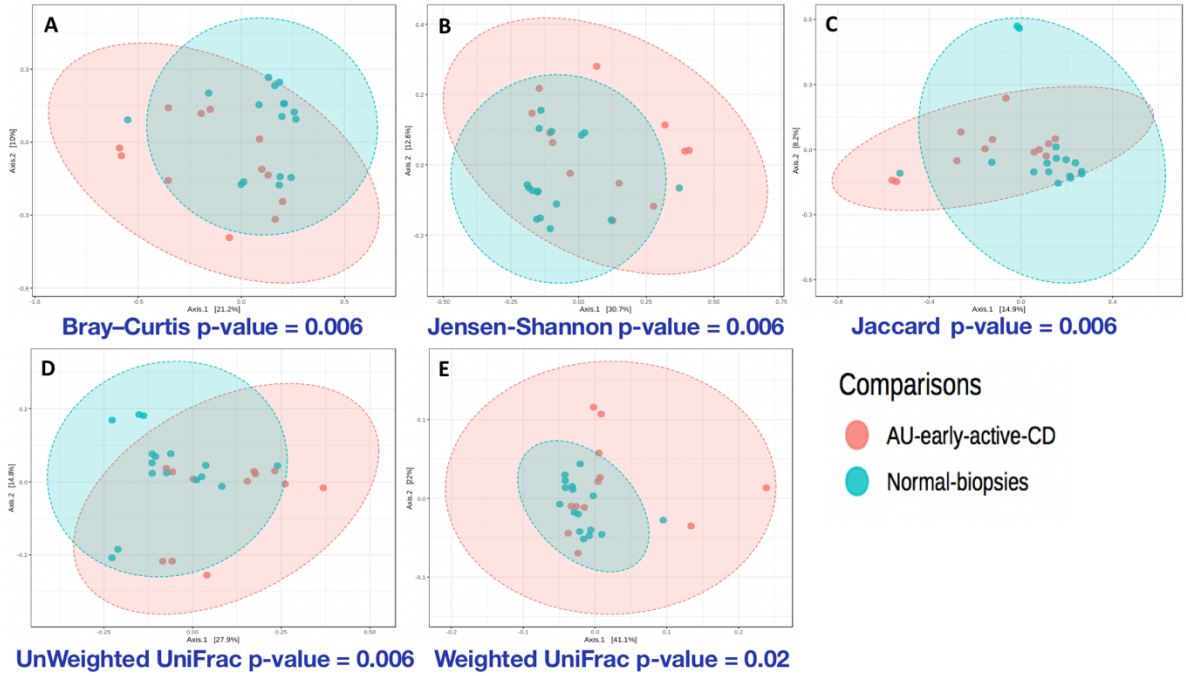


Figure 2.7. Principal component analysis (PCoA 2D) plots at genus level based on the disease stages in CD patients and control using microbial analysist. Each data point represents an individual sample. Red colour indicates aphthous ulcers (AU) and green colour indicates the normal biopsies from controls. Between-sample dissimilarities were measured by **A**) Bray-Curtis distance, **B)** Jensen-Shannon divergence, **C**) Jaccard distances, **D**) Unweighted UniFrac, and **E**) Weighted UniFrac. Permutational multivariate analysis of variance (PERMANOVA) was performed to analyse statistical significance. P-value ($p<0.05$) is indicated at the bottom of each plot.

2.4.5.2 Inactive small bowel CD vs controls

We used pairwise PERMANOVA to evaluate the gut microbiota diversity (beta diversity) between inactive small bowel CD and controls and visualised using MicrobialAnalyst (Figure 2.8 and Supplemental Table 2.12-2.15). Except for unweighted UniFrac distance metrics, all other beta diversity comparison indices (Bray-Curtis distance, Jensen-Shannon divergence, Jaccard distances, and weighted UniFrac distance metrics) using pairwise PERMANOVA for gut microbiome suggested

statistically significant (*P* <0.01) community differences between inactive-small bowel-CD and normal biopsies in controls (Figure 2.8). Three samples (72-1, 77-1, 95-1) from this inactive small bowel had active disease in the colon of the CD patients. We also observed a significant difference in the microbial community based on beta diversity comparisons index (Bray-Curtis distance, Jensen-Shannon divergence, and Jaccard distances) in those samples (inactive-SB-CD with active disease in the colon) compared with the normal biopsies in controls (Supplemental Figure 2.12).

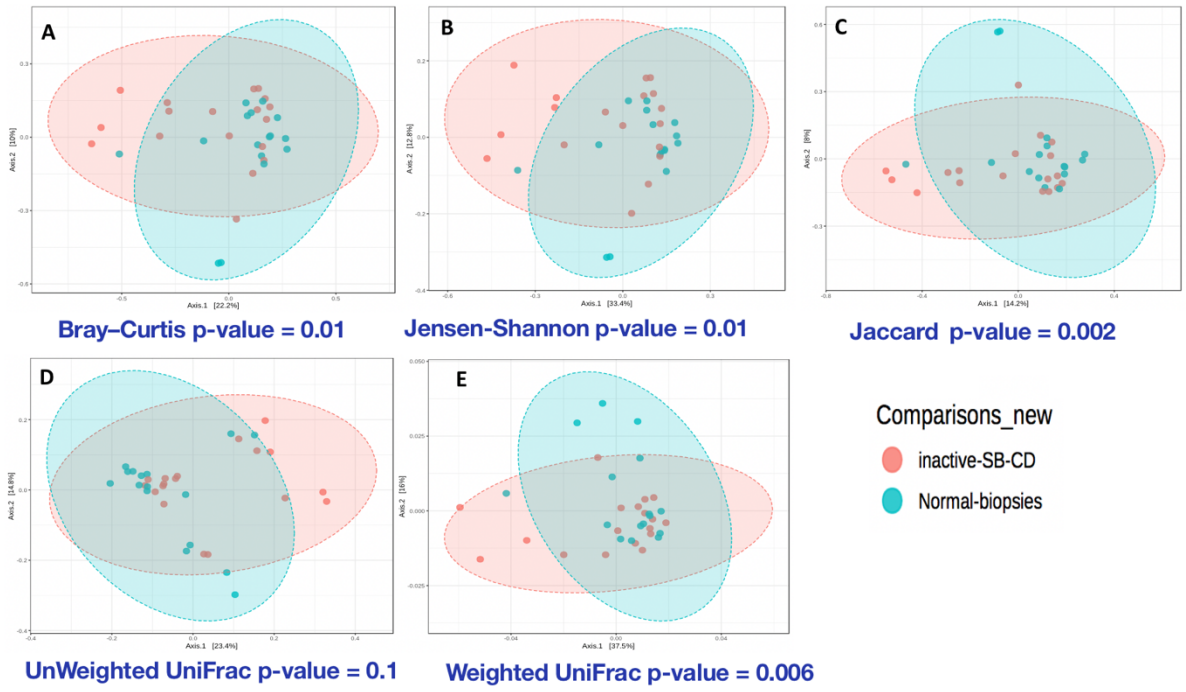


Figure 2.8. Principal component analysis (PCoA 2D) plots at genus level based on the disease stages in CD patients and controls using microbial analysist. Each data point represents an individual sample. Red colour indicates inactive stage-small bowel CD, and green colour indicates the controls. Between-sample dissimilarities were measured by **A**) Bray-Curtis distance, **B)** Jensen-Shannon divergence, **C**) Jaccard distances, **D**) Unweighted UniFrac, and **E**) Weighted UniFrac. Permutational multivariate analysis of variance (PERMANOVA) was performed to analyse statistical significance. P-value (p<0.05) is indicated in the bottom of each plot. *Here, SB indicates small bowel tissue.

2.4.5.3 Active small bowel CD vs controls

We used pairwise PERMANOVA analysis to evaluate the gut microbiota diversity (beta diversity) between active small bowel CD and controls using QIIME-2 and microbial analyst (Figure 2.9 and Supplemental Table 2.12-2.15). Beta diversity comparisons based on all the index (Bray-Curtis distance, Jensen-Shannon divergence, Jaccard distances, and unweighted UniFrac distance metrics) using pairwise PERMANOVA for gut microbiome suggested statistically significant ($p < 0.01$) community differences between active-small bowel-CD and normal biopsies in controls (Figure 2.9 and Supplemental Table 2.12-2.15).

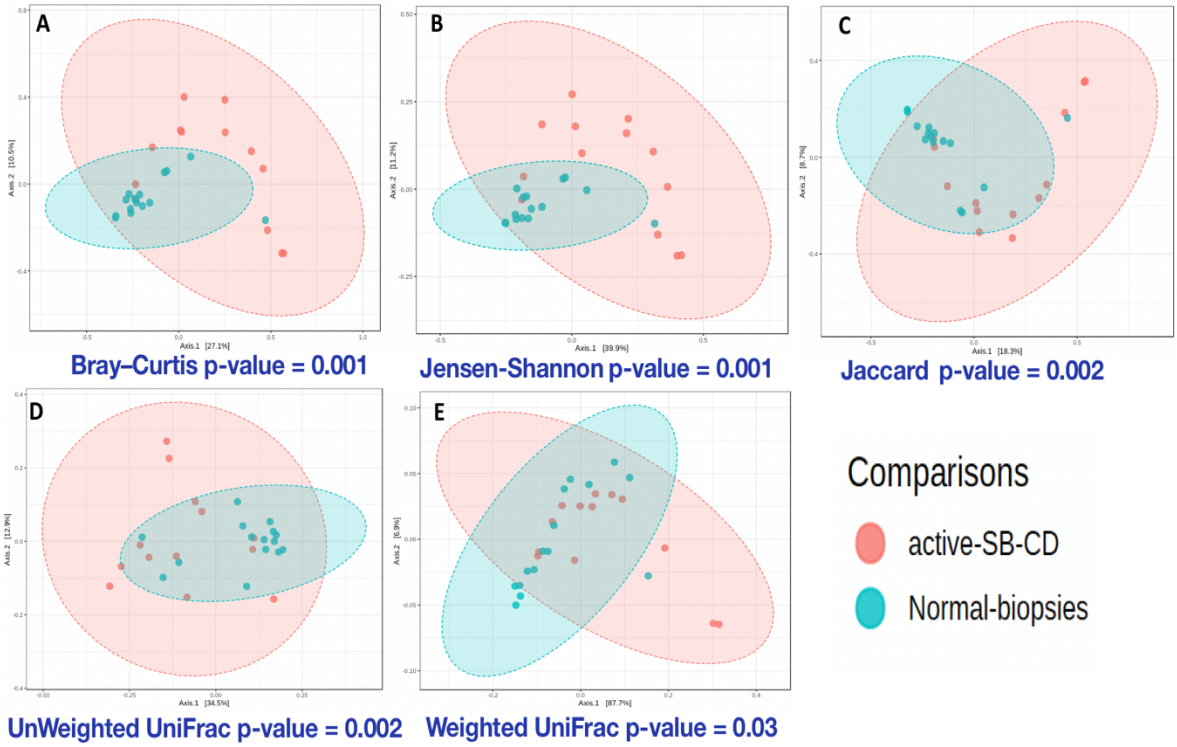


Figure 2.9. Principal component analysis (PCoA 2D) plots at feature level based on the active small bowel CD and control using microbial analysist. Each data point represents an individual sample. Red colour indicates normal biopsies from the controls, and green colour indicates both active and inactive stage-small bowel CD. Between-sample dissimilarities were measured by **A**) Bray-Curtis distance, **B)** Jensen-Shannon divergence, **C**) Jaccard distances, **D**) Unweighted UniFrac, and **E**) Weighted UniFrac. Permutational multivariate analysis of variance (PERMANOVA) was performed to analyse statistical significance. P-value ($p<0.05$) is indicated in the bottom of each plot. *Here, SB indicates small bowel tissue.

2.4.5.4 Inactive colonic CD vs controls

Beta diversity comparisons based on all the indices (Bray-Curtis distance, Jensen-Shannon divergence, Jaccard distances, unweighted UniFrac and weighted UniFrac distance metrics) using pairwise PERMANOVA analysis for gut microbiome suggested statistically significant (P <0.001) community differences between inactive colon CD and normal biopsies in controls (Figure 2.10 and Supplemental Table 2.12-2.15).

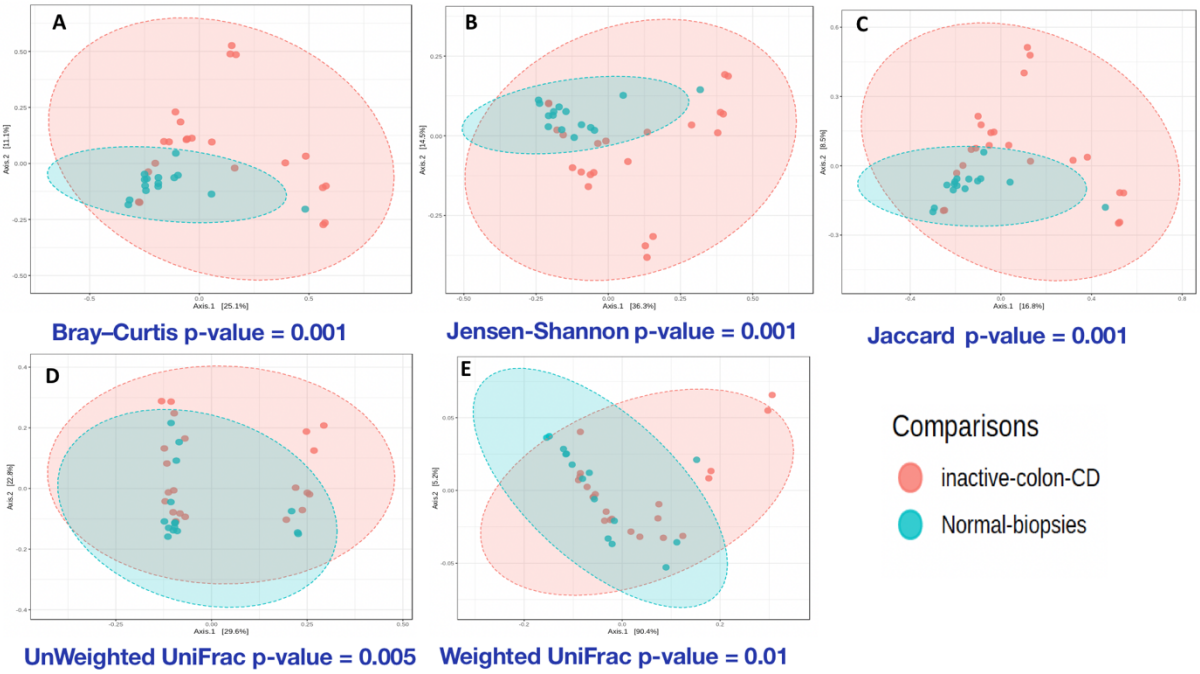


Figure 2.10. PCoA 3D plots of beta diversity analysis of CD patients (red dots) and control (blue dots). Between-sample dissimilarities were measured by **A**) Bray-Curtis distance, **B)** Jensen-Shannon divergence, **C**) Jaccard distances, **D**) Unweighted UniFrac, and **E**) Weighted UniFrac. Permutational multivariate analysis of variance (PERMANOVA) was performed to analyse statistical significance. P-value (p<0.05) is indicated at the bottom of each plot.

2.4.5.5 Active colonic CD vs controls

Beta diversity comparisons based on all the indices (Bray-Curtis distance, Jensen-Shannon divergence, Jaccard distances, unweighted UniFrac, and weighted UniFrac distance metrics) using pairwise PERMANOVA for gut microbiome suggested

statistically significant (p <0.001) community differences between active colon CD and normal biopsies in controls (Figure 2.11 and Supplemental Table 2.12-2.15).

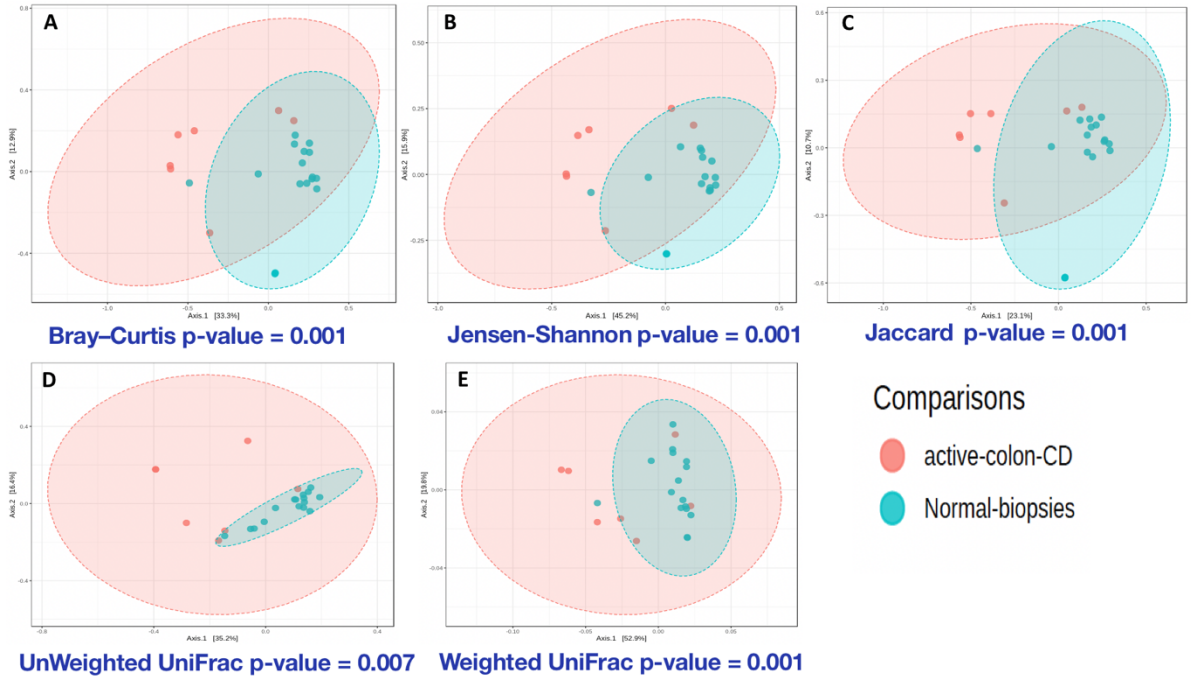


Figure 2.11. Principal component analysis (PCoA 2D) plots at genus level based on the disease stages in CD patients and control using microbial analysist. Each data point represents an individual sample. Red colour indicates active-colon-CD and green colour indicates the normal biopsies from the controls. Between-sample dissimilarities were measured by **A**) Bray-Curtis distance, **B)** Jensen-Shannon divergence, **C**) Jaccard distances, **D**) Unweighted UniFrac, and **E**) Weighted UniFrac. Permutational multivariate analysis of variance (PERMANOVA) was performed to analyse statistical significance. P-value (p<0.05) is indicated at the bottom of each plot.

2.4.5.6 Bacterial diversity comparisons between active and inactive small bowel CD

Beta diversity comparisons based on different indices (Bray-Curtis distance, Jensen-Shannon divergence, and Jaccard distances) using pairwise PERMANOVA for gut microbiome suggested statistically significant (p <0.02) community differences between active SB and inactive SB in patients with CD (Supplemental Figure 2.13). Only unweighted UniFrac distance did not show any significantly difference of microbial pattern between active SB and inactive SB in CD.

2.4.5.7 Beta microbial diversity comparisons between active and inactive colonic CD

Beta diversity comparisons based on different indices (Bray-Curtis distance, Jensen-Shannon divergence, Jaccard distances, and unweighted UniFrac distance metrics) using pairwise PERMANOVA for gut microbiome suggested no statistically significant ($P > 0.05$) microbial community differences between active colon and inactive colon in patients with CD.

2.4.6 The gut microbial abundance at taxonomic levels phylum and genus in Crohn's disease patients and controls

To test whether CD in different stages of the disease and tissue types harbour distinct microbial communities, we compared the relative abundance of taxa between each of our groups of samples (CD and controls). CD and controls show differences in taxonomic distribution at different taxonomic levels.

2.4.6.1 At phylum level

Compared to healthy controls, CD patients demonstrated a microbial imbalance at the phylum level (Figure 2.12A). Top 5 bacteria based on relative abundance at the phylum level (L2) in CD and control were: Firmicutes, Bacteroidetes, Proteobacteria, Fusobacteria and Actinobacteria respectively. We observed that in CD, this five-phylum changed in abundance relative to healthy controls (Figure 2.12A). In CD, the abundance of Bacteroidetes, which is often associated with a healthy gut, was deceased ($p = 0.03$). In contrast, the abundance of Proteobacteria, a phylum associated including a wide variety of pathogens, was increased ($p = 0.01$) (Figure 2.12B and Supplemental Table 2.16 and Supplemental Figure 2.14). In different disease stages and tissue types in CD and controls, Firmicutes was the most abundant

phylum in the gut in all disease stages in CD and controls. We observed that in CD, these top five most abundant phyla changed in abundance relative to healthy controls (Figure 2.12A). The abundance of Proteobacteria was significantly increased both in inactive colonic CD and active-small bowel CD ($p = 0.008$) from that of the normal biopsies in controls (Supplemental Table 2.17 and Supplemental Figure 2.10). Similarly, Proteobacteria was also found in higher abundance in active colon CD compared with normal biopsies in controls (Supplemental Table 2.17 and Supplemental Figure 2.15). The abundance of Firmicutes did not change significantly in disease stages and tissue types in CD compared to controls (Supplemental Table 2.18 and Supplemental Figure 2.16). However, Bacteroides phylum was reduced in abundance in active colonic CD ($p = 0.004$) and inactive small bowel CD ($p = 0.02$) compared with normal biopsies in controls (Supplemental Table 2.19 and Supplemental Figure 2.17). The abundance profiles for the most dominant bacterial phyla (Proteobacteria, Firmicutes and Bacteroidetes) varied across disease stages and tissue types (Figure 2.12B)

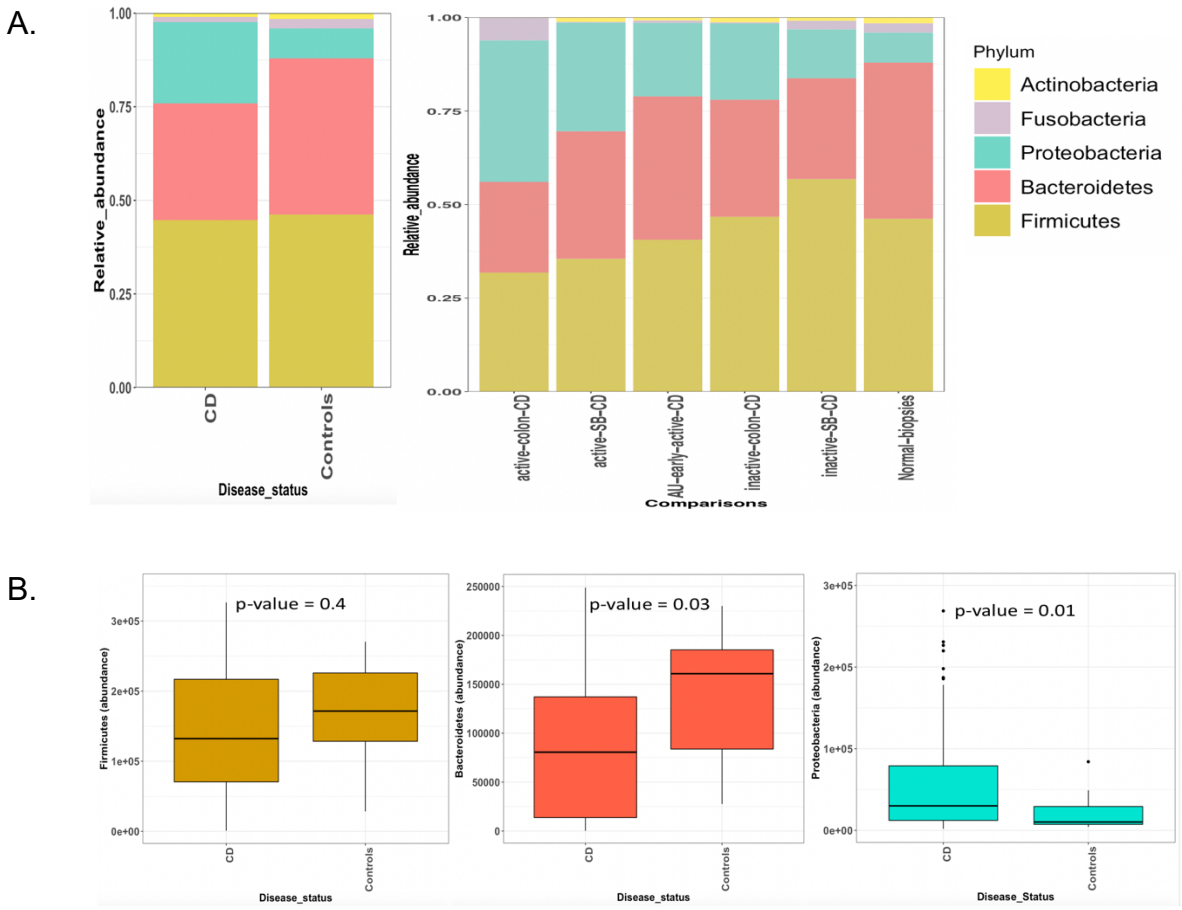


Figure 2.12. A) Relative abundance of top 5 most abundant bacterial phyla between CD and controls **B)** Box plots depicting difference in abundances of the top three most abundant gut bacterial genera between CD and controls using Kruskal-Wallis multiple comparison tests. Statistically significant values are represented as $p < 0.05$ at the top of the boxplot. * CD= Crohn's disease, SB= small bowel and AU= aphthous ulcers.

2.4.6.2 At genus level

Top bacteria at the genus level (L6) in CD, controls and disease stages were *Bacteroides*, *Faecalibacterium*, *Escherichia-shigella*, *Blautia* and *Lachnoclostridium*, respectively (Figure 2.13A). We further investigated the top three important genera (*Bacteroides*, *Faecalibacterium*, *Escherichia-shigella*) belonging to each of the top-most abundant phyla (Firmicutes, Proteobacteria, Bacteroidetes) to identify if the abundance of any genus was significantly different between disease status CD and controls, and among disease stages and tissue types. The Kruskal-Wallis test results showed that abundances of bacterial genera *Faecalibacterium* and *Escherichia-*

shigella were significantly altered in different disease stages and tissue types (Figure 2.13 and Supplemental Table 2.20). Classical univariate statistical analysis also revealed a total of 48 genera either positively or negatively associated with CD compared with controls (Supplemental Table 2.20). Intestinal microbiomes at the genus level that were significantly different between CD and controls as determined by the Kruskal-Wallis test are presented in Figure 13B. *Faecalibacterium* genus was significantly reduced ($p = 0.02$) in CD. In CD, the Bacteroides genus reduced in abundance, but the difference in abundance was not significantly associated (Figure 2.13B). Finally, for the Proteobacteria phylum, we observed an imbalance in *Escherichia-shigella* genera, with a significant increase in the abundance of *Escherichia-shigella* (Figure 2.13B) in CD patients compared to the controls. In disease stages, abundance of *Escherichia-shigella* increased significantly in active colonic CD ($p = 0.005$), active small bowel CD ($p = 0.009$) and inactive colon-CD ($p = 0.02$) patients compared to normal biopsies in controls (Figure 2.13C and Supplemental Table 2.21). The abundance level of *Faecalibacterium* genus was reduced significantly in the active colon CD ($p = 0.03$), active small bowel CD ($p = 0.02$) and inactive colonic CD ($p = 0.01$) relative to normal biopsies in controls (Figure 2.13D and Supplemental Table 2.22). The abundance of *Bacteroides* was only reduced significantly in the active colon CD ($p = 0.008$) relative to normal biopsies in controls (Figure 2.13E and Supplemental Table 2.23). In summary, we demonstrate that multiple phyla and genera differ in abundance between CD and controls, as well as in disease stages of CD and controls.

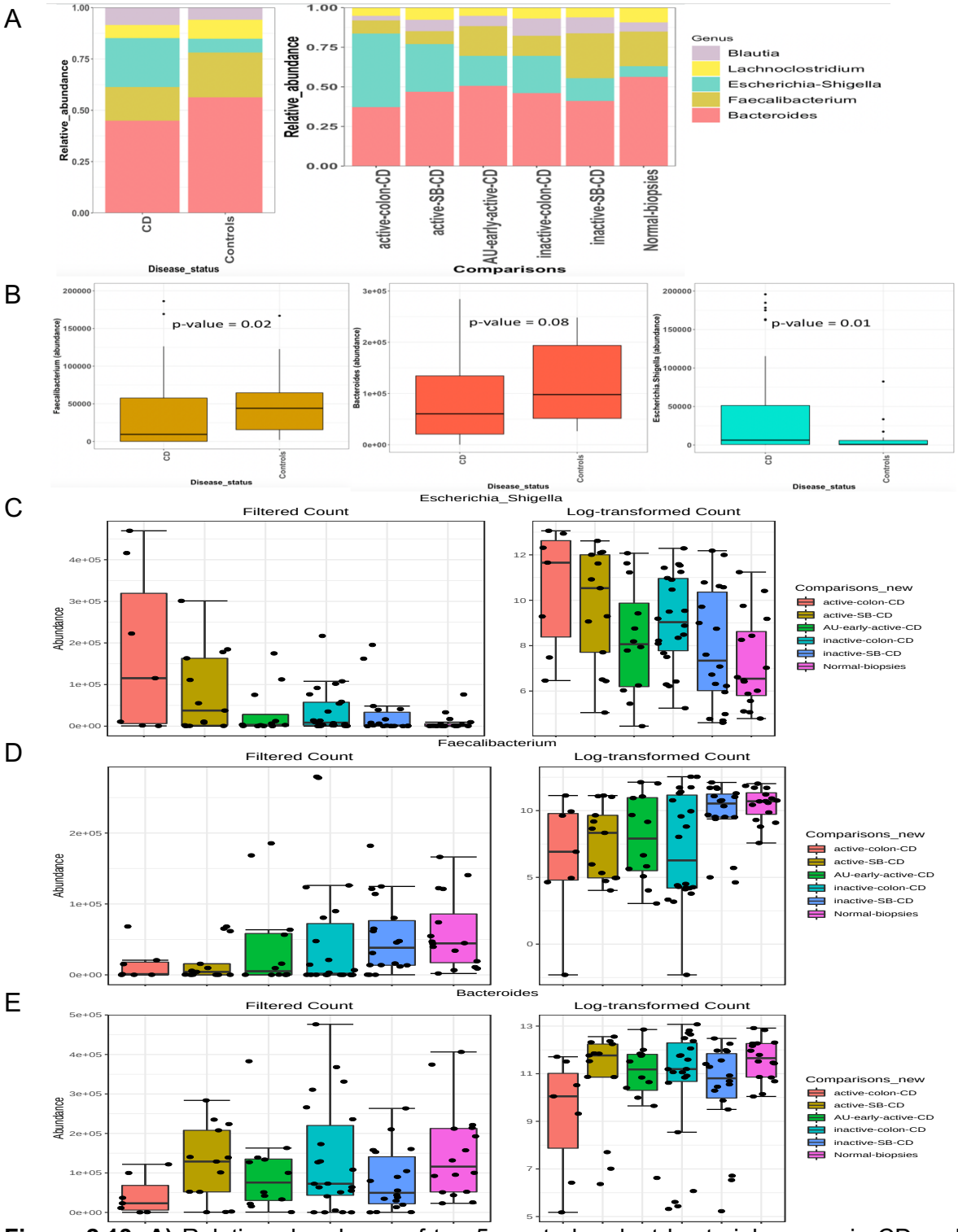
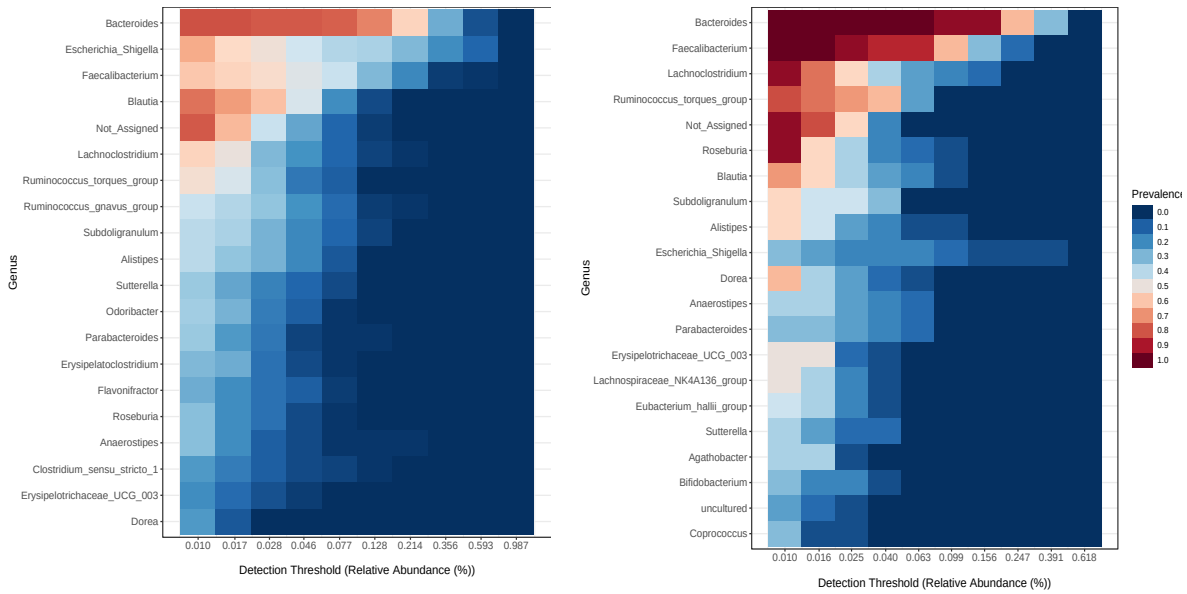


Figure 2.13. A) Relative abundance of top 5 most abundant bacterial genera in CD and controls, **B)** Box plots depicting the difference in abundances of the top three most abundant gut bacterial genus between CD and controls. **C)** Box plots showing the difference in the abundance of *Escherichia_Shigella* in the disease stages and tissue types in CD and controls. **D)** Box plots showing the difference in the abundance of *Faecalibacterium* in the disease stages and tissue types in CD and controls and **E)** Box plots showing the difference in the abundance of *Bacteroides* in the disease stages and tissue types in CD and controls. Kruskal-Wallis multiple comparisons were made, and statistically significant values were represented as $p < 0.05$ at the top of the boxplot. *Here, CD= Crohn's disease, SB= small bowel and AU= aphthous ulcers.

2.4.7 Comparison of the core intestinal microbiome in disease status and different disease stages in Crohn's disease and controls

A heatmap based on the core microbiome in CD and controls samples showed that some disease-specific genera were increased, and some genera were decreased in abundance in CD. No disease-specific clustering was observed in CD. CD and controls could not accurately be discriminated by their microbial profiles, although the abundance of some core genera was either increased or decreased in CD (Figure 2.14). The main characterising feature in CD was the reduction in the abundance of *Bacteroides*, *Faecalibacterium*, *Lachnoclostridium*, *Ruminococcus- torques group*, *and Roseburia,* which are abundant in healthy controls. On the other hand, *Escherichia-Shigella* were the overrepresented core microbiome in CD. We also compared the core gut microbiota in different disease stages and tissue types in CD with controls. Notably, the core microbiome differs distinctly in different disease activity (active and inactive CD) and tissue types (aphthous ulcers, small bowel CD and colon CD) in CD compared with controls (Supplemental Figure 2.18-2.21).



Core microbiome in CD Core microbiome in controls

Figure 2.14. Core gut microbiome in CD and controls

2.4.8 Microbial biomarkers in Crohn's disease compared with controls

Using linear discriminant analysis effect size (LEfSe), we found biomarkers in the gut microbiota in patients with CD, particularly *Escherichia_Shigella*, *Ruminococcus_gnavus*_group, *Flavonifractor*, *Clostridium_innocuum*_group, *Klebsiella*, *Megamonas*, *CAG_352, Christensenellaceae_R_7_group* and *Butyricimonas*, which were positively associated with CD and different disease stages in CD. In contrast, *Faecalibacterium, Ruminococcus_torques*_group, *Roseburia*, *Erysipelotrichaceae_UCG_003*, *Dorea*, *Eubacterium_hallii_group*, *Lachnospiraceae_NK4A136_group*, *Lachnospiraceae_UCG_004*, *Agathobacter*, *Phascolarctobacterium*, and *Fusicatenibacter* were negatively associated with CD while overrepresented in the control group (Figure 2.15, and Table 2.3). Furthermore, significant negative correlations were found in the genera depleted in the CD patients, abundant in healthy controls, suggesting their synergy as commensal bacteria in maintaining gut microbiota homeostasis (Figure 2.15). On the other hand, the genera enriched in CD showed negative correlations with those depleted in the healthy controls.

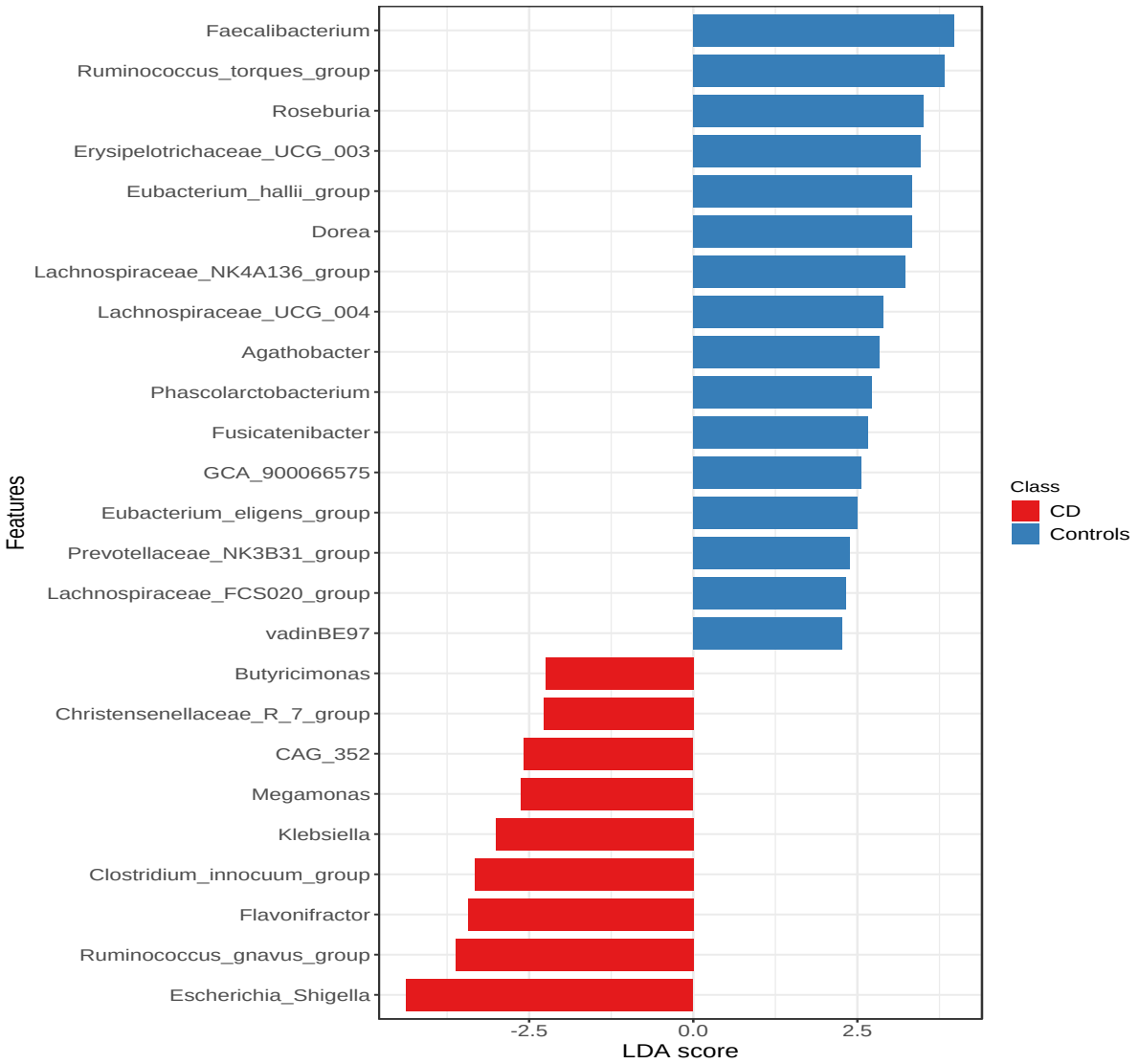


Figure 2.15. Top 25 microbiota biomarkers for CD patients compared to controls, calculated with linear discriminant analysis effect size (LEfSe) using MicrobiomeAnalyst.

Table 2.3. Microbiota biomarker for Crohn's disease patients compared to controls calculated with linear discriminant analysis effect size (LEfSe)

Genus	P-value	FDR	CD	Controls	LDA score
Faecalibacterium	0.02	0.17	39604	58507	3.98
Ruminococcus_torques_group	0.006	0.09	9029.6	22231	3.82
Roseburia	0.0003	0.01	3936.3	10238	3.5
Erysipelotrichaceae_UCG_003	0.007	0.09	3241.5	9047.4	3.46
Dorea	0.0004	0.01	2219	6472.2	3.33
Eubacterium_hallii_group	0.03	0.2	1512.5	5742.8	3.33
Lachnospiraceae_NK4A136_group	0.0002	0.01	1581.6	4991.6	3.23
Lachnospiraceae_UCG_004	0.0002	0.012	1276	2827	2.89

Agathobacter	0.003	0.05	1514.5	2911.8	2.84
Phascolarctobacterium	0.01	0.13	2183.3	3234.1	2.72
Fusicatenibacter	0.02	0.20	1491.6	2411.8	2.66
GCA_900066575	0.002	0.04	451.18	1177.1	2.56
Eubacterium_eligens_group	0.04	0.28	349.44	979.88	2.5
Prevotellaceae_NK3B31_group	0.0002	0.012	2.0278	495.69	2.39
Lachnospiraceae_FCS020_group	0.0007	0.02	242.11	653.75	2.32
vadinBE97	0.008	0.10	36.389	409.69	2.27
Lachnospiraceae_ND3007_group	0.003	0.05	806.12	1045.4	2.08
Holdemanella	0.01	0.1	458.4	697.31	2.08
Clostridia_UCG_014	0.0002	0.012	141.1	319.81	1.96
Muribaculaceae	0.004	0.06	17.681	169.88	1.89
Ruminococcus_gauvreauii_group	0.01	0.13	551.1	680.06	1.82
Lachnospiraceae_UCG_001	0.003	0.05	23.597	101	1.6
Marvinbryantia	0.002	0.04	108.18	184.12	1.59
Allisonella	0.01	0.12	35.694	107.31	1.57
Lachnospiraceae_UCG_008	0.04	0.26	15.25	76.188	1.5
Blastococcus	4.16E-10	1.17E-07	0	19.25	1.03
Howardella	0.01	0.15	10.75	26.25	0.942
Acidocella	0.0002	0.01	0	6.9375	0.65
Helicobacter	0.02	0.20	1.0833	7.5625	0.627
Shuttleworthia	0.002	0.04	0	5.6875	0.585
UCG_008	0.002	0.04	0	1.625	0.258
Eubacterium	0.02	0.20	66.986	0	-1.54
UCG_003	0.04	0.28	1138.7	857.94	-2.15
CAG_56	0.002	0.048	1262.3	951.56	-2.19
Butyricimonas	0.03	0.24	745.67	391.88	-2.25
Christensenellaceae_R_7_group	0.02	0.17	1297.4	917.44	-2.28
CAG_352	0.008	0.10	1195.9	443.88	-2.58
Megamonas	0.04	0.28	823.11	0	-2.62
Klebsiella	0.01	0.17	2031.9	8.0625	-3.01
Clostridium_innocuum_group	0.02	0.20	5110.7	823.19	-3.33
Flavonifractor	0.01	0.16	5978.6	586.81	-3.43
Ruminococcus_gnavus_group	0.003	0.05	9678.9	1412.4	-3.62
Escherichia_Shigella	0.01	0.16	55753	9341.8	-4.37

FDR-False discovery rate; LDA-Linear discriminant analysis

2.4.9 Co-abundance of different bacterial genera in the gut

A correlation network, which included only significant genera–genera relationships based on their median of the abundances using Pearson's correlation coefficient and p-value threshold<0.05), was constructed using gut microbiota abundance profiles in CD and controls. This network has 76 nodes and 329 edges, where only 12 had positive, and the remaining 317 had negative correlations with CD (Figure 2.16). We then visualised the connection of the top three most abundant genera in CD (*Bacteroides*, *Escherichia-Shigella and Faecalibacterium*), which were connected with a total of 14 nodes (Figure 2.16 A). We observed that in the top three most abundant bacterial genera, *Escherichia-Shigella* is correlated negatively with *Faecalibacterium* (p = 1.00E-04) and *Bacteroides* (p = 0) in the network based on Pearson's product-moment correlation method (Figure 2.16 B). In addition to that, *Escherichia-Shigella* genus associated nodes also connected negatively with several genera (Table 2.4). Genus *Bacteroides* only formed a network with *Escherichia-Shigella* (p = 0). *Faecalibacterium* genus was positively associated with UCG_002 (p = 0), *Ruminococcus_gauvreauii*_group (p = 0), *Christensenellaceae*_R_7_group (p = 0), *Fusicatenibacter* (p =6.00E-04) and Family_XIII_AD3011_group (p = 8.00E-04). This suggests that a change in the abundance of one genus is likely to change the abundance of the correlated genera collectively. This could arise from a change in the microenvironment due to the inflammatory response producing mediators, cytokines, diarrhoea, increased motility, and changes in mucus production.

.

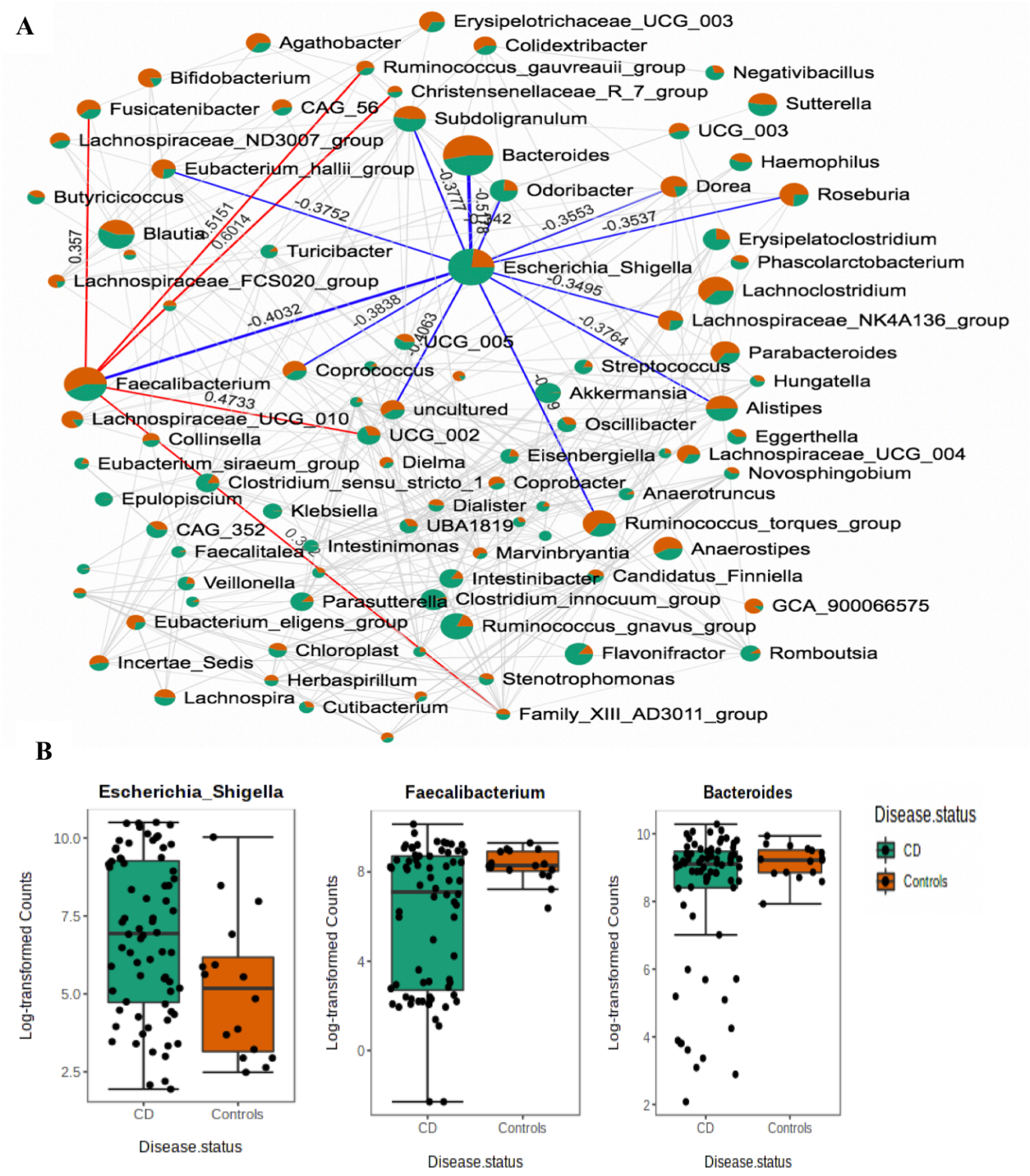


Figure 2.16. A. The network plot represents the co-coexistence of bacterial genera. In this network, the correlation of the top three most abundant genera (*Bacteroides*, *Faecalibacterium*, and *Escherichia-Shigella*) was highlighted. The network connections are based on Pearson's correlation coefficient test using the median of abundance. The P-value threshold was < 0.05, and the correlation threshold was 0.33. **B**. The box plots represent the difference in the abundance of *Bacteroides, Faecalibacterium, and Escherichia-Shigella* between CD and controls.

Table 2.4. Correlation of *Escherichia-Shigella* with other bacterial genera

Taxon1	Taxon2	Correlation	P value
Escherichia_Shigella	*Bacteroides*	-0.5178	0
Escherichia_Shigella	*uncultured*	-0.4063	1.00E-04
Escherichia_Shigella	*Faecalibacterium*	-0.4032	1.00E-04
Coprococcus	*Escherichia_Shigella*	-0.3838	2.00E-04
Escherichia_Shigella	*Ruminococcus_torques_group*	-0.379	3.00E-04
Escherichia_Shigella	*Subdoligranulum*	-0.3777	3.00E-04
Alistipes	*Escherichia_Shigella*	-0.3764	3.00E-04
Escherichia_Shigella	*Eubacterium_hallii_group*	-0.3752	3.00E-04
Dorea	*Escherichia_Shigella*	-0.3553	7.00E-04
Escherichia_Shigella	*Roseburia*	-0.3537	7.00E-04
Escherichia_Shigella	*Lachnospiraceae_NK4A136_group*	-0.3495	8.00E-04
Escherichia_Shigella	*Odoribacter*	-0.342	0.0011

2.5 Discussion

CD is one of the most studied human diseases associated with gut microbiota. Gut microbiota and its association with the gut immune system is assumed to play a major role in the pathogenesis of CD; however, many aspects of this correlation are not well-defined yet. In this study, we demonstrated differences with respect to mucosal microbiota from the gut between CD patients and controls based on 16S rRNA (V3-V4) sequencing analysis. In addition, the dominant dysbiotic pattern was characterised by the statistically significant decreased diversity of gut mucosal microbiota in CD compared to controls. Many previous studies have also shown reduced diversity of gut microbiota in patients with CD as compared to that in healthy controls (Ott et al., 2004; Manichanh et al., 2006; Kowalska-Duplaga et al., 2019; Boland et al., 2020). However, most of those studies did not consider different disease stages and tissue types, which could better explain the influence on gut microbiota in patients with CD. And some of those studies were only based on faecal samples. Our study collected biopsy and

tissue samples from different stages (early to late) in patients with CD along with the different gut locations. In addition, we also studied the effect of disease activity (active and inactive disease) on the gut microbiota in CD compared with controls.

We examined the bacterial community structure (using PCoA) of microbiota in CD patients and controls and extended these observations to show that there were differences between the early stage of active CD (aphthous ulcers) and controls which are structurally distinguishable. Detailed compositional alterations of bacterial phyla and genera in gut mucosa-associated microbiota in CD patients were identified in this study. Around 99% of intestinal bacteria consists of four major phyla—Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria (Eckburg et al., 2005; Ley et al., 2008). The two most dominating phyla in the intestinal microbiota of healthy adults are Bacteroidetes and Firmicutes (Sheehan et al., 2015; Andoh, 2016). Bacterial abundance and composition vary in different parts of the intestinal tract. Few species of bacteria with lower abundance reside in the stomach and the upper portion of the small intestine and the abundance of bacteria progressively increases in the lower intestine (Sartor, 2008). A decrease in Bacteroidetes and Firmicutes and an increase in Gammaproteobacteria were the most frequently observed alterations in the CD microbiome (Scarpellini et al., 2015). In this study, *Proteobacteria* was the significantly increased phylum in CD patients, which agreed with a consistent finding across many studies (Man et al., 2011; Vester-Andersen et al., 2019a). In contrast, a significantly lower number of *Bacteroidetes* phylum was observed in patients with CD compared to that of controls. These results corroborated the finding of Alam *et al., 2020* (Alam et al., 2020), where they found a 2.4-fold decreased abundance of Bacteroidetes in CD compared with a healthy gut.

In this present study, the abundance of *Escherichia-Shigella* was significantly increased ($p = 0.01$) in CD compared to controls. We also observed that the abundance of *Escherichia-Shigella and Faecalibacterium* was significantly different among some disease stages in CD and controls. In our study, the abundance of *Escherichia-shigella* increased significantly in active colonic CD, active small bowel CD and inactive colonic CD compared to normal biopsies in controls.

In core microbiome analysis at different disease stages and tissue types, we have found that, *Escherichia-shigella* was the most predominant core microbiome in both active colonic CD and active small bowel CD. In contrast, *Escherichia-shigella* was the fourth most abundant core microbiome in aphthous ulcers (early active CD) and inactive small bowel CD, whereas third most abundant core microbiome in the inactive colonic CD. Interestingly, *Escherichia-shigella* was the most predominant core microbiome in inactive small bowel CD with active disease in colon. In many previous studies, *Escherichia-Shigella* genus was also found to be notably higher in CD patients, as compared to that in controls. *Escherichia coli*, particularly AIEC, is an important pathotype hypothesized to play a critical role in the pathogenesis of CD (Darfeuille-Michaud et al., 2004). Studies based on the murine models also suggested an important role of *E. coli* in induction and maintaining intestinal inflammation in patients with CD (Carvalho et al., 2009; Schmitz et al., 2019).

Other disease-associated microbiome shifts were characterised by a depletion of health-associated bacteria in CD patients relative to controls. *Faecalibacterium* represented a higher average proportion (13%) in the mucosal samples of the controls, and 10% of the mucosal samples in the patients with CD. Our study also observed a significant reduction in the abundance of the *Faecalibacterum* genus in active and inactive colonic mucosa and active small bowel

in CD patients. *Faecalibacterium* was reported to play a critical role in maintaining intestinal homeostasis by restoring the intestinal barrier structure (Martinez-Medina et al., 2006; Xu et al., 2020). Numerous studies demonstrated a decrease of *Faecalibacterium* in both newly diagnosed and chronic CD patients (Sokol et al., 2008a; Willing et al., 2010; Fujimoto et al., 2013; Gevers et al., 2014; Lopez-Siles et al., 2014b). Moreover, Willing et al. (Willing et al., 2010) reported an increase in *F. prausnitzii* in colonic CD in twins with inflammatory bowel disease but a decrease in *F. prausnitzii* in ileal CD. Sokol et al. (Sokol et al., 2008a) also found that *F. prausnitzii* presented a reduction in resected ileal mucosa and was associated with endoscopic recurrence in CD.

In addition to *Faecalibacterium*, among Firmicutes, some other genera were overrepresented in controls relative to the CD. In this studies, CD patient microbiomes were dominated by depletion of genera belonging to Clostridiales relative to controls, particularly *Faecalibacterium*, *Ruminococcus_torques_group*, *Roseburia*, *Erysipelotrichaceae_UCG_003*, *Dorea*, *Eubacterium_hallii_group*, *Lachnospiraceae_NK4A136_group*, *Lachnospiraceae_UCG_004*, *Agathobacter*, *Phascolarctobacterium and Fusicatenibacter.* In addition to that, the beta diversity results showed a significant difference in the microbiome composition in CD compared with controls. This may indicate the potential role of dysbiosis in the pathogenesis of CD. Future studies should explore the connection of microbial alteration with immunological and metabolic changes in active and inactive disease stages to explain the complex relationship between microbiota and CD, specifically at the species level.

Several studies have reported the gut microbial imbalances for CD; however, the imbalance at the level of different phylum is variable across studies (Frank et al., 2007; Nishida et al., 2018; Franzosa et al., 2019; Vester-Andersen et al., 2019a). This

imbalance could likely reflect heterogeneity in the microbiome associated with the disease. A strength of this study is the focus on the mucosa-associated microbiome, which has definite advantages compared to other studies. This mucosa-associated microbiota may play a more crucial role than faecal microbiota in CD pathogenesis (Willing et al., 2009). Most studies used faecal sampling for measuring the microbial diversity and composition in CD patients due to its high availability and stability; however, the bacteria associated with the mucosa may be more likely to play a role in CD. The faecal microbiota does not represent the adherent microbial population and mucosa-associated bacteria.

In contrast, mucosal biopsy samples and bowel resected mucosal tissue samples harbour adherent microbial populations, mucosa-associated bacteria and luminal bacteria. And the previous study has revealed a significant difference between faecal and mucosal microbial community composition (Eckburg et al., 2005). he most positive side of our analysis is mucosal samples instead of faecal materials. Another major strength of our study is the demonstration of the disease along the course, stages and activity of disease such as early active aphthous ulcers, active and inactive stage of CD.

However, this study was limited by the relatively small sample size compared among the disease stages and tissue types in CD and controls, which might have resulted in a poor statistical power to identify minimal but possibly significant microbial alterations in different disease stages and tissue types. In addition to that, due to the high inconsistency of microbial community in some disease stages (CD and controls) and tissue types, the relative abundances of specific bacteria between groups could not reach statistical significance after adjusting for FDR. Prospective cohort studies with a bigger sample size on treatment-naive patients may provide a clear view of this

particular point. We have been able to show the shifts in bacterial diversity and abundance at different disease stages between CD and controls. Nevertheless, we have also observed microbial alterations associated with changes in disease stages, highlighting the importance of microbiome analysis in monitoring the disease activity in CD.

2.6 Conclusion

In conclusion, we present a comprehensive analysis of mucosal microbiota in different disease stages and tissue types in patients with CD and controls. There was no single dominant bacterium found in this study. The negative correlation of *Faecalibacterum* along with the positive association of *Escherichia-Shigella* with CD suggested that the inflamed tissues and the diarrhoeal fluid with all its inflammatory mediators make it difficult for *Faecalibacterum* to survive and create a favourable niche for *Escherichia-Shigella* that might play a pathogenic role for the initiation of CD lesions. As *Escherichia-Shigella* changed the most in CD, we intended to investigate it further, particularly *E. coli*.

2.7 References

Abraham, C., and Cho, J. (2009). Interleukin-23/Th17 pathways and inflammatory bowel disease. *Inflammatory bowel diseases* 15(7), 1090-1100.

Ajayi, T.A., Innes, C.L., Grimm, S.A., Rai, P., Finethy, R., Coers, J., et al. (2019). Crohn's disease IRGM risk alleles are associated with altered gene expression in human tissues. *American journal of physiology. Gastrointestinal and liver physiology* 316(Goodhand et al.), G95-G105. doi: 10.1152/ajpgi.00196.2018.

Alam, M.T., Amos, G.C.A., Murphy, A.R.J., Murch, S., Wellington, E.M.H., and Arasaradnam, R.P. (2020). Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels. *Gut Pathogens* 12(Goodhand et al.), 1. doi: 10.1186/s13099-019-0341-6.

Andoh, A. (2016). Physiological role of gut microbiota for maintaining human health. *Digestion* 93(3), 176-181.

Atarashi, K., Tanoue, T., Ando, M., Kamada, N., Nagano, Y., Narushima, S., et al. (2015). Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial Cells. *Cell* 163(Polito et al.), 367-380. doi: 10.1016/j.cell.2015.08.058.

Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., et al. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature genetics* 40(8), 955-962. doi: 10.1038/ng.175.

Boland, K., Bedrani, L., Turpin, W., Kabakchiev, B., Stempak, J., Borowski, K., et al. (2020). Persistent Diarrhea in Patients With Crohn's Disease After Mucosal Healing Is Associated With Lower Diversity of the Intestinal Microbiome and Increased Dysbiosis. *Clin Gastroenterol Hepatol*. doi: 10.1016/j.cgh.2020.03.044.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37(8), 852-857. doi: 10.1038/s41587-019-0209-9.

Boukercha, A., Mesbah-Amroun, H., Bouzidi, A., Saoula, H., Nakkemouche, M., Roy, M., et al. (2015). NOD2/CARD15 gene mutations in North Algerian patients with inflammatory bowel disease. *World journal of gastroenterology* 21(25), 7786-7794. doi: 10.3748/wjg.v21.i25.7786.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13(7), 581-583. doi: 10.1038/nmeth.3869.

Carrière, J., Darfeuille-Michaud, A., and Nguyen, H.T.T. (2014). Infectious etiopathogenesis of Crohn's disease. *World journal of gastroenterology* 20(34), 12102-12117. doi: 10.3748/wjg.v20.i34.12102.

Carvalho, F.A., Barnich, N., Sivignon, A., Darcha, C., Chan, C.H., and Stanners, C.P. (2009). Crohn's disease adherent-invasive Escherichia coli colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM. *J Exp Med* 206. doi: 10.1084/jem.20090741.

Chong, J., Liu, P., Zhou, G., and Xia, J. (2020). Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nature Protocols* 15(3), 799-821. doi: 10.1038/s41596-019-0264-1.

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

Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., et al. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *science* 314(5804)**,** 1461-1463.

Dunn, O.J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics* 6(3)**,** 241-252. doi: 10.1080/00401706.1964.10490181.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., et al. (2005). Diversity of the human intestinal microbial flora. *Science* 308(5728)**,** 1635-1638. doi: 10.1126/science.1110591.

Ellinghaus, D., Jostins, L., Spain, S.L., Cortes, A., Bethune, J., Han, B., et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nature genetics* 48(5)**,** 510-518. doi: 10.1038/ng.3528.

Fava, F., and Danese, S. (2011). Intestinal microbiota in inflammatory bowel disease: friend of foe? *World journal of gastroenterology: WJG* 17(5)**,** 557.

Frank, D.N., St. Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences* 104(34)**,** 13780-13785. doi: 10.1073/pnas.0706625104.

Franzosa, E.A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H.J., Reinker, S., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature microbiology* 4(Polito et al.)**,** 293-305.

Fujimoto, T., Imaeda, H., Takahashi, K., Kasumi, E., Bamba, S., Fujiyama, Y., et al. (2013). Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease. *J Gastroenterol Hepatol* 28(4)**,** 613-619. doi: 10.1111/jgh.12073.

Galazzo, G., Tedjo, D.I., Wintjens, D.S.J., Savelkoul, P.H.M., Masclee, A.A.M., Bodelier, A.G.L., et al. (2019). Faecal Microbiota Dynamics and their Relation to Disease Course in Crohn's Disease. *J Crohns Colitis* 13(10)**,** 1273-1282. doi: 10.1093/ecco-jcc/jjz049.

Gevers, D., Kugathasan, S., Denson, L.A., Vazquez-Baeza, Y., Treuren, W., and Ren, B. (2014). The treatment-naive microbiome in new-onset Crohn’s disease. *Cell Host Microbe* 15. doi: 10.1016/j.chom.2014.02.005.

Hall, L.J., Walshaw, J., and Watson, A.J.M. (2014). Gut Microbiome in New-Onset Crohn's Disease. *Gastroenterology* 147(4)**,** 932-934. doi: https://doi.org/10.1053/j.gastro.2014.08.014.

He, Q., Gao, Y., Jie, Z., Yu, X., Laursen, J.M., Xiao, L., et al. (2017). Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. *Gigascience* 6(7)**,** gix050.

Hedin, C.R., McCarthy, N.E., Louis, P., Farquharson, F.M., McCartney, S., Taylor, K., et al. (2014). Altered intestinal microbiota and blood T cell phenotype are shared by patients with Crohn's disease and their unaffected siblings. *Gut* 63(10)**,** 1578. doi: 10.1136/gutjnl-2013-306226.

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G.D., et al. (2013). Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PloS one* 8(6)**,** e66019.

Iida, T., Onodera, K., and Nakase, H. (2017). Role of autophagy in the pathogenesis of inflammatory bowel disease. *World journal of gastroenterology* 23(11)**,** 1944-1953. doi: 10.3748/wjg.v23.i11.1944.

John, G., Hegarty, J.P., Yu, W., Berg, A., Pastor, D.M., Kelly, A.A., et al. (2011). NKX2-3 variant rs11190140 is associated with IBD and alters binding of NFAT. *Mol Genet Metab* 104(1-2)**,** 174-179. doi: 10.1016/j.ymgme.2011.06.023.

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., et al. (2011a). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60(5), 631-637.

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., et al. (2011b). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60(5), 631-637. doi: 10.1136/gut.2010.223263.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *nature* 491(7422), 119.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30(14), 3059-3066. doi: 10.1093/nar/gkf436.

Kennedy, N.A., Lamb, C.A., Berry, S.H., Walker, A.W., Mansfield, J., Parkes, M., et al. (2018). The Impact of NOD2 Variants on Fecal Microbiota in Crohn's Disease and Controls Without Gastrointestinal Disease. *Inflammatory bowel diseases* 24(3), 583-592. doi: 10.1093/ibd/izx061.

Kim, S., Eun, H.S., and Jo, E.-K. (2019). Roles of Autophagy-Related Genes in the Pathogenesis of Inflammatory Bowel Disease. *Cells* 8(Goodhand et al.), 77. doi: 10.3390/cells8010077.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic acids research* 41(Goodhand et al.), e1-e1.

Kowalska-Duplaga, K., Gosiewski, T., Kapusta, P., Sroka-Oleksiak, A., Wędrychowicz, A., Pieczarkowski, S., et al. (2019). Differences in the intestinal microbiome of healthy children and patients with newly diagnosed Crohn's disease. *Scientific Reports* 9(Goodhand et al.), 18880. doi: 10.1038/s41598-019-55290-9.

Lassen, K.G., and Xavier, R.J. (2017). Genetic control of autophagy underlies pathogenesis of inflammatory bowel disease. *Mucosal immunology* 10(3), 589-597. doi: 10.1038/mi.2017.18.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., et al. (2008). Evolution of mammals and their gut microbes. *science* 320(5883), 1647-1651.

Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics* 47(9), 979-986. doi: 10.1038/ng.3359.

Loh, G., and Blaut, M. (2012). Role of commensal gut bacteria in inflammatory bowel diseases. *Gut Microbes* 3(6), 544-555. doi: 10.4161/gmic.22156.

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., Duncan, S.H., Flint, H.J., et al. (2014a). Mucosa-associated Faecalibacterium prausnitzii and Escherichia coli co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int J Med Microbiol* 304(3-4), 464-475. doi: 10.1016/j.ijmm.2014.02.009.

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., Duncan, S.H., Flint, H.J., et al. (2014b). Mucosa-associated Faecalibacterium prausnitzii and Escherichia coli co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *International journal of medical microbiology* 304(3-4), 464-475.

Lozupone, C., and Knight, R. (2005). UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology* 71(12), 8228-8235. doi: 10.1128/aem.71.12.8228-8235.2005.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 73(5)**,** 1576-1585. doi: 10.1128/aem.01996-06.

Ma, H.-Q., Yu, T.-T., Zhao, X.-J., Zhang, Y., and Zhang, H.-J. (2018). Fecal microbial dysbiosis in Chinese patients with inflammatory bowel disease. *World journal of gastroenterology* 24(13)**,** 1464-1477. doi: 10.3748/wjg.v24.i13.1464.

Man, S.M., Kaakoush, N.O., and Mitchell, H.M. (2011). The role of bacteria and pattern-recognition receptors in Crohn's disease. *Nature reviews gastroenterology and hepatology* 8(3)**,** 152-168.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., et al. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55(Polito et al.)**,** 205-211. doi: 10.1136/gut.2005.073817.

Martinez-Medina, M., Aldeguer, X., Gonzalez-Huix, F., Acero, D., and Garcia-Gil, J.L. (2006). Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction-denaturing gradient gel electrophoresis. *Inflammatory bowel diseases* 12(12)**,** 1136-1145.

Martinez-Medina, M.X., Aldeguer, M., Lopez-Siles, F., González-Huix, C., López-Oliu, G., and Dahbi, G. (2009). molecular diversity of Escherichia coli in the human gut: new ecological evidence supporting the role of adherent-invasive E. coli (AIEC) in Crohn's disease. *Inflamm Bowel Dis* 15. doi: 10.1002/ibd.20860.

Matsuoka, K., and Kanai, T. (Year). "The gut microbiota and inflammatory bowel disease", in: *Seminars in immunopathology*: Springer), 47-55.

Meggyesi, N., Kiss, L.S., Koszarska, M., Bortlik, M., Duricova, D., Lakatos, L., et al. (2010). NKX2-3 and IRGM variants are associated with disease susceptibility to IBD in Eastern European patients. *World Journal of Gastroenterology: WJG* 16(41)**,** 5233.

Nedelsky, N., Kuballa, P., Castoreno, A.B., and Xavier, R.J. (2019). Inflammatory Bowel Disease at the Intersection of Autophagy and Immunity: Insights from Human Genetics. *Molecular Genetics of Inflammatory Bowel Disease***,** 305-328. doi: 10.1007/978-3-030-28703-0_14.

Nishida, A., Inoue, R., Inatomi, O., Bamba, S., Naito, Y., and Andoh, A. (2018). Gut microbiota in the pathogenesis of inflammatory bowel disease. *Clinical journal of gastroenterology* 11(Goodhand et al.)**,** 1-10.

O'Brien, C.L., Pavli, P., Gordon, D.M., and Allison, G.E. (2014). Detection of bacterial DNA in lymph nodes of Crohn's disease patients using high throughput sequencing. *Gut*. doi: 10.1136/gutjnl-2013-305320.

Ott, S., Musfeldt, M., Wenderoth, D., Hampe, J., Brant, O., Fölsch, U., et al. (2004). Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 53(5)**,** 685-693.

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., et al. (2017). A microbial signature for Crohn's disease. *Gut* 66(5)**,** 813-822.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* 5(3)**,** e9490. doi: 10.1371/journal.pone.0009490.

Prosberg, M., Bendtsen, F., Vind, I., Petersen, A.M., and Gluud, L.L. (2016). The association between the gut microbiota and the inflammatory bowel disease activity: a systematic review and meta-analysis. *Scandinavian Journal of Gastroenterology* 51(12)**,** 1407-1415. doi: 10.1080/00365521.2016.1216587.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41(D1)**,** D590-D596. doi: 10.1093/nar/gks1219.

Ringel, Y., Maharshak, N., Ringel-Kulka, T., Wolber, E.A., Sartor, R.B., and Carroll, I.M. (2015). High throughput sequencing reveals distinct microbial populations within the mucosal and luminal niches in healthy individuals. *Gut Microbes* 6(3), 173-181. doi: 10.1080/19490976.2015.1044711.

Sartor, R.B. (2008). Microbial influences in inflammatory bowel diseases. *Gastroenterology* 134(Polito et al.), 577-594. doi: 10.1053/j.gastro.2007.11.059.

Scaldaferri, F., Gerardi, V., Lopetuso, L.R., Del Zompo, F., Mangiola, F., Boškoski, I., et al. (2013). Gut microbial flora, prebiotics, and probiotics in IBD: their current usage and utility. *Biomed Res Int* 2013, 435268. doi: 10.1155/2013/435268.

Scarpellini, E., Ianiro, G., Attili, F., Bassanelli, C., De Santis, A., and Gasbarrini, A. (2015). The human gut microbiota and virome: Potential therapeutic implications. *Digestive and Liver Disease* 47(12), 1007-1012.

Schmitz, J.M., Tonkonogy, S.L., Dogan, B., Leblond, A., Whitehead, K.J., Kim, S.C., et al. (2019). Murine adherent and invasive E. coli induces chronic inflammation and immune responses in the small and large intestines of monoassociated IL-10-/-mice independent of long polar fimbriae adhesin A. *Inflammatory bowel diseases* 25(5), 875-885.

Seksik, P., Rigottier-Gois, L., Gramet, G., Sutren, M., Pochart, P., Marteau, P., et al. (2003). Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* 52(Polito et al.), 237-242. doi: 10.1136/gut.52.2.237.

Sha, S., Xu, B., Wang, X., Zhang, Y., Wang, H., Kong, X., et al. (2013). The biodiversity and composition of the dominant fecal microbiota in patients with inflammatory bowel disease. *Diagnostic Microbiology and Infectious Disease* 75(3), 245-251. doi: https://doi.org/10.1016/j.diagmicrobio.2012.11.022.

Sheehan, D., Moran, C., and Shanahan, F. (2015). The microbiota in inflammatory bowel disease. *Journal of gastroenterology* 50(5), 495-507.

Sidiq, T., Yoshihama, S., Downs, I., and Kobayashi, K.S. (2016). Nod2: A Critical Regulator of Ileal Microbiota and Crohn's Disease. *Frontiers in Immunology* 7(367). doi: 10.3389/fimmu.2016.00367.

Simon, T.G., Van Der Sloot, K.W.J., Chin, S.B., Joshi, A.D., Lochhead, P., Ananthakrishnan, A.N., et al. (2018). IRGM Gene Variants Modify the Relationship Between Visceral Adipose Tissue and NAFLD in Patients With Crohn's Disease. *Inflammatory bowel diseases* 24(10), 2247-2257. doi: 10.1093/ibd/izy128.

Sivanesan, D., Beauchamp, C., Quinou, C., Lee, J., Lesage, S., Chemtob, S., et al. (2016). IL23R (Interleukin 23 Receptor) Variants Protective against Inflammatory Bowel Diseases (IBD) Display Loss of Function due to Impaired Protein Stability and Intracellular Trafficking. *J Biol Chem* 291(16), 8673-8685. doi: 10.1074/jbc.M116.715870.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L.G., Gratadoux, J.-J., et al. (2008a). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences* 105(43), 16731-16736.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L.G., Gratadoux, J.J., et al. (2008b). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 105(43), 16731-16736. doi: 10.1073/pnas.0804812105.

Sokol, H., Seksik, P., Furet, J.P., Firmesse, O., Nion-Larmurier, I., Beaugerie, L., et al. (2009). Low counts of Faecalibacterium prausnitzii in colitis microbiota. *Inflamm Bowel Dis* 15(8), 1183-1189. doi: 10.1002/ibd.20903.

Spalinger, M.R., McCole, D.F., Rogler, G., and Scharl, M. (2016). Protein tyrosine phosphatase non-receptor type 2 and inflammatory bowel disease. *World journal of gastroenterology* 22(3)**,** 1034-1044. doi: 10.3748/wjg.v22.i3.1034.

Tlaskalová-Hogenová, H., Štěpánková, R., Kozáková, H., Hudcovic, T., Vannucci, L., Tučková, L., et al. (2011). The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. *Cellular & molecular immunology* 8(Polito et al.)**,** 110-120.

Toiyama, Y., Araki, T., Yoshiyama, S., Hiro, J.-i., Miki, C., and Kusunoki, M. (2006). The expression patterns of Toll-like receptors in the ileal pouch mucosa of postoperative ulcerative colitis patients. *Surgery today* 36(3)**,** 287-290.

Vester-Andersen, M., Mirsepasi-Lauridsen, H., Prosberg, M., Mortensen, C., Träger, C., Skovsen, K., et al. (2019a). Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. *Scientific reports* 9(Goodhand et al.)**,** 1-10.

Vester-Andersen, M.K., Mirsepasi-Lauridsen, H.C., Prosberg, M.V., Mortensen, C.O., Träger, C., Skovsen, K., et al. (2019b). Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. *Scientific reports* 9(Goodhand et al.)**,** 13473-13473. doi: 10.1038/s41598-019-49833-3.

Walker, A.W., Sanderson, J.D., Churcher, C., Parkes, G.C., Hudspith, B.N., Rayment, N., et al. (2011). High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiology* 11(Goodhand et al.)**,** 7. doi: 10.1186/1471-2180-11-7.

Wang, W., Chen, L., Zhou, R., Wang, X., Song, L., Huang, S., et al. (2014). Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. *Journal of clinical microbiology* 52(Polito et al.)**,** 398-406. doi: 10.1128/JCM.01500-13.

Wang, Y., Gao, X., Ghozlane, A., Hu, H., Li, X., Xiao, Y., et al. (2018). Characteristics of Faecal Microbiota in Paediatric Crohn's Disease and Their Dynamic Changes During Infliximab Therapy. *J Crohns Colitis* 12(3)**,** 337-346. doi: 10.1093/ecco-jcc/jjx153.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5(Goodhand et al.)**,** 27. doi: 10.1186/s40168-017-0237-y.

Weiss, S.J., Xu, Z., Amir, A., Peddada, S., Bittinger, K., Gonzalez, A., et al. (2015). "Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data". PeerJ PrePrints).

Willing, B., Halfvarson, J., Dicksved, J., Rosenquist, M., Jarnerot, G., Engstrand, L., et al. (2009). Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis* 15(5)**,** 653-660. doi: 10.1002/ibd.20783.

Willing, B.P., Dicksved, J., Halfvarson, J., Andersson, A.F., Lucio, M., Zheng, Z., et al. (2010). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 139(6)**,** 1844-1854. e1841.

Wright, E.K., Kamm, M.A., Wagner, J., Teo, S.-M., Cruz, P.D., Hamilton, A.L., et al. (2017). Microbial Factors Associated with Postoperative Crohn's Disease Recurrence. *Journal of Crohn's and Colitis* 11(Polito et al.)**,** 191-203. doi: 10.1093/ecco-jcc/jjw136.

Xu, J., Liang, R., Zhang, W., Tian, K., Li, J., Chen, X., et al. (2020). Faecalibacterium prausnitzii-derived microbial anti-inflammatory molecule regulates intestinal integrity in diabetes mellitus mice via modulating tight junction protein expression. *Journal of diabetes* 12(3)**,** 224-236. doi: 10.1111/1753-0407.12986.

Xu, W.-D., Xie, Q.-B., Zhao, Y., and Liu, Y. (2015). Association of Interleukin-23 receptor gene polymorphisms with susceptibility to Crohn's disease: A meta-analysis. *Scientific Reports* 5(Goodhand et al.), 18584. doi: 10.1038/srep18584.

Yilmaz, B., Spalinger, M.R., Biedermann, L., Franc, Y., Fournier, N., Rossel, J.-B., et al. (2018). The presence of genetic risk variants within PTPN2 and PTPN22 is associated with intestinal microbiota alterations in Swiss IBD cohort patients. *PloS one* 13(7), e0199664-e0199664. doi: 10.1371/journal.pone.0199664.

Zhao, H., Xu, H., Chen, S., He, J., Zhou, Y., and Nie, Y. (2021). Systematic review and meta-analysis of the role of Faecalibacterium prausnitzii alteration in inflammatory bowel disease. *Journal of Gastroenterology and Hepatology* 36(Polito et al.), 320-328.

Zhernakova, A., Festen, E.M., Franke, L., Trynka, G., van Diemen, C.C., Monsuur, A.J., et al. (2008). Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *Am J Hum Genet* 82(5), 1202-1210. doi: 10.1016/j.ajhg.2008.03.016.

2.8 Supplemental Materials

Supplemental tables-

Supplemental Table 2.1. Overview and data integrity check of the QIIME data:

Data Integrity Check	
Data type:	OTU abundance table
File format:	text
OTU annotation:	SILVA
OTU number:	14497
OTUs with ≥ 2 counts:	3171
Number of experimental factors:	5 [discrete: 5 continuous: 0]
Contain metadata with replicates:	Yes
Total read counts:	39163446
Average counts per sample:	445039
Maximum counts per sample:	1807708
Minimum counts per sample:	38739
Phylogenetic tree uploaded:	Yes
Number of samples in metadata:	88
Number of samples in OTU table:	88
Sample names match (metadata vs. OTU table):	Yes
Number of sample names matched (metadata vs. OTU table):	88
Number of samples that will be processed:	88

Supplemental Table 2.2. Alpha diversity (observed OTUs): CD vs controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test

Group 1	**Group 2**	**H**	**p-value**	**q-value**
CD (n=72)	Controls (n=16)	16.0674064	6.11E-05	6.11E-05

Supplemental Table 2.3. Alpha diversity (Shannon diversity index): CD vs controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test.

Group 1	Group 2	H	p-value	q-value
CD (n=72)	Controls (n=16)	4.225	0.0398	0.0398

Supplemental Table 2.4. Alpha diversity (observed OTUs): different disease stages in CD vs normal biopsies in controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test

Group 1	Group 2	H	p-value	q-value
AU-early-active-CD (n=12)	Normal biopsies (n=16)	8.8276	0.003	0.01
AU-early-active-CD (n=12)	Active-SB-CD (n=13)	0.0740	0.785	0.84
AU-early-active-CD (n=12)	Active-colon-CD (n=7)	0.9455	0.330	0.45
AU-early-active-CD (n=12)	Inactive-SB-CD (n=18)	0.4881	0.484	0.55
AU-early-active-CD (n=12)	Inactive-colon-CD (n=22)	1.3301	0.248	0.37
Normal biopsies (n=16)	Active-SB-CD (n=13)	13.2481	0.0003	0.004
Normal biopsies (n=16)	Active-colon-CD (n=7)	9.2468	0.002	0.01
Normal biopsies (n=16)	Inactive-SB-CD (n=18)	7.2440	0.007	0.02
Normal biopsies (n=16)	Inactive-colon-CD (n=22)	8.7422	0.003	0.01
Active-SB-CD (n=13)	Active-colon-CD (n=7)	0.6923	0.405	0.50
Active-SB-CD (n=13)	Inactive-SB-CD (n=18)	1.7992	0.179	0.29
Active-SB-CD (n=13)	Inactive-colon-CD (n=22)	2.2574	0.133	0.24
Active-colon-CD (n=7)	Inactive-SB-CD (n=18)	2.4771	0.115	0.24
Active-colon-CD (n=7)	Inactive-colon-CD (n=22)	3.5567	0.059	0.14
Inactive-SB-CD (n=18)	Inactive-colon-CD (n=22)	0.0067	0.935	0.93

Supplemental Table 2.5. Alpha diversity (Shannon) at different disease stages in CD compared to controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test

Group 1	Group 2	H	p-value	q-value
AU-early-active-CD (n=12)	Normal biopsies (n=16)	3.448	0.063	0.278
AU-early-active-CD (n=12)	Active-SB-CD (n=13)	0.296	0.586	0.586
AU-early-active-CD (n=12)	Active-colon-CD (n=7)	0.579	0.447	0.521
AU-early-active-CD (n=12)	Inactive-SB-CD (n=18)	1.507	0.220	0.366
AU-early-active-CD (n=12)	Inactive-colon-CD (n=22)	0.812	0.368	0.521
Normal biopsies (n=16)	Active-SB-CD (n=13)	5.608	0.018	0.268
Normal biopsies (n=16)	Active-colon-CD (n=7)	4.290	0.038	0.278
Normal biopsies (n=16)	Inactive-SB-CD (n=18)	0.430	0.512	0.549
Normal biopsies (n=16)	Inactive-colon-CD (n=22)	1.850	0.174	0.326
Active-SB-CD (n=13)	Active-colon-CD (n=7)	0.567	0.452	0.521
Active-SB-CD (n=13)	Inactive-SB-CD (n=18)	2.827	0.093	0.278
Active-SB-CD (n=13)	Inactive-colon-CD (n=22)	1.865	0.172	0.326
Active-colon-CD (n=7)	Inactive-SB-CD (n=18)	3.081	0.079	0.278
Active-colon-CD (n=7)	Inactive-colon-CD (n=22)	2.184	0.139	0.326
Inactive-SB-CD (n=18)	Inactive-colon-CD (n=22)	0.579	0.447	0.521

Supplemental Table 2.6. Alpha diversity (observed OTUs) in different tissue types in CD compared with controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test

Group 1	Group 2	H	p-value	q-value
AU-early-active-CD (n=12)	Colon-CD (n=29)	0.34	0.55	0.83
AU-early-active-CD (n=12)	Normal biopsies (n=16)	8.82	0.003	0.005
AU-early-active-CD (n=12)	SB-CD (n=31)	0.09	0.75	0.90
Colon-CD (n=29)	Normal biopsies (n=16)	11.90	0.0006	0.001
Colon-CD (n=29)	SB-CD (n=31)	0.004	0.94	0.946
Normal biopsies (n=16)	SB-CD (n=31)	13.06	0.0003	0.001

Supplemental Table 2.7. Alpha diversity (Shannon): different tissue types in CD vs normal biopsies in controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test

Group 1	Group 2	H	p-value	q-value
AU-early-active-CD (n=12)	Colon-CD (n=29)	0.210	0.64	0.69
AU-early-active-CD (n=12)	Normal biopsies (n=16)	3.448	0.06	0.20
AU-early-active-CD (n=12)	SB-CD (n=31)	0.265	0.60	0.69
Colon-CD (n=29)	Normal biopsies (n=16)	3.333	0.06	0.22
Colon-CD (n=29)	SB-CD (n=31)	0.154	0.69	0.69
Normal biopsies (n=16)	SB-CD (n=31)	2.686	0.10	0.20

Supplemental Table 2.8. Microbial patterns (beta diversity) in Crohn's disease and control groups based on Bray-Curtis index using Pairwise PERMANOVA analysis

Group 1	Group 2	Sample size	pseudo-F	p-value	q-value
CD	Controls	88	3.44885989	0.001	0.001

Supplemental Table 2.9. Microbial patterns (beta diversity) in Crohn's disease and control groups based on Jaccard index using Pairwise PERMANOVA analysis

Group 1	Group 2	Sample size	pseudo-F	p-value	q-value
CD	Controls	88	2.09795891	0.001	0.001

Supplemental Table 2.10. Microbial patterns (beta diversity) in Crohn's disease and control groups based on unweighted Unifrac index using Pairwise PERMANOVA analysis.

Group 1	Group 2	Sample size	pseudo-F	p-value	q-value
CD	Controls	88	2.22463036	0.003	0.003

Table 2.11. Microbial patterns (beta diversity) in Crohn's disease and controls based on weighted Unifrac index using Pairwise PERMANOVA analysis.

Group 1	Group 2	Sample size	Permutations	pseudo-F	p-value	q-value
CD	Controls	88	999	3.40383087	0.016	0.016

Supplemental Table 2.12. Microbial patterns (beta diversity) in different disease stages and tissue types in Crohn's disease and control groups based on Bray-Curtis index using Pairwise PERMANOVA analysis.

Group 1	Group 2	Sample size	pseudo-F	p-value	q-value
AU-early-active-CD	Normal biopsies	28	1.982	0.006	0.0225
AU-early-active-CD	Active-SB-CD	25	0.908	0.515	0.638
AU-early-active-CD	Active-colon-CD	19	1.554	0.072	0.135
AU-early-active-CD	Inactive-SB-CD	30	0.909	0.596	0.638
AU-early-active-CD	Inactive-colon-CD	34	0.883	0.593	0.638
Normal biopsies	Active-SB-CD	29	3.104	0.001	0.005
Normal biopsies	Active-colon-CD	23	3.841	0.001	0.005
Normal biopsies	Inactive-SB-CD	34	1.769	0.018	0.054
Normal biopsies	Inactive-colon-CD	38	2.878	0.001	0.005
Active-SB-CD	Active-colon-CD	20	1.179	0.241	0.361
Active-SB-CD	Inactive-SB-CD	31	1.643	0.047	0.100
Active-SB-CD	Inactive-colon-CD	35	0.761	0.71	0.71
Active-colon-CD	Inactive-SB-CD	25	2.127	0.03	0.075
Active-colon-CD	Inactive-colon-CD	29	1.520	0.098	0.163
Inactive-SB-CD	inactive-colon-CD	40	0.973	0.457	0.623

Supplemental Table 2.13. Microbial patterns (beta diversity) in different disease stages and tissue types in Crohn's disease and control groups based on Jaccard index using Pairwise PERMANOVA analysis.

Group 1	Group 2	Sample size	pseudo-F	p-value	q-value
AU-early-active-CD	Normal biopsies	28	1.60	0.006	0.018
AU-early-active-CD	Active-SB-CD	25	1.04	0.279	0.321
AU-early-active-CD	Active-colon-CD	19	1.34	0.032	0.053
AU-early-active-CD	Inactive-SB-CD	30	1.05	0.226	0.282
AU-early-active-CD	Inactive-colon-CD	34	1.11	0.212	0.282
Normal biopsies	Active-SB-CD	29	1.74	0.002	0.007
Normal biopsies	Active-colon-CD	23	1.98	0.001	0.007
Normal biopsies	Inactive-SB-CD	34	1.47	0.002	0.007
Normal biopsies	Inactive-colon-CD	38	1.93	0.001	0.007
Active-SB-CD	Active-colon-CD	20	1.34	0.018	0.033
Active-SB-CD	Inactive-SB-CD	31	1.09	0.139	0.208
Active-SB-CD	Inactive-colon-CD	35	0.95	0.598	0.598
Active-colon-CD	Inactive-SB-CD	25	1.42	0.016	0.033
Active-colon-CD	Inactive-colon-CD	29	1.51	0.009	0.022
Inactive-SB-CD	inactive-colon-CD	40	0.98	0.513	0.549

Supplemental Table 2.14. Microbial patterns (beta diversity) in different disease stages and tissue types in Crohn's disease and control groups based on unweighted Unifrac index using Pairwise PERMANOVA analysis.

Group 1	Group 2	Sample size	pseudo-F	p-value	q-value
AU-early-active-CD	Normal biopsies	28	1.96	0.006	0.02
AU-early-active-CD	Active-SB-CD	25	0.93	0.496	0.496
AU-early-active-CD	Active-colon-CD	19	1.10	0.309	0.370
AU-early-active-CD	Inactive-SB-CD	30	1.08	0.321	0.370
AU-early-active-CD	Inactive-colon-CD	34	1.28	0.139	0.23
Normal biopsies	Active-SB-CD	29	2.56	0.002	0.02
Normal biopsies	Active-colon-CD	23	2.04	0.007	0.026
Normal biopsies	Inactive-SB-CD	34	1.31	0.118	0.22
Normal biopsies	Inactive-colon-CD	38	1.88	0.005	0.026
Active-SB-CD	Active-colon-CD	20	1.67	0.046	0.100
Active-SB-CD	Inactive-SB-CD	31	1.84	0.018	0.054
Active-SB-CD	Inactive-colon-CD	35	1.55	0.047	0.100
Active-colon-CD	Inactive-SB-CD	25	1.26	0.18	0.245
Active-colon-CD	Inactive-colon-CD	29	1.29	0.163	0.244
Inactive-SB-CD	Inactive-colon-CD	40	1.00	0.446	0.477

Supplemental Table 2.15. Microbial patterns (beta diversity) in different disease stages and tissue types in Crohn's disease and control groups based on weighted Unifrac index using Pairwise PERMANOVA analysis.

Group 1	Group 2	Sample size	pseudo-F	p-value	q-value
AU-early-active-CD	Normal biopsies	28	2.52	0.02	0.06
AU-early-active-CD	Active-SB-CD	25	1.18	0.309	0.356
AU-early-active-CD	Active-colon-CD	19	1.97	0.09	0.141
AU-early-active-CD	Inactive-SB-CD	30	2.09	0.06	0.112
AU-early-active-CD	Inactive-colon-CD	34	1.41	0.218	0.297
Normal biopsies	Active-SB-CD	29	2.89	0.037	0.092
Normal biopsies	Active-colon-CD	23	5.24	0.001	0.015
Normal biopsies	Inactive-SB-CD	34	2.85	0.006	0.04
Normal biopsies	Inactive-colon-CD	38	2.84	0.018	0.06
Active-SB-CD	Active-colon-CD	20	0.77	0.481	0.515
Active-SB-CD	Inactive-SB-CD	31	2.54	0.045	0.096
Active-SB-CD	Inactive-colon-CD	35	0.63	0.594	0.594
Active-colon-CD	Inactive-SB-CD	25	3.75	0.008	0.04
Active-colon-CD	Inactive-colon-CD	29	1.92	0.094	0.141
Inactive-SB-CD	Inactive-colon-CD	40	1.19	0.271	0.338

Supplemental Table 2.16. Classical univariate statistical comparisons based on the comparative abundance of phylum in CD and controls (using Kruskal Wallis test and FDR correction)

Phylum	P-values	FDR	Statistics
Not Assigned	0.003	0.026	848
Proteobacteria	0.012	0.0654	808
Vertebrata	0.02	0.0903	784.5
Bacteroidota/ Bacteroidetes	0.03	0.107	379

Supplemental Table 2.17. Relative abundances of phylum Proteobacteria at disease stages (AU-early-CD, active and inactive-CD) and different tissue types (small bowel, aphthous ulcers, colon) between CD and controls.

Comparison (Proteobacteria)	Z	P.unadj	P.adj
active-colon-CD vs active-SB-CD	0.039	0.969	1
active-colon-CD vs AU-early-active-CD	1.333	0.183	1
active-SB-CD vs AU-early-active-CD	1.537	0.124	1
active-colon-CD vs inactive-colon-CD	0.319	0.750	1
active-SB-CD vs inactive-colon-CD	0.343	0.732	1
AU-early-active-CD vs inactive-colon-CD	-1.380	0.168	1
Active-colon-CD vs inactive-SB-CD	1.516	0.130	1
Active-SB-CD vs inactive-SB-CD	1.804	0.071	1
AU-early-active-CD vs inactive-SB-CD	0.111	0.912	1
inactive-colon-CD vs inactive-SB-CD	1.688	0.091	0.3990664
active-colon-CD vs normal biopsies	2.217	0.027	0.1238383
Active-SB-CD vs normal biopsies	2.641	0.008	1
AU-early-active-CD vs normal biopsies	0.972	0.331	0.1256474
Inactive-colon-CD vs normal biopsies	2.637	0.008	1
Inactive-SB-CD vs normal biopsies	0.956	0.337	1

Supplemental Table 2.18. Relative abundances of phylum Firmicutes at disease stages (AU-early-CD, active and inactive-CD) and different tissue types (small bowel, aphthous ulcers, colon) between CD and controls.

Comparison (Firmicutes)	Z	P.unadj	P.adj
Active-colon-CD vs active-SB-CD	-0.57	0.57	1
active-colon-CD vs AU-early-active-CD	-0.55	0.58	1
Active-SB-CD vs AU-early-active-CD	0.01	0.99	1
Active-colon-CD vs inactive-colon-CD	-1.54	0.12	1
Active-SB-CD vs inactive-colon-CD	-1.15	0.25	1
AU-early-active-CD vs inactive-colon-CD	-1.13	0.26	1
Active-colon-CD vs inactive-SB-CD	-1.44	0.15	1

Active-SB-CD vs inactive-SB-CD	-1.02	0.31	1
AU-early-active-CD - inactive-SB-CD	-1.01	0.31	1
Inactive-colon-CD vs inactive-SB-CD	0.09	0.93	1
Active-colon-CD vs normal biopsies	-1.55	0.12	1
Active-SB-CD vs normal biopsies	-1.16	0.25	1
AU-early-active-CD vs normal biopsies	-1.15	0.25	1
Inactive-colon-CD vs normal biopsies	-0.10	0.92	1
Inactive-SB-CD vs normal biopsies	-0.18	0.86	1

Supplemental Table 2.19. Relative abundances of phylum Bacteroidetes/ Bacteroidota at disease stages (AU-early-CD, active and inactive-CD) and different tissue types (small bowel, aphthous ulcers, colon) between CD and controls.

Comparison (Bacteroidetes)	**Z**	**P.unadj**	**P.adj**
Active-colon-CD vs active-SB-CD	-1.857	0.063	0.947
Active-colon-CD vs AU-early-active-CD	-1.691	0.090	1
Active-SB-CD vs AU-early-active-CD	0.166	0.867	1
Active-colon-CD vs inactive-colon-CD	-1.90	0.056	0.843
Active-SB-CD vs inactive-colon-CD	0.121	0.903	1
AU-early-active-CD vs inactive-colon-CD	-0.067	0.945	1
Active-colon-CD vs inactive-SB-CD	-1.214	0.224	1
Active-SB-CD vs inactive-SB-CD	0.906	0.364	1
AU-early-active-CD vs inactive-SB-CD	0.706	0.480	1
Inactive-colon-CD vs inactive-SB-CD	0.904	0.365	1
Active-colon-CD vs normal biopsies	-2.852	0.004	0.065
Active-SB-CD vs normal biopsies	-1.129	0.258	1
AU-early-active-CD vs normal biopsies	-1.279	0.200	1
Inactive-colon-CD vs normal biopsies	-1.412	0.157	1
Inactive-SB-CD vs normal biopsies	-2.187	0.028	0.430

Supplemental Table 2.20. Classical univariate statistical comparisons based on the comparative abundance of genera in CD and controls (using Kruskal Wallis test)

Serial	**Genus**	**P-values**	**FDR**	**Statistics**
1	Ruminococcus_gnavus_group	0.004	0.060	842
2	Flavonifractor	0.017	0.163	796
3	Escherichia_Shigella	0.018	0.169	794
4	Teleostei	0.022	0.181	784.5
5	Clostridium_innocuum_group	0.026	0.208	780
6	Klebsiella	0.019	0.174	762
7	Eubacterium	0.027	0.210	720
8	Peptostreptococcus	0.033	0.229	712
9	Asinibacterium	0.040	0.257	704

10	Megamonas	0.048	0.291	696
11	Phyllobacterium	0.030	0.213	512.5
12	Helicobacter	0.028	0.213	512
13	S5_A14a	0.002	0.051	504
14	Shuttleworthia	0.002	0.051	504
15	UCG_008	0.002	0.051	504
16	Howardella	0.015	0.159	486
17	Acidocella	0.0002	0.012	468
18	Muribaculaceae	0.0050	0.072	454
19	Allisonella	0.011	0.127	441
20	Prevotellaceae_NK3B31_group	0.0002	0.012	440
21	Holdemanella	0.016	0.163	427
22	vadinBE97	0.009	0.107	422
23	Bacteroides	0.086	0.405	417
24	Eubacterium_eligens_group	0.049	0.291	411.5
25	UCG_003	0.049	0.291	401
26	Butyricimonas	0.038	0.252	398.5
27	Eubacterium_hallii_group	0.035	0.238	383
28	Lachnospiraceae_UCG_001	0.003	0.060	379.5
29	Fusicatenibacter	0.030	0.213	378.5
30	Christensenellaceae_R_7_group	0.021	0.174	375
31	CAG_352	0.008	0.107	367
32	Ruminococcus_gauvreauii_group	0.013	0.139	365
33	Faecalibacterium	0.020	0.174	362
34	Phascolarctobacterium	0.012	0.136	350.5
35	CAG_56	0.002	0.051	331
36	Erysipelotrichaceae_UCG_003	0.007	0.096	330
37	Ruminococcus_torques_group	0.006	0.093	326
38	Marvinbryantia	0.002	0.051	319.5
39	Lachnospiraceae_ND3007_group	0.003	0.054	315
40	GCA_900066575	0.002	0.051	313
41	Agathobacter	0.003	0.054	313
42	Clostridia_UCG_014	0.0002	0.012	312
43	Blastococcus	4.46E-10	1.25E-07	288
44	Lachnospiraceae_FCS020_group	0.0008	0.025	283
45	Dorea	0.0004	0.017	255
46	Lachnospiraceae_NK4A136_group	0.0003	0.012	245
47	Roseburia	0.0003	0.012	245
48	Lachnospiraceae_UCG_004	0.0002	0.012	244

Supplemental Table 2.21. Relative abundances of *Escherichia-Shigella* genera at disease stages (AU-early-CD, active and inactive-CD) and different tissue types (small bowel, aphthous ulcers, colon) between CD and controls.

Comparison (Escherichia-Shigella)	Z	P.unadj	P.adj
Active-colon-CD vs active-SB-CD	0.620	0.535	1
Active-colon-CD vs AU-early-active-CD	1.867	0.062	0.92
Active-SB-CD vs AU-early-active-CD	1.491	0.136	1
Active-colon-CD vs inactive-colon-CD	1.269	0.205	1
Active-SB-CD vs inactive-colon-CD	0.743	0.458	1
AU-early-active-CD vs inactive-colon-CD	-0.940	0.347	1
Active-colon-CD vs inactive-SB-CD	2.298	0.022	0.32
Active-SB-CD vs inactive-SB-CD	2.013	0.044	0.66
AU-early-active-CD vs inactive-SB-CD	0.365	0.715	1
Inactive-colon-CD vs inactive-SB-CD	1.489	0.137	1
Active-colon-CD vs normal biopsies	2.790	0.005	0.078
Active-SB-CD vs normal biopsies	2.608	0.009	0.13
AU-early-active-CD vs normal biopsies	0.987	0.324	1
Inactive-colon-CD vs normal biopsies	2.173	0.030	0.446
Inactive-SB-CD vs normal biopsies	0.701	0.483	1

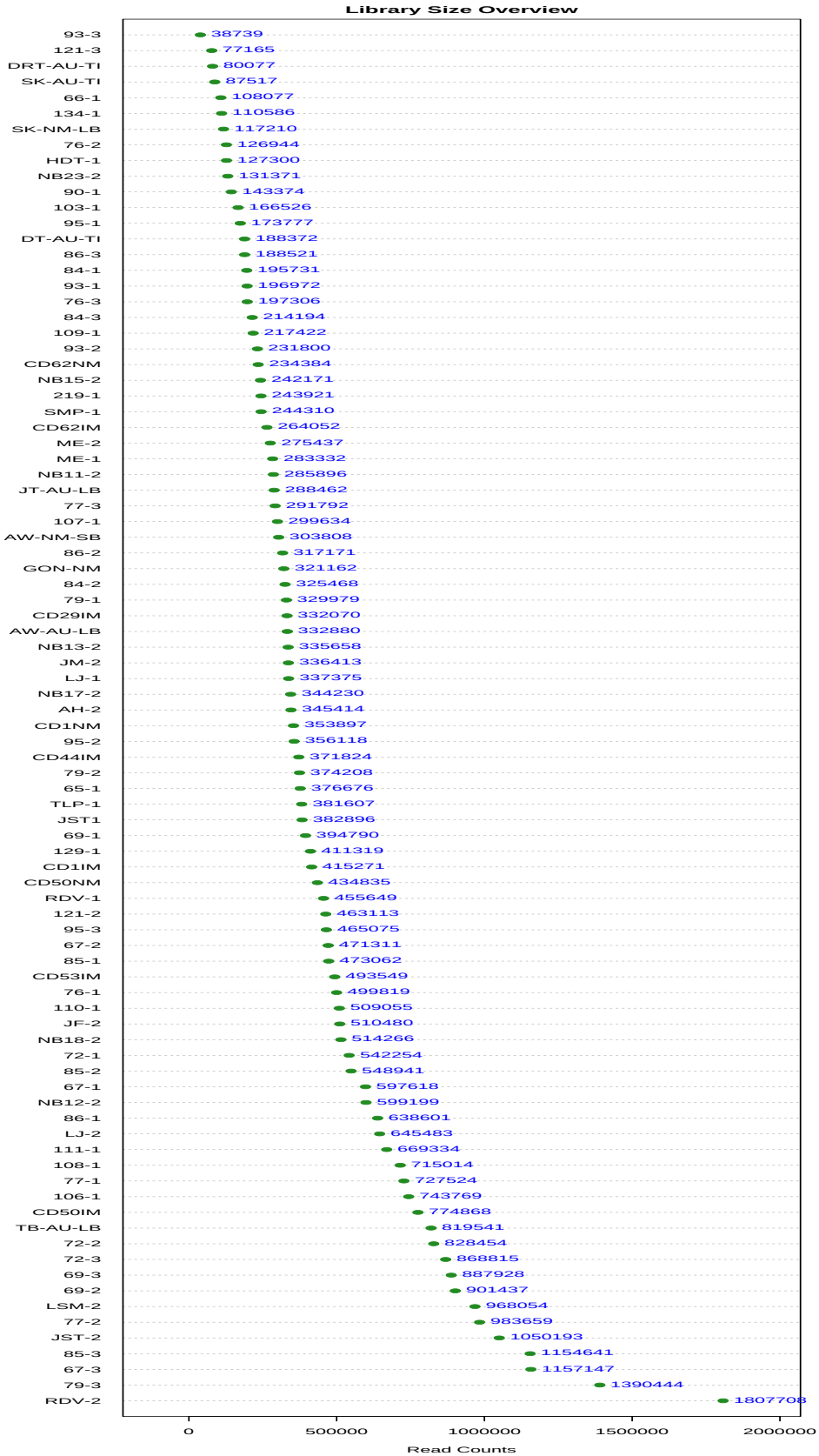
Supplemental Table 2.22. Relative abundances of *Faecalibacterium* at disease stages (AU-early-CD, active and inactive-CD) and different tissue types (small bowel, aphthous ulcers, colon) between CD and controls.

Comparison (*Faecalibacterium*)	Z	P.unadj	P.adj
Active-colon-CD vs active-SB-CD	-0.274	0.784	1
Active-colon-CD vs AU-early-active-CD	-0.684	0.494	1
Active-SB-CD vs AU-early-active-CD	-0.492	0.622	1
Active-colon-CD vs inactive-colon-CD	-0.315	0.753	1
Active-SB-CD vs inactive-colon-CD	-0.024	0.981	1
AU-early-active-CD vs inactive-colon-CD	0.526	0.599	1
Active-colon-CD vs inactive-SB-CD	-1.867	0.062	0.928
Active-SB-CD vs inactive-SB-CD	-1.932	0.053	0.800
AU-early-active-CD vs inactive-SB-CD	-1.358	0.174	1.000
Inactive-colon-CD vs inactive-SB-CD	-2.186	0.029	0.432
Active-colon-CD vs normal biopsies	-2.147	0.032	0.477
Active-SB-CD vs normal biopsies	-2.262	0.024	0.356
AU-early-active-CD vs normal biopsies	-1.696	0.090	1
Inactive-colon-CD vs normal biopsies	-2.545	0.011	0.164
Inactive-SB-CD vs normal biopsies	-0.411	0.681	1.000

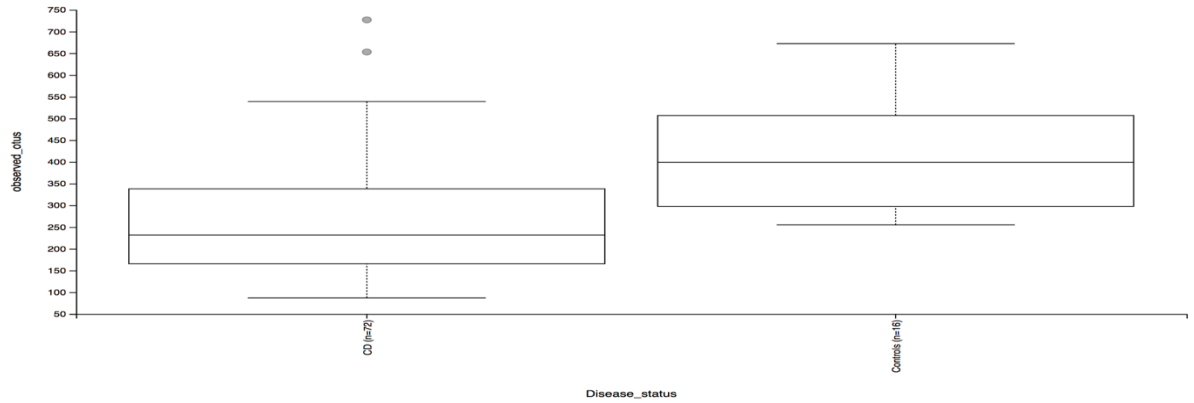
Supplemental Table 2.23. Relative abundances of *Bacteroides* at disease stages (AU-early-CD, active and inactive-CD) and different tissue types (small bowel, aphthous ulcers, colon) between CD and controls.

Comparison (Bacteroides)	Z	P.unadj	P.adj
Active-colon-CD vs active-SB-CD	-2.050	0.040	0.606
Active-colon-CD vs AU-early-active-CD	-1.593	0.111	1
Active-SB-CD vs AU-early-active-CD	0.508	0.612	1
Active-colon-CD vs inactive-colon-CD	-2.013	0.044	0.662
Active-SB-CD vs inactive-colon-CD	0.250	0.802	1
AU-early-active-CD vs inactive-colon-CD	-0.322	0.747	1
Active-colon-CD vs inactive-SB-CD	-1.291	0.197	1
Active-SB-CD vs inactive-SB-CD	1.060	0.289	1
AU-early-active-CD vs inactive-SB-CD	0.490	0.624	1
Inactive-colon-CD vs inactive-SB-CD	0.939	0.348	1
Active-colon-CD vs normal biopsies	-2.617	0.009	0.133
Active-SB-CD vs normal biopsies	-0.602	0.547	1
AU-early-active-CD vs normal biopsies	-1.121	0.262	1
Inactive-colon-CD vs normal biopsies	-0.951	0.342	1
Inactive-SB-CD vs normal biopsies	-1.778	0.075	1

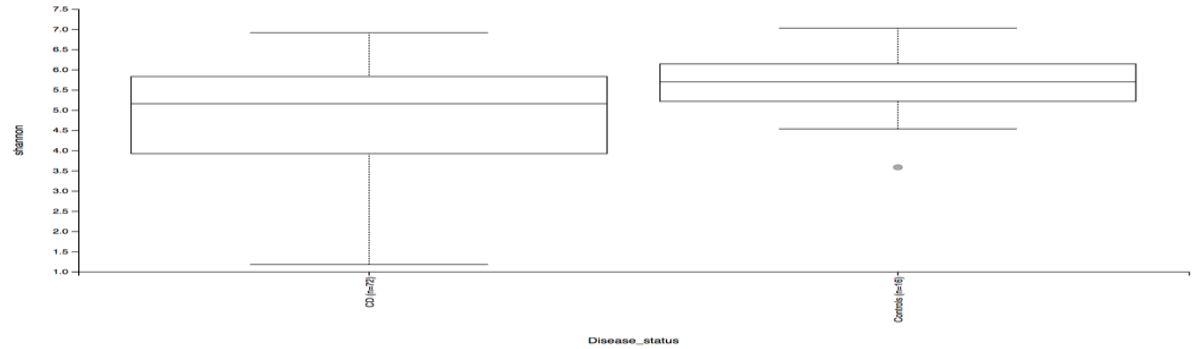
Supplemental Figures-



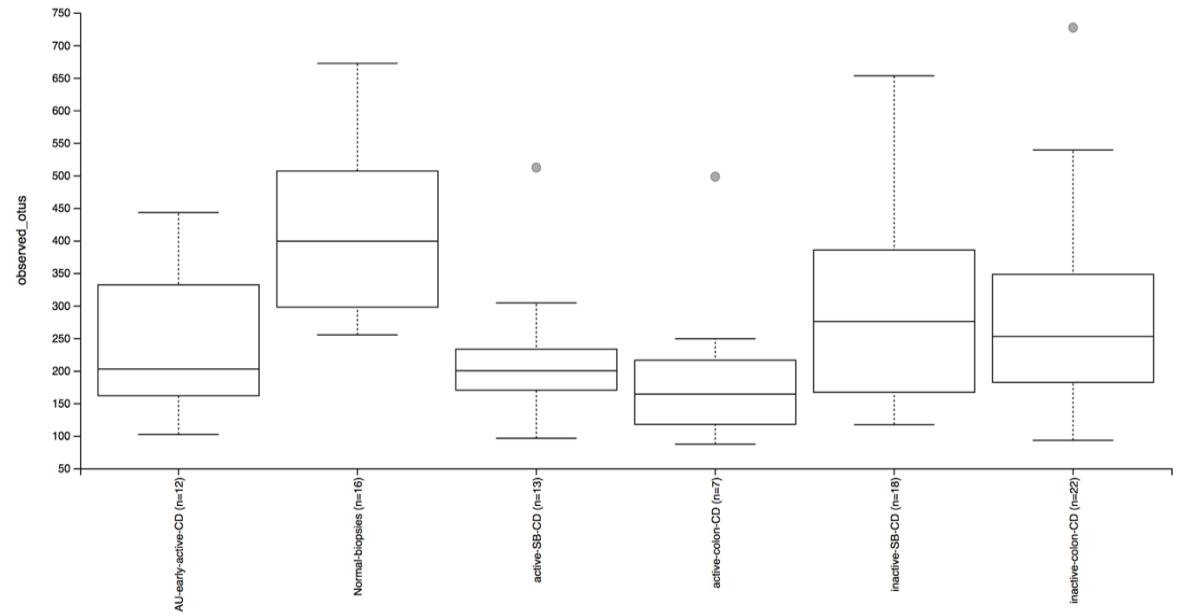
Supplemental Figure 2.1. Library size overview of the 16s QIIME data for microbiome analysis



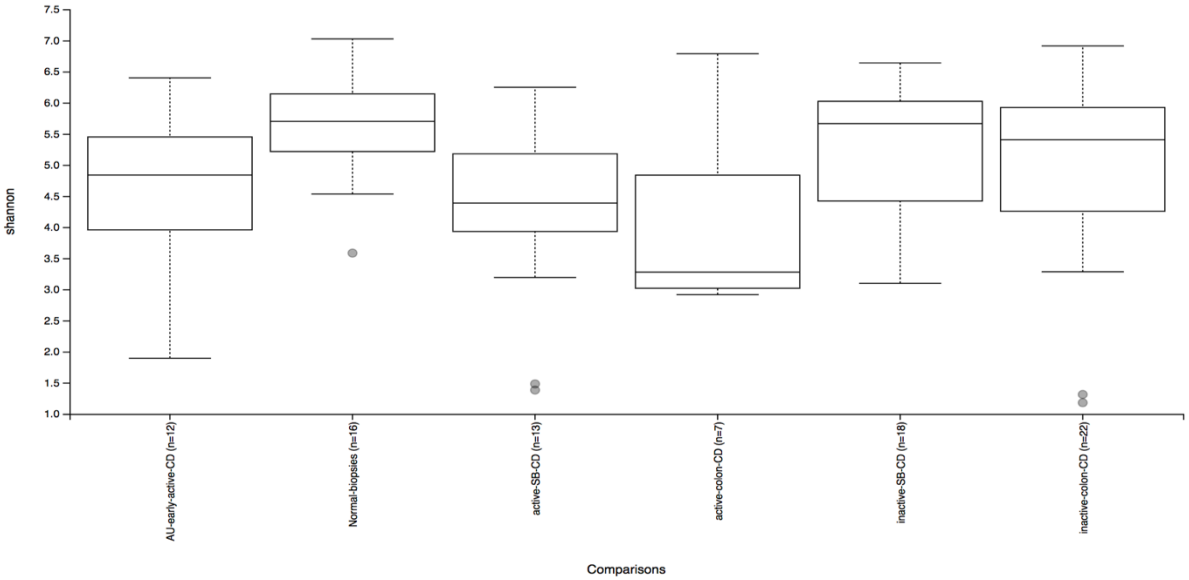
Supplemental Figure 2. 2. Alpha diversity between CD vs controls (based on QIIME analysis) based on observed OTUs Kruskal-Wallis (pairwise) test



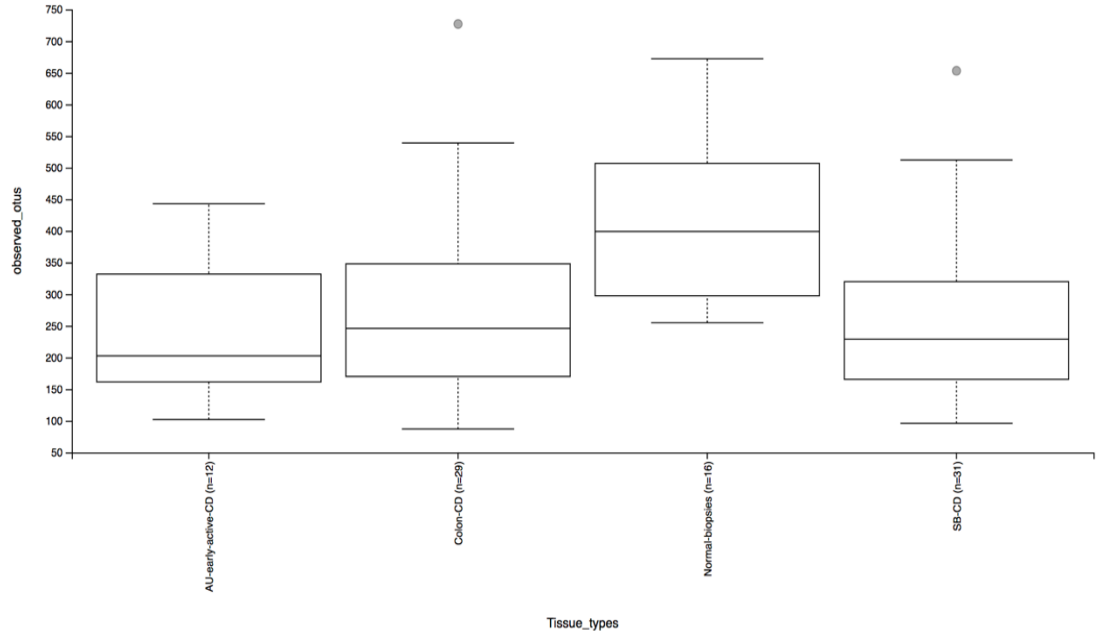
Supplemental Figure 2.3. Alpha diversity between CD vs controls based on Shannon diversity index using Kruskal-Wallis (pairwise) test



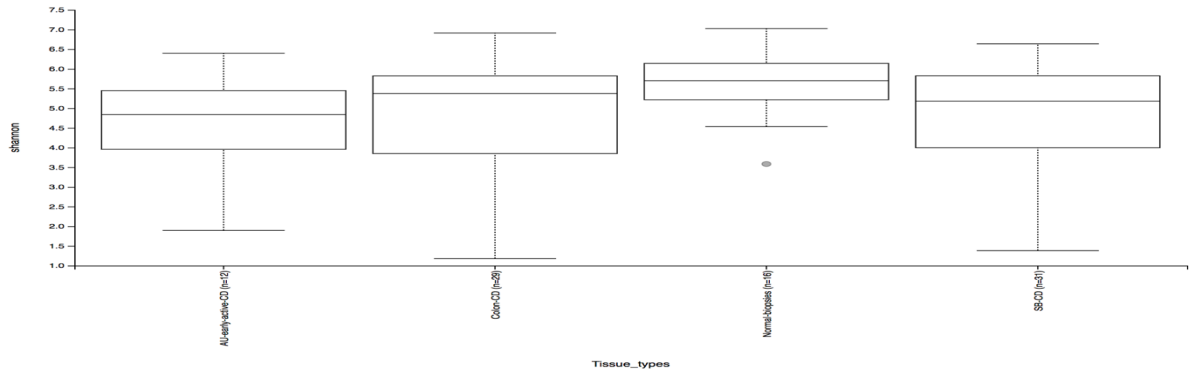
Supplemental Figure 2.4. Alpha diversity (observed otus) at different disease stages in CD compared to controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test.



Supplemental Figure 2.5. Alpha diversity (Shannon) at different disease stages in CD compared with controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test

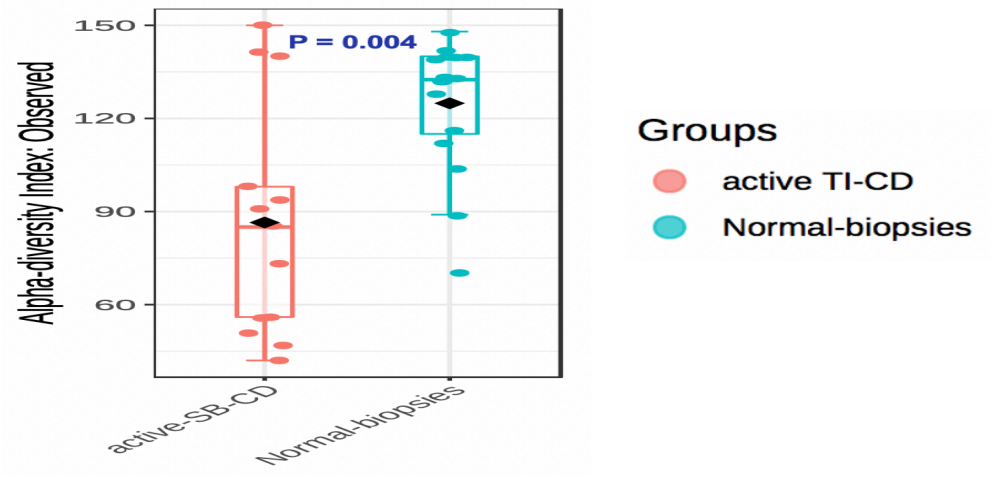


Supplemental Figure 2.6. Alpha diversity (observed_otus): different tissue types (AU, colon and small bowel) in CD vs normal biopsies in controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test.

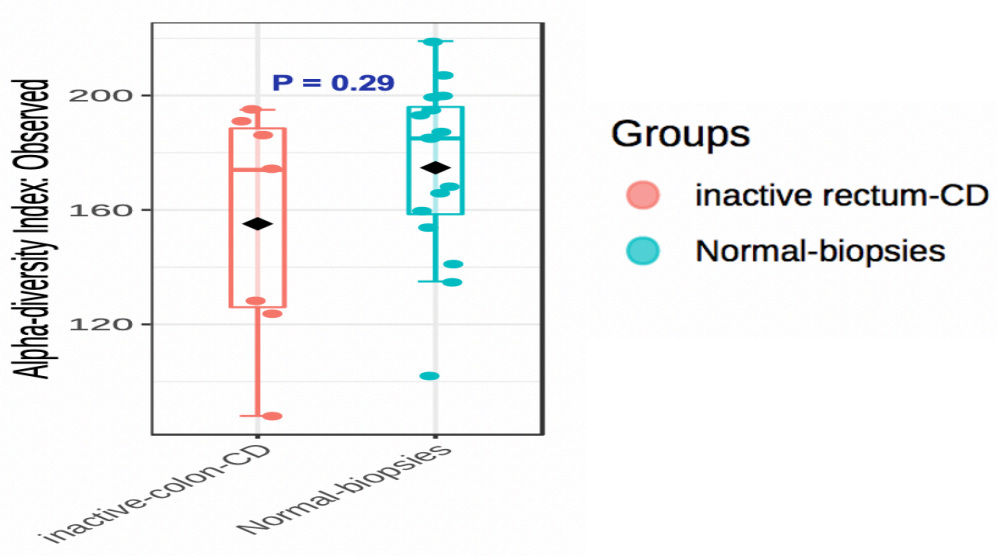


Supplemental Figure 2.7. Alpha diversity (Shannon) in different tissue types in CD compared to controls (based on QIIME analysis) based on Kruskal-Wallis (pairwise) test

A

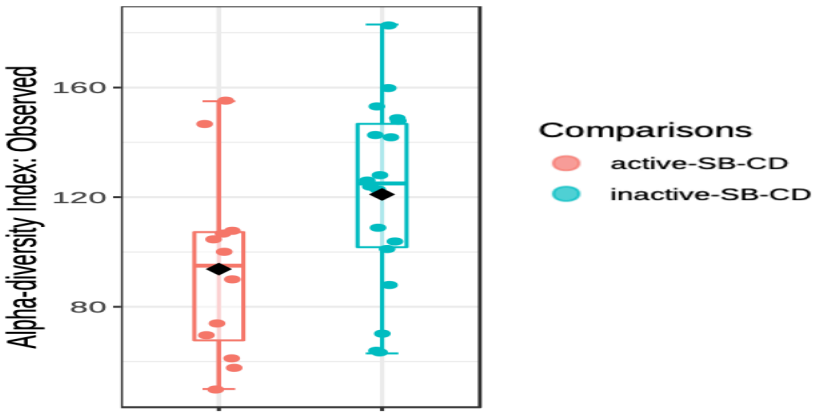


B

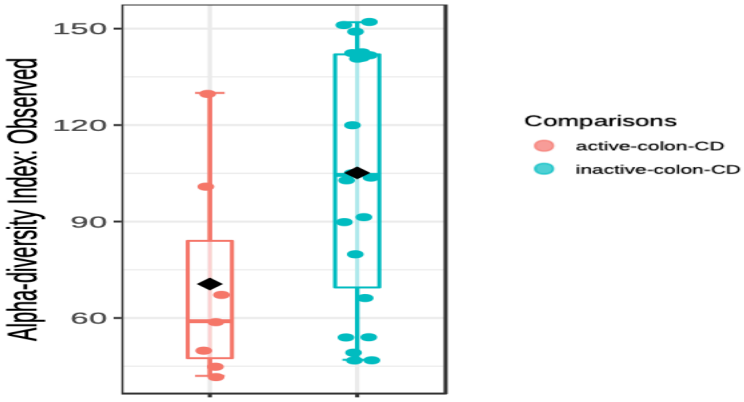


Supplementary Figure 2.8 A) Alpha diversity based on observed OTUs group significance between active TI in CD and normal biopsies in controls and **B)** Alpha diversity based on observed OTUs group significance between inactive rectum in CD and normal biopsies in controls. These box plots are showing the difference at the sequencing depth of 38739 reads per sample. Here, TI is terminal ileum

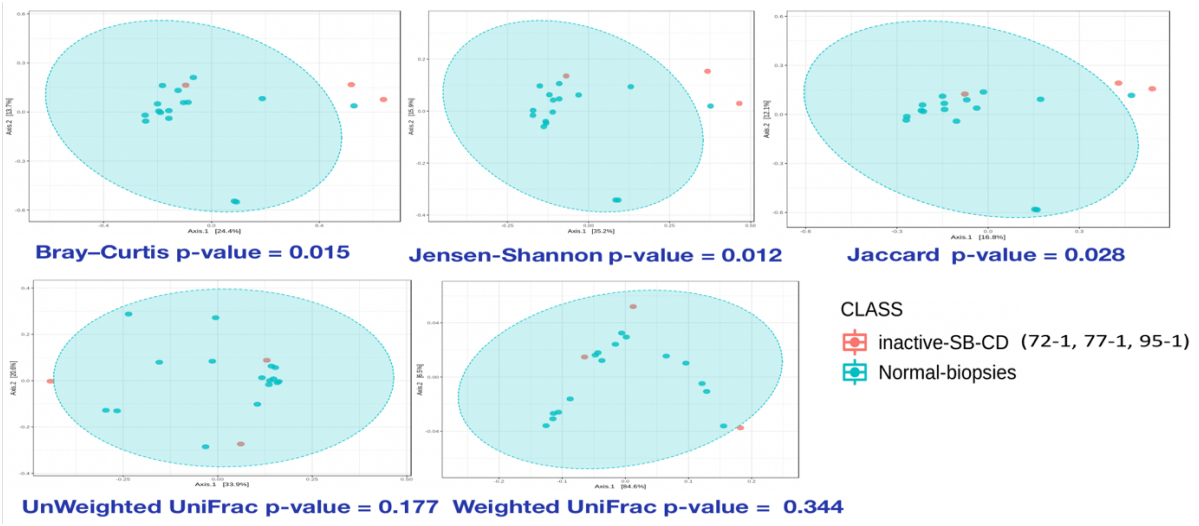
Supplemental Figure 2.9 Alpha diversity based on observed OTUs group significance between active aphthous ulcers and inactive small bowel in CD. and These box plots are showing the difference at the sequencing depth of 38739 reads per sample. Here, AU is aphthous ulcer and SB is small bowel



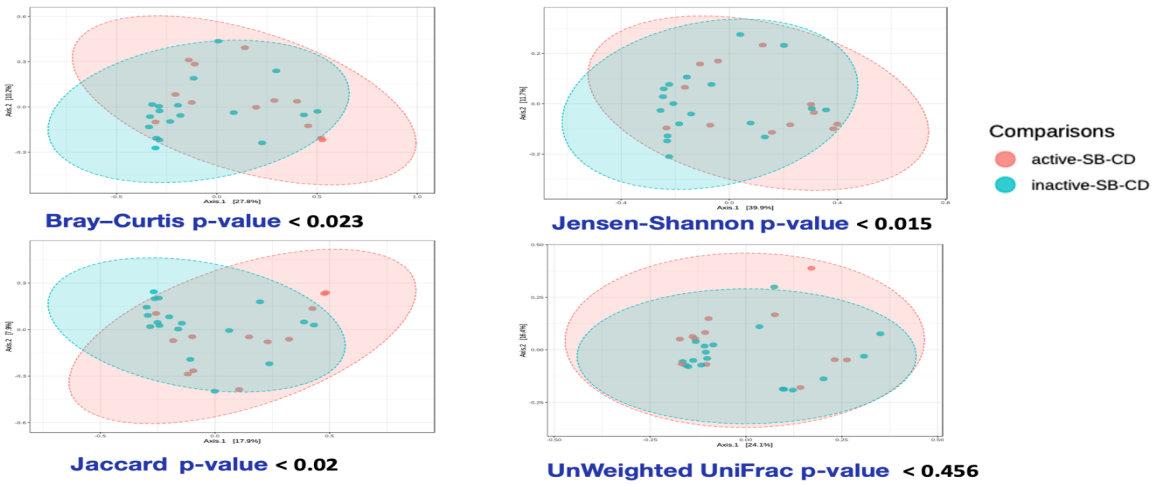
Supplemental Figure 2.10 Alpha diversity based on observed OTUs group significance between active small bowel and inactive small bowel in CD. and These box plots are showing the difference at the sequencing depth of 38739 reads per sample. Here, SB is small bowel.



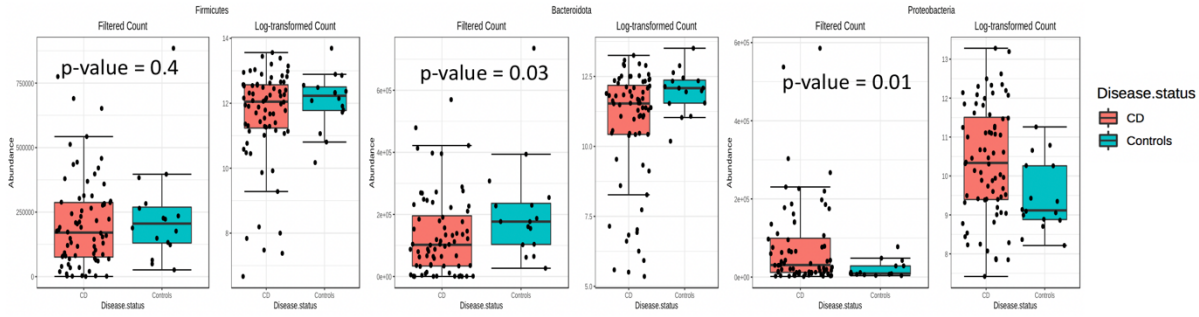
Supplemental Figure 2.11 Alpha diversity based on observed OTUs group significance between active colon and inactive colon in CD. and These box plots are showing the difference at the sequencing depth of 38739 reads per sample.



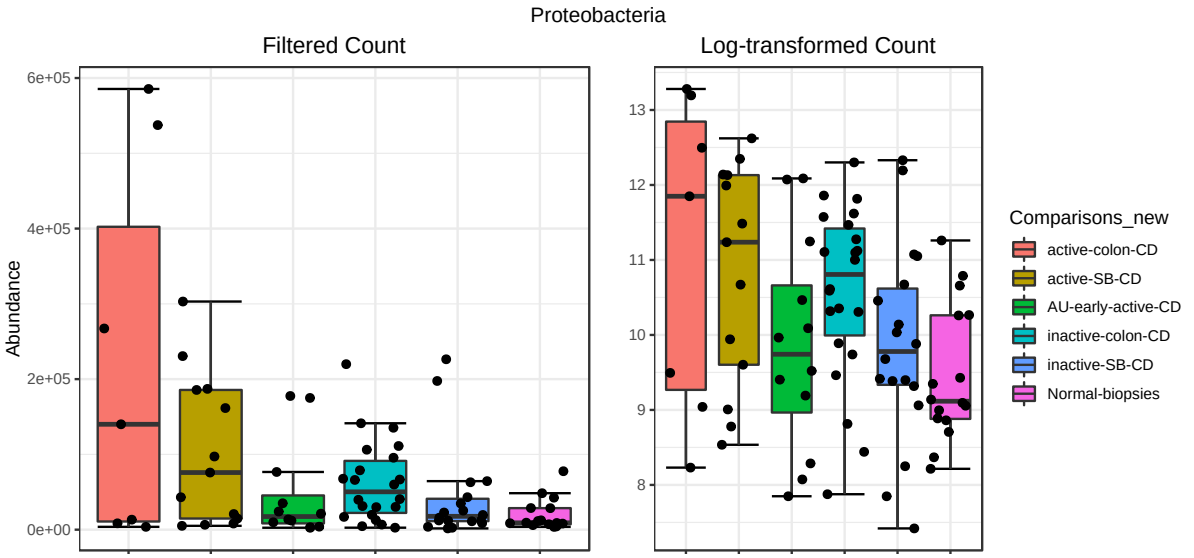
Supplemental Figure 2.12. Beta diversity analysis between inactive-SB-CD (n=3) (but have active disease in colon) and normal biopsies (n=16) using pairwise PERMANOVA analysis.



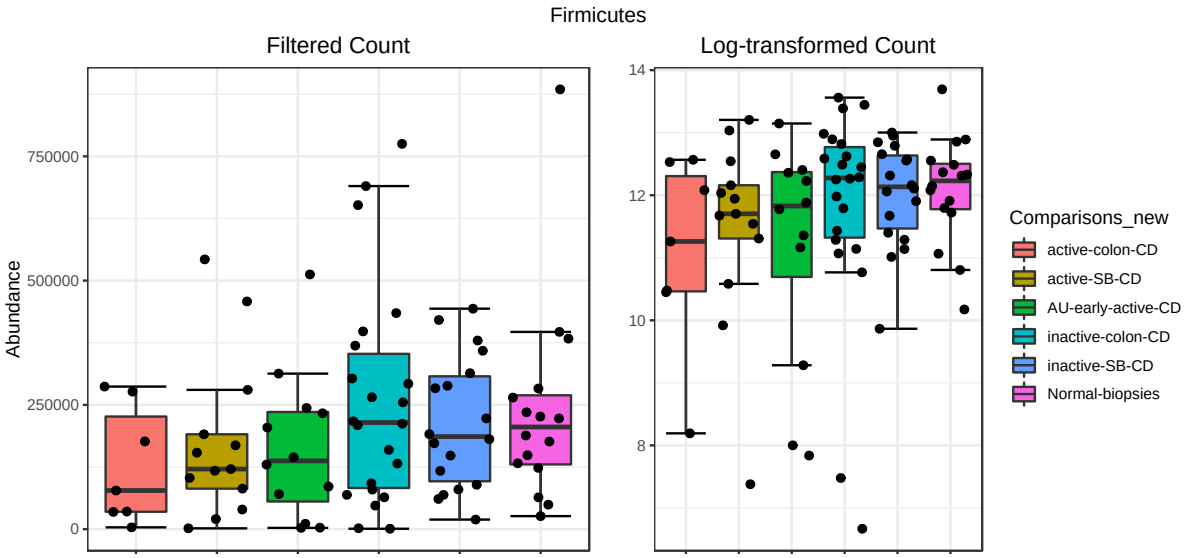
Supplemental Figure 2.13. Beta diversity analysis between active small bowel CD (n=13) and inactive small bowel CD (n=18) using pairwise PERMANOVA analysis.



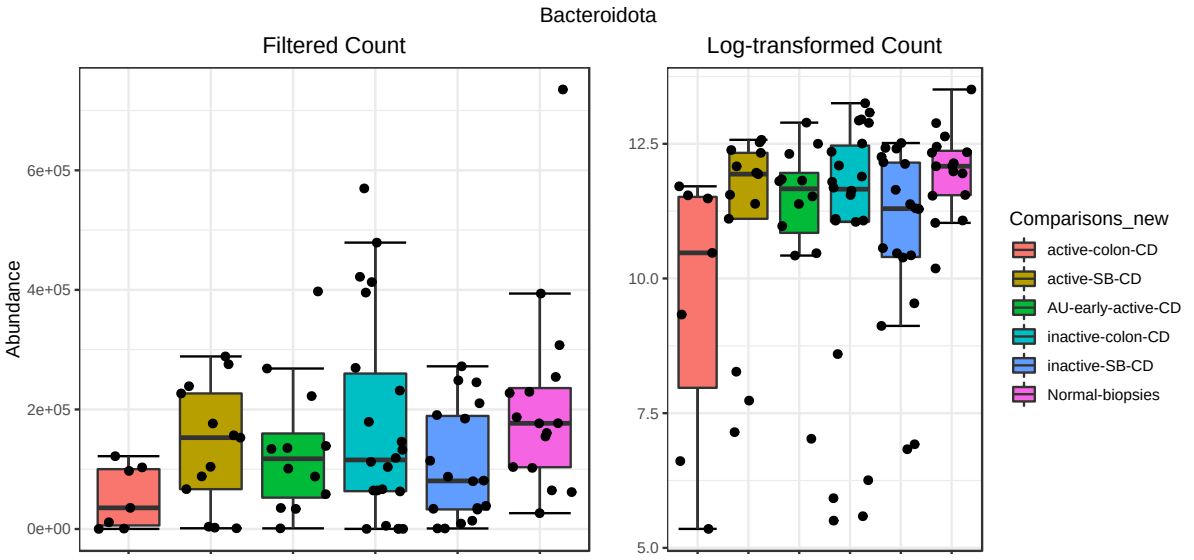
Supplemental Figure 2.14. Box plots depicting difference in abundances of the top three most abundant gut bacterial genus between CD and controls. Kruskal-Wallis multiple comparison was done, and the p-values adjusted with the Bonferroni method by Post-hoc Dunn test. Statistically significant values after FDR correction are represented as P ($p < 0.05$) in the top the boxplot. (This test was also done in R using the genera abundance result directly from QIIME output)



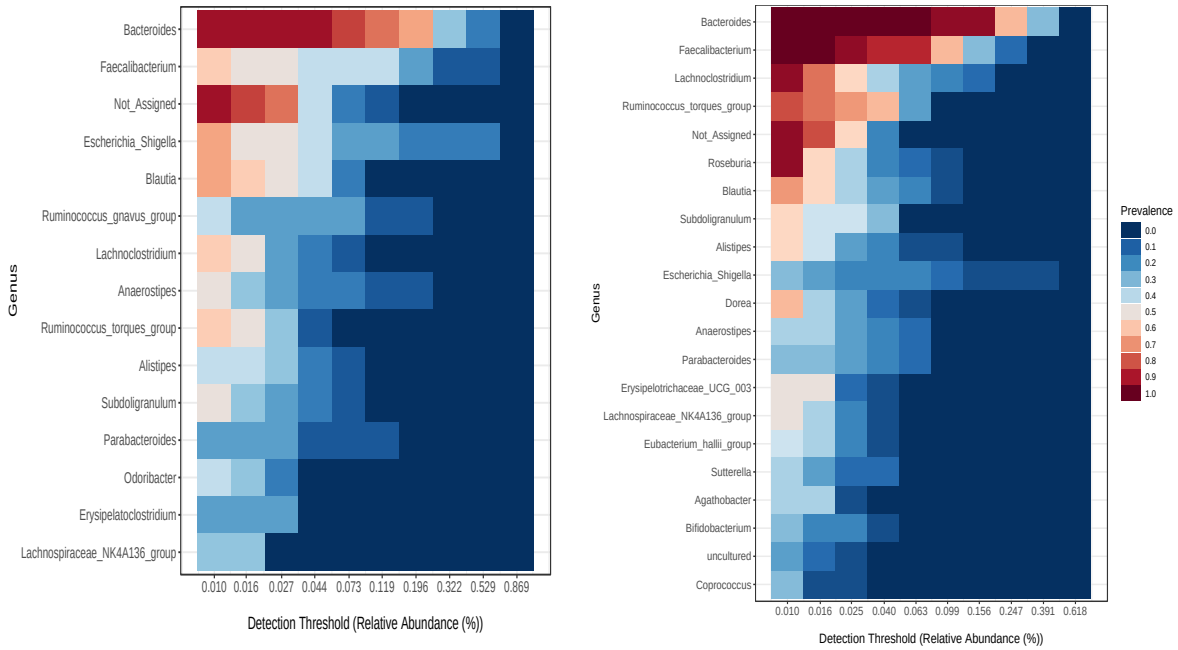
Supplemental Figure 2.15. A) Box plots depicting difference in relative abundances of phylum Proteobacteria at disease stages (AU-early-CD, active and inactive-CD) and different tissue types (small bowel, aphthous ulcers, colon) between CD and controls. Kruskal-Wallis multiple comparison was done, and the p-values adjusted with the Bonferroni method by Post-hoc Dunn test. Statistically significant values after Bonferroni correction are represented as P ($p < 0.05$) in the top the boxplot. *Here N indicates controls



Supplemental Figure 2.16. Box plots depicting difference in relative abundances of phylum Firmicutes at disease stages (AU-early-CD, active and inactive-CD) and different tissue types (small bowel, 'aphthous ulcers, colon) between CD and controls. Kruskal-Wallis multiple comparison was done, and the p-values adjusted with the Bonferroni method by Post-hoc Dunn test. Statistically significant values after Bonferroni correction are represented as P ($p < 0.05$) in the top the boxplot. *Here SB indicates small bowel



Supplemental Figure 2.17. Box plots depicting difference in relative abundances of phylum Bacteroidetes at disease stages (AU-early-CD, active and inactive-CD) and different tissue types (small bowel, aphthous ulcers, colon) between CD and controls Kruskal-Wallis multiple comparison was done and the p-values adjusted with the Bonferroni method by Post-hoc Dunn test. Statistically significant values after Bonferroni correction are represented as P ($p < 0.05$) in the top the boxplot. *Here SB indicates small bowel.

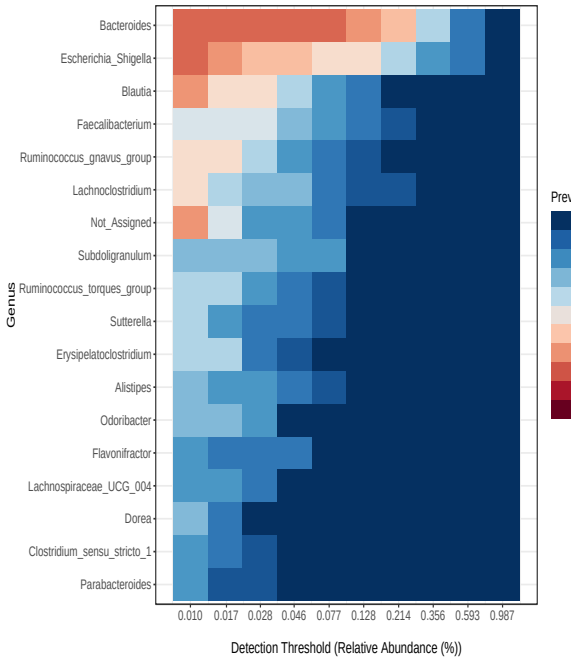


Core microbiome in aphthous ulcers

Core microbiome in controls

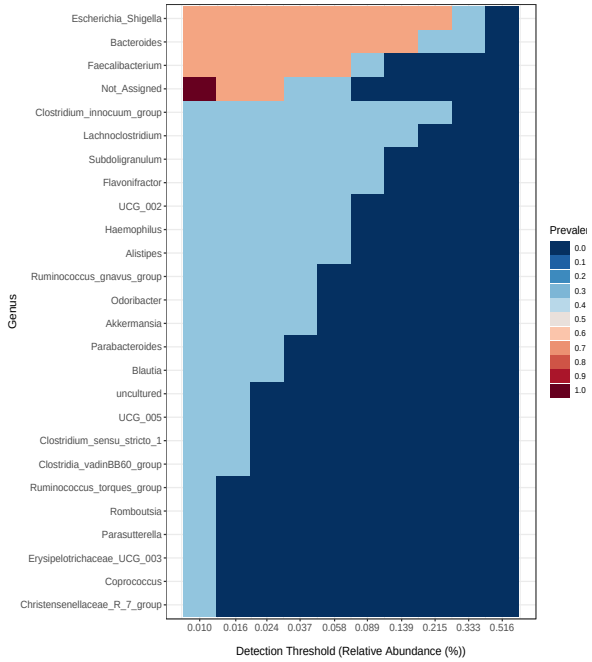
Supplemental Figure 2.18. Core gut microbiome in aphthous ulcers in early CD and controls

Core microbiome in inactive small bowel CD

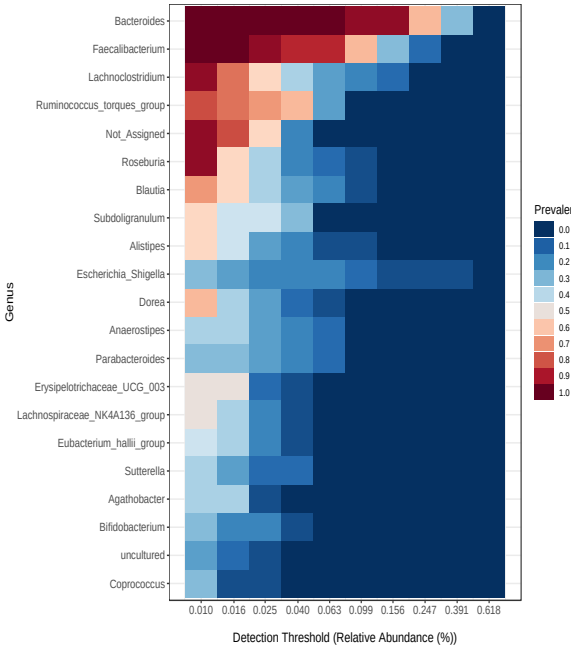


Core microbiome of active small bowel CD

Supplemental Figure 2.19. Core gut microbiome in inactive and active small bowel CD

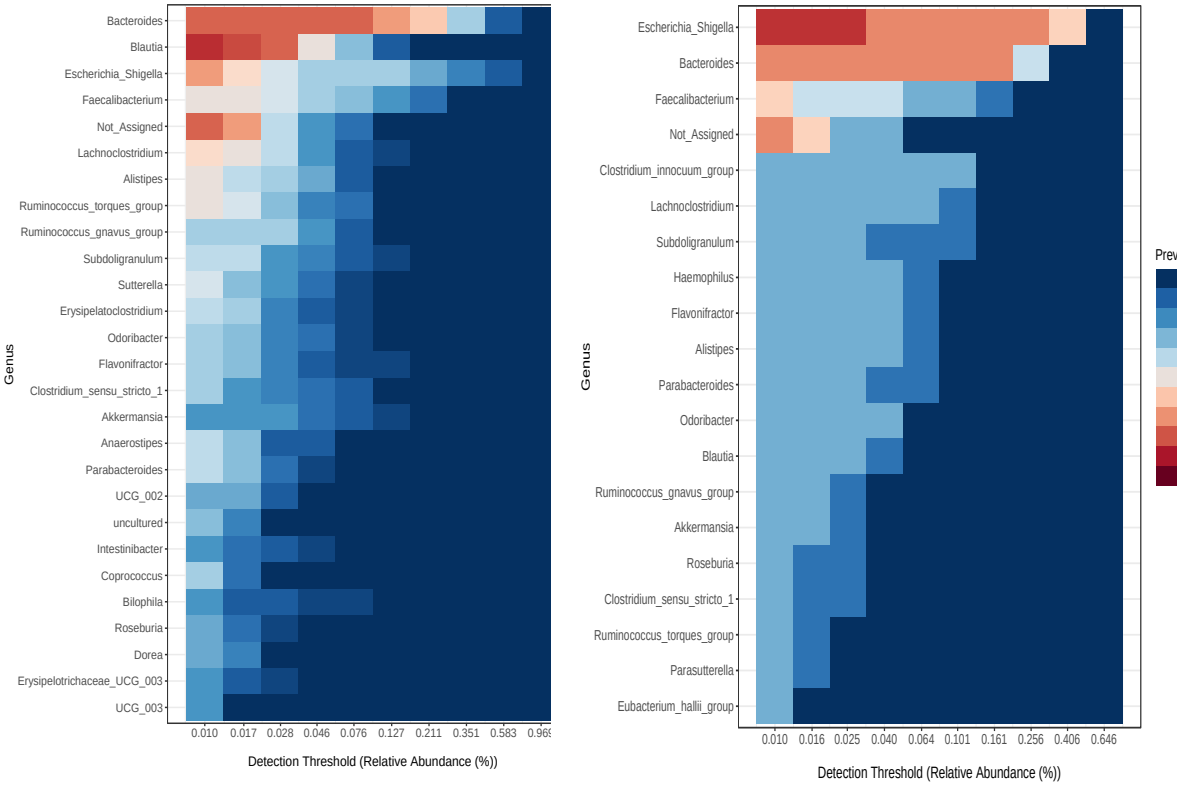


Core microbiome in inactive small bowel CD (with active disease in colon)



Core microbiome of normal biopsies in controls

Supplemental Figure 2.20. Core gut microbiome in inactive small bowel CD with active disease in colon and normal biopsies in controls



Core microbiome in inactive colon CD

Core microbiome in active colon CD

Supplemental Figure 2.21. Core gut microbiome in inactive and active colon CD

Chapter 3: Comparative Genomics of *E. coli* Isolated from Patients with Inflammatory Bowel Disease and Controls

3.1 Abstract

The inflammatory bowel diseases (IBD) are chronic inflammatory conditions of the digestive tract. The two major forms of IBD are Crohn's disease (CD) and ulcerative colitis (UC). Several lines of evidence implicate bacteria in the pathogenesis of IBD, and *Escherichia coli* is one of the leading candidate triggers. Numerous studies demonstrate an expansion of *E. coli* in IBD mucosa, although mechanistic insight remains lacking. The aim of this study was to determine if particular *E. coli* genes or molecular properties are associated with tissue type, including lymph nodes and the initial CD lesion (aphthous ulcer), or adult and paediatric isolates. This study involved whole-genome comparisons of 179 *E. coli* strains, isolated from 62 patients with CD (n = 23 paediatric CD), 18 with UC, and 20 healthy controls (HC). The source of the isolates included lymph nodes (n = 13), aphthous ulcers (n = 7), intestinal mucosa (n = 133), and other sources (n = 26). Following assembly and annotation, pan-genome and genome-wide association studies were conducted using various tools. The phylogroup, serotype, sequence type (ST), virulence genes, plasmids, bacteriophage, CRISPRs, capsules, bacteriocins, and antibiotic resistance profiles were determined for each strain. The CD, UC and HC strains were phylogenetically diverse. Overall, phylogroup B2 and ST 95 were most common. Different molecular properties were significantly associated with paediatric CD and lymph node CD isolates compared to HC (Table 3.1). Based on total gene content, CD isolates were more likely to have genes associated with adhesion, the toxin-antitoxin system, plasmid partitioning, and conjugation compared with HC. Lymph node CD isolates were associated with genes involved in adhesion, phosphate binding and catalase-peroxidase activity compared to HC (Table 3.2).

Table 3.1. Different molecular properties significantly associated with paediatric CD and lymph node CD *E. coli* isolates.

Molecular property	Comparison	Significance
Bottromycin bacteriocin	Paediatric CD vs HC	P <0.0001
ABC capsule		P = 0.009
E. coli pUTI89 plasmid		P = 0.003
Group IV e capsule	Lymph node CD vs HC	P = 0.022
Bottromycin bacteriocin		P = 0.023

Table 3.2. Genes associated with *E. coli* of lymph nodes in CD.

Gene name	Functional groups	Annotation (function)	Significance
group_3641/ htrE	Adhesion	Fimbrial assembly protein similar to papC	P = 0.001
group_14340/ yadV	Adhesion	Putative fimbrial chaperone YadV	P = 0.006
group_1161/ yadM	Adhesion	Putative fimbrial chaperone YadM	P = 0.006
yraK	Adhesion	Outer membrane usher protein	P = 0.01
group_7386/ papC	Adhesion	Outer membrane usher protein PapC	P = 0.01
group_7785/ katG	Catalase-peroxidase	Catalase-peroxidase	P = 0.004
pstS1	Phosphate binding	Phosphate-binding protein	P = 0.006

This study suggests that *E. coli* strains isolated from different IBD tissues, as well as adult and paediatric patients, harbour diverse molecular properties. Catalase-peroxidase activity may allow *E. coli* to survive within lymph nodes, consequently providing an on-going stimulus to the immune system. The functional significance of this association should be explored further.

3.2 Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are the two major types of inflammatory bowel disease (IBD). Colitis, unclassified is the term used when pathologic findings are not sufficient to distinguish CD and UC. The IBDs are chronic conditions, which significantly affect the quality of life of patients, and place significant demands on the healthcare system. There is no cure for the IBDs, and their incidence is increasing globally (Shivashankar et al., 2017). Although significant progress has been made in the treatment of these diseases, not all patients respond to therapy. Treatment is solely aimed at controlling the clinical manifestations of disease. CD and UC share many clinical features, but are distinguished based on their microscopic and macroscopic appearances (Sartor, 2006). Inflammatory lesions in UC are limited to the superficial mucosa and the colon, while CD lesions are transmural and can occur anywhere along the gastrointestinal tract (Baumgart and Carding, 2007). CD most frequently affects the terminal ileum and colon (Stange et al., 2006).

This disease is thought to result from a dysregulated immune response to intestinal microbiota in genetically predisposed individuals. Environmental factors such as diet, psychological stress, smoking, antibiotics, nonsteroidal anti-inflammatory drugs (NSAIDs), as well as epigenetics (Ananthakrishnan, 2013; Loddo and Romano, 2015; Dutta and Chacko, 2016) may also contribute to CD pathogenesis (Molodecky and Kaplan, 2010). The concordance of CD in monozygotic twins is between 20 to 50%, whereas, in dizygotic twins raised in the same environment, it is below 10% (Halfvarson et al., 2003) (Yang et al., 1993; Thompson et al., 1996; Orholm et al., 2000). Apart from host-derived factors, several lines of evidence implicate bacteria in the pathogenesis of CD. The terminal ileum and colon, the predominant site for the CD lesion, harbour the highest concentrations and diversity of bacteria in the gut (Flint

et al., 2012); diversion of the faecal stream reduces pathology, whilst reinfusion reactivates inflammation (Rutgeerts et al., 1991; D'Haens et al., 1998); oral antibiotics improve the clinical manifestations in some CD patients (Feller et al., 2010; Su et al., 2015; Nitzan et al., 2016); antibiotics affect recurrence and remission rates (Doherty et al., 2010); gene variants most significantly associated with CD are involved in bacterial recognition [NOD2 mutation (Kennedy et al., 2018)]; autophagy [ATG16L1 (Salem et al., 2015; Wildenberg et al., 2017)],; and signalling in response to bacteria [IRGM] (Naser et al., 2012; Glas et al., 2013)]; multiple species of bacteria can be cultured from 18-48% of mesenteric lymph nodes in CD patients compared with 5-15% in controls (Ambrose et al., 1984; O'Boyle et al., 1998; Takesue et al., 2002a; O'Brien et al., 2014). O'Brien *et al.* (2014) sampled mesenteric lymph nodes from CD patients and detected bacterial DNA in 28% of lymph nodes. In a follow-up study, where they used RNA instead of DNA showed that the microbial communities detected in lymph nodes are viable (Kiely et al., 2018). In addition to that, granulomas (focal aggregates of immune cells forming tiny nodules at the site of the infection or inflammation) are found in the intestinal mucosa of 30% – 68.9 % of bowel resected samples from patients with CD (Kleer and Appelman, 2001; Pierik et al., 2005). This granuloma possesses bacterial DNA (Ryan et al., 2004) which suggests an immune reaction to microbes.

Studies also show that bacterial load is comparatively higher in patients with CD than healthy individuals, and microbial community composition varies significantly between CD and controls (Swidsinski et al., 2002; Baumgart and Carding, 2007; Frank et al., 2007). Although a single causative microorganism has not been identified (Eckburg and Relman, 2007; Knights et al., 2014), many reports suggest the potential association of microbial pathogens with CD. For example, many studies have shown

an expansion of mucosa-associated *E. coli* population with CD (Darfeuille-Michaud et al., 1998a; Darfeuille-Michaud et al., 2004; Martin et al., 2004; Kotlowski et al., 2007a; Willing et al., 2009; de Souza et al., 2012; Lopez-Siles et al., 2014; Proctor et al., 2019). The increased abundance of *E. coli* adherent to inflamed mucosa of patients with CD has driven great interest in these bacteria over the last couple of decades. The high prevalence of adherent *E. coli* (AIEC) in mucosa of CD patient led to the comparison of several Crohn's disease-associated strains with strains isolated from patients without IBD, and failed to detect any virulence genes unique to the CD-associated strains (Darfeuille-Michaud et al., 1998a; O'Brien et al., 2016). However, many of those strains have the property of adhering and invading intestinal epithelial cells, as well as the ability to replicate within macrophages; hence they are categorized as a new specific pathogenic group of *E. coli* named adherent-invasive *E. coli*. AIEC strains are able to adhere and invade the intestinal epithelial cell as well as survive and replicate within the macrophage without killing the phagocytes (Boudeau et al., 1999; Glasser et al., 2001; Meconi et al., 2007; Lapaquette et al., 2010) and influence the release of the pro-inflammatory cytokine, TNFa (Rolhion and Darfeuille-Michaud, 2007). Several studies have reported a higher prevalence of the AIEC pathovar in CD patients compared to healthy subjects suggesting the plausible association of *E. coli* invasiveness with the pathogenesis of CD (Boudeau et al., 2001; Darfeuille-Michaud et al., 2004; Baumgart et al., 2007; Martinez-Medina et al., 2009; Agus et al., 2014; Martinez-Medina and Garcie-Gil, 2014).

The initial lesion in CD, the aphthous ulcer (AU), can be observed macroscopically during colonoscopy. They overlie the follicle-associated epithelium (FAE) of Peyer's patches (PPs) in the small intestine and lymphoid follicles (LFs) in the large intestine (Hizawa et al., 1994; Fujimura et al., 1996; Kinjo et al., 2004). M cells

(Membranous/microfold cells) are specialized epithelial cells present in the FAE, that transcytose antigens and microorganisms across the epithelial barrier, delivering them to underlying immune cells for processing (Kraehenbuhl and Neutra, 2000). Chassaing and colleagues investigated the ability of AIEC bacteria to interact with PPs and the virulence factors associated with this phenotype. They found that the prevalence of AIEC strains possessing the lpf operon was significantly higher in isolates from patients with CD than controls, and showed that LPF plays a critical role in the attachment of AIEC to M cells (Chassaing et al., 2011). When AIEC gained access to the lamina propria via M cells, they were taken up by macrophages, where they are able to resist killing within the phagolysosome and replicate (Bringer et al., 2005). Macrophages can also mediate bacterial translocation to draining lymph nodes (Hooper et al., 2012; O'Brien et al., 2014). In several studies, *E.coli* has been cultured from lymph nodes in CD patients (Ambrose et al., 1984; Laffineur et al., 1992; Takesue et al., 2002b; O'Brien et al., 2016) and is the dominant species in this location. The possible reason behind this is, AIEC delay apoptosis in infected macrophages, supporting their own survival in CD patients, by a mechanism involving increase of S-nitrosylation and proteasomal degradation of caspase-3 (Dunne et al., 2013). In addition to that, AIEC also favours its intracellular replication by modulating the ubiquitin proteasome system in inflammatory bowel disease (Cleynen et al., 2014). Certainly, defects in the autophagy associated genes (ATG16L1, IRGM, or NOD2) supported the intra-macrophage replication of AIEC leading to enhanced secretion of proinflammatory cytokines in patients with CD (Lapaquette et al., 2012). However, AIEC was also isolated from healthy individuals suggesting that AIEC are opportunistic pathogens that can cause disease in susceptible hosts.

A study suggested that mucosa-associated, but not luminal, *E. coli* potentially confer an adaptive advantage associated with the initiation and perpetuation of inflammation in CD (de Souza et al., 2012). Although a specific genotype of *E. coli* was not found in Crohn's ileal mucosa, some genotypes were more commonly associated with chronic or early recurrent ileal CD. Pathogenic *E. coli* belonging to phylogroup B2 + D are significantly more prevelent in patients with IBD than in controls (Kotlowski et al., 2007b) and virulence genes of pathogenic *E. coli* were more commonly isolated from B2 strains in IBD patients with active disease than patients with inactive disease (Petersen et al., 2009) suggesting that a novel group of pathogenic *E. coli* associated with IBDs.

Paediatric CD is considered to be more complex, with a distinct features compared with adult CD in onset, disease course, extensive intestinal involvement, presentation and progression (Haberman et al., 2014; Ledder et al., 2014; Rigoli and Caruso, 2014; Herzog et al., 2017) and the incidence of CD in childhood is rising worldwide (Benchimol et al., 2017; Sýkora et al., 2018). Previous reports have identified distinct differences in paediatric CD phenotypes compared to adult CD (Van Limbergen et al., 2008; Sagiv-Friedgut et al., 2010; Lovasz et al., 2013; Ostrowski et al., 2016). Recent studies revealed that, very-early-onset IBD (VEO-IBD) that occurs in children less than 6 years of age and infantile IBD that develops in children less than 2 years of age differs greatly in several aspects comparing with adult CD. Very early onset IBD, particularly infantile onset IBD (VEO-IBD, <2 yr) is more likely to have a genetic cause. For example, a study on VEO-IBD has identified distinct homozygous mutations of IL10RA, IL10RB in children caused hyperinflammatory immune responses in the intestine (Glocker et al., 2009).

Numerous genetic defects that disturb intestinal epithelial barrier function or affect immune regulations have been found in patients from a very young age. In addition to that, neonatal IBD or infantile-onset IBD patients have higher rates of affected first-degree relatives indicating more genetic association than older-onset IBD patients (Ashton et al., 2017; Kelsen et al., 2019; Shim, 2019). The similarities and differences in the *E. coli* genome between the adult and paediatric CD may provide the link to a better understanding of the pathogenesis.

The present study aimed to identify unique/signature genes of *E. coli* isolated in patients with CD compared with UC and healthy control, adult vs paediatric, and lymph node (LN) and aphthous ulcer (AU) isolates. Our aim was to determine if particular *E. coli* genes or molecular properties are associated with tissue type, including lymph nodes and initial CD lesion (aphthous ulcer), or adult and paediatric isolates.

3.3 Materials and Methods

3.3.1 Sample information

This study was based on a collection of 179 *E. coli* strains isolated from biopsies of aphthous ulcers (AU) and intestinal mucosa; lymph nodes (LN) and involved and uninvolved mucosa from surgically resected specimens; and stool samples of paediatric patients. Control isolates were obtained from intestinal mucosal and bladder samples, as well as stool. The total list of samples along with their source is given in the Table 3.3 and Supplemental Table 3.1.

Table 3.3. Numbers of samples from different sources

Name of sample	Source	Number of patients	Number of samples	Reference

Crohn's disease (CD)	Paediatric patient	23	23	This study
	Aphthous ulcers (AU)	7	7	Dixit et al., 2018, and this study
	Lymph nodes (LN)	7	13	This study and O'Brien et al., 2016
	Intestinal mucosa	31	71	O'Brien et al., 2016; Dixit et al., 2018
Ulcerative colitis (UC)	Intestinal Mucosa	17	33	O'Brien et al., 2016; Dixit et al., 2018
Healthy control (HC)	stool	2	2	O'Brien et al., 2016; Dixit et al., 2018
	Urinary bladder	1	1	O'Brien et al., 2016; Dixit et al., 2018
	Intestinal mucosa	17	29	O'Brien et al., 2016; Dixit et al., 2018

In this study, we collected the biopsy samples from adult patients (age between 18-71) who underwent colonoscopy procedures in Canberra Hospital, ACT, Australia, in 2010-2011. These samples were collected from five different gut locations: rectum, descending colon (DC), transverse colon (TC), ascending colon (AC), and terminal ileum (Ti) in a sequential order starting from the rectum to prevent possible contamination of the bacteria from other gut locations. In some patients, aphthous ulcers were also found, and tissue samples from those ulcers were collected. In each step of the sample collection, we used sterile equipment and took the sample into LB broth-glycerol solution. Then, we preserved the biopsy samples at -80^{o}C until processing (Gordon et al., 2015; O'Brien et al., 2016; A. et al., 2018).

In addition to those biopsy samples, 20 *E. coli* (13 from lymph nodes (LN), five from normal mucosa and two from involved mucosa) isolates from seven (7) bowel resected samples were collected from Crohn's disease patient during surgery at the Canberra Hospital. From bowel resected samples, we collected three different kinds of tissue samples - lymph nodes (LN), normal mucosa, and involved mucosa, in that sequence,

with sterile forceps for each tissue sample. Each tissue sample was streaked onto MacConkey agar plate and put into LB-glycerol (lysogenic broth). MacConkey plates were incubated for 18 hours at 37°C and LB-glycerol tubes containing tissue samples were placed at -80°C. Professor Georgina Hold at Microbiome Research Centre, St George Hospital, UNSW, Sydney kindly provided the *E. coli* genomes from paediatric CD patients.

3.3.2 Isolation and identification of *E. coli* isolates from the biopsy tissues and bowel resected samples

Biopsy and bowel resected samples were streaked onto MacConkey agar plates, and the plates were incubated at 37°C overnight. Up to 50 lactose-positive (red) colonies were inoculated onto Simmon's citrate agar plates and incubated overnight at 37°C. Up to 50 probable *E. coli* isolates (lactose positive and citrate negative) were randomly selected from each sample and transferred to LB plates. These LB plates were incubated at 37°C for 18 hours and kept at 4°C for longer preservation.

3.3.3 Characterization of *E. coli* by Rep-PCR and Clermont PCR

We extracted the *E. coli* DNA using a protocol described by Dixit and colleagues (A. et al., 2018). The repetitive element palindromic-polymerase chain reaction (rep-PCR) method amplifies intergenic regions of the *E. coli* genome resulting in a unique fingerprint for each strain. We used two sets of REP-PCR primers, N6(CGG)4 (Adamus-Bialek et al., 2009) and ERIC primers (Wei et al., 2004) to identify differences in the banding patterns of the 50 *E. coli* colonies extracted for each tissue sample. Isolates that produced the same banding patterns using N6 (CGG)4 and ERIC primers were considered as same strain. A representative for each fingerprint, for each tissue

type, was used to determine the phylogroup of the strain, using the Clermont phylo-typing method (Clermont et al., 2013).

3.3.4 Whole genome sequencing using high throughput sequencing (HTS) technology

E.coli isolates used in this study were previously sequenced as a part of a study assessing the gene content of AIEC and control isolates (n = 15 strains) (O'Brien et al., 2016), as part of a study assessing *E.coli* along the length of the gastrointestinal tract (n = 122 strains) (Gordon et al., 2015), and strains newly sequenced as part of this study (bowel resected isolates, n = 18 strains; aphthous ulcers, n = 2 and paediatric CD isolates, n = 23).

We sequenced 43 *E. coli* isolates from paediatric patients (23) and bowel resected samples, including involved and uninvolved mucosa, as well as lymph nodes. Unique Rep-types of *E. coli* were considered from each patient for whole-genome sequencing. We extracted each isolate using an ISOLATE II Genomic DNA Kit (Bioline) according to the manufacturer's instructions. The DNA concentration was measured using a Qubit, and DNA samples were diluted to a final concentration of 0.2 ng/ml. We used A Nextera XT DNA library preparation kit (Illumina) and Nextera XT index kit (Illumina) for DNA library preparation according to the manufacturer's instructions. A MiSeq Reagent kit V3 (600 cycles) was used for sequencing the strains on a MiSeq (Illumina) platform.

3.3.5 *E. coli* genome contig assembly

We used the A5_miseq_macOS_20150522 pipeline to assemble *E. coli* genomes

using the raw fastq.gz sequence files produced by the MiSeq platform according to

the code described by Coil et al (2015).

3.3.6 WGS data analysis for molecular characterization of the *E. coli* isolates

We used Centre for Genomic Epidemiology (CGE) website (http://www.genomicepidemiology.org/) to characterize the strains by MLST 1.8 tool (MultiLocus Sequence Typing) (Larsen et al., 2012) , ResFinder 3.0 tool (Zankari et al., 2012), SerotypeFinder 1.1 tool (Joensen et al., 2015), PlasmidFinder 1.3 tool (Carattoli et al., 2014), and VirulanceFinder 1.5 (Joensen et al., 2014) tools. BAGEL4, a web-based bacteriocin genome mining tool (de Jong et al., 2006) was used to identify the antimicrobial peptides produced by bacteria called bacteriocins (http://bagel.molgenrug.nl/) and CRISPRfinder program online (http://crispr.i2bc.paris-saclay.fr/) to determine clustered regularly interspaced short palindromic repeats (CRISPRs) . To identify prophage sequences within the *E. coli* genome, we used the web server PHASTER (http://phaster.ca/) (Arndt et al., 2016). CapsuleFinder tool (Rendueles et al., 2017) was used to determine the capsule type and ConjScan tool (Cury et al., 2020) was used to identify the conjugative elements in the strains in Galaxy web server https://galaxy.pasteur.fr/.

3.3.7 Plasmid analysis

For plasmid analysis, we assembled the read files for all strains using Spades (Bankevich et al., 2012). We had to assemble the read files again using Spades assembly because the plascope tool needed Spades assembly file for plasmid analysis. The Spades assemblies were then inputted into three plasmid analysis programs: Plascope (Royer et al., 2018), MLplasmids (Arredondo-Alonso et al., 2018), and MobTyper (Robertson and Nash, 2018). These software programs do not yield identical results, and no one program provides 100% accuracy for any given analysis.

The specificity and sensitivity of each program also differs. Therefore, the predicted plasmid contigs from all three programs were aligned to reference plasmid sequences within Mauve (Darling et al., 2004b), in order to improve our chances of identifying complete plasmid contigs. Plasmid contigs were considered to be true if they were predicted as plasmid associated contigs and matched the mauve alignment perfectly. All candidate plasmid contigs were also aligned against the NCBI nucleotide database using blastn. Contigs with at least 80% nucleotide identity and 75% length coverage of a reference plasmid sequence were considered as potential plasmid contigs. Candidate plasmid contigs were realigned with reference plasmid sequences using Mauve 2.4.0 (Darling et al., 2004b). A new computational method, which uses the CONJscan module of MacSyFinder (Cury et al., 2020), was used to identify conjugative systems in plasmids and chromosomes. This method identifies the protein composition of the system by hidden Markov model and determines if the organism possessing a conjugative system based on the components and genetic organization (Guglielmini et al., 2014). These methods enable the identification of conjugative elements in bacterial genomes (Cury et al., 2020).

3.3.8 Phylogenetic Analysis

Molecular evolutionary genetics analysis software, MEGA7 (Kumar et al., 2016) was used for phylogenetic analysis of genes and variants significantly associated with *E. coli* isolates from patients with CD compared to other groups (UC and healthy controls). We performed this phylogenetic analysis based on the maximum likelihood (ML) method (Tamura and Nei, 1993). To visualise the phylogenetic tree of all *E. coli* genomes, along with the different properties (disease status, gender, serotypes, phylogroups, antibiotic resistance, capsule types and bacteriocins) we used a web-

based tool, iTOL, accessible online http://itol.embl.de/login.cgi (Letunic and Bork, 2016).

3.3.9 Genome annotation and pan-genome association studies

The Prokka rapid prokaryotic genome annotation tool version 1.12 (Seemann, 2014) was used for annotating genomic DNA sequences of *E. coli* (n = 146). The input file was the contig file of each isolate. Roary (v.3.11.2) was used for the pan-genome analysis of all *E. coli* genomes (Page et al., 2015). Annotated assembly outputs from PROKKA (.gff files) were used as the input file for Roary. Roary is a pan-genome tool that can rapidly calculate large-scale pan genomes identifying the core and accessory genes. Then, Scoary tool was used for pan-genome association studies. Scoary is a python-based script tool which determines whether or not a particular gene in the pan-genome is associated with a particular treatment or disease state of the host (Brynildsrud et al., 2016). Scoary provides p values for each trait vs phenotype association, with a significant association considered present if $p < 0.05$.

3.3.10 Principal component analysis (PCA) and statistical methods

Past 3 software was used to conduct the principal component analysis (Hammer et al., 2001). To determine whether or not genes found to be significantly over-represented in the *E. coli* genomes were more likely to be associated with a specific trait (disease status, tissue/sample type, adult/paediatric), a similarity of percentages (SIMPER) analysis was conducted within the statistical tool PAST3 (Hammer et al., 2001). JMP 4.1 software was also used for other statistical analysis. In JMP software we have done analysis of distribution and the modelling type determine analysis/ fit Y/X analysis. In fit Y/X analysis distribution of Y for each X was determined.

Contingency analysis was performed in JMP software to get the Fisher exact test results and to generate a mosaic plot comparing the groups.

3.3.11 Bacterial genome-wide association study (GWAS) using Pyseer

Pyseer uses linear models with fixed or mixed effects to estimate the contribution of genetic variation in a bacterial population or phenotype of interest, while taking into consideration population structure, such as phylogenetic differences. It is therefore possible to conduct GWAS in clonal organisms such as bacteria. Pyseer also uses k-mers and SNP (single nucleotide polymorphism) data to identify gene variants that are significantly associated with a particular bacterial population or phenotype (Lees et al., 2018). k-mers are a specific number of nucleotide sequences that overlap. It allows us to generate a unique sequence from reads that represent genomic regions with different small sequences. k-mers helps to eliminate sequences repeated more than once because of unequal sequence coverage.

3.3.12 Genome alignment and visualization

We performed genome rearrangement (alignment) of each genome using Mauve (Darling et al., 2004a) against the reference genome Ed1a. Mauve is a whole-genome alignment visualisation tool that enables genome comparisons among different strains. Genes of interest identified using Roary (pan-genome pipeline) were validated by visualizing their alignment to the reference, to determine their presence/absence and completeness.

3.4 Results

Only one representative isolate of each unique sequence type (ST) was considered per patient and the resulting 146 strains used for all downstream analyses, except where strains from the same patient were compared.

3.4.1 Phylogroups of isolates with unique sequence types (STs)

In some cases, more than one *E. coli* isolate was collected from each patient/control, and both or more strains had the same sequence type (ST). They are therefore likely to be the same strain. To avoid any bias caused by subjects having multiple similar strains, thereby gene content, only one representative of each unique ST per patient was considered for all further downstream analysis except where strains from the same patient were compared. For example, three of the 179 isolates belonged to patient CD70, were isolated from involved mucosa, uninvolved mucosa, and lymph node, and all three strains (CD70NM, CD70IM, CD70LN) had the same ST. In this case, we only included the lymph node strain in the analysis (Supplemental Table 3.1). After excluding duplicate STs, 146 isolates remained for further analysis. The most abundant phylogroup in all groups is B2 (43% in CD, 70% in paediatric CD, 56% in HC and 52% in UC), followed by group D (19% in CD, 13% in paediatric CD, 22% in HC and 23% in UC) and B1 (21% in CD, 13% in paediatric CD, 11% in HC and 16% in UC) (Figure 3.1). Phylogroup C and Clades are the rarest phylogroups in this collection. Compared to HC (9%) and UC (15%), the proportion of B1 phylogroup was higher in CD isolates (paediatric CD, adult and lymph node-associated isolates all together) (23%) and lymph node-associated CD isolates (34%). We observed a higher proportion of B2 phylogroup in AIEC (80%) and paediatric CD isolates (70%)

compared with lymph node isolates (22%), non-AIEC (63%) and healthy controls (56%) although these differences were not significant (Table 3.4).

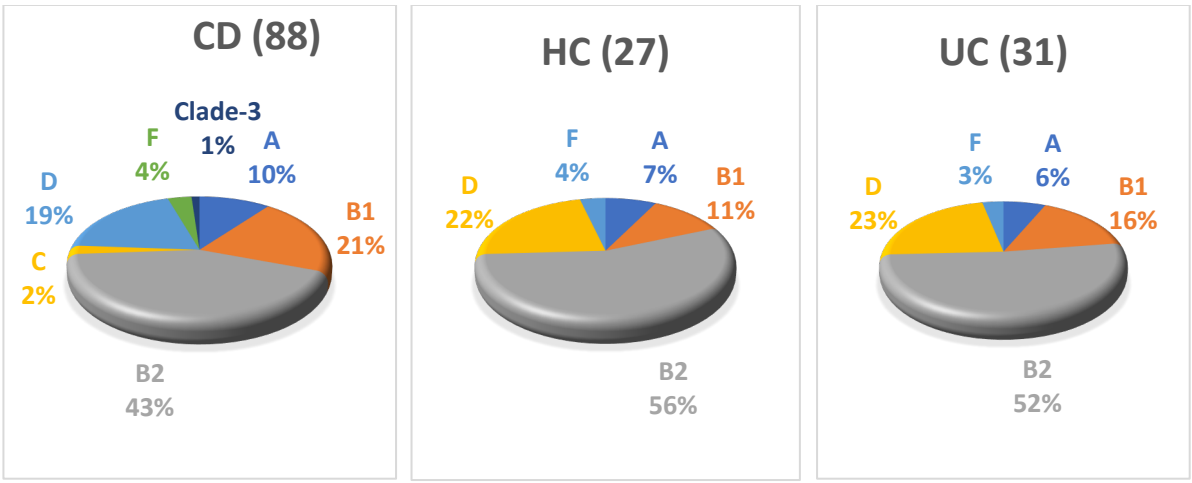


Figure 3.1. Phylogroup distribution of *E. coli* for the 146 strains representing unique STs for each subject (CD, HC, and UC). ∗ Here CD indicates Crohn's disease, HC indicates healthy controls and UC indicates ulcerative colitis samples.

Table 3.4. Distribution of phylogroups for patient groups, various types of tissue, and *E. coli* phenotypes

Phylogroup	CD (88)	Adult CD (65)	Paediat-ric CD (23)	UC (31)	AU (7)	LN (9)	AIEC (15)	non AIEC (19)	HC (27)
A	9 (10%)	9 (14%)	0	2 (6%)	0	2 (22%)	1 (7%)	1 (5%)	2 (7%)
B1	18 (21%)	15 (23%)	3 (13%)	5 (16%)	1 (14%)	3 (34%)	2 (13%)	3 (16%)	3 (11%)
B2	38 (43%)	22 (34%)	16 (70%)	16 (52%)	3 (43%)	2 (22%)	12 (80%)	12 (63%)	15 (56%)
C	2 (2%)	1 (2%)	1(4%)	0	0	1 (11%)	0	1 (5%)	0
D	17 (19%)	14 (22%)	3 (13%)	7 (23%)	2 (29%)	0	0	1 (5%)	6 (22%)
Clade-3	1 (1%)	1 (2%)	0	0	0	1 (11%)	0	1 (5%)	0
F	3 (4%)	3 (5%)	0	1 (3%)	1 (14%)	0	0	0	1 (4%)

*AIEC: Adherent invasive *E. coli*, non-AIEC: non-Adherent invasive *E. coli*, AU: Aphthous ulcer, LN: Lymph node, and HC: Healthy controls

3.4.1.1 Phylogenetic relationships among the strains in different groups

Eighty-eight (88) *E. coli* isolates in CD represented 52 different STs, whereas 31 isolates of UC and 27 isolates of HC represented 22 and 21 different STs, respectively (Table 3.5). We found fewer ST types (14 out of 23) in paediatric CD isolates. In the AIEC group, 15 isolates have ten different ST types compared with 13 STs in 19 non-AIEC isolates. All aphthous ulcers isolates are of different ST types. We observed seven different STs in nine (9) lymph node isolates, where one isolate could not be typed. The distribution of STs is random and not related to the disease (CD and UC). ST95 is the most common ST type, represented by 12 (14%) isolates in CD (adult and paediatric), 5 (22%) in paediatric CD, 7 (23%) isolates in UC group, 5 (33%) AIEC, 4 (21%) non-AIEC, 2 (22%) lymph node associated CD isolates and 3 (11%) HC isolates. We found that no ST was overrepresented in CD, UC, AIEC, lymph node and aphthous ulcer isolates compared with HC.

Interestingly we found two AIEC and non-AIEC pairs from the same patients with the same sequence type and fewer gene differences (150-200 genes). They are 12-1-Ti12 and 12-2-Ti13 with ST-127 in healthy control; and 20-2-TC13 and 20-3-TC12 with ST-1304 in CD. It indicates that isolates with high genetic similarities from the same individual could be phenotypically different like AIEC and non-AIEC pairs.

Table 3.5. Distribution of sequence types (ST) among the *E. coli* isolates belonging to different disease groups, tissue types and phenotypes

ST types	CD (88)	HC (27)	UC (31)	Paedia -tric CD (23)	AU (7)	LN (9)	AIEC (15)	non-AIEC (19)
ST-95	12	3	7 (23%)	5 (22%)	0	2	5 (33%)	4 (21%)
ST-131	3	2	2	2	1	0	2	1
ST-362	3	1	3	1	0	0	0	0

ST-73	6	2	1	4	1	0	0	1
ST-132	2	0	1	0	1	0	0	0
ST-69	2	0	1	0	0	0	0	1
ST-10	2	0	0	0	0	0	0	0
ST-127	0	2	0	1	0	0	1	1
ST-409	2	0	0	0	0	1	0	0
ST-569	2	0	0	0	1	0	0	3
ST-58	2	0	0	1	1	0	0	0
ST-59	2	0	0	0	1	0	0	0
total ST type	52 ST types	22 ST types	21 ST types	14 ST types	7 ST types	7 ST types & 1 unknown ST	10 ST types	12 ST types & 1 unknown ST

*AIEC: Adherent invasive *E. coli*, non-AIEC: non-Adherent invasive E. coli, AU: Aphthous ulcer, LN: Lymph node, and HC: Healthy controls. In the first row of the table, numbers within the bracket indicates the total number of the samples within each group.

3.4.2 Serotype diversity among the *E. coli* isolates

Using an *in silico* approach, 83% (74) of the 88 CD isolates, 85% (23) of the 27 HC isolates and 61% (19) of the 31 UC isolates could be assigned an O type. We identified twenty-nine (29) different O types in the CD isolates, with O1 being the most common (11%), followed by O8 (8%), whereas in HC and UC isolates we identified 15 and 16 different O types respectively, with O18 (11%), O6 (11%) and O2 (7%), being the most common in HC compared with O1 (6%), O7 (6%) and O15 (6%) in UC. Twenty-four (24), fifteen (15) and sixteen (16) H types were identified in CD, HC, and UC isolates, respectively. H7 is the most common in all three groups (17% in CD, 15% in HC and 32% in UC), followed by H4 (13%) and H1 (10%) in CD; H5 (11%) and H6 (11%) in HC and H1(10%) in UC isolates. The most common serotype in CD isolated is O1: H7 (8/88; 9%) (5 out of these eight isolates belong to ST-95). In HC, the most abundant serotype is O18: H7 (10%) all of which belong to AIEC and ST-95. H7 (13%) is the most common serotype in UC isolates. When we compared the O type of AIEC

(n = 15) and non-AIEC *E. coli* isolates (n = 19), there was a significant difference of O18 type that comprised 27% of AIEC strains (p = 0.029) and 0% of the non-AIEC strains. In the AIEC group, the most abundant serotype is O18: H7 (4/15; 27%) all of which belong to ST-95 and are significantly associated with AIEC (p = 0.029) compared with non-AIEC. The serotypes of lymph node isolates were diverse, although two isolates with different STs from the same patient belonged to O8: H9. Twenty per cent (20%) lymph node isolates had serotype O8: H49. Serotypes of *E. coli* in aphthous ulcer were also highly diverse (O11: H45, O149: H1, O17/O44: H18, O25: H4, O46: H31, O50/O2: H1 and O9: H25) and there was no serotype in common to the 7 aphthous ulcer isolates.

3.4.3 Virulence properties

In silico analysis revealed that all the isolates in CD, HC and UC carried at least one virulence gene. The list of virulence genes carried by the isolates in three groups is shown in Table 3.5. Only *cnf1* (cytotoxic necrotizing factor) was significantly associated with HC isolates (p = .0107) (in these isolates 4/7 were AIEC) compared to CD isolates and *ireA* (siderophore receptor) was significantly associated with CD (combined adult and paediatric) (p = 0.0417), paediatric CD (p = 0.017) and UC (p = 0.0288) compared to HC. Virulence genes present in >10% of the CD isolates included gad (94%), *iss* (66%), *lpfA* (33%), *senB* (34%), vat (34%), *air* (16%), *iroN* (23%), *eilA* (16%), *iha* (17%), *ireA* (20%), *sat* (16%), *mchF* (18%) and *astA* (10%). None of the CD isolates harboured *eae*, *espA*, *espJ*, *espF*, *sfaS*, *tir*, *nleA* and *nleB*. Virulence factor *sfaS* (S-fimbriae minor subunit) was significantly associated (p = 0.0294) with AIEC compared to non-AIEC isolates. No virulence factor was overrepresented in aphthous ulcer-associated *E. coli* in CD patient although air (43%), *eilA* (43%), *iroN* (43%), *iss*

(85%), *lpfA* (43%) and *mchF* (43%) in CD *E. coli* was higher in proportion compared to air (22%), *eilA* (22%), *iroN* (26%), *iss* (52%), *lpfA* (25%) and *mchF* (11%) in *E. coli* from HC. Similarly, we did not find any significant association of virulence factors to lymph node isolates; however, we found a higher proportion of *lpfA* in lymph node isolates (56%) compared with HC (26%) isolates (p = 0.12). The 100% stack graph (Figure 3.2) shows the relative abundance of different virulent genes in CD, HC, and UC. Details information of the presence and absence of virulence genes in *E. coli* isolates are also shown in Table 3.6.

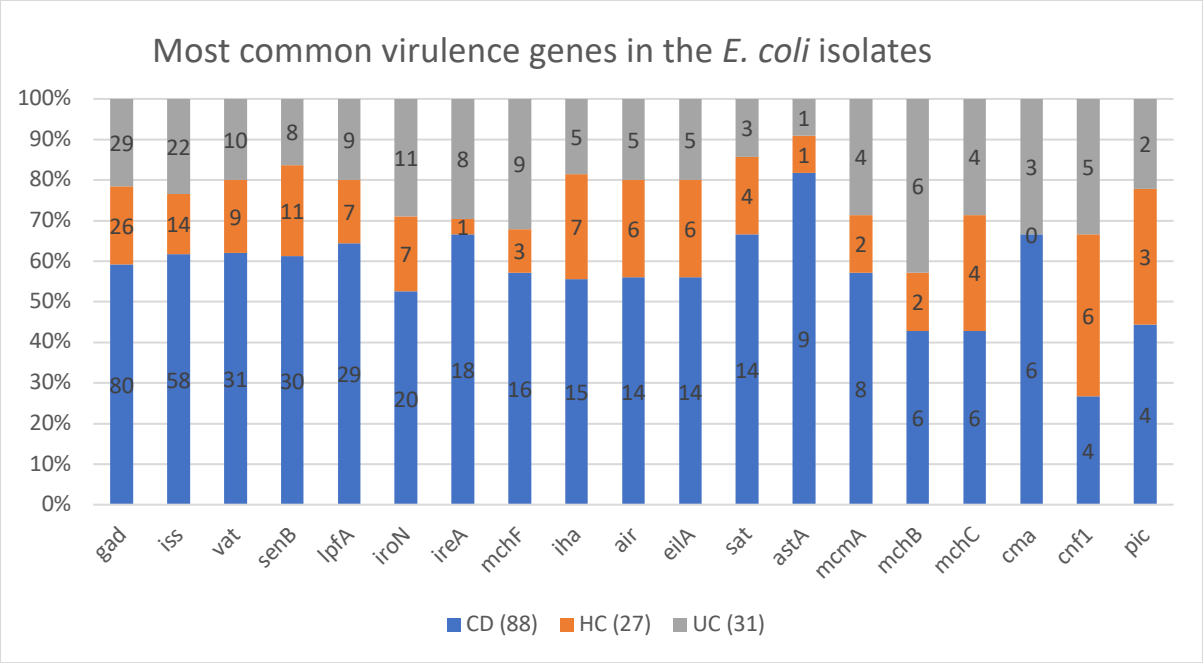


Figure 3.2. Stacked bar plot showing the abundance of virulence factors of *E. coli* isolates in in each group (CD, HC, and UC).

Table 3.6. Virulence factors of *E. coli* isolates in lymph nodes, aphthous ulcers, AIEC, non-AIEC isolates and HC

Virulence factor	LN (n = 9)	Aphthous Ulcer (n = 7)	AIEC (n = 15)	Non-AIEC (n = 19)	HC (n = 27)
iroN	2 (22%)	3 (43%)	6 (40%)	7 (37%)	7 (26%)
lpfA	5 (56%)	3 (43%)	3 (20%)	4 (21%)	7 (26%)
gad	9 (100%)	5 (71%)	15 (100%)	15 (79%)	26 (96%)
iss	4 (44%)	6 (86%)	9 (60%)	14 (74%)	14 (52%)

senB	1 (11%)	1 (14%)	6 (40%)	8 (42%)	11(41%)
vat	2 (22%)	2 (29%)	7 (47%)	11 (58%)	10 (37%)
iha	0	2 (29%)	1 (7%)	4 (21%)	7 (26%)
cnf1	0	0	5 (33%)	2 (10%)	6 (22%)
eilA	0	3 (43%)	0	1 (5%)	6 (22%)
air	1 (11%)	3 (43%)	0	1 (5%)	6 (22%)
mcmA	0	2 (29%)	2 (13%)	0	4 (13%)
sat	0	2 (29%)	1 (7%)	4 (21%)	4 (13%)
mchC	0	2 (29%)	2 (13%)	1 (5%)	4 (13%)
mchF	2 (22%)	3 (43%)	2 (13%)	5 (26%)	3 (11%)
pic	0	0	2 (13%)	1 (5%)	3 (11%)
sfaS	0	0	4 (27%)	0	3 (11%)
mchB	0	2 (29%)	3 (20%)	1 (5%)	2 (7%)
astA	1 (11%)	1 (14%)	0	1 (5%)	1 (4%)
ireA	2 (22%)	0	2 (13%)	6 (32%)	1 (4%)
tir	0	0	0	0	1 (4%)
nleB	0	0	0	0	1 (4%)

3.4.4 Plasmid analysis of the *E. coli* isolates

We have taken two different approaches for plasmid analysis. First approach is based on plasmid finder tool in Centre for Genomic Epidemiology (CGE) and the second approach is by using a combined method of three different tools: Plascope, MLplasmids, and MobTyper.

3.4.4.1 Plasmid profiles of *E. coli* isolates based on plasmid finder tool

No significant difference was seen in CD isolates (combined adult CD and paediatric CD) compared to HC isolates. However, Col156 plasmid was significantly associated with healthy control isolates ($p = 0.022$) and paediatric CD isolates ($p = 0.0033$) compared to adult CD. Plasmid IncFII (29) was also significantly associated with

paediatric CD isolates compared to healthy control (p = 0.0214). In healthy controls, this IncFII (29) plasmid was found only on three isolates out of 27, and two of them were an AIEC/non-AIEC pair from same patients with the same ST and varying by only 150-200 gene. So, if we consider these two as the same, the significance will be higher (p = 0.0065). Plasmid IncFIB (AP001918) was significantly associated (p = 0.0143) with non-AIEC compared to AIEC. The plasmid content was not significantly associated with lymph node and aphthous ulcer isolates in CD compared to the HC or UC isolates. However, IncFIB (AP001918) was higher in proportion in AU (71%) compared with HC (48%) although the difference was not statistically significant. Distribution of plasmids in each *E. coli* strain is shown in Table 3.7 and Table 3.8.

Table 3.7. Plasmid distribution in *E. coli* isolates among different groups

Plasmid	**CD (all) (n = 88)**	**Adult CD (n = 65)**	**Paediat -ric CD (n = 23)**	**HC (n = 27)**	**UC (n = 31)**	**p value**
Col (BS512)	3	3	0	0	0	NS
Col (MG828)	4	2	2	1	1	NS
Col156	27	14	13	13	8	0.022 (associated with HC compared to adult CD) and 0.0033 (associated with paediatric CD compared to adult CD)
ColpVC	2	1	1	0	3	NS
IncB/O/K/Z	9	5	4	2	2	NS
IncFIA	4	1	3	2	2	NS
IncFIB (AP001918)	48	33	15	13	19	NS
IncFIB (K)	2	2	0	1	1	NS
IncFIB (pB171)	4	2	2	2	1	NS
IncFIC (FII)	5	2	3	1	2	NS

IncFII	17	11	6	10	9	NS
IncFII (29)	24	14	10	3	6	0.0214 (associated with paediatric CD compared to HC)
IncFII (pHN7A8)	3	3	0	0	2	NS
IncFII (pRSB107)	2	0	2	2	2	NS
IncI1	12	12	0	2	5	NS
IncI2	7	7	0	0	1	NS
IncX4	5	5	0	0	1	NS
IncY	2	2	1	1	5	NS
no plasmid	13	13	0	6	7	NS
p0111	1	1	0	1	1	NS
IncFII (pCoo)	2	2	0	0	0	NS
IncFII (pCRY)	1	1	0	0	1	NS
IncN	2	2	0	0	0	NS
IncP1	1	1	0	0	1	NS
IncQ1	2	1	1	2	0	NS
IncHI1A	1	1	0	0	0	NS
IncHI1B (R27)	1	1	0	0	0	NS
IncL/M (pOXA-48)	1	1	0	0	0	NS
ncFIB (AP001918)	1	1	0	0	0	NS
Col440I	3	0	3	0	0	NS
Col (IMGS31)	0	0	0	1	0	NS

NS-Not significant ($p > 0.05$)

Table 3.8. Distribution of plasmid in AIEC and non-AIEC isolates

Plasmids	**AIEC (15)**	**non-AIEC (18)**	**P value**
Col156	6	8	NS
IncFIB (AP001918)	6	15	0.0143 (significantly associated with non-AIEC compared to AIEC)
IncFII	6	7	NS

IncFII (29)	2	5	NS
IncI1	2	3	NS
IncX4	0	3	NS
no plasmid	2	1	NS

NS-Not significant ($p > 0.05$)

3.4.4.2 Identified plasmids in *E. coli* isolates using Plascope, MLplasmids, and MobTyper

Bacterial plasmids often possess genes associated with pathogenicity. To search for candidate plasmid contigs we investigated different isolates from all groups (adult CD, paediatric CD, UC, and healthy control) independently. Of the 146 strains (65 adult CD, 23 paediatric CD, 31 UC and 27 control) we detected 11 plasmids of various lengths (2–293 kb) and origins. Some strains harboured plasmids whose closest homolog are pathogenic bacteria, e.g. uro-pathogenic and enterohemorrhagic *E. coli*, Klebsiella pneumoniae, Shigella and Salmonella, with sequence coverage of more than 80% and similarity between 95 –100 % (For example *Escherichia coli* UTI89 plasmid pUTI89 is highly similar to *Escherichia coli* Eco889 plasmid pECO-fce, *Escherichia coli* LBV005/17 plasmid pR005, *Escherichia coli* ECONIH2 plasmid pECO-bc6, and *Escherichia coli* SF-166 plasmid pSF-166-1, therefore we counted those plasmids as the same for further analysis).

3.4.4.2.1 Plasmids in healthy controls

A total of 54 different types of plasmids were found. Among these, plasmids mostly originated from *E. coli* (42), followed by *Salmonella* (5), *shigella* (3), *Klebsiella* (Polito et al.) and *Enterobacter* (Polito et al.). The most common plasmids were the UPEC 26-1 plasmid and the PA45B plasmid pPA45B, which are highly similar (99.92% identity and 87% query cover). Plasmid pPA45B is an IncF-type conjugative plasmid

(147,172 bp; 51.6% GC content) contains 177 predicted protein-coding genes and three genes associated with antibiotic resistance (aadA1, blaTEM-1B, and sul1). These plasmids were found in 7 (26%, n = 7) HC strains from 5 different individuals. Other common plasmid in healthy control isolates are *Shigella flexneri* 1a 0228 plasmid (n = 4; 15%), *Escherichia coli* HUSEC2011 plasmid pHUSEC2011-1/ *Escherichia coli* plasmid Plm (n = 3; 11%) *Escherichia coli* JJ1886 plasmid pJJ1886_3 (n = 3, 11%), *Escherichia coli* UTI89 plasmid pUTI89/ *Escherichia coli* UM146 plasmid pUM146/ *Escherichia coli* SF-166 plasmid pSF-166-1(n = 2, 7%), *Escherichia coli* AR_0162 plasmid tig00002623 (n = 2; 7%) *Shigella sonnei* FORC_011 plasmid pFORC11.2 (n = 2;7%), *Escherichia coli* Ecol_743 plasmid pEC743_1/ *Escherichia coli* MNCRE44 plasmid pMNCRE44_6 (n = 2; 7%), *Escherichia coli* plasmid pEC3I (n = 2; 7%), *Escherichia coli* plasmid p838C-R1(n = 2, 7%).

3.4.4.2.2 Plasmids in adult CD isolates

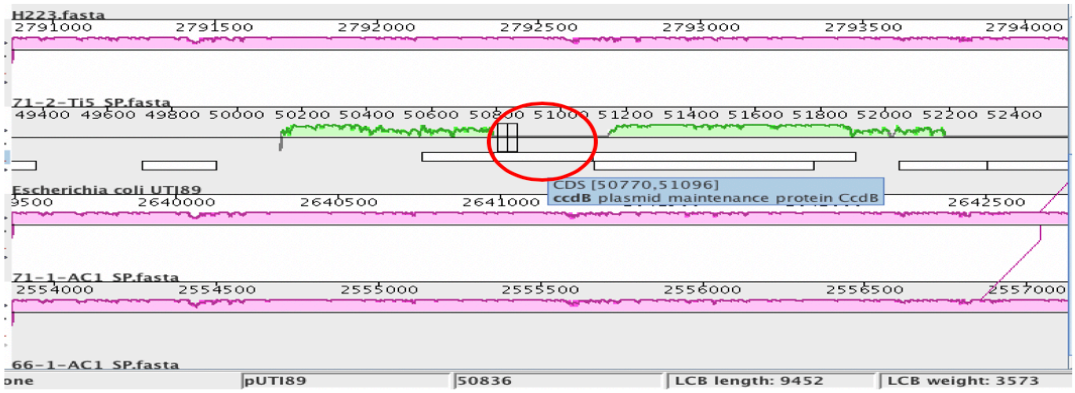
We identified 150 different types of plasmids in the 65 adult CD strains. However, we found that six plasmids were very similar: 99% identity and 97% query cover for *Escherichia coli* UTI89 plasmid pUTI89. the similar plasmid of *Escherichia coli* UM146 plasmid pUM146, *Escherichia coli* Eco889 plasmid pECO-93a are *Escherichia coli* LBV005/17 plasmid pR005, *Escherichia coli* SF-166 plasmid pSF-166-1 and *Escherichia coli* UMN026 plasmid p1ESCUM. Therefore, we considered those plasmids to be the same during this analysis. We found that *Escherichia coli* UTI89 plasmid pUTI89 and its closely similar plasmids are significantly associated (p = 0.032) with adult CD isolates (n = 15; 23%) compared to healthy controls (n = 1; 4%). The other common plasmids in adult CD are *Escherichia coli* 1/23 plasmid pKST23 (n = 10; 15%), *Escherichia coli* HUSEC2011 plasmid pHUSEC2011-1/ *Escherichia coli* plasmid Plm (n = 9; 14%), *Escherichia coli* AR_0162 plasmid tig00002623 (n = 9;14%),

Escherichia coli pTC plasmid (n = 8; 12%), *Escherichia coli* M11 plasmid B (n = 6; 9%), *Escherichia coli* JJ1886 plasmid pJJ1886_3 (n = 5; 8%), *Escherichia coli* plasmid pRK100 RepFIB replication region (n = 5; 8%), and *Shigella flexneri* 1a 0228 plasmid (n = 5; 8%). However, these plasmids are not significantly associated with adult CD isolates.

3.4.4.2.3 Plasmids in paediatric CD isolates

We found 51 different types of plasmids in the 23 paediatric CD strains. However, some plasmids share more than 97% query cover and 99% identity. Like adult CD isolates, paediatric CD isolates were found to be significantly associated (n = 9; p = 0.0034) with the *Escherichia coli* UTI89 plasmid pUTI89 and its closely similar plasmids like *Escherichia coli* UM146 plasmid pUM146, *Escherichia coli* Eco889 plasmid pECO-93a, *Escherichia coli* LBV005/17 plasmid pR005, *Escherichia coli* SF-166 plasmid pSF-166-1 and *Escherichia coli* UMN026 plasmid p1ESCUM. Other common plasmids are *Escherichia coli* pTC plasmid (n = 4, 17%), *Escherichia coli* HUSEC41 plasmid pHUSEC41-1(n = 3; 13%), and *Escherichia coli* SF-088 plasmid pSF-088-1 (n = 3; 13%).

Mauve alignment of healthy control isolates with pUTI89 plasmid



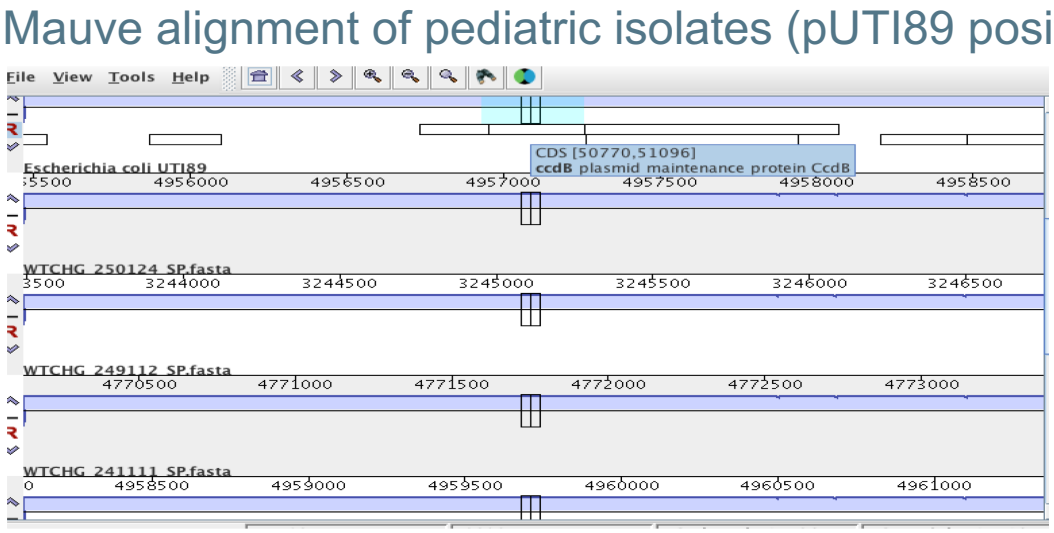


Figure 3.3. Mauve alignment showing that *Escherichia coli* UTI89 plasmid pUTI89 in the paediatric CD isolates. In healthy controls this plasmid was absent in most of the isolates. Mauve alignment also showing that *ccdB* gene that was found significantly associated with the CD isolates was also present within this pUTI89 plasmids in paediatric CD isolates.

Escherichia coli UTI89 plasmid pUTI89 is of 114230 bp length and contains 145 CDs. This plasmid mostly represented by replication proteins, ABC transporter, stable plasmid inheritance protein, F fimbriae pilin precursor pilus assembly protein, DNA binding protein, fertility inhibitor protein, site specific recombinase, proteins involved in DNA maintenance and conjugation, and possible virulence determinants such as enterotoxin protein, outer membrane colicin receptors, thioredoxin family protein, resistance proteins, etc.

3.4.4.2.4 Plasmid in ulcerative colitis isolates

This plasmid analysis revealed 78 different types of plasmids in ulcerative colitis strains. We did not find any plasmid significantly associated with the ulcerative colitis isolates compared to the healthy control. However, some plasmids are found in higher percentage such as, *Escherichia coli* Ecol_422 plasmid pEC422_4 (n = 5, 16%), *Escherichia coli* PSUO2 plasmid pPSUO2 (n = 4; 13%), *Escherichia coli* SF-166

plasmid pSF-166-1/ *Escherichia coli* UTI89 plasmid pUTI89 (n = 4; 13%) and *Escherichia coli* M11 plasmid B (n = 3; 10%).

3.4.4.2.5 Plasmids in lymph node CD isolates

In silico analysis revealed that 37 different types of plasmids are found in lymph node isolates. *Escherichia coli* HUSEC2011 plasmid pHUSEC2011-1 / *Escherichia coli* plasmid Plm (closely similar) was found in 33% of the LN isolates. However, this plasmid is not significantly associated with LN isolates compared to healthy control.

3.4.4.2.6 Plasmids in aphthous ulcers (AU) isolates

Plasmid analysis revealed 37 different plasmid types in AU isolates. *Escherichia coli* 1/23 plasmid pKST23 was significantly associated with AU isolates (n = 3, 43%; p = 0.021) compared to HC isolates (n = 1, 4%). Similarly, *Escherichia coli* AR_0162 plasmid tig00002623 was also found to be overrepresented in AU isolates (n = 3, 43%; p = 0.046) compared to healthy control (n = 2, 7%). In addition to those, *Escherichia coli* HUSEC2011 plasmid pHUSEC2011-1/ *Escherichia coli* plasmid Plm and *Escherichia coli* 2009C-3133 plasmid was also found in higher proportion (n = 3, 29%) in AU isolates. However. these plasmids are not significantly associated with AU isolates.

3.4.4.2.7 Plasmid in AIEC and non AIEC isolates

We found different plasmids in 15 AIEC and 18 non-AIEC isolates. *Escherichia coli* PA45B plasmid pPA45B/ *Escherichia coli* UPEC 26-1 plasmid was found in 27% (n = 4 out of 15) of the AIEC isolates and 11% (n = 2 out of 18) of the non-AIEC isolates. *Escherichia coli* HUSEC2011 plasmid pHUSEC2011-1/ *Escherichia coli* plasmid Plm, *Escherichia coli* M11 plasmid B and *Escherichia coli* SF-166 plasmid pSF-166-1/

Escherichia coli UTI89 plasmid pUTI89 was found in 13% (n = 2 out of 15) of AIEC isolates and 17% (n = 3 out of 18), 5% (n = 1 out of 18) and 22% (n = 4 out of 18) of non-AIEC isolates respectively. *Escherichia coli* JJ1886 plasmid pJJ1886_3, *Escherichia coli* 1/23 plasmid pKST23 and *Escherichia coli* PSUO2 plasmid pPSUO2 was found in 22% (n = 4 out of 18) of non-AIEC isolates compared to 7%, 7% and 0% of the AIEC isolates respectively. Other plasmids were found in very few AIEC and non-AIE isolates. Among the healthy control isolates 43% are AIEC and 29% are non-AIEC and 29% strains are from unknown phenotype (AIEC/non-AIEC). *Escherichia coli* PA45B plasmid pPA45B/ *Escherichia coli* UPEC 26-1 plasmid was found both in the AIEC and non-AIEC pair of the same individual that differ only by 150 genes (12-1-Ti12 and 12-2-Ti13) and are therefore not associated with AIEC.

3.4.5 Presence of conjugative elements in different groups of isolates

CONJScan a conjugative elements detection tool, allows the detection of conjugative systems in genomes (Guglielmini et al., 2014). We used this tool to identify the conjugative system within the isolates and to compare between different groups. A conjugative system is defined by the presence of a relaxase, a type 4 coupling protein and a type 4 secretion system. There are 8 different types, corresponding to the phylogeny of the sole ubiquitous protein of the T4SS, Virb4. The 8 different types may be associated with a particular clade: such as – type B: Bacteroidetes – type C: Cyanobacteria – types T/ I / F/ G: Proteobacteria – type FA: Firmicutes / Actinobacteria – type FATA: Firmicutes / Actinobacteria / Tenericutes / Archea.

Our focus of interest was conjugative systems Type T, I, F and G, as those are particularly associated with *E. coli*. In this collection of *E. coli* different types of conjugative elements were identified. These are listed in the table below. Aphthous

ulcer-associated CD isolated were found to be significantly associated with type T conjugative elements compared with healthy isolates (p = 0.0035). However, no significant difference was observed in the other types of conjugative element among the different groups (Table 3.9 and Figure 3.4).

Table 3.9. Different types of conjugative elements in *E. coli* isolates

Disease (total number of samples)	Type T	Type I	Type F	Type G	No conjugative element
Adult CD + paediatric CD (n = 88)	19	16	39	1	29
Adult CD (n = 65)	16	11	25	1	22
Paediatric CD (n = 23)	3	5	14	0	7
HC (n = 27)	1	4	11	0	13
UC (n = 31)	5	5	11	0	12
LN-CD (n = 9)	2	2	3	0	4
AU-CD (n = 7)	3	1	4	0	1
AIEC (n = 15)	0	1	7	1	5

*AIEC: Adherent invasive *E. coli*, non-AIEC: non-Adherent invasive *E. coli*, AU: Aphthous ulcer, LN: Lymph node, and HC: Healthy controls.

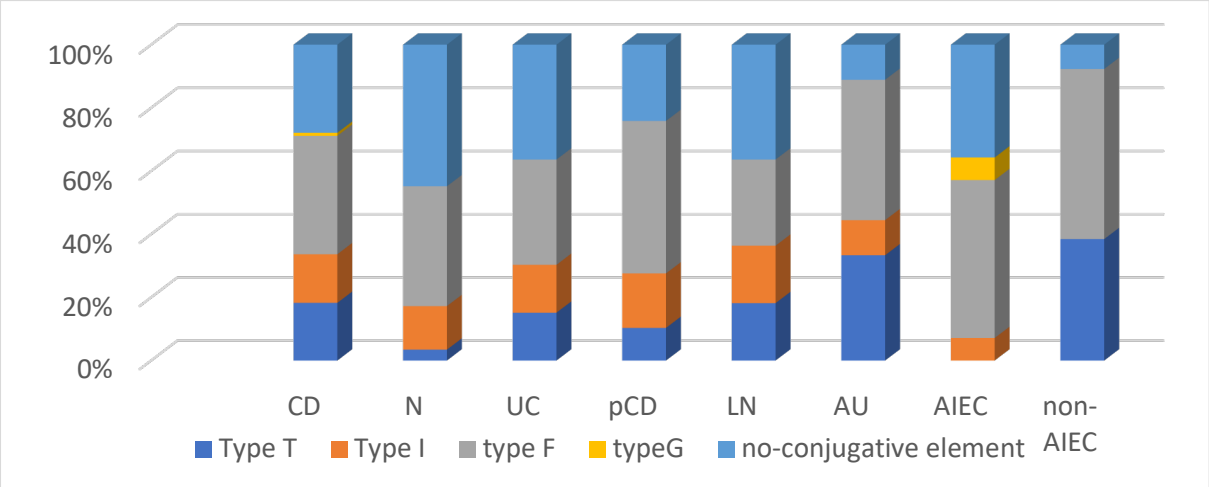


Figure 3.4. Stacked bar plot showing various conjugative elements (Type T, I, F, G) in different groups of *E. coli* isolates.

3.4.6 Antibiotic resistance in *E. coli* isolates

To ensure a comprehensive analysis of all known resistance determinants, we used the publicly available resistance gene database, ResFinder [Centre for Genomic Epidemiology (CGE), DTU]. We found a total of 23 genes specific to 15 different groups of antimicrobials. Based on gene presence/absence, the vast majority of the 146 strains were sensitive to all antimicrobials (66%; n = 58 in CD isolates, 74%; n = 20 in HC isolates; 61%, n = 19 in UC isolates; 78%, n = 7 in lymph-node isolates and 43%, n = 3 in aphthous ulcers isolates). In CD (adult CD + paediatric CD) isolates, the most common resistance genes were for beta-lactam antibiotics (20%; n = 18) and aminoglycosides (20%; n = 18), followed by sulphonamide (15%; n = 13), tetracycline (14%; n = 12), trimethoprim (13%, n = 11), macrolide (3%; n = 3) and phenolics (2%; n = 2). Similarly, in UC isolates, the highest antimicrobial resistance was observed against beta-lactam antibiotics (23%; n = 7) followed by aminoglycosides (19%; n = 6), sulphonamides (16%; n = 5), tetracyclines and trimethoprim (10%; n = 3), and macrolides (10%; n = 2).Very few paediatric CD isolates (39%; n = 9) were sensitive to antibiotics compared with healthy controls (HC) (74%; n = 20) and the difference was statistically significant (p = 0.02. Higher resistance to beta-lactam (48%; n = 11), aminoglycoside (48%, n = 11), tetracycline (26%, n = 6), sulphonamide (35%, n = 8) and trimethoprim (22%, n = 5) antibiotics has been observed in paediatric CD isolates than in controls, but the difference was not statistically significant. Of the lymph node isolates, seven out of nine were not resistant to any antimicrobials, and two of them were resistant to antimicrobials; one against beta-lactam, the other against nalidixic acid and ciprofloxacin). Among the aphthous ulcer isolates, two were multidrug-resistant against beta-lactam, aminoglycoside, tetracycline, and trimethoprim; two were resistant to quinolone and fluoroquinolones and the remaining three isolates

were not resistant to any of the antibiotics. A list of the resistance genes and their frequencies are shown in Table 3.10. The most frequent gene among the isolates is blaTEM-1B (14% of adult CD, 31% of paediatric CD, 30% of HC and 23% of UC isolates) followed by aph (3")-Ib (9% of adult CD, 35% of paediatric CD, 22% of HC and 10% of UC patients) and sul2 (11% of adult CD, 26% of paediatric CD and HC, and 14% of UC isolates). Gene aph (3")-Ib, blaTEM-1B and sul2 are found in higher proportion in 35 %, 30% and 26% of paediatric CD isolates compared with 9%, 14% and 10% of adult CD isolates respectively.

Table 3.10. Antibiotic resistance properties for IBD and non-IBD isolates and different tissue types

Resistance gene	**Resistance against**	**CD (n = 88)**	**HC (n = 27)**	**UC (n = 31)**	**P- CD (n = 23)**	**LN (n = 9)**	**AU (n = 7)**
No resistance	-	58	20	19	9	7	3
tet(A)	Tetracycline	6	2	2	3	0	1
tet(B)	Tetracycline	5	3	1	3	0	1
tet(D)	Tetracycline	1	0	0	0	0	0
sul1	Sulphonamide	4	7	2	2	0	1
sul2	Sulphonamide	13	4	4	6	0	1
aadA1	Aminoglycoside	3	6	3	2	0	0
aadA5	Aminoglycoside	2	0	1	0	0	1
aph(3')-Ia	Aminoglycoside	1	2	0	0	0	0
aph (6)-Id	Aminoglycoside	11	4	3	5	0	1
aph (3")-Ib	Aminoglycoside	14	6	3	8	0	1
blaCTX-M-15	Beta-lactam	1	0	0	0	0	0
blaOXA-1	Beta-lactam	0	1	0	0	0	0
blaTEM-1A	Beta-lactam	2	0	0	2	0	0
blaTEM-1B	Beta-lactam	16	8	7	7	1	2
blaTEM-1C	Beta-lactam	3	2	0	2	0	0
erm(B)	Macrolide	1	0	0	0	0	0
mph(A)	Macrolide	1	0	2	0	0	0
dfrA1	Trimethoprim	3	0	0	2	0	0
dfrA14	Trimethoprim	1	0	2	0	0	0
dfrA17	Trimethoprim	2	0	1	0	0	1
dfrA5	Trimethoprim	3	1	0	2	0	1
dfrA7	Trimethoprim	2	1	0	1	0	0

catA1	Phenicol	2	2	0	1	0	0
Mutation par E p.I355T par C p. S57T	Nalidixic acid and Ciprofloxacin	1	0	0	0	1	0
Mutation in parEp.I529L parEp.D475E	Quinolones and fluoroquinolones	1	0	0	0	0	1

3.4.7 Capsule typing

Capsules are surface-enveloping structures synthesized by diverse pathways. Most of the capsules are high molecular weight polysaccharides produced by repeat units of oligosaccharides. In Group I or Wzx/Wzy-dependent pathway capsule synthesis, the oligosaccharide repeat units are associated with a phosphate acceptor in the cytoplasm by membrane-bound glycosyltransferases. This precursor is then transferred across the inner membrane by the Wzx flippase and by the Wzy polymerase. The Group II and Group III capsule polysaccharides are linked to a phospholipid acceptor and transported to the inner membrane by the ATP-binding cassette (ABC) transporter. Group II and III capsules are referred to as ABC-dependent capsules. Polysaccharides group IV capsule synthesis is associated with Wzy polymerase but not with Wzx flippase. The export machinery of group IV capsules is also very diverse, which in some cases associated with proteins homologous to those of Group I. Synthase-dependent pathway can also produce polysaccharide capsules by unique enzyme systems that are responsible for the initiation, polymerization and translocation of the capsule (Rendueles et al., 2017).

In the present study, the most common capsule type in all three groups was Wzy_stricte found in 97% (n = 63) of adult CD strains, 100% (n = 23) of paediatric CD patients, 100% (n = 31) of UC stains, and 96% (n = 26) of control strains (Table 3.11).

ABC-dependent capsules (groups II and III) were significantly associated with paediatric CD compared with adult CD (p = 0.0001), UC (0.0001) and the HC (p = 0.0119). Similarly, Group IV_e capsule was significantly associated with lymph node CD isolates compared to control isolates (p = 0.0126) and paediatric CD isolates (p = 0.0362). The Wzy-dependent pathway is used for groups 1 and 4 capsules, which are often found in isolates causing gastrointestinal disease (Whitfield, 2006). No clear difference was observed among the different groups of isolates based on Group IV_s_stricte and Wzy_stricte group of capsules

Table 3.11. Distribution of capsules in different groups of *E. coli* strains

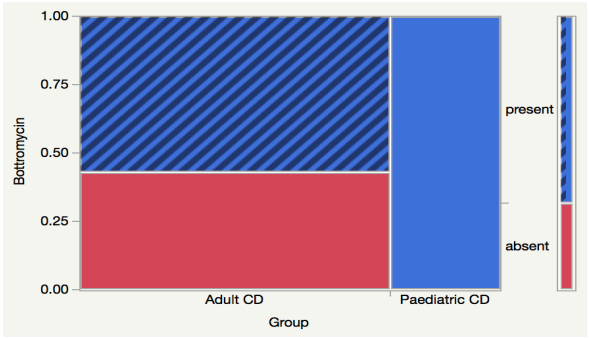
Capsule	**Adult CD (65)**	**UC (31)**	**HC (27)**	**Paediatric CD (23)**	**CD-all (88)**	**Lymph node (9)**	**Aphthous Ulcers (7)**
ABC	10 (15%)	6 (19%)	10 (37%)	17 (74%)	27 (31%)	2 (22%)	1 (14%)
GroupIV_e_stricte	29 (45%)	10 (32%)	6 (22%)	4 (17%)	33 (38%)	6 (67%)	2 (29%)
GroupIV_s_stricte	36 (55%)	15 (48%)	12 (44%)	11 (48%)	47 (53%)	5 (56%)	4 (57%)
Wzy_stricte	63 (97%)	31 (100%)	26 (96%)	23 (100%)	86 (98%)	9 (100%)	6 (86%)

*Note: In the first row of the table, numbers mentioned within the brackets indicate the total numbers of isolates in each group.

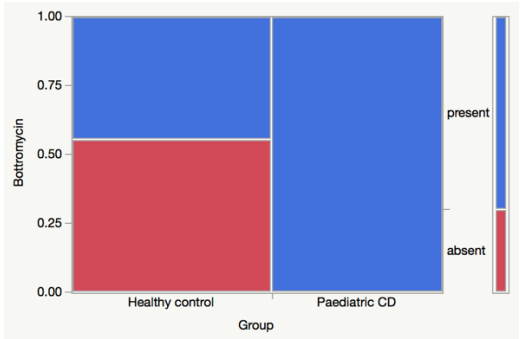
3.4.8 Bacteriocin production

Of the 146 strains, 127 strains (87%) possess at least one bacteriocin, including 85% of adult CD strains (n = 55); 100% of paediatric CD strains (n = 23); 81% of HC strains (n = 22); 90% of UC strains (n = 28), 100% of the LN strains (n = 8) and 86% of the aphthous ulcer isolates (n = 6). We found thirteen (13) different types of colicins and five different types of microcins in this study. In addition, we found other types of bacteriocins: Linocin_M18 bacteriocin, Mccl47_mchS2, sactipeptides, lasso-peptide and Bottromycin. Bottromycin was significantly associated with CD compared with healthy controls (p = 0.0396). However, bottromycin is strongly associated with

paediatric CD compared to adult CD (p = <0.0001), healthy controls (p = <0.0001) (Figure 3.5). and UC (p = <0.0001) (Figure 3.6) Bottromycin was also found in 88% and 85% of the lymph node, and aphthous ulcer strains respectively, and significantly associated with lymph node isolates compared to HC (p = 0.023) (Figure 3.6) and ulcerative colitis (p = 0.0152). Carocin_D was also significantly associated with paediatric CD (17/23; 74%) isolates compared to adult CD (p = 0.0015), HC (p = 0.0185), lymph node isolates (p = 0.0114) and aphthous ulcers isolates (p = 0.04).

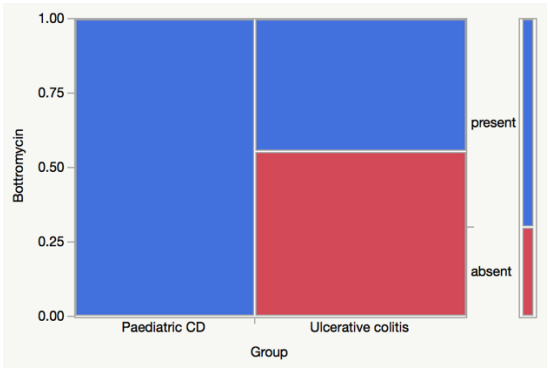


P = <0.0001; Probability of presence of bottromycin is greater for paediatric CD than adult CD

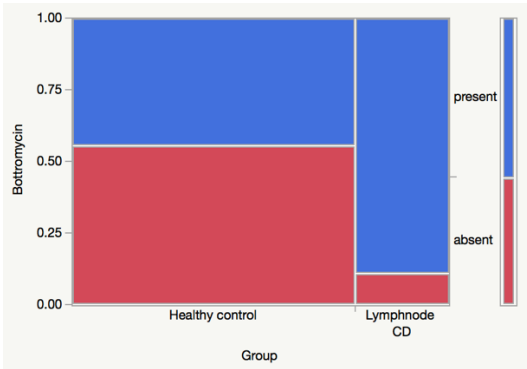


P = <0.0001; Probability of presence of bottromycin is greater for paediatric CD than healthy controls

Figure 3.5. Bottromycin bacteriocins are significantly associated with paediatric CD isolates.



P = <0.0001; Probability of presence of bottromycin greater for paediatric CD than UC



P = 0.0232; Probability of presence of bottromycin is greater for lymph node CD than healthy controls

Figure 3.6. Bottromycin bacteriocins are significantly associated with paediatric CD compared with ulcerative colitis and lymph node CD isolates compared with healthy controls isolates.

3.4.9 Prophage analysis

A total of 99 different prophages were identified in the 146 isolates. The most common prophage was PHAGE_Entero_lambda_NC_001416, found in 63% of adult CD (n = 42); 61% of the paediatric CD (n = 14); 44% of lymph node CD (n = 4); 29% of aphthous ulcers (n = 2); 70% of healthy control (n = 19) and 68% of the ulcerative colitis isolates (n = 21). The second most common prophage was PHAGE_Entero_mEp460_NC_019716, found in 48% of adult CD (n = 31); 57% of the paediatric- CD (n = 13); 44% of lymph node CD (n = 4); 29% of aphthous ulcers (n = 2); 67% of healthy controls (n = 18) and 55% of the ulcerative colitis isolates (n = 17). However, no prophage was significantly associated with IBD isolates compared to HC.

3.4.10 CRISPR analysis

CRISPR is ranging from 24–50 bp in length found in the DNA of many bacteria and archaea. They form the CRISPR-Cas System by CRISPRs and associated proteins (Cas) (Couvin et al., 2018).We identified one to three different CRISPR-associated (CRISP-Cas) protein systems in each isolate: CAS- type I-E, I-F and CAS as well as different combinations of the Cas genes in 57% (13/23) of the paediatric CD isolates, 71% (46/65) of adult CD isolates, 89% (8/9) of lymph node CD, 57% (4/7) of the aphthous ulcers isolates, 56% (15/27) of healthy control isolates and 68% (21/31) of UC isolates. However, no CAS type was significantly associated with any of the groups of IBD isolates. In addition to the CAS system, we identified 137 CRISPRs in 146 isolates ranging from 1-15 crisprs per isolate. No single isolate was devoid of a CRISPR. Only one CRISPR (sequence: CCACCTTTTTTACCTGCTTCAGATGC) was associated with paediatric CD (17/23, 74%) isolates compared to adult CD ($p = 0.0001$), UC ($p = 0.0008$) and HC ($p = 0.047$).

3.4.11 Phylogenetic Analysis

The phylogenetic tree was constructed to check if the isolates from similar groups (CD, UC, HC) formed any distinct cluster in the tree, or they are phylogenetically diverse in nature. Alongside using ITOL visualization of tree and molecular property together, we wanted to see if any molecular property is concentrated in a particular location of the tree or associated with a particular group of isolates in the tree. The phylogenetic tree shows that the IBD isolates are highly heterogenous and did not form any distinct cluster. Only a small cluster formed by the CD isolates (marked by a red triangle in the phylogenetic tree) belonged to B1, D and C phylogroups (Figure 3.7). However, even within this cluster there is still variety in phylogroup, serotype and plasmid etc (Figure 3.7). Overall, based on the phylogenetic analysis along with the molecular properties of the *E. coli* isolates, we unable to identify a CD profile based on different properties of the isolates (Figure 3.7, Figure 3.8 and Figure 3.9).

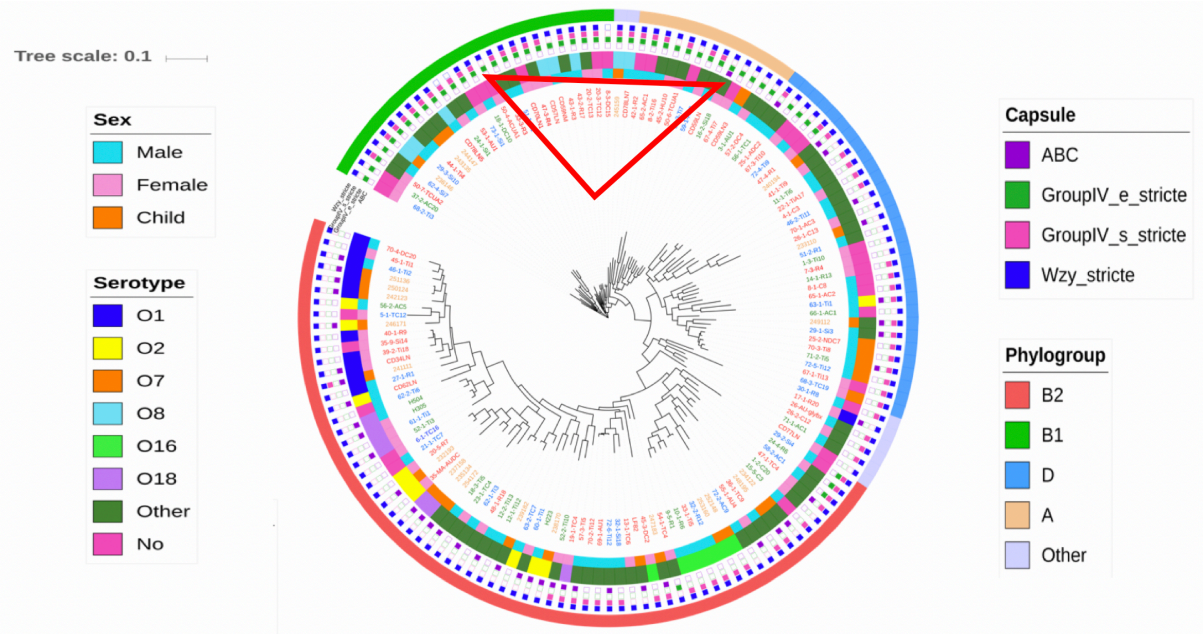


Figure 3.7. Phylogenetic tree of 146 *E. coli* isolates and their molecular properties (serotypes, capsules and phylogroups). The SNP tree was created with CSI phylogeny. Label colours of isolates indicates different groups; red: adult CD, orange: paediatric CD, green: healthy

controls, blue: UC. The innermost ring shows the sex of the patient, the second ring marks different serotypes, the third ring specifies the capsular types and the outermost ring indicates the phylogroups of the isolates.

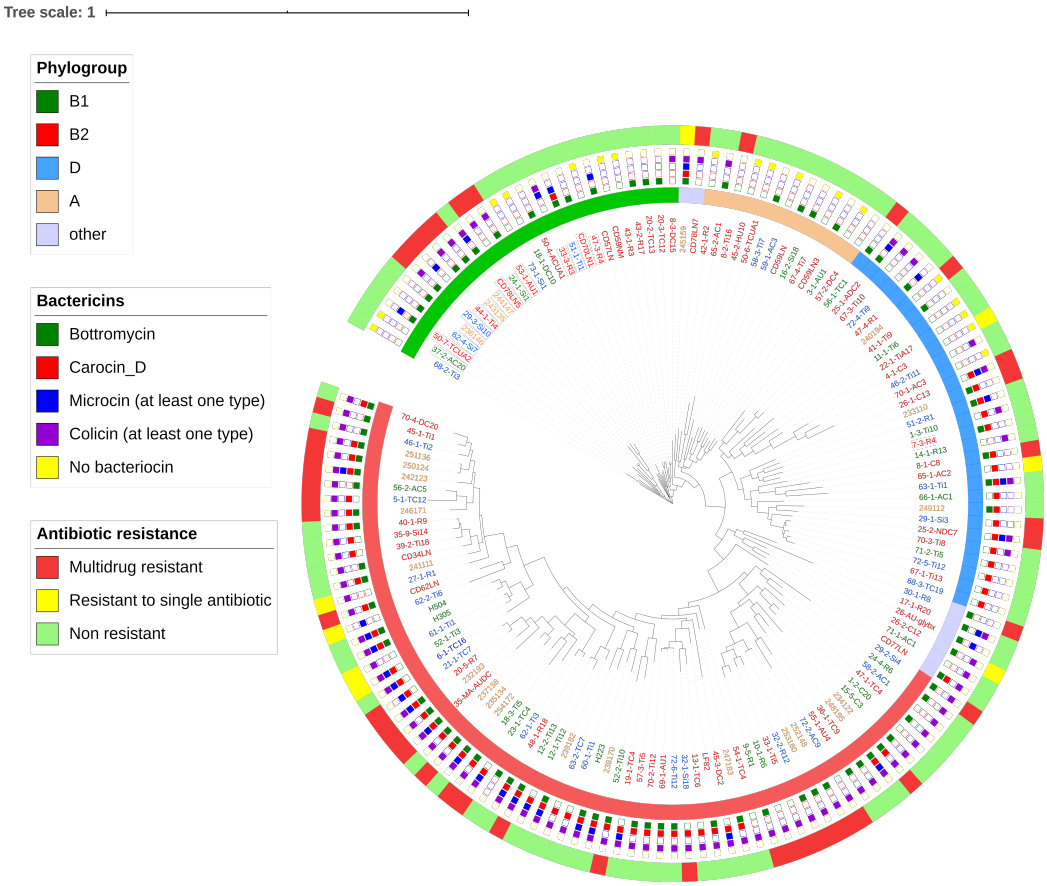


Figure 3.8. Phylogenetic analysis of 146 *E. coli* isolates and their molecular properties (phylogroup, bacteriocins and antibiotic resistance). The SNP tree was created with CSI phylogeny. Label colours of isolates indicates different groups; red: adult CD, orange: paediatric CD, green: healthy controls, blue: UC. The innermost ring specifies different phylogroups identified. The middle ring indicates bacteriocins, and the outer ring marks the antibiotic resistance property of the isolates.

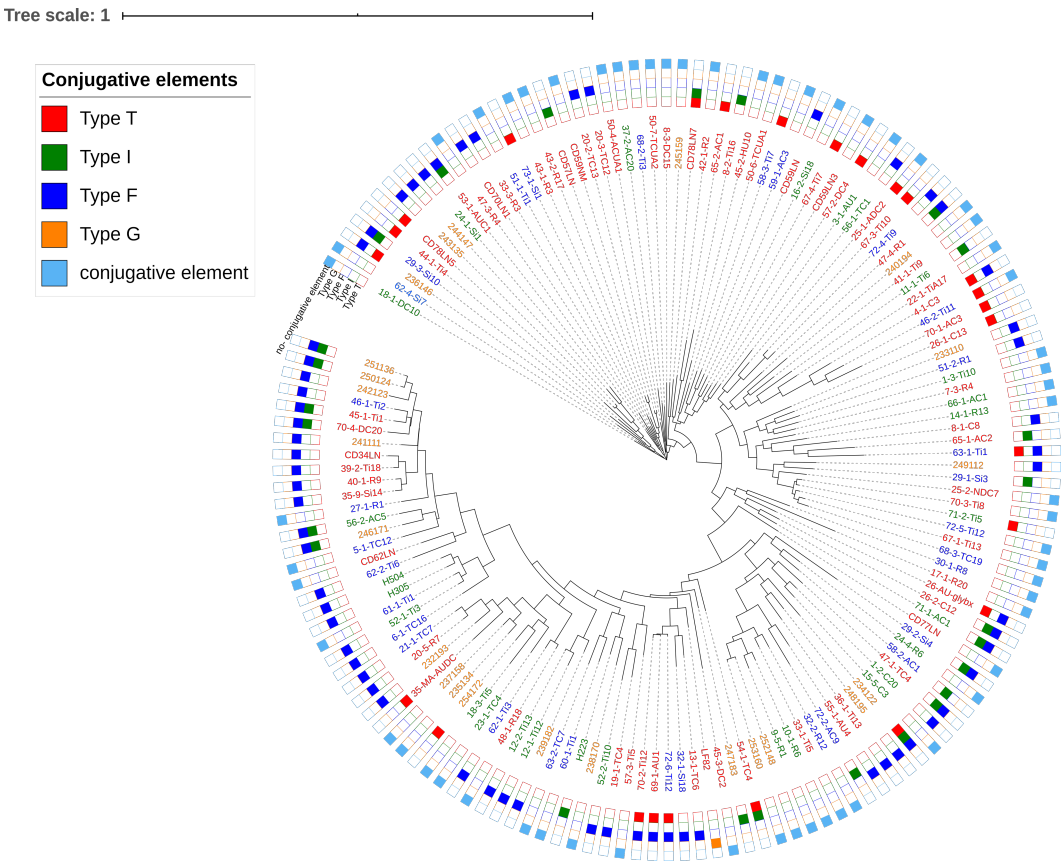


Figure 3.9. Phylogenetic analysis of 146 *E. coli* isolates and presence of conjugative elements. The SNP tree was created with CSI phylogeny. Label colours of isolates indicates different groups; red: adult CD, orange: paediatric CD, green: healthy controls, and blue: UC. The rings around the phylogenetic tree from inner to outer indicates the presence or absence of different conjugative elements.

3.4.12 Principal component analysis

Principal component analysis (PCA) a dimension-reduction method that transforms a broad set of correlated variables into a small set of uncorrelated variables called principal components. PCA is commonly used as a tool in exploratory data analysis and for making predictive models. We used PCA to visualize genetic distance, and relatedness between *E. coli* strains to belong to different groups (CD, UC paediatric- CD, lymph node CD, aphthous ulcers, and healthy controls). We performed the PCA based on the accessory gene content in the pan-genome data from all 146 *E. coli* isolates. Principal component analysis of all isolates revealed that

isolates from all the different groups (adult CD, paediatric CD, healthy controls, and ulcerative colitis) are scattered throughout the PCA plot and form three different clusters each containing isolates from all different groups. These clusters represent the phylogroups of the isolates. Cluster 1 includes Phylogroup D and F, Cluster 2 consists of B1, C and A, and Cluster 3 represents B2. The CD, UC and HC isolates are highly diverse and did not form any distinct cluster based on the disease phenotypes (Figure 3.10).

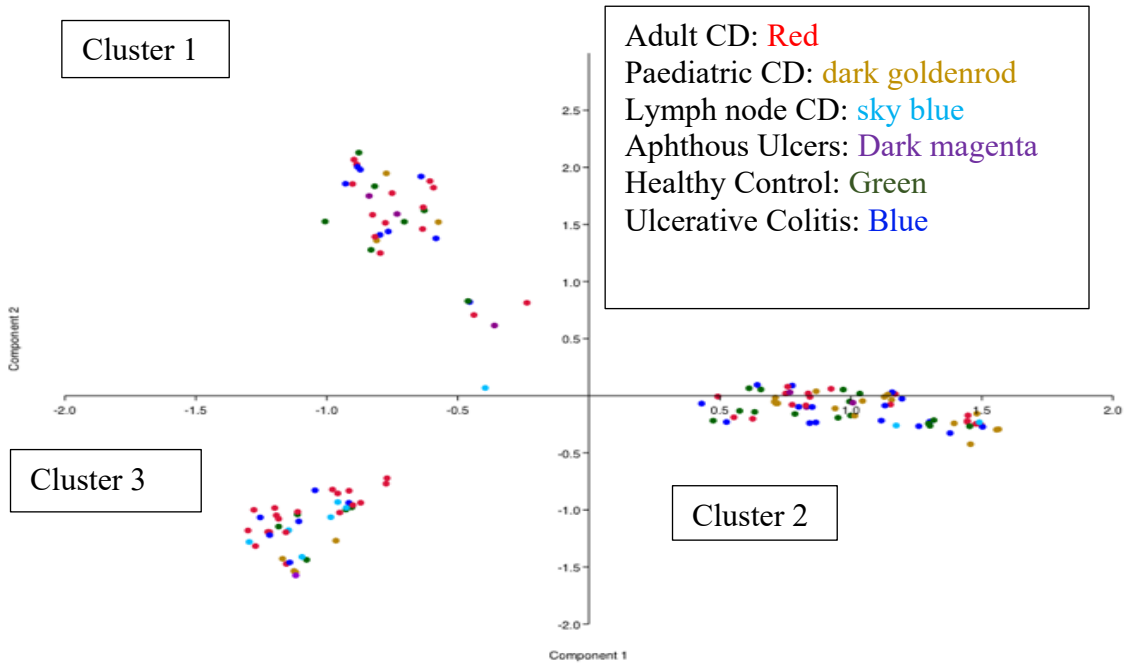


Figure 3.10. Principal component analysis (PCA) plot of *E. coli* isolates. Different coloured dots indicate groups of isolates, such as, red: adult CD, golden: paediatric CD, green: healthy control, blue: UC, sky blue: lymph node (LN) and dark magenta: aphthous Ulcers (AU).

3.4.13 Pan-genome analysis

The pan-genome is classified into a "core genes" that consists of genes present in all individuals, "shell genes" that are composed of genes present in two or more strains, and "cloud genes" that contains genes found only in a single strain. The Roary pangenome analysis revealed a total of 27,515 gene clusters. This total 27515 gene

clusters containing 21565 cloud genes, present in fewer than 21 strains; 2649 core genes, present in at least 144 strains; softcore genes, present in at least 138 strains but not more than 144 strains; and shell genes, present in at least 21 strains but not more than 138 strains.

3.4.14 Phenotype and genotype association study using Scoary

A Scoary analysis was performed to identify genes associated with different traits/phenotypes. This analysis revealed that there were some genes significantly associated with CD, UC, lymph node, aphthous ulcers, and paediatric CD compared with healthy control isolates.

3.4.14.1 Between CD and Healthy controls

Only genes significantly associated with CD isolates were considered, using a cut-off value of > 2.5 odds ratio (OR) and p value <0.05. A list of 72 genes were found to be significantly associated with CD isolates compared to healthy control (OR > 2.5 and naive P value < 0.05). Here most of the genes are associated with plasmid, fimbriae, horizontal gene transfer and hypothetical proteins. Among the 72 genes, we identified several virulence genes typically supporting adhesion, invasion and colonization of the epithelial cell, serum survival, and various plasmid associated genes, including genes associated with the toxin-antitoxin system, plasmid maintenance, conjugal transfer of DNA between bacterial cells, and post-segregational killing of plasmid-free cells. In addition, there were several chromosome and plasmid-associated conserved hypothetical genes, but the role of these genes is unknown. Of note, we found six genes associated with CD isolates belonging to operons encoding fimbrial structures, including type1 pili and long polar fimbriae. These genes include a variant of lpfA, lpfB, lpfC, papG, htrE, and variant of fimA. In addition to the fimbriae-associated genes,

there were 20 plasmid-associated genes associated with the CD isolates: repA, traJ, ccdB, ccdA, srp54/ ParA, parM/ stbA, stbB, ydeA/ ykfF, ydiA, traM, traP, traD, finO and some genes encoding hypothetical proteins. These genes are involved in plasmid maintenance/toxin anti-toxin system/ post-segregational killing of plasmid-free cells, conjugal transfer of DNA between bacterial cells, plasmid partitioning, plasmid replication and some hypothetical proteins (uncharacterized protein). Interestingly most of these genes are present with a perfect match in *E. coli* UTI89 plasmid pUTI89 isolated from the urinary bladder of patient with urinary tract infection (UTI). Other genes significantly associated with CD isolates are associated with phage tail protein, type III secretion apparatus protein OrgA, increased serum survival lipoprotein Iss, Lysozyme RrrD, Penicillin G acylase, IS2 element protein InsA, chromosomal replication initiation protein DnaA, H repeat-associated protein yhhI, ATPase and hypothetical proteins (Table 3.12).

Table 3.12. Significantly associated genes in CD compared with HC

Gene	Annotation	Number in adult CD	Number in HC	Odds ratio	Naive p
group_1084	Hypothetical protein	18	0	inf	0.006
yfjQ_5	Hypothetical protein	18	0	inf	0.006
group_3838	Hypothetical protein	18	0	inf	0.006
group_5263	Hypothetical protein	17	0	inf	0.01
yjbB	Hypothetical protein	88	24	inf	0.01
group_6236	Hypothetical protein	16	0	inf	0.02
group_4352	Hypothetical protein	16	0	inf	0.02
lpfB	Putative fimbrial chaperone LpfB	16	0	inf	0.02
lpfA_1	putative major fimbrial subunit LpfA	16	0	inf	0.02
lpfC'	putative outer membrane usher protein LpfC'	16	0	inf	0.02
ydjO_2	Hypothetical protein	13	0	inf	0.03
group_696	Hypothetical protein	13	0	inf	0.03
group_1282	Hypothetical protein	13	0	inf	0.03

group_2041	Hypothetical protein	13	0	inf	0.03
group_4405	Oxygen-regulated invasion protein OrgA	13	0	inf	0.03
group_15014	5-carboxymethyl-2-hydroxymuconate Delta-isomerase	14	0	inf	0.03
group_6737	Hypothetical protein	14	0	inf	0.03
group_2455	Lysozyme RrrD	14	0	inf	0.03
group_195	Outer membrane protein X	14	0	inf	0.03
ecnA	Hypothetical protein	87	24	10.87	0.03
yeiS	Hypothetical protein	87	24	10.87	0.03
pac	Penicillin G acylase	24	1	9.75	0.007
yjbE	hypothetical protein / uncharacterized propien	81	15	9.25	0.00005
group_9158	hypothetical protein	21	1	8.14	0.02
insC-1_1	hypothetical protein	21	1	8.14	0.02
orgA	hypothetical protein	21	1	8.14	0.02
group_5728	PTS system 2-O-alpha-mannosyl-D-glycerate-specific EIIABC component	21	1	8.14	0.02
leuC	3-isopropylmalate dehydratase large subunit	85	21	8.09	0.005
hokA_3	Protein HokA	85	21	8.09	0.005
yigG	Inner membrane protein YigG	20	1	7.64	0.02
yjgL_1	Putative protein YjgL	20	1	7.64	0.02
group_2367	Hypothetical protein	84	20	7.35	0.003
papG	Fimbrial adhesin PapG	19	1	7.15	0.04
group_2979	Outer membrane usher protein HtrE	19	1	7.15	0.04
traJ	Protein TraJ	32	2	7.14	0.003
ccdB	Toxin CcdB	32	2	7.14	0.003
group_4444	Hypothetical protein	18	1	6.68	0.04
ybfD	Hypothetical protein	30	2	6.46	0.006
group_13501	Hypothetical protein	30	2	6.46	0.006
srp54	Signal recognition particle 54 kDa protein	30	2	6.46	0.006
selB	Selenocysteine-specific elongation factor	85	22	6.43	0.01
group_13502	Hypothetical protein	29	2	6.14	0.01
parM	Plasmid segregation protein ParM	29	2	6.14	0.01
stbB	Protein StbB/ control of plasmid pertitioning	29	2	6.14	0.01
group_10602	Hypothetical protein	27	2	5.53	0.02

traD	Coupling protein TraD involved in Conjugative DNA transfer	26	2	5.24	0.02
ykfF_2	Hypothetical protein	26	2	5.24	0.02
group_13534	Hypothetical protein	25	2	4.96	0.03
group_3978	Hypothetical protein	25	2	4.96	0.03
umuC_3	Protein UmuC	25	2	4.96	0.03
ygjK	Glucosidase YgjK	84	22	4.77	0.03
mdaB	Modulator of drug activity B	84	22	4.77	0.03
yeeN	Putative transcriptional regulatory protein YeeN	84	22	4.77	0.03
ydiV	Putative anti-FlhC(Polito et al.)FlhD(4) factor YdiV	83	21	4.74	0.01
group_10575	Hypothetical protein	24	2	4.68	0.03
yqeL	Putative protein YqeL	24	2	4.68	0.03
traM	Relaxosome protein TraM	24	2	4.68	0.03
group_3938	Hypothetical protein	31	3	4.35	0.01
group_4922	hypothetical protein	36	4	3.98	0.01
focC	Chaperone protein FocC	82	21	3.90	0.03
dsdX	D-serine transporter DsdX	82	21	3.90	0.03
yhaB	Hypothetical protein	46	6	3.83	0.007
group_8719	Hypothetical protein	35	4	3.79	0.01
fimA_2	Fimbrial subunit type 1	51	8	3.27	0.01
group_4190	Hypothetical protein	42	6	3.19	0.02
ccdA_2	Antitoxin CcdA	37	5	3.19	0.03
xerD_5	Tyrosine recombinase XerD	36	5	3.04	0.03
ydhV	Putative oxidoreductase YdhV	75	18	2.88	0.04
group_12466	Hypothetical protein	47	8	2.72	0.04
group_4659	Hypothetical protein	46	8	2.60	0.04
finO	Fertility inhibition protein	53	10	2.57	0.04

Inf-Infinity

3.4.14.2 Between adult CD and Healthy controls

We identified a total of 89 genes overrepresented in adult CD isolates based on a cut-off value of > 2.5 in odds ratio (OR) and p value <0.05. These genes are involved with conjugative transfer of DNA, plasmid maintenance, outer membrane protein, CRISPR, ion transport, small toxic polypeptide, adhesion to and invasion of epithelial cell lines. We also found some genes associated with hypothetical proteins. Of note, we found seven genes belonging to operons encoding fimbrial structures, including type1 pili

and long polar fimbriae. These genes are lpfB, variant of lpfA, lpfC, fimD, htrE, ybgD and variant of fimA. In addition to the fimbriae associated genes we also found at least 11 plasmid-associated genes. They are yfjQ/ yubP, repA, traR, ccdB, traJ, srp54, stbB, parM, yihA and some genes encoding hypothetical proteins. These genes are involved in plasmid maintenance/toxin anti-toxin system/ post-segregational killing of plasmid-free cells, conjugal transfer of DNA between bacterial cells, plasmid partitioning, plasmid replication and some hypothetical proteins. Of note, most of the plasmid-associated genes are present with a perfect match in *E. coli* UTI89 plasmid pUTI89 isolated from the urinary bladder of patient with UTI. Other significantly associated genes in adult CD isolates are involved in CRISPR-associated endoribonuclease Cas2, transcriptional regulation, ABC transporter ATP binding, potassium-efflux system, oxidoreductase activity, putative membrane-associated lysozyme RrrD, penicillin G acylase, ATPase, transposon and hypothetical proteins (Supplemental Table 3.2).

3.4.14.3 Between adult CD and paediatric CD

Strongly overrepresented genes in paediatric CD isolates are predominantly involved with fimbriae proteins, general stress protein A, elongation factor, transcriptional activator and DNA repair, DNA recombination, toxin-antitoxin biofilm protein (tabA_2, tabA_1), toxin-antitoxin system, actin cross-linking toxin, outer membrane usher protein, outer membrane protein, HTH-type transcriptional regulator, other transcriptional regulatory proteins, energy metabolism, mobility and motility specially in flagellar biosynthesis and activity operons, biosynthesis and transcriptional regulation, colicin, major exported protein, DNA-binding protein, iron binding and import, prophage, catalase HPII/ catalase and peroxidase activity, type II secretion system protein, multidrug export protein, tRNA-Arg, multidrug resistance protein, [Fe-

S]-dependent transcriptional repressor, lysozyme, phage shock protein G, chemotaxis protein, inner membrane proteins, sensor protein, ATP-binding protein, and hypothetical proteins (Figure 3. 11 and Supplemental Table 3.3)

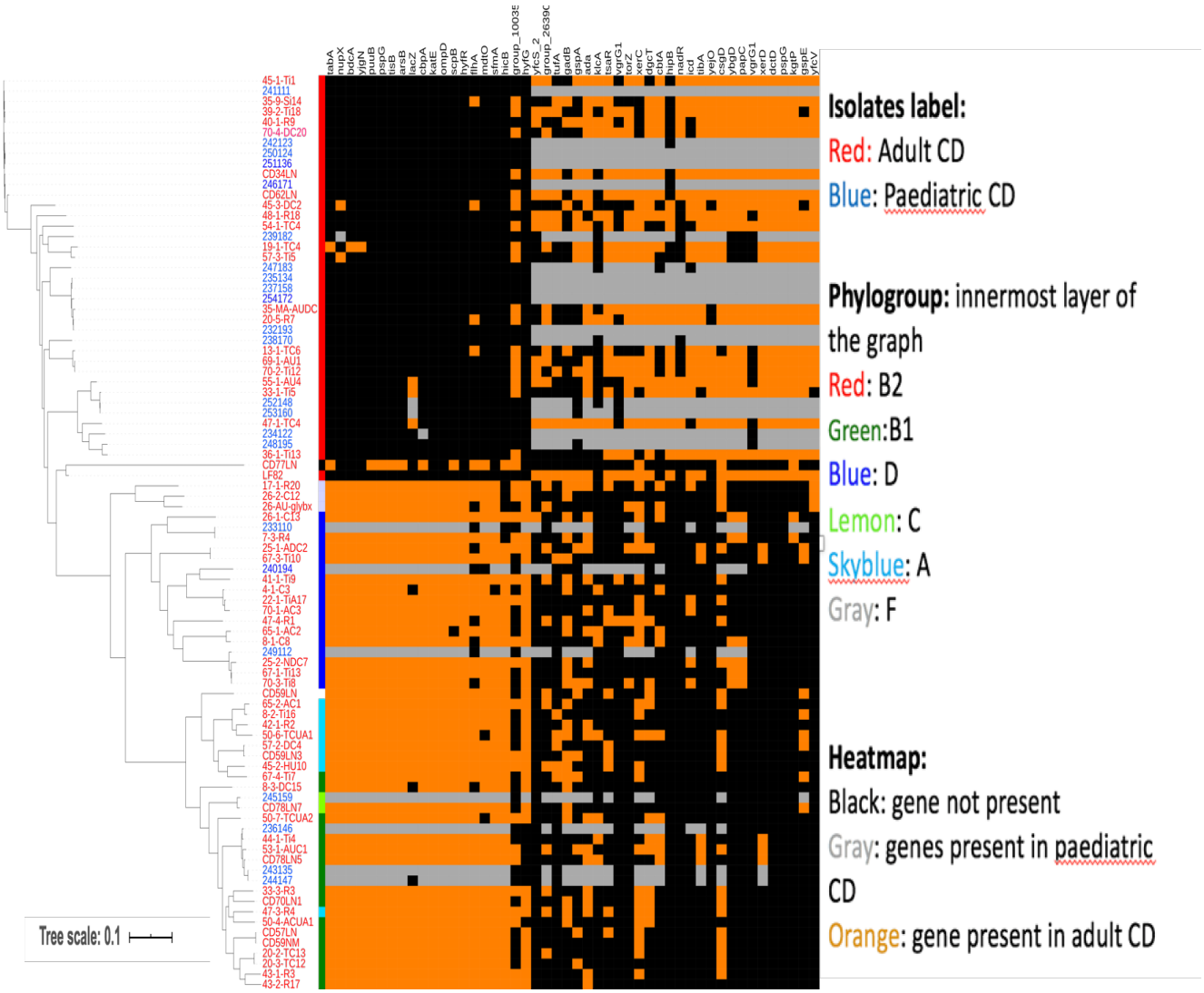


Figure 3.11. Heatmap showing significantly associated genes in adult CD and paediatric CD. The SNP tree of 88 *E. coli* isolates from adult and paediatric CD was created with CSI phylogeny. Label colours of isolates indicates different groups; red: adult CD, and blue: paediatric CD. In heatmap, the innermost row of blocks indication the phylogroups of the isolates. red: B2, green: B1, blue: D, lemon: C, sky-blue: A and grey: F. Genes are label at the top of the blocks. Inside the heatmap black coloured blocks indicate the absence of the gene, grey colour shows the presence of the genes in paediatric CD and orange colour marks the presence of the genes in adult CD isolates.

3.4.14.5 Genes associated with paediatric CD compared with HC isolates

Scoary analysis was conducted, highly significant genes were listed by sorted naive p value of <0.03. The strongest association identified (naïve p = 3.55E-07, Bonferroni p = 0.0047, Benjamini H p = 0.0023) was group_26390, which codes for transfer RNA [tRNA-Asn(gtt)] and was over-represented in the paediatric CD isolates. Enhanced tRNA levels have been shown to protect *E. coli* against oxidative stress caused by hydrogen peroxide (Zhong et al., 2015). Greater numbers of plasmid-associated genes were found in paediatric CD isolates. Most paediatric CD-enriched genes are located on plasmids and some genes and operons with various functions (stress response, metabolism, transcription and translation, flagella and fimbriae, toxin and antitoxin and DNA binding and recombination etc) are also associated with paediatric-CD (Figure 3.12 and Supplemental table 3.4).

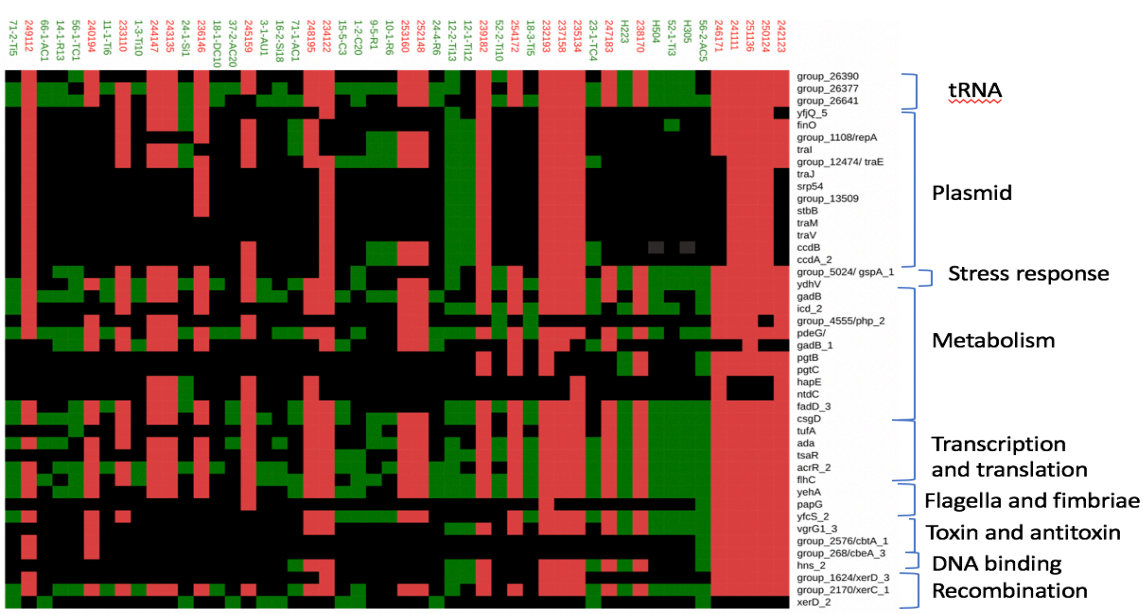


Figure 3.12. Genes overrepresented in paediatric CD compared with healthy controls (HC).

Heat map showing significantly associated genes with paediatric CD isolates compared with HC. Isolates labelled green indicate healthy controls and red colour

showing paediatric CD. Inside the heatmap, black coloured bocks imply the absence of the corresponding gene, green indicates the presence of the gene in HC, and red indicates the presence of the gene in the paediatric CD isolates.

3.4.14.6 Scoary analysis of ulcerative colitis (UC) and healthy controls (HC) strains

Highly significant genes associated with ulcerative colitis were sorted based on a naive p-value of <0.03. We found that 24 genes have a strong positive association with ulcerative colitis (UC) isolates (p<0.03 and odds ratio >3.0). Among these 24 genes, five are associated with plasmid: fertility inhibition protein (finO), conjugal transfer of DNA (traJ), plasmid maintenance post-segregational killing of plasmid-free cell/ toxin-antitoxin system (ccdA_2 and ccdB) and a hypothetical protein (yfjQ_5). The rest of the genes are associated with putative allantoin permease (ybbW), HTH-type transcriptional regulation (chbR), glutamine transport system permease protein (glnP), cardiolipin synthase B (group_9272/ clsB), putative signal peptide peptidase (sppA), flagellar hook-associated protein 2 (fliD), colicin I receptor (cirA_4), phage tail protein (group_2624, group_2631) and hypothetical protein (yjbE, insD-1, yncH etc) (Table 3.13).

Table 3.13. Significantly associated gene in *E. coli* isolates in UC compared with HC

Gene	Annotation	No. in UC	No. in HC	Naive p
ybbW	Putative allantoin permease	31	19	0.001
yjbE	Hypothetical protein	28	15	0.005
yfjQ_5	Hypothetical protein	12	2	0.006
traJ	Protein TraJ	11	2	0.01
finO	Fertility inhibition protein	16	5	0.01
group_2769	hypothetical protein/ lytic transglycosylase/plasmid associated	13	3	0.01
chbR	HTH-type transcriptional regulator ChbR	31	22	0.01
glnP	Glutamine transport system permease protein GlnP	31	22	0.01

group_2422	Hypothetical protein / VWA domain-containing protein	30	20	0.01
group_13510	Hypothetical protein	10	2	0.02
group_9272	Cardiolipin synthase B	10	2	0.02
group_4555	Phosphotriesterase homology protein	10	2	0.02
group_2624	Hypothetical protein	6	0	0.02
group_2631	Hypothetical protein	6	0	0.02
group_2238	Putative signal peptide peptidase SppA	6	0	0.02
group_4212	hypothetical protein/ phage tail protein	6	0	0.02
ccdB	Toxin CcdB	15	5	0.02
ccdA_2	Antitoxin CcdA	15	5	0.02
fliD	Flagellar hook-associated protein 2	25	14	0.02
insC-1_1	Hypothetical protein	8	1	0.02
cirA_4	Colicin I receptor	8	1	0.02
insD-1	Hypothetical protein	8	1	0.02
yncH	Hypothetical protein	8	1	0.02

3.4.14.7 Scoary analysis of Crohn's disease (CD) and ulcerative colitis (UC) isolates

Using a p value of <0.03, this analysis revealed that 137 genes were overrepresented in UC and only one gene, xseA, involved with exodeoxyribonuclease 7 large subunit, was overrepresented in CD. However, no gene was found to be strongly associated with UC compared to CD. Genes overrepresented in UC are listed in the table 3.14.

Table 3.14. List of genes significantly associated with *E. coli* in UC compared with CD

Gene	Annotation	No. present in UC	No. present in CD	Naive p
xseA	Exodeoxyribonuclease 7 large subunit	72	17	0.007
group_4260	Exodeoxyribonuclease 7 large subunit	17	13	0.01
ybcC	Hypothetical protein	25	17	0.01
rem	Hypothetical protein	13	11	0.01
group_2798	Hypothetical protein	54	26	0.02
group_2558	Hypothetical protein	15	13	0.007
group_603	Hypothetical protein	10	10	0.01
group_5797	Arylsulfatase	8	9	0.01
higA	Antitoxin HigA	53	27	0.007

group_10431	Hypothetical protein	4	6	0.01
group_7106	Hypothetical protein	4	6	0.01
group_13357	Hypothetical protein	4	6	0.01
group_17367	Hypothetical protein	4	6	0.01
stpA_2	Hypothetical protein	4	6	0.01
group_11119	Hypothetical protein	3	5	0.02
group_2078	Hypothetical protein	3	5	0.02
group_8594	Hypothetical protein	3	5	0.02
group_2031	Hypothetical protein	3	5	0.02
group_26496	tRNA-Ala(tgc)	3	5	0.02

3.4.14.8 Scoary analysis of Crohn's disease lymph node (LN) isolates and healthy control (HC) isolates

Scoary analysis between lymph node (LN) isolates in CD and HC isolates revealed that, 321 genes were significantly associated with either LN or HC isolates. However, we considered only those genes (n = 106) which have greater positive association with LN isolates only ($p<0.03$ and odds ratio >7.0) to find those genes that have strong association with the LN isolates (Figure 3.13 and Supplemental Table 3.5).

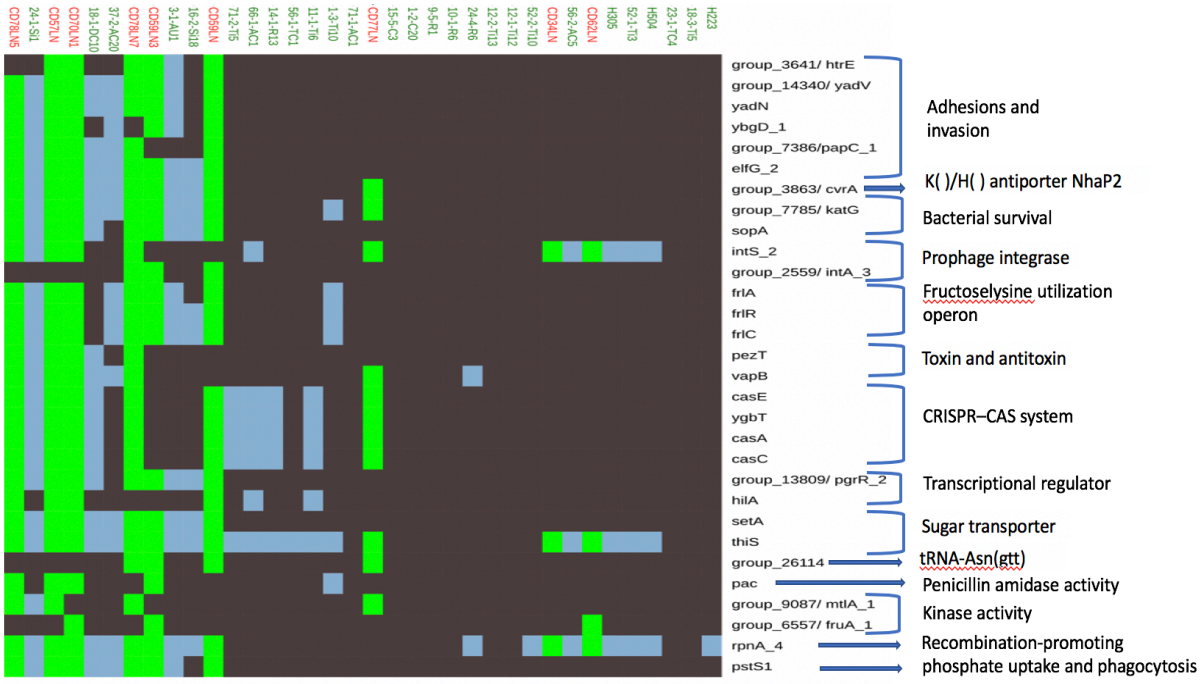


Figure 3.13. Heat map showing the significantly associated genes in LN (red) compared with healthy controls (green). *E. coli* isolates ordered based on their SNP tree. Inside the heat map green colour indicates the presence of corresponding gene in the lymph node isolates, grey

colour indicates presence of gene in healthy controls and black indicates absence of corresponding gene.

3.4.14.9 Scoary analysis between aphthous ulcer (AU) and healthy control (HC) isolates

Scoary analysis between aphthous ulcer (AU) isolates in Crohn's disease and healthy control (N) isolates revealed that some genes were associated with AU isolates. Genes overrepresented in AU are mostly associated with plasmids, more specifically conjugative plasmids (group_16142/ traM, group_1711/ TraX, traJ, group_5929/ TrbA, group_6366/ stbB, traI, traA). Other genes are involved with DNA-binding transcription factor activity (group_10646, group_15450), HTH-type transcriptional regulator PrtR (prtR), replication protein (group_997/ yfdO), cell division inhibition (group_12496/ kilR), DNA adenine methyltransferase activity (group_1890/yhdJ_2), tRNA (group_7578, group_26390 and group_26502), endonuclease (yhcR), aspartate ammonia-lyase (aspA_1) and hypothetical proteins (Supplemental Table 3.6).

3.4.14.10 Scoary comparison of bacterial isolates from aphthous ulcers and resected tissues from patients with CD

Scoary analysis between aphthous ulcer (AU) isolates and late stage bowel resected isolates in CD revealed that some genes were associated with AU isolates and bowel resected isolates. Genes overrepresented in AU are mostly associated with DNA base-flipping protein YbcN (ybcN), toxin YdaT domain-containing protein (group_2424), complement resistance protein TraT (ylpA), glycoside hydrolase family 127 protein (hypBA1), helix-turn-helix domain-containing protein (group_9312), and HNH nuclease family protein (group_820). Genes overrepresented in bolwel reseced isolates in CD are mostly associated with recombination-promoting nuclease RpnC (yadD), flagellar biosynthesis protein FlhA (group_4174), prophage integrase IntS (intS_2), lysozyme inhibitor LprI family protein (yecT), PLP-dependent

aminotransferase family protein (yjiR), Cytochrome b (yodB), and penicillin acylase family protein (pac).

3.4.14.11 Significantly associated gene in AIEC compared with non-AIEC isolates

Scoary analysis revealed a total of 163 genes significantly associated with either AIEC or non-AIEC. Most of the genes/ operons significantly associated with AIEC are involved with fimbrial protiens (group_5695/ fimA_2, fimD_3, lpfA_4, papB_2, sfaG, group_9129/ sfaH, group_15228), prophage integrase IntA (group_15152/ intA_1), GTPase Era (era_3), secretory immunoglobulin A-binding protein EsiB (esiB), chaperone protein FocC (group_15225), transposase for transposon Tn5 (tnpA_1), putative metal chaperone YciC (group_7036), anti-FlhC(Polito et al.)FlhD(4) factor YdiV (ydiV_2), major exported protein (group_13316), elongation factor Tu 2 (group_1761), ribosomal RNA small subunit methyltransferase I (rsmI_2), outer membrane protein (group_3122), and putative nuclease YbcO (group_3422) (Supplemental table 3.7).

3.4.15 Genome-wide association study (GWAS) by Pyseer

Using pyseer tool we have done both clusters of orthologous genes COGs (genes) and k-mer analysis. Pyseer uses the original seer model, using MDS components as fixed effects to control for the population structure. Then, gene presence/absence Rtab file from roary output was used to test COG presence-absence. Filtered P-value considers only the presence and absence of genes in two different groups and gives a raw p-value. However, lrt p-value considered the phylogenetic differences and relationship, as well as the gene presence-absence among the strains hence it gives different but more accurate results. The mixed-effect model includes a linear mixed

model (LMM), which uses random effects to control for population structure. It allows us to look at the association in linear time using a kinship matrix obtained from a subset of variants or a phylogeny.

3.4.15.1 Adult CD vs paediatric CD

We listed the most significantly associated genes/COGs in adult CD compared with paediatric CD (considering lrt P-value) based on fixed effect model (Supplemental table 3.8) and mixed effect model (Supplemental table 3.9). Here, it is found that genes associated with flagellar biosynthesis protein FlhA (group_4994), recombination-promoting nuclease RpnA (rpnA_2), universal stress protein UspG (group_8511), sensor protein TorS (group_2748), hydrogenase-4 component G (hyfG) and, hypothetical proteins (yjiC, group_10291) were overrepresented in adult CD isolates compared th**e** paediatric CD isolates. Genes associated with paediatric CD isolates compared with adult CD are listed in the Supplemental table 3.10 and Supplemental table 3.11. The most important genes in paediatric CD isolates are associated with fimbrial chaperone (yfcS_2), elongation factor Tu1(tufA), AlpA family transcriptional regulator (group-2557), inovirus Gp2 family protein (group_857), general stress protein A (group_5024/ gspA_1), inner membrane ABC transporter ATP-binding protein YddA (yddA), anti-restriction protein KlcA (klcA_1) and type IV toxin-antitoxin system toxin CbtA (yeeW_1) and hypothetical proteins (group_1543, group_7835, group_4547, group_2391, group_6211). In mixed effect k-mer analysis between adult CD and the paediatric CD we found a total of 97 filtered k-mer (significantly associated) associated with either adult CD or paediatric CD. Then we did the annotation against a list of reference genomes and all the filtered k-mers were annotated. Here, we found that two genes are associated with adult CD. According to k-mers analysis, the most

important gene in adult CD was associated with flagellar biosynthesis protein FlhB (Table 3.15 and Figure 3.14) this finding corroborates the previous COGs findings.

Table 3.15. Significantly associates k-mers in adult CD *E. coli* isolates

Gene	Hits	Max p	Avg. AF	Avg. MAF	Avg. beta	Function
cds-GJ12_RS08340	1	8.135	0.17	0.17	0.644	
cds-WP_077631359.1	48	19.723	0.45585417	0.4558	0.5819791	Flagellar biosynthesis protein FlhB

Max p- Maximum -log10(p-value), Avg. AF- average allele frequency, Avg. MAF- average minor allele frequency, Avg. beta- effect size for the variant.

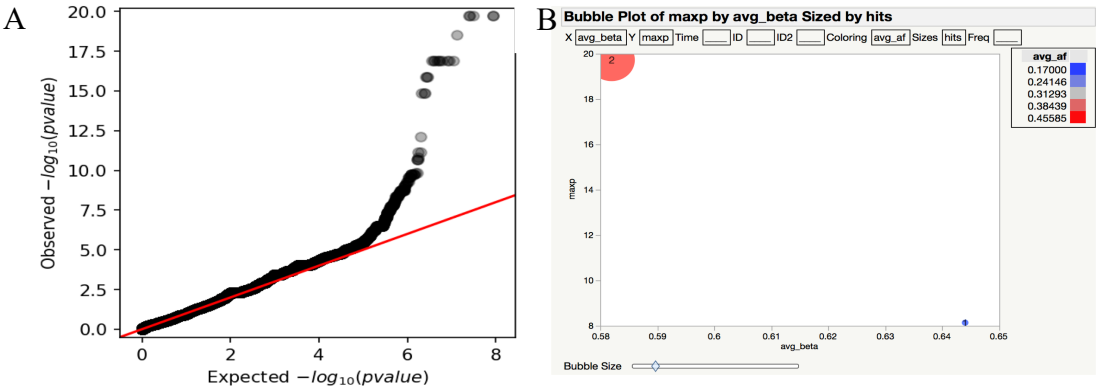


Figure 3.14. k-mers associated with adult CD compared with paediatric CD. A) Q-Q plot for observed and expected p-values for K-mers between adult CD and paediatric CD. B) Bubble plot for significantly associated k-mers in adult CD compared to paediatric CD. Red bubble at upper left corner indicates highly significant k-mer associated with adult CD. Avg. AF indicates the average allele frequency.

3.4.15.2 CD vs healthy control (HC)

Genes significantly associated with CD (combined adult CD and paediatric CD) based on fixed effect and mixed effect model are listed in the Supplemental table 3.12 and 3.13. This shows that based on fixed effect model and mixed effect model the most

important genes are associated with conjugative transfer pilus assembly protein (traI, group_1711, group_12474, traJ), lytic transglycosylase (group_2769), fertility inhibition protein (finO), phage integrase family protein/Tyrosine recombinase XerD (group_1624), fimbrial chaperone protein FimC (focC), tRNA-Asn (group_26390), selenocysteinyl-tRNA-specific translation elongation factor SelB (selB), D-serine transporter DsdX (dsdX), type I toxin-antitoxin system toxin HokA (hokA_3), signal recognition particle 54 kDa protein (srp54), type II toxin-antitoxin system toxin CcdB (ccdB), replication protein (group_1108), major type 1 subunit fimbrin (fimA_2) and hypothetical proteins (yjbE, group_4184, yehM, group_2422, yhaB, group_13509). In mixed effect analysis comparing between CD and healthy controls we have found additional genes significantly associated with CD isolates compared with healthy controls. They are associated with DUF945 domain-containing protein (yfjQ_5), alpha-glucosidase (yhak) and D-serine transporter DsdX/ recombinase RecT (dsdX). In the mixed effect k-mer analysis between CD and HC we found a total of 284 k-mers associated with either CD or HC. We annotated 214 k-mers against the reference strains and 70 remained unannotated. Among these annotated k-mers, we found 113 hits. Greater number of hits (17) were found for the plasmid-associated gene codes for hypothetical protein/ IS4 family transposase protein. The second most hits (13) were found for the genes associated with NADPH-dependent aldehyde reductase Ahr and type IV secretion system protein TraC as shown in table 3.16 and Supplemental Figure 3.1.

Table 3.16. Significantly associated k-mers in CD compared with HC

Gene	Hits	Max p	Avg. AF	Avg. MAF	Avg. beta	Function
cds-WP_065228126.1	13	7.52	0.10	0.10	0.66	NADPH-dependent aldehyde reductase Ahr

cds-E3725_RS05290	17	8.80	0.10	0.10	0.67	Hypothetical protein/ IS4 family transposase (plasmid_associated)
traC	13	7.68	0.07	0.070	0.81	Type IV secretion system protein TraC

Max p- Maximum -log10(p-value), Avg. AF- average allele frequency, Avg. MAF- average minor allele frequency, Avg. beta- effect size for the variant.

3.4.15.3 Adult CD vs healthy control (HC)

Genes significantly associated with adult CD based on fixed effect and mixed effect model are listed in the Supplemental table 3.14 and 3.15. According to pyseer fixed effect and mixed effect model the top significantly associated genes with the adult CD isolates are associated with phage integrase family protein (group_1624), lytic transglycosylase (group_2769), transcriptional activator/ PAS domain-containing protein (traJ), DUF945 domain-containing protein (yfjQ_5/ yubP), signal recognition particle 54 kDa (srp54), CopG family transcriptional regulator (group_13509), plasmid stability protein (stbB), type II toxin-antitoxin system antitoxin CcdB and CcdA (ccdB and ccdA_2), type-F conjugative transfer system pilin acetylase TraX (group_1711), translation error-prone DNA polymerase V subunit UmuC (umuC_3), phage head morphogenesis protein (selB, focC), fertility inhibition protein (finO), multifunctional conjugation protein (Cottone et al.), type I toxin-antitoxin system toxin HokA (hokA_3), VWA domain-containing protein (group_2422), alpha-glucosidase (ygjK), and hypothetical proteins (yjbE, group_13510, group_10588, yehM, yhaB, group_1624).

In k-mer analysis between adult CD and healthy control we found a total of 219 k-mers and 206 were annotated and 13 were unannotated. Among the annotated k-mers 65 gene hits were observed. The highest hits (44) were found for the gene that encodes a hypothetical protein and the second highest hit (42 hits) was found for a gene

associated with DUF945 domain-containing protein (Table 3.17 and Supplemental Figure 3.2).

Table 3.17. Significantly associated k-mers in adult CD isolates based on mixed effect model

Gene	Hits	Max p	Avg. AF	Avg. MAF	Avg. beta	Function
cds_WP_065228127.1	44	6.372	0.0978	0.0978	0.77	Hypothetical protein
cds-WP_065227645.1	41	6.818	0.152	0.152	0.683	DUF945 domain-containing protein

Max p- Maximum -log10(p-value), Avg. AF- average allele frequency, Avg. MAF- average minor allele frequency, Avg. beta- effect size for the variant.

3.4.15.4 Paediatric CD vs healthy controls (HC)

In both fixed effect and mixed effect model list of gene found to associated with paediatric CD isolates compared to healthy controls are similar and the lists of genes with the function of associated proteins show in the Supplemental Table 3.16 and Supplemental Table 3.17. Significantly associated gene are associated with elongation factor Tu 1 (tufA), DUF945 domain-containing protein (yfjQ_5), type VI secretion system PAAR protein (group_3296), lytic transglycosylase (group_2769), GGDEF domain-containing protein (dgcT), replication protein (group_1108), FliC/FljB family flagellin (group_14671), IncF plasmid conjugative transfer protein TraI (Cottone et al.), regulator with HipB (group_3260), peptidase M35 (group_14755), fertility inhibition protein FinO (finO), phage integrase family protein (group_1624), general stress protein A (group_5024), TraE protein of plasmid F (group_12474), GGDEF domain-containing protein (dgcT), transcriptional regulator TdcA (tsaR) and hypothetical protein (group_10994, group_1543, group_4184, group_7835).

A total 367 k-mer were observed to be significantly associated with paediatric CD isolates. Among this k-mers we were able to annotate 358 k-mers against the reference genome list. Remaining 8 were unannotated. From their annotated k-mers I found a list of 55 gene hits. The highest gene hits (62) were found for a gene encoding DUF945 domain-containing protein followed by gene hits (57) that encodes type-F conjugative transfer system secretin TraK protein. 31 hits were found for a gene encoding IS66-like element ISEc8 family transposase (Table 3.18 and Supplemental Figure 3.3).

Table 3.18. Significantly associated k-mers in paediatric CD isolates based on mixed effect model

Gene	Hits	Max p	Avg. AF	Function
cds-WP_085959725.1	62	6.236	0.48	DUF945 domain-containing protein
traK	57	6.236	0.42	Type-F conjugative transfer system secretin TraK
cds-WP_000998025.1	31	6.183	0.44	IS66-like element ISEc8 family transposase
cds-WP_046201784.1	11	6.643	0.50	DUF945 domain-containing protein

Max p- Maximum -log10(p-value), Avg. AF- average allele frequency.

3.4.15.5 UC vs Healthy controls (HC)

In COGs analysis gene overrepresented significantly are listed in the Supplemental Table 3.18 and Supplemental Table 3.19. According to this analysis, genes associated with plasmid-associated DUF945 domain-containing protein (yfjQ_5), plasmid-associated lytic transglucosylase(group_2769), Fertility inhibition protein (finO), Protein TraJ (traJ), Flagellar hook-associated protein 2 (fliD), fimbrial assembly protein

(group_5332), toxin CcdB (ccdB), antitoxin CcdA (ccdA_2), AIDA-I family autotransporter YfaL (group_1823), helix-turn-helix domain-containing protein, L-rhamnonate dehydratase (rhmD), putative fimbrial-like adhesin protein (yehA) and hypothetical proteins (yjbE, yhaB)

In k-mer analysis between UC and healthy control we found a total of 77 k-mers and after annotation we got 72 annotated k-mer and 5 remaining unannotated k-mers. 43 Gene hits found in UC vs HC k-mer analysis. Top hits were observed against genes that code for conjugal transfer protein TraP, and hypothetical proteins as shown in the table below (Table 3.19 and Supplemental Figure 3.4).

Table 3.19. Significantly associated k-mers in UC isolates compared with HC

Gene	Hits	Max p	Avg. AF	Function
trap	14	3.978	0.362	Conjugal transfer protein TraP
cds-E0725_RS04340/ EC725_04340 in organism *E. coli* 13E0725	9	4.756	0.354	Hypothetical protein
cds-E0725_RS05035/ EC725_05035 in organism *E. coli* 13E0725	8	4.756	0.340	Hypothetical protein
cds-WP_065227197.1	6	4.431	0.577	M-antigen undecaprenyl disphosphate flippase

Max p- Maximum -log10(p-value), Avg. AF- average allele frequency.

3.4.15.6 Lymph nodes (LN) vs healthy control (HC)

Based on fixed and mixed effect model a few genes significantly associated with lymph node CD isolates as listed in the Supplemental table 3.20 and Supplemental table 3.21. Among those genes two genes were associated with catalase-peroxidase HPI (group_7945) and hypothetical protein (group_6734). In mixed effect model, only few genes were found to be significantly associated with LN isolates. They are

group_26282 (tRNA-Phe (gaa)), group_3725 (fructose-bisphosphatase class II), and group_7945 (catalase-peroxidase HPI, katG). In k-mer analysis a total of 232 k-mers were identified. Among there k-mers 228 k-mers were annotated against the reference strains and three k-mers remained unannotated. From the annotated k-mers we found a list of 71 gene hits and the topmost hits (76) were found for the gene that is associated with threonine/serine exporter (Table 3.20 and Supplemental Figure 3.5.).

Table 3.20. K-mers associated with LN isolates compared with HC

Gene	Hits	Max p	Avg. AF	Function
cds-WP_000538189.1	76	6.638	0.250	Threonine/serine exporter
cds-EC725_RS24030	26	8.382	0.203	Penicillin acylase
cds-WP_077775211.1	13	6.455	0.194	Penicillin G acylase

Max p- Maximum -log10(p-value), Avg. AF- average allele frequency

3.4.15.7 Aphthous ulcers (AU) vs Healthy controls (HC)

Gene associated with aphthous ulcer isolates according to fixed effect and mixed effect are listed in the Supplemental Table 3.22 and Supplemental Table 3.23. The most significant genes are associated with helix-turn-helix domain-containing protein (group_10646, group_15450), helix-turn-helix transcriptional regulator/ HTH-type transcriptional regulator PrtR (prtR), cell division inhibitor protein group_12496/ kilR, regulator/ bacteriophage regulatory protein CII (ydaT), tRNA-Asn(gtt) (group_26390), relaxosome protein TraM (group_16142), and DUF4102 domain-containing protein (group_15336).

In mixed effect k-mer analysis we found that 581 k-mers were identified in AU isolates compared to healthy control. After annotation 531 k-mers and 50 remained unannotated. Among these annotated k-mers highest hits (112) were found against the gene encoding helix-turn-helix domain-containing protein, region: YdaS_antitoxin/ phage protein Cro-UMN026. This is shown in the table 3.21 and Supplemental Figure 3.6.

Table 3.21. K-mers associated with AU isolates compared with HC

Gene	Hits	Max p	Avg. AF	Function
cds-WP_001067459.1 YP_002411352.1 NZ_CP015228.1	112	6.396	0.178	Helix-turn-helix domain-containing protein, region: YdaS_antitoxin/ phage protein Cro-UMN026
cds-WP_000858975.1	35	5.764	0.176	Helix-turn-helix transcriptional regulator, region: Phage repressor protein C, contains Cro/C1-type HTH
cds-WP_001182868.1	46	6.396	0.142	Hypothetical protein, region_name: YdaT_toxin, "Putative bacterial toxin ydaT are neutralized by the putative antitoxin Yda

Max p- Maximum -log10(p-value), Avg. AF- average allele frequency

3.4.15.8 AIEC vs non-AIEC

In fixed and mixed effect models, AIEC isolates significantly associated with genes listed in Supplemental Table 3.24 and Supplemental Table 3.25. They are elongation factor Tu (tufB), helix-turn-helix domain-containing protein (higA-2_1), phage head morphogenesis protein addiction module toxin RelE (higB-2_1), LysR family transcriptional regulator (group_1777), hypothetical protein (group_13437,

group_2878, group_2878), DUF1364 domain-containing protein (ybcO), and AIDA-I family autotransporter YfaL (yfaL). In k-mer analysis we found 560 k-mers associated with AIEC. We could annotate 549 k-mers against the reference genome and the remaining 11 k-mers were unannotated. Based on these annotated k-mer we then found 284 gene hits and the highest hits (101) were for a gene associated with type 1 fimbrial protein followed by histidinol-phosphate transaminase (40 hits), multidrug efflux RND transporter subunit MdtA (29 hits), type 1 fimbria D-mannose specific adhesin FimH (26 hits), and type 1 fimbrial protein fimA (Table 3.21 and Supplemental Figure 3.7).

Table 3.21. k-mers associated with AIEC compared with non-AIEC based on mixed effect model

Gene	Hits	Max p	Avg. AF	Function
WP_000824114.1	101	4.301	0.333	Type 1 fimbrial protein
WP_065227179.1	40	4.540	0.335	Histidinol-phosphate transaminase
cds-WP_000679006.1	29	4.301	0.333	Multidrug efflux RND transporter subunit MdtA
cds-WP_000832227.1	26	4.301	0.333	Type 1 fimbria D-mannose specific adhesin FimH
fimA	11	4.712	0.273	Type 1 fimbrial protein

Max p- Maximum -log10(p-value), Avg. AF- average allele frequency

3.4.16 Comparison of SNPs and genes of *E. coli* strains in bowel resected samples

In Canberra Hospital we collected 7 bowel resected sample from CD patients. Taking aseptic measures, we collected three different tissues from each bowel resected samples: normal mucosa, involved mucosa and lymph nodes (one to several in number). Then we isolated *E. coli* isolates from those tissue samples. We wanted check if the *E. coli* isolates from different tissue samples have any genomic similarities

of differences. Genome assemblies were obtained for 20 *E. coli* isolates from bowel resected samples of 7 Crohn's disease patients. At first, we utilized CSI Phylogeny 1.4 interface (Kaas et al., 2014) in CGE (https://cge.cbs.dtu.dk/ services/CSIPhylogeny/) to examine relationships among the sequenced strains from bowel resected samples. This tool calls and filters SNPs, identifies the high-quality SNPs and creates a phylogeny. Here, all-vs-all mapping was done using assemblies of individual isolates and comparing the SNP distances with one another. A SNPs matrix was downloaded, extracted with Excel and plotted along with the phylogenetic tree using Itol (Letunic and Bork, 2016).

SNP analysis in these isolates showed that isolates from one patient with similar sequence type tend to cluster together, even when these isolates are isolated from different tissues (NM-normal mucosa, IM= involved mucosa, L N = lymph node) of the bowel resected sample. The number of SNPs within a patient was negligible (less than 50) as compared to the interpatient diversity (maximum difference more than 67,000) (Supplemental Figure 3.8). Alignment of genome sequences from tissue of the bowel resected samples revealed minor heterogeneity in Mauve alignment. This suggested that the whole bowel resected sample of a CD patient is colonized mostly by a single strain of *E. coli (*except CD59 patient where we found three different strains and CD78 patient where we have found 2 different strains with greater SNP differences).

When we checked the gene differences, similar results came out. Roary analysis revealed the gene differences among the different *E. coli* isolates in bowel resected samples from the same patient were less (minimum: 71). However, in some patients with several strains in different tissue from the same bowel resected specimen, the gene difference was up to 1941 (Supplemental Figure 3.9). Interpatient diversity of the genes varies from a minimum of 982 to a maximum of 3105 genes.

3.5 Discussion

Some commensal *E. coli* strains acquire virulence factors associated with enterotoxins, adhesion and invasion to intestinal mucosa, converted to pathogenic strains and subsequently involved in intestinal diseases (Nataro and Kaper, 1998). *E coli* is also implicated in the pathogenesis of CD, and it has been proposed that *E coli* strains in patients with inflammatory bowel disease are phenotypically distinctive from those found in normal faeces from healthy individuals (Dickinson et al., 1980; Burke and Axon, 1988; Giaffer et al., 1992; Sartor, 1995; Darfeuille-Michaud et al., 1998b). There is an increased abundance of *E. coli* is implicated in CD in comparison with healthy controls (Baumgart et al., 2007; Lopez-Siles et al., 2014; Martinez-Medina and Garcie-Gil, 2014). Several attempts were also made to identify whether the intestines of CD patients harbour a single or multiple strains of *E. coli* (Martinez-Medina and Garcie-Gil, 2014). Indeed, previous studies in our group showed the clonal diversity of *E. coli* in the intestinal mucosa of terminal ileum, ascending, transverse and descending colon, and rectum of 69 patients, 32 of whom were CD patients (Gordon et al., 2015),

To identify the association of *E. coli* with the active late stage of CD (those requiring surgery for bowel resection), we also investigated *E. coli* isolated from bowel resected samples. We observed that in an individual patient *E. coli* isolates in involved mucosa, normal mucosa and lymph nodes from the bowel resected sample have fewer SNP and gene differences, confirming genetic similarity of *E. coli* within the same intestine. However, in two bowel resected samples, we found different *E. coli* strains from different tissue samples (patient 59: lymph nodes and normal mucosa and patient 78: Lymph nodes), which makes the identification of potential trigger of CD more difficult since either one could cause initiation of inflammation.

From the time when CD-associated *E. coli* was defined, its phylogeny has been questioned whether it was evolved from a common commensal ancestor by acquiring virulence genes or having different ancestral origin (Rolhion et al., 2007; Miquel et al., 2010). In this study we suggest that CD-associated *E. coli* are phylogenetically diverse with high interpatient and inpatient heterogeneity in serotypes, phylogroups and MLSTs indicating only functional /phenotypic similarities between CD, and no common genotype of *E. coli* and no common origin. However, some independently isolated strains (39-2-Ti18, 40-1-R9, 45-1-Ti1, 70-4-DC20, CD34LN, 241111, 242123, 250124, 251136) from different Crohn's disease patients had same sequence type (ST-95), serotype (either O1 type or O1:H7 type) and highly similar molecular properties (plasmids, virulent genes, bacteriocin) and may share a common origin. This also leads to the fact that, CD-associated phenotype could arise by horizontal gene transfer possibly mediated through plasmids or phage from other bacteria.

The association of *E. coli* and Crohn's disease raises the question whether it is a pathogen, or a survivor. Virulence property distribution among the CD isolates and healthy control isolates indicates that there was no significant difference in the presence of virulence gene among these two groups. However, in bacteriocin analysis we found that bottromycin bacteriocin was significantly associated with CD compared with healthy controls ($p = 0.0396$). This may indicate that the bottromycin bacteriocin may help the CD-associated bacteria to survive in the inflamed intestine by the bactericidal activity possibly by outcompeting other bacterial. In genome-wide association studies, we found that genes enriched in CD *E. coli* are associated with conjugation and pilus assembly, plasmid maintenance and post-segregational killing of plasmid free cells, conjugal transfer of DNA between bacterial cells, fimbrial protein and type I and type II toxin antitoxin system proteins in plasmids. These genes are

commonly associated with pathogenic *E. coli* (like plasmids of uro-pathogenic *Escherichia coli* UTI89) than commensals suggesting a pathogenic role of CD *E. coli.* This also provides further support that CD-specific features in *E. coli* strains are not specific genes, but combined function of many genes, possibly obtained by horizontal gene transfer.

In terms of molecular properties there were no significant differences between *E. coli* isolates from adult CD and HC. However, we identified some genes overrepresented in adult CD isolates compared to healthy control typically involved with conjugative transfer of DNA, plasmid maintenance, fertility inhibition protein, phage integrase family protein, signal recognition particle, transcriptional regulator, plasmid stability protein, lytic transglycosylase, type II toxin-antitoxin system toxin CcdB and antitoxin CcdA, multifunctional conjugation protein and fimbrial chaperone protein. Of note, most of the plasmid-associated genes are present with a perfect match in *E. coli* UTI89 plasmid pUTI89 isolated from the urinary bladder of patient with urinary tract infections (UTI). k-mer analysis between adult CD and healthy controls revealed that genes that encode a hypothetical protein, DUF945 domain-containing protein, NADPH-dependent aldehyde reductase Ahr and IS4 family transposase are associated significantly with adult CD. This result could be interpreted that the function of some genes associated with adult CD is still unknown and needs to be explored further.

On the other hand, different molecular properties were found to be significantly associated with paediatric CD compared with HC. Such as virulence factor-associated gene ireA was (siderophore receptor) was significantly associated with paediatric CD ($p = 0.0174$) compared with HC. Similarly, plasmid IncFII (29) was also significantly associated with paediatric CD isolates compared to healthy controls ($p = 0.0214$). In healthy controls, this IncFII (29) plasmid was found only on three isolates out of 27,

and two of them were AIEC and non-AIEC pair from same patients with same ST which varies only by 150-200 gene. So, if we consider these two as one, the significance will be higher (p = 0.0065). In another plasmid analysis using three different tools we also had similar findings. Paediatric CD isolates were found to be significantly associated (n = 9; p = 0.0034) with the *Escherichia coli* UTI89 plasmid pUTI89 and closely similar plasmids like *Escherichia coli* UM146 plasmid pUM146, *Escherichia coli* Eco889 plasmid pECO-93a, *Escherichia coli* LBV005/17 plasmid pR005, *Escherichia coli* SF-166 plasmid pSF-166-1 and *Escherichia coli* UMN026 plasmid p1ESCUM. Plasmid contigs identified in isolates from 9 paediatric patients out of 23 were highly similar to the previously published reference uro-pathogenic strains of *Escherichia coli* UTI89 plasmid pUTI89 (Additional files). This plasmid was also found in both AIEC and non-AIEC pairs of one healthy individual. *Escherichia coli* UTI89 plasmid pUTI89 is 114230 bp length. This plasmid mostly represented by replication proteins, ABC transporter, stable plasmid inheritance protein, F fimbriae pilin precursor pilus assembly protein, DNA binding protein, fertility inhibitor protein, site specific recombinase, proteins involved in DNA maintenance and conjugation, and possible virulence determinants such as enterotoxin protein, outer membrane colicin receptors, thioredoxin family protein, resistance proteins, etc. In capsule analysis we also found that, ABC-dependent capsules (groups II and III) were significantly associated with paediatric CD compared to healthy controls (P = 0.0119). In addition to that, bottromycin is strongly associated (p = <0.0001) with paediatric CD compared to healthy controls. Similarly, carocin_D was also significantly associated (p = 0.0185) with paediatric CD (17/23; 74%) isolates compared to healthy controls.

Genome-wide association studies and pan-genome analysis found some significantly associated genes in paediatric CD isolates compared with HC. Those genes are

associated with elongation factor Tu 1 (235ufa), DUF945 domain-containing protein, type VI secretion system PAAR protein, lytic transglycosylase, GGDEF domain-containing protein, replication protein, IncF plasmid conjugative transfer protein TraI, fertility inhibition protein FinO, phage integrase family protein, general stress protein A, TraE protein of plasmid F, and hypothetical protein. Greater numbers of plasmid-associated genes were found in paediatric CD isolates compared with HC. K-mer analysis suggest that paediatric CD *E. coli* are associated with DUF945 domain-containing protein, type-F conjugative transfer system secretin TraK protein, and IS66-like element ISEc8 family transposase. These molecular properties and genes presumably provide *E. coli* with an increased ability for intestine colonization and increase survival and fitness to play a pathogenic role to produce a chronic inflammatory condition in the gut of the paediatric CD patients.

Paediatric CD *E. coli* isolates showed significant differences in molecular properties when compared with adult CD isolates. Col156 plasmid was significantly associated ($p = 0.0033$) with paediatric CD *E. coli* compared with adult CD. ABC-dependent capsules (groups II and III) were significantly associated with paediatric CD compared with adult CD ($P = 0.0001$). In bacteriocin analysis we also found that bottromycin is strongly associated with paediatric CD compared to adult CD ($p = <0.0001$) and carocin_D was also significantly associated with paediatric CD (17/23; 74%; $p = 0.0015$) isolates compared to adult CD. However, these *E. coli* strains come from another laboratory and come from faecal samples. This difference may be due to the different geographical location and sample type. In addition to that, CRISPR (sequence: CCACCTTTTTTACCTGCTTCAGATGC) was associated with paediatric CD (17/23, 74%) isolates compared to adult CD ($p = 0.0001$). In genome-wide association studies, it was found that genes associated with flagellar biosynthesis

protein FlhA, recombination-promoting nuclease RpnA, universal stress protein UspG, sensor protein TorS, hydrogenase-4 component G and hypothetical proteins were overrepresented in adult CD isolates compared with paediatric CD isolates. On the other hand, genes associated with fimbrial chaperone, elongation factor Tu1(tufA), AlpA family transcriptional regulator, inovirus Gp2 family protein, general stress protein A, inner membrane ABC transporter, ATP-binding protein YddA, anti-restriction protein KlcA, type IV toxin-antitoxin system toxin CbtA and some hypothetical proteins were overrepresented significantly in paediatric CD isolates. In mixed effect K-mer analysis between adult CD and the paediatric shows that a gene associated with flagellar biosynthesis protein FlhB (with 48 hits) was significantly associated with adult CD and supports the COGs analysis results. This distinct difference between adult CD and paediatric CD may be due to geographical distribution and type of sample, as paediatric CD isolates were obtained from patients in the UK and isolated from fecal samples, whereas adult isolates were obtained from mucosa and from Australian patients. However, previous studies found distinct phenotypic differences in paediatric CD compared with adult CD and significant differences (de Bie, 2013; Pigneur, 2010) in the molecular properties and genes could explain this difference.

When we compared the molecular properties and genes of lymph nodes isolates with HC isolates, we observed some significant differences. In capsule analysis, Group iV_e capsule was significantly associated with lymph node CD isolates compared to control isolates ($p = 0.0126$). In addition, bottromycin bacteriocin was found to be significantly associated ($n = 8/9$; 88%; $p = 0.023$) with lymph node isolates compared to healthy controls. Genome-wide association studies revealed that genes overrepresented in LN are associated with tRNA-Phe (gaa), fructose-bisphosphatase class II protein, and catalase-peroxidase HPI, katG. KatG functions to alter host cell

physiology and promote bacterial survival in host tissues. In a new study researchers found that Crohn's disease associated *E. coli* survival within the macrophage correlates with resistance to superoxide (Tawfik et al., 2019). When I did the visualization of the nucleotide sequences in these variants of katG in MEGA and translated both variants to protein, I found that in position 29 amino acid G (glycine) in KatG is replaced by A(alanine) in group_7945 variant. This group 7045 variant of katG is overrepresented in lymph-node isolates. Under normal circumstances, activated macrophages secrete reactive oxygen and nitrogen species as a defence mechanism against microorganisms during infection. Macrophages use several oxygen-dependent and oxygen -independent bactericidal pathways (Bennett et al., 2015). The toxic effects of reactive oxygen and nitrogen species have been determined in a previous study (Storz and Imlayt, 1999). Most bacteria are rapidly killed and degraded in the phagolysosome, but a few bacteria survive in macrophages. Inside the phagosome, bacteria are killed in two different ways. One is mediated through the generation of reactive oxygen species following a respiratory burst. The second mechanism is independent of oxygen where proteolytic enzymes in preformed granules kill microbes (Flaherty, 2012). NADPH oxidase generates major reactive oxygen species such as singlet oxygen, hydrogen peroxide, and hydroxyl radicals, which are toxic to microbes inside the phagosome. The intracellular H2O2 accumulates to toxic levels, and microbial scavenging systems are prone to be inadequate (Seaver and Imlay, 2001). This toxic H2O2 causes damage to bacterial DNA, mutations or lethal lesions to bacteria (Park et al., 2005). H2O2 is highly toxic, and therefore to keep its concentration minimal, bacteria create scavenging enzymes. Mutants that cannot secrete these enzymes grow poorly, undergo excessive mutagenesis, or die. Several enzymes involved in H2O2 degradation enable bacteria

to survive under oxidative stress. In addition, catalases and NADH peroxidase (Ahp) are primary scavengers of *E. coli* and many other bacteria. A study has demonstrated their importance through phenotypic assays and directly measuring the H2O2 concentration in vivo (Mishra and Imlay, 2012). Although many other enzymes were reported to perform similar functions and degrade H2O2 in vitro, their roles in vivo have not been confirmed.

Catalase and hydroperoxides enzymes are an important component of bacterial defensive mechanisms to oxidative stress, mostly by converting highly reactive oxygen species to less reactive molecules (Schellhorn, 1995). Therefore, katG and katE genes, which encode for an enzyme with catalase and peroxidase activity are essential for bacterial survival in oxidative stress by converting hydrogen peroxide to oxygen and water (Chaithawiwat et al., 2016).

The above observations suggest that CD-enriched genomic features of lymph node-associated CD *E. coli* presumably provide bacteria with an increased ability for survival inside the macrophage resulting in a chronic stimulus to the immune system that causes chronic inflammation.

Hence it seems that lymph node-associated *E. coli* are not just commensal strains but can survive within the inflammatory condition. In addition, they possess some features of pathogenic *E. coli*. So, it is likely that some *E. coli* strains acquire genes via horizontal gene transfer and may convert from commensal to pathogenic strains.

Based on molecular properties, Aphthous ulcer-associated CD isolated were found to be significantly associated with the type T conjugative elements compared with healthy isolates ($p = 0.0035$). Genome-wide association studies revealed some significantly associated genes in AU isolates compared with healthy controls. The most significant

genes are associated with helix-turn-helix domain-containing protein, helix-turn-helix transcriptional regulator/ HTH-type transcriptional regulator PrtR, cell division inhibitor protein, regulator/ bacteriophage regulatory protein CII, relaxosome protein TraM, and DUF4102 domain-containing protein. In k-mer analysis the highest hits (112) were found against the gene encoding helix-turn-helix domain-containing protein, region: YdaS_antitoxin/ phage protein Cro-UMN026 that also supports the COGs (gene) finding. The role of this gene in AU associated *E. coli* isolates need to be explored further.

We did not observe any molecular properties to be significantly associated with UC. However, in genome-wide association studies we found some genes were significantly associated with *E. coli* isolates in UC. Genes overrepresented significantly in UC are associated with plasmid-associated DUF945 domain-containing protein, plasmid-associated lytic transglucosylase, fertility inhibition protein, protein TraJ, flagellar hook-associated protein, fimbrial assembly protein, toxin CcdB, antitoxin CcdA, AIDA-I family autotransporter YfaL, helix-turn-helix domain-containing protein, putative fimbrial-like adhesin protein (yehA) and hypothetical proteins. In k-mer analysis between UC and healthy controls we found k-mer with top hits were associated with genes that code for conjugal transfer protein TraP, and hypothetical proteins. In Scoary phenotype genotype association studies, 24 genes were found associated with UC, among these genes, five are plasmid-associated genes; fertility inhibition protein (finO), conjugal transfer of DNA (traJ), plasmid maintenance post segregational killing of plasmid free cell/ toxin antitoxin system (ccdA_2 and ccdB) and hypothetical protein that supports the genome wide association study conducted by Pyseer. genes are associated with putative allantoin permease, HTH-type transcriptional regulation, glutamine transport system permease protein, cardiolipin synthase B, putative signal

peptide peptidase, flagellar hook-associated protein 2, colicin I receptor, phage tail protein and hypothetical protein. The role of these genes in the pathogenesis of UC needs to the investigated in the future studies.

In the AIEC group, the most abundant serotype was O18: H7 (4/15; 27%) all of which belonged to ST-95 and were significantly associated with AIEC ($p = 0.029$) compared with non-AIEC. Virulence factor sfaS (S-fimbriae minor subunit) was significantly associated ($p = 0.0294$) with AIEC compared to non-AIEC isolates. Plasmid IncFIB (AP001918) was significantly associated ($p = 0.0143$) with non-AIEC compared to AIEC. In fixed and mixed effect models of the genome-wide association study, it was found that genes significantly associated with AIEC are associated with elongation factor Tu (tufB), helix-turn-helix domain-containing protein, phage head morphogenesis protein addiction module toxin RelE, LysR family transcriptional regulator, hypothetical protein, DUF1364 domain-containing protein, and AIDA-I family autotransporter YfaL. In k-mer analysis we found 560 k-mers associated with AIEC. Based on these annotated k-mer we found highest hits (101) were for a gene associated with type 1 fimbrial protein followed by histidinol-phosphate transaminase (40 hits), multidrug efflux RND transporter subunit MdtA (29 hits), type 1 fimbria D-mannose specific adhesin FimH (26 hits), and type 1 fimbrial protein fimA (11 hits). Association of type 1 fimbriae found in this k-mer analysis is supported by previous studies (Boudeau et al., 2001; Barnich et al., 2003; Barnich et al., 2007; Chassaing et al., 2011; Dreux et al., 2013).

3.6 Conclusion

Overall, we identified properties of *E. coli* associated with paediatric CD and adult CD isolates, as well as lymph nodes. These properties may allow *E. coli* to evade the host

immune system by conferring the ability to survive in the phagolysosome in patients with CD.

These need to be viewed in the context of the genetic background of an individual to see how specific bacteria cause IBD in that individual. In other words, we may need to be testing a specific isolate in a single person to find the reasons why that individual develops IBD. This will obviously be very difficult and time consuming. Future studies in our lab will involve isolating macrophage from patient bowels and co-culturing them with these *E. coli* strains to determine their response compared to healthy controls.

3.7 References

Adamus-Bialek, W., Wojtasik, A., Majchrzak, M., Sosnowski, M., and Parniewski, P. (2009). (CGG)(4)-Based PCR as a Novel Tool for Discrimination of Uropathogenic Escherichia coli Strains: Comparison with Enterobacterial Repetitive Intergenic Consensus-PCR. *Journal of Clinical Microbiology* 47(12), 3937-3944. doi: 10.1128/JCM.01036-09.

Ambrose, N.S., Johnson, M., Burdon, D.W., and Keighley, M.R. (1984). Incidence of pathogenic bacteria from mesenteric lymph nodes and ileal serosa during Crohn's disease surgery. *Br J Surg* 71(8), 623-625.

Ananthakrishnan, A.N. (2013). Environmental Risk Factors for Inflammatory Bowel Disease. *Gastroenterology & Hepatology* 9(6), 367-374.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44(W1), W16-21. doi: 10.1093/nar/gkw387.

Arredondo-Alonso, S., Rogers, M.R.C., Braat, J.C., Verschuuren, T.D., Top, J., Corander, J., et al. (2018). mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. *Microb Genom* 4(11). doi: 10.1099/mgen.0.000224.

Ashton, J.J., Ennis, S., and Beattie, R.M. (2017). Early-onset paediatric inflammatory bowel disease. *The Lancet Child & Adolescent Health* 1(Polito et al.), 147-158. doi: 10.1016/S2352-4642(17)30017-2.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology* 19(5), 455-477. doi: 10.1089/cmb.2012.0021.

Barnich, N., Boudeau, J., Claret, L., and Darfeuille-Michaud, A. (2003). Regulatory and functional co-operation of flagella and type 1 pili in adhesive and invasive abilities of AIEC strain LF82 isolated from a patient with Crohn's disease. *Molecular microbiology* 48(3), 781-794.

Barnich, N., Carvalho, F.A., Glasser, A.-L., Darcha, C., Jantscheff, P., Allez, M., et al. (2007). CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa

colonization in Crohn disease. *Journal of Clinical Investigation* 117(6), 1566-1574. doi: 10.1172/JCI30504.

Baumgart, D.C., and Carding, S.R. (2007). Inflammatory bowel disease: cause and immunobiology. *The Lancet* 369(9573), 1627-1640. doi: https://doi.org/10.1016/S0140-6736(07)60750-8.

Baumgart, M., Dogan, B., Rishniw, M., Weitzman, G., Bosworth, B., and Yantiss, R. (2007). Culture independent analysis of ilealmucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *The ISME Journal* 1. doi: 10.1038/ismej.2007.52.

Benchimol, E.I., Bernstein, C.N., Bitton, A., Carroll, M.W., Singh, H., Otley, A.R., et al. (2017). Trends in Epidemiology of Pediatric Inflammatory Bowel Disease in Canada: Distributed Network Analysis of Multiple Population-Based Provincial Health Administrative Databases. *Am J Gastroenterol* 112(7), 1120-1134. doi: 10.1038/ajg.2017.97.

Bennett, J.E., Dolin, R., and Blaser, M.J. (2015). Mandell, Douglas, and Bennett's principles and practice of infectious diseases. *Philadelphia, PA: Saunders* 2858.

Boudeau, J., Barnich, N., and Darfeuille-Michaud, A. (2001). Type 1 pili-mediated adherence of Escherichia coli strain LF82 isolated from Crohn's disease is involved in bacterial invasion of intestinal epithelial cells. *Molecular microbiology* 39(5), 1272-1284.

Boudeau, J., Glasser, A.-L., Masseret, E., Joly, B., and Darfeuille-Michaud, A. (1999). Invasive Ability of an Escherichia coliStrain Isolated from the Ileal Mucosa of a Patient with Crohn's Disease. *Infection and immunity* 67(9), 4499-4509.

Bringer, M.A., Barnich, N., Glasser, A.L., Bardot, O., and Darfeuille-Michaud, A. (2005). HtrA stress protein is involved in intramacrophagic replication of adherent and invasive Escherichia coli strain LF82 isolated from a patient with Crohn's disease. *Infect Immun* 73(Polito et al.), 712-721. doi: 10.1128/iai.73.2.712-721.2005.

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biology* 17(Goodhand et al.), 238. doi: 10.1186/s13059-016-1108-8.

Burke, D., and Axon, A. (1988). Adhesive Escherichia coli in inflammatory bowel disease and infective diarrhoea. *Bmj* 297(6641), 102-104.

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial Agents and Chemotherapy* 58(7), 3895-3903. doi: 10.1128/AAC.02412-14.

Chaithawiwat, K., Vangnai, A., McEvoy, J.M., Pruess, B., Krajangpan, S., and Khan, E. (2016). Role of oxidative stress in inactivation of Escherichia coli BW25113 by nanoscale zero-valent iron. *Science of The Total Environment* 565, 857-862. doi: https://doi.org/10.1016/j.scitotenv.2016.02.191.

Chassaing, B., Rolhion, N., de Vallée, A., Salim, S.a.Y., Prorok-Hamon, M., Neut, C., et al. (2011). Crohn disease–associated adherent-invasive E. coli bacteria target mouse and human Peyer's patches via long polar fimbriae. *The Journal of Clinical Investigation* 121(3), 966-975. doi: 10.1172/JCI44632.

Clermont, O., Christenson, J.K., Denamur, E., and Gordon, D.M. (2013). The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5(Goodhand et al.), 58-65. doi: 10.1111/1758-2229.12019.

Cleynen, I., Vazeille, E., Artieda, M., Verspaget, H.W., Szczypiorska, M., Bringer, M.-A., et al. (2014). Genetic and microbial factors modulating the ubiquitin proteasome system in inflammatory bowel disease. *Gut* 63(8), 1265-1274.

Coil, D., Jospin, G., and Darling, A.E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31(4)**,** 587-589. doi: 10.1093/bioinformatics/btu661.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Neron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* 46(W1)**,** W246-w251. doi: 10.1093/nar/gky425.

Cury, J., Abby, S.S., Doppelt-Azeroual, O., Néron, B., and Rocha, E.P. (2020). "Identifying Conjugative Plasmids and Integrative Conjugative Elements with CONJscan," in *Horizontal Gene Transfer*. Springer), 265-283.

de Bie, C. I., Paerregaard, A., Kolacek, S., Ruemmele, F. M., Koletzko, S., Fell, J. M., Escher, J. C., & EUROKIDS Porto IBD Working Group of ESPGHAN (2013). Disease phenotype at diagnosis in pediatric Crohn's disease: 5-year analyses of the EUROKIDS Registry. Inflammatory bowel diseases, 19(2), 378–385. https://doi.org/10.1002/ibd.23008

D'Haens, G.R., Geboes, K., Peeters, M., Baert, F., Penninckx, F., and Rutgeerts, P. (1998). Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded ileum. *Gastroenterology* 114(Polito et al.)**,** 262-267.

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

Darfeuille-Michaud, A., Neut, C., Barnich, N., Lederman, E., Di Martino, P., Desreumaux, P., et al. (1998a). Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease. *Gastroenterology* 115(6)**,** 1405-1413.

Darfeuille-Michaud, A., Neut, C., Barnich, N., Lederman, E., Martino, P., and Desreumaux, P. (1998b). Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease. *Gastroenterology* 115. doi: 10.1016/s0016-5085(98)70019-8.

Darling, A.C., Mau, B., Blattner, F.R., and Perna, N.T. (2004a). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14. doi: 10.1101/gr.2289704.

Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004b). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research* 14(7)**,** 1394-1403. doi: 10.1101/gr.2289704.

de Jong, A., van Hijum, S.A.F.T., Bijlsma, J.J.E., Kok, J., and Kuipers, O.P. (2006). BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Research* 34(Web Server issue)**,** W273-W279. doi: 10.1093/nar/gkl237.

de Souza, H.L., de Carvalho, V.R., Romeiro, F.G., Sassaki, L.Y., Keller, R., and Rodrigues, J. (2012). Mucosa-associated but not luminal Escherichia coli is augmented in Crohn's disease and ulcerative colitis. *Gut Pathogens* 4**,** 21-21. doi: 10.1186/1757-4749-4-21.

Dickinson, R., Varian, S., Axon, A., and Cooke, E. (1980). Increased incidence of faecal coliforms with in vitro adhesive and invasive properties in patients with ulcerative colitis. *Gut* 21(9)**,** 787-792.

Dixit O.V.A, O'Brien, C., Paul, P., and Gordon, D.M. (2018). Within-host evolution versus immigration as a determinant of Escherichia coli diversity in the human gastrointestinal tract. Environmental Microbiology 20(3), 993-1001. doi: doi:10.1111/1462-2920.14028.

Doherty, G.A., Bennett, G.C., Cheifetz, A.S., and Moss, A.C. (2010). Meta-analysis: targeting the intestinal microbiota in prophylaxis for post-operative Crohn's disease. *Aliment Pharmacol Ther* 31(8)**,** 802-809. doi: 10.1111/j.1365-2036.2010.04231.x.

Dreux, N., Denizot, J., Martinez-Medina, M., Mellmann, A., Billig, M., Kisiela, D., et al. (2013). Point mutations in FimH adhesin of Crohn's disease-associated adherent-invasive Escherichia coli enhance intestinal inflammatory response. *PLoS pathogens* 9(Goodhand et al.)**,** e1003141.

Dunne, K.A., Allam, A., McIntosh, A., Houston, S.A., Cerovic, V., Goodyear, C.S., et al. (2013). Increased S-nitrosylation and proteasomal degradation of caspase-3 during infection contribute to the persistence of adherent invasive Escherichia coli (AIEC) in immune cells. *Plos one* 8(7)**,** e68386.

Dutta, A.K., and Chacko, A. (2016). Influence of environmental factors on the onset and course of inflammatory bowel disease. *World Journal of Gastroenterology* 22(3)**,** 1088-1100. doi: 10.3748/wjg.v22.i3.1088.

Eckburg, P.B., and Relman, D.A. (2007). The Role of Microbes in Crohn's Disease. *Clinical Infectious Diseases* 44(Polito et al.)**,** 256-262. doi: 10.1086/510385.

Feller, M., Huwiler, K., Schoepfer, A., Shang, A., Furrer, H., and Egger, M. (2010). Long-Term Antibiotic Treatment for Crohn's Disease: Systematic Review and Meta-Analysis of Placebo-Controlled Trials. *Clinical Infectious Diseases* 50(4)**,** 473-480. doi: 10.1086/649923.

Flaherty, D.K. (2012). *Immunology for pharmacy.* St. Louis, Mo.: Elsevier.

Flint, H.J., Scott, K.P., Louis, P., and Duncan, S.H. (2012). The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* 9(10)**,** 577-589. doi: 10.1038/nrgastro.2012.156.

Frank, D.N., St. Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences* 104(34)**,** 13780-13785. doi: 10.1073/pnas.0706625104.

Fujimura, Y., Kamoi, R., and Iida, M. (1996). Pathogenesis of aphthoid ulcers in Crohn's disease: correlative findings by magnifying colonoscopy, electron microscopy, and immunohistochemistry. *Gut* 38(5)**,** 724-732.

Giaffer, M., Holdsworth, C., and Duerden, B. (1992). Virulence properties of Escherichia coli strains isolated from patients with inflammatory bowel disease. *Gut* 33(5)**,** 646-650.

Glas, J., Seiderer, J., Bues, S., Stallhofer, J., Fries, C., Olszak, T., et al. (2013). IRGM variants and susceptibility to inflammatory bowel disease in the German population. *PLoS One* 8(Goodhand et al.)**,** e54338. doi: 10.1371/journal.pone.0054338.

Glasser, A.-L., Boudeau, J., Barnich, N., Perruchot, M.-H., Colombel, J.-F., and Darfeuille-Michaud, A. (2001). Adherent invasive Escherichia coli strains from patients with Crohn's disease survive and replicate within macrophages without inducing host cell death. *Infection and immunity* 69(9)**,** 5529-5537.

Glocker, E.-O., Kotlarz, D., Boztug, K., Gertz, E.M., Schäffer, A.A., Noyan, F., et al. (2009). Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *New England Journal of Medicine* 361(21)**,** 2033-2045.

Gordon, D.M., O'brien, C.L., and Pavli, P. (2015). Escherichia coli diversity in the lower intestinal tract of humans. *Environmental microbiology reports* 7(4)**,** 642-648.

Guglielmini, J., Néron, B., Abby, S.S., Garcillán-Barcia, M.P., de la Cruz, F., and Rocha, E.P.C. (2014). Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic acids research* 42(9)**,** 5715-5727. doi: 10.1093/nar/gku194.

Haberman, Y., Tickle, T.L., Dexheimer, P.J., Kim, M.-O., Tang, D., Karns, R., et al. (2014). Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *The Journal of Clinical Investigation* 124(8)**,** 3617-3633. doi: 10.1172/JCI75436.

Halfvarson, J., Bodin, L., Tysk, C., Lindberg, E., and Jarnerot, G. (2003). Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 124(7), 1767-1773.

Hammer, Ø., Harper, D., and Ryan, P. (2001). Paleontological statistics software: package for education and data analysis. *Palaeontologia Electronica* (4).

Herzog, D., Fournier, N., Buehr, P., Rueger, V., Koller, R., Heyland, K., et al. (2017). Prevalence of intestinal complications in inflammatory bowel disease: a comparison between paediatric-onset and adult-onset patients. *Eur J Gastroenterol Hepatol* 29(8), 926-931. doi: 10.1097/meg.0000000000000896.

Hizawa, K., Iida, M., Kohrogi, N., Kuroki, F., Yao, T., Sakamoto, K., et al. (1994). Crohn disease: early recognition and progress of aphthous lesions. *Radiology* 190(Polito et al.), 451-454.

Hooper, L.V., Littman, D.R., and Macpherson, A.J. (2012). Interactions Between the Microbiota and the Immune System. *Science* 336(6086), 1268.

Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol* 52(5), 1501-1510. doi: 10.1128/jcm.03617-13.

Joensen, K.G., Tetzschner, A.M., Iguchi, A., Aarestrup, F.M., and Scheutz, F. (2015). Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 53(8), 2410-2426. doi: 10.1128/jcm.00008-15.

Kaas, R.S., Leekitcharoenphon, P., Aarestrup, F.M., and Lund, O. (2014). Solving the Problem of Comparing Whole Bacterial Genomes across Different Sequencing Platforms. *PLOS ONE* 9(8), e104984. doi: 10.1371/journal.pone.0104984.

Kelsen, J.R., Conrad, M.A., Dawany, N., Patel, T., Shraim, R., Merz, A., et al. (2019). The Unique Disease Course of Children with Very Early onset-Inflammatory Bowel Disease. *Inflammatory Bowel Diseases*. doi: 10.1093/ibd/izz214.

Kennedy, N.A., Lamb, C.A., Berry, S.H., Walker, A.W., Mansfield, J., Parkes, M., et al. (2018). The Impact of NOD2 Variants on Fecal Microbiota in Crohn's Disease and Controls Without Gastrointestinal Disease. *Inflammatory Bowel Diseases* 24(3), 583-592. doi: 10.1093/ibd/izx061.

Kiely, C.J., Pavli, P., and O'Brien, C.L. (2018). The microbiome of translocated bacterial populations in patients with and without inflammatory bowel disease. *Internal Medicine Journal* 48(11), 1346-1354. doi: https://doi.org/10.1111/imj.13998.

Kinjo, F., Tomiyama, R., Hokama, A., Matayoshi, R., Yonamine, Y., Sunagawa, T., et al. (2004). Gastrointestinal: Aphthous lesions to overt Crohn's disease. *Journal of gastroenterology and hepatology* 19(6), 707-707.

Kleer, C.G., and Appelman, H.D. (2001). Surgical pathology of Crohn's disease. *Surg Clin North Am* 81(Goodhand et al.), 13-30, vii.

Knights, D., Silverberg, M.S., Weersma, R.K., Gevers, D., Dijkstra, G., Huang, H., et al. (2014). Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome medicine* 6(12), 107.

Kotlowski, R., Bernstein, C.N., Sepehri, S., and Krause, D.O. (2007a). High prevalence of Escherichia coli belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut* 56(5), 669-675. doi: 10.1136/gut.2006.099796.

Kotlowski, R., Bernstein, C.N., Sepehri, S., and Krause, D.O. (2007b). High prevalence of Escherichia coli belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut* 56. doi: 10.1136/gut.2006.099796.

Kraehenbuhl, J.P., and Neutra, M.R. (2000). Epithelial M cells: differentiation and function. *Annu Rev Cell Dev Biol* 16, 301-332. doi: 10.1146/annurev.cellbio.16.1.301.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7), 1870-1874. doi: 10.1093/molbev/msw054.

Laffineur, G., Lescut, D., Vincent, P., Quandalle, P., Wurtz, A., and Colombel, J.F. (1992). [Bacterial translocation in Crohn disease]. *Gastroenterol Clin Biol* 16(10), 777-781.

Lapaquette, P., Bringer, M.A., and Darfeuille-Michaud, A. (2012). Defects in autophagy favour adherent-invasive Escherichia coli persistence within macrophages leading to increased pro-inflammatory response. *Cellular microbiology* 14(6), 791-807.

Lapaquette, P., Glasser, A.L., Huett, A., Xavier, R.J., and Darfeuille-Michaud, A. (2010). Crohn's disease-associated adherent-invasive E. coli are selectively favoured by impaired autophagy to replicate intracellularly. *Cellular microbiology* 12(Goodhand et al.), 99-113.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 50(4), 1355-1361. doi: 10.1128/jcm.06094-11.

Ledder, O., Catto-Smith, A.G., Oliver, M.R., Alex, G., Cameron, D.J., and Hardikar, W. (2014). Clinical patterns and outcome of early-onset inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 59(5), 562-564. doi: 10.1097/mpg.0000000000000465.

Lees, J.A., Galardini, M., Bentley, S.D., Weiser, J.N., and Corander, J. (2018). pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* 34(24), 4310-4312. doi: 10.1093/bioinformatics/bty539.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44(W1), W242-W245. doi: 10.1093/nar/gkw290.

Loddo, I., and Romano, C. (2015). Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. *Frontiers in Immunology* 6, 551. doi: 10.3389/fimmu.2015.00551.

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., Duncan, S.H., Flint, H.J., et al. (2014). Mucosa-associated Faecalibacterium prausnitzii and Escherichia coli co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int J Med Microbiol* 304(3-4), 464-475. doi: 10.1016/j.ijmm.2014.02.009.

Lovasz, B.D., Lakatos, L., Horvath, A., Szita, I., Pandur, T., Mandel, M., et al. (2013). Evolution of disease phenotype in adult and pediatric onset Crohn's disease in a population-based cohort. *World J Gastroenterol* 19(14), 2217-2226. doi: 10.3748/wjg.v19.i14.2217.

Martin, H.M., Campbell, B.J., Hart, C.A., Mpofu, C., Nayar, M., and Singh, R. (2004). Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology* 127. doi: 10.1053/j.gastro.2004.03.054.

Martinez-Medina, M., and Garcie-Gil, L.G. (2014). Escherichia coli In chronic inflammatory bowel diseases: an update on adherent invasive Escherichia coli pathogenicity. *World J Gastrointest Pathophysiol* 5.

Meconi, S., Vercellone, A., Levillain, F., Payré, B., Al Saati, T., Capilla, F., et al. (2007). Adherent-invasive Escherichia coli isolated from Crohn's disease patients induce granulomas in vitro. *Cellular microbiology* 9(5), 1252-1261.

Miquel, S., Peyretaillade, E., Claret, L., de Vallée, A., Dossat, C., Vacherie, B., et al. (2010). Complete Genome Sequence of Crohn's Disease-Associated Adherent-Invasive E. coli Strain LF82. *PLOS ONE* 5(9), e12714. doi: 10.1371/journal.pone.0012714.

Mishra, S., and Imlay, J. (2012). Why do bacteria use so many enzymes to scavenge hydrogen peroxide? *Archives of biochemistry and biophysics* 525(Polito et al.), 145-160. doi: 10.1016/j.abb.2012.04.014.

Molodecky, N.A., and Kaplan, G.G. (2010). Environmental Risk Factors for Inflammatory Bowel Disease. *Gastroenterology & Hepatology* 6(5), 339-346.

Naser, S.A., Arce, M., Khaja, A., Fernandez, M., Naser, N., Elwasila, S., et al. (2012). Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. *World journal of gastroenterology* 18(5), 412-424. doi: 10.3748/wjg.v18.i5.412.

Nataro, J.P., and Kaper, J.B. (1998). Diarrheagenic escherichia coli. *Clinical microbiology reviews* 11(Goodhand et al.), 142-201.

Nitzan, O., Elias, M., Peretz, A., and Saliba, W. (2016). Role of antibiotics for treatment of inflammatory bowel disease. *World Journal of Gastroenterology* 22(3), 1078-1087. doi: 10.3748/wjg.v22.i3.1078.

O'Brien, C.L., Bringer, M.-A., Holt, K.E., Gordon, D.M., Dubois, A.L., Barnich, N., et al. (2016). Comparative genomics of Crohn's disease-associated adherent-invasive Escherichia coli. *Gut*. doi: 10.1136/gutjnl-2015-311059.

O'Brien, C.L., Pavli, P., Gordon, D.M., and Allison, G.E. (2014). Detection of bacterial DNA in lymph nodes of Crohn's disease patients using high throughput sequencing. *Gut*. doi: 10.1136/gutjnl-2013-305320.

O'Boyle, C., MacFie, J., Mitchell, C., Johnstone, D., Sagar, P., and Sedman, P. (1998). Microbiology of bacterial translocation in humans. *Gut* 42(Goodhand et al.), 29-35.

Orholm, M., Binder, V., Sorensen, T.I., Rasmussen, L.P., and Kyvik, K.O. (2000). Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand J Gastroenterol* 35(10), 1075-1081.

Ostrowski, J., Paziewska, A., Lazowska, I., Ambrozkiewicz, F., Goryca, K., Kulecka, M., et al. (2016). Genetic architecture differences between pediatric and adult-onset inflammatory bowel diseases in the Polish population. *Scientific Reports* 6, 39831. doi: 10.1038/srep39831

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22), 3691-3693. doi: 10.1093/bioinformatics/btv421.

Park, S., You, X., and Imlay, J.A. (2005). Substantial DNA damage from submicromolar intracellular hydrogen peroxide detected in Hpx⁻ mutants of <em>Escherichia coli</em&gt. *Proceedings of the National Academy of Sciences of the United States of America* 102(26), 9317. doi: 10.1073/pnas.0502051102.

Petersen, A.M., Nielsen, E.M., Litrup, E., Brynskov, J., Mirsepasi, H., and Krogfelt, K.A. (2009). A phylogenetic group of Escherichia coli associated with active left-sided inflammatory bowel disease. *BMC Microbiol* 9. doi: 10.1186/1471-2180-9-171.

Pierik, M., De Hertogh, G., Vermeire, S., Van Assche, G., Van Eyken, P., Joossens, S., et al. (2005). Epithelioid granulomas, pattern recognition receptors, and phenotypes of Crohn's disease. *Gut* 54(Polito et al.), 223. doi: 10.1136/gut.2004.042572.

Pigneur, B., Seksik, P., Viola, S., Viala, J., Beaugerie, L., Girardet, J. P., Ruemmele, F. M., & Cosnes, J. (2010). Natural history of Crohn's disease: comparison between childhood- and adult-onset disease. Inflammatory bowel diseases, 16(6), 953–961. https://doi.org/10.1002/ibd.21152

Proctor, L.M., Creasy, H.H., Fettweis, J.M., Lloyd-Price, J., Mahurkar, A., Zhou, W., et al. (2019). The Integrative Human Microbiome Project. *Nature* 569(7758), 641-648. doi: 10.1038/s41586-019-1238-8.

Rendueles, O., Garcia-Garcera, M., Neron, B., Touchon, M., and Rocha, E.P.C. (2017). Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. *PLoS Pathog* 13(7), e1006525. doi: 10.1371/journal.ppat.1006525.

Rigoli, L., and Caruso, R.A. (2014). Inflammatory bowel disease in pediatric and adolescent patients: a biomolecular and histopathological review. *World journal of gastroenterology* 20(30)**,** 10262-10278. doi: 10.3748/wjg.v20.i30.10262.

Robertson, J., and Nash, J.H.E. (2018). MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom* 4(8). doi: 10.1099/mgen.0.000206.

Rolhion, N., Carvalho, F.A., and Darfeuille-Michaud, A. (2007). OmpC and the sigma(E) regulatory pathway are involved in adhesion and invasion of the Crohn's disease-associated Escherichia coli strain LF82. *Mol Microbiol* 63. doi: 10.1111/j.1365-2958.2007.05638.x.

Rolhion, N., and Darfeuille-Michaud, A. (2007). Adherent-invasive Escherichia coli in inflammatory bowel disease. *Inflammatory bowel diseases* 13(10)**,** 1277-1283.

Royer, G., Decousser, J.W., Branger, C., Dubois, M., Médigue, C., Denamur, E., et al. (2018). PlaScope: a targeted approach to assess the plasmidome from genome assemblies at the species level. *Microbial genomics* 4(9)**,** e000211. doi: 10.1099/mgen.0.000211.

Rutgeerts, P., Goboes, K., Peeters, M., Hiele, M., Penninckx, F., Aerts, R., et al. (1991). Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum. *Lancet* 338(8770)**,** 771-774.

Ryan, P., Kelly, R.G., Lee, G., Collins, J.K., O'Sullivan, G.C., O'Connell, J., et al. (2004). Bacterial DNA within Granulomas of Patients with Crohn's Disease[mdash]Detection by Laser Capture Microdissection and PCR. *Am J Gastroenterol* 99(8)**,** 1539-1543.

Sagiv-Friedgut, K., Karban, A., Weiss, B., Shaoul, R., Shamir, R., Bujanover, Y., et al. (2010). Early-onset Crohn disease is associated with male sex and a polymorphism in the IL-6 promoter. *J Pediatr Gastroenterol Nutr* 50(Goodhand et al.)**,** 22-26. doi: 10.1097/MPG.0b013e3181b7a6a4.

Salem, M., Ammitzboell, M., Nys, K., Seidelin, J.B., and Nielsen, O.H. (2015). ATG16L1: A multifunctional susceptibility factor in Crohn disease. *Autophagy* 11(4)**,** 585-594. doi: 10.1080/15548627.2015.1017187.

Sartor, R. (1995). Microbial agents in the pathogenesis, differential diagnosis, and complications of inflammatory bowel disease. *Infections of the gastrointestinal tract***,** 435-458.

Sartor, R.B. (2006). Mechanisms of Disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nature Clinical Practice Gastroenterology &Amp; Hepatology* 3**,** 390. doi: 10.1038/ncpgasthep0528.

Schellhorn, H. (1995). Regulation of hydroperoxidase (catalase) expression in Escherichia coli. *FEMS microbiology letters* 131(Polito et al.)**,** 113-119.

Seaver, L.C., and Imlay, J.A. (2001). Hydrogen peroxide fluxes and compartmentalization inside growing Escherichia coli. *Journal of bacteriology* 183(24)**,** 7182-7189.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30. doi: 10.1093/bioinformatics/btu153.

Shim, J.O. (2019). Recent Advance in Very Early Onset Inflammatory Bowel Disease. *Pediatric gastroenterology, hepatology & nutrition* 22(Goodhand et al.)**,** 41-49. doi: 10.5223/pghn.2019.22.1.41.

Shivashankar, R., Tremaine, W.J., Harmsen, W.S., and Loftus, E.V., Jr. (2017). Incidence and Prevalence of Crohn's Disease and Ulcerative Colitis in Olmsted County, Minnesota From 1970 Through 2010. *Clin Gastroenterol Hepatol* 15(6)**,** 857-863. doi: 10.1016/j.cgh.2016.10.039.

Stange, E.F., Travis, S.P.L., Vermeire, S., Beglinger, C., Kupcinskas, L., Geboes, K., et al. (2006). European evidence based consensus on the diagnosis and management of

Crohn's disease: definitions and diagnosis. *Gut* 55(Suppl 1), i1-i15. doi: 10.1136/gut.2005.081950a.

Storz, G., and Imlayt, J.A. (1999). Oxidative stress. *Current opinion in microbiology* 2(Polito et al.), 188-194.

Su, J.W., Ma, J.J., and Zhang, H.J. (2015). Use of antibiotics in patients with Crohn's disease: a systematic review and meta-analysis. *J Dig Dis* 16(Polito et al.), 58-66. doi: 10.1111/1751-2980.12216.

Swidsinski, A., Ladhoff, A., Pernthaler, A., Swidsinski, S., Loening–Baucke, V., Ortner, M., et al. (2002). Mucosal flora in inflammatory bowel disease. *Gastroenterology* 122(Goodhand et al.), 44-54.

Sýkora, J., Pomahačová, R., Kreslová, M., Cvalínová, D., Štych, P., and Schwarz, J. (2018). Current global trends in the incidence of pediatric-onset inflammatory bowel disease. *World journal of gastroenterology* 24(25), 2741.

Takesue, Y., Ohge, H., Uemura, K., Imamura, Y., Murakami, Y., Yokoyama, T., et al. (2002a). Bacterial translocation in patients with Crohn's disease undergoing surgery. *Dis Colon Rectum* 45(12), 1665-1671. doi: 10.1097/01.Dcr.0000034515.66401.8b.

Takesue, Y., Ohge, H., Uemura, K., Imamura, Y., Murakami, Y., Yokoyama, T., et al. (2002b). Bacterial translocation in patients with Crohn’s disease undergoing surgery. *Diseases of the colon & rectum* 45(12), 1665-1671.

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10(3), 512-526. doi: 10.1093/oxfordjournals.molbev.a040023.

Tawfik, A., Knight, P., Duckworth, C.A., Pritchard, D.M., Rhodes, J.M., and Campbell, B.J. (2019). Replication of Crohn’s Disease Mucosal E. coli Isolates inside Macrophages Correlates with Resistance to Superoxide and Is Dependent on Macrophage NF-kappa B Activation. *Pathogens* 8(Polito et al.), 74.

Thompson, N.P., Driscoll, R., Pounder, R.E., and Wakefield, A.J. (1996). Genetics versus environment in inflammatory bowel disease: results of a British twin study. *Bmj* 312(7023), 95-96. doi: 10.1136/bmj.312.7023.95.

Van Limbergen, J., Russell, R.K., Drummond, H.E., Aldhous, M.C., Round, N.K., Nimmo, E.R., et al. (2008). Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease. *Gastroenterology* 135(4), 1114-1122. doi: 10.1053/j.gastro.2008.06.081.

Wei, G., Pan, L., Du, H., Chen, J., and Zhao, L. (2004). ERIC-PCR fingerprinting-based community DNA hybridization to pinpoint genome-specific fragments as molecular markers to identify and track populations common to healthy human guts. *Journal of Microbiological Methods* 59(Goodhand et al.), 91-108. doi: https://doi.org/10.1016/j.mimet.2004.06.007.

Whitfield, C. (2006). "Biosynthesis and assembly of capsular polysaccharides in Escherichia coli", in: *Annual Review of Biochemistry.*).

Wildenberg, M.E., Koelink, P.J., Diederen, K., te Velde, A.A., Wolfkamp, S.C.S., Nuij, V.J., et al. (2017). The ATG16L1 risk allele associated with Crohn’s disease results in a Rac1-dependent defect in dendritic cell migration that is corrected by thiopurines. *Mucosal Immunology* 10(Polito et al.), 352-360. doi: 10.1038/mi.2016.65.

Willing, B., Halfvarson, J., Dicksved, J., Rosenquist, M., Jarnerot, G., Engstrand, L., et al. (2009). Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis* 15(5), 653-660. doi: 10.1002/ibd.20783.

Yang, H., McElree, C., Roth, M.P., Shanahan, F., Targan, S.R., and Rotter, J.I. (1993). Familial empirical risks for inflammatory bowel disease: differences between Jews and non-Jews. *Gut* 34(4)**,** 517-524. doi: 10.1136/gut.34.4.517.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy* 67(11)**,** 2640-2644. doi: 10.1093/jac/dks261.

Zhong, J., Xiao, C., Gu, W., Du, G., Sun, X., He, Q.-Y., et al. (2015). Transfer RNAs Mediate the Rapid Adaptation of Escherichia coli to Oxidative Stress. *PLOS Genetics* 11(6)**,** e1005302. doi: 10.1371/journal.pgen.1005302.

3.8 Supplemental Materials

Supplemental tables-

Supplemental Table 3.1. List of all *E. coli* strains used in this comparative genomics study (Total 179)

Strain name	Phylogroup	Disease Status	Source of the sample and phenotypes (AIEC/non-AIEC)	Sequence type
1-2-C20	B2	HC	Biopsy	ST-91
1-3-Ti10	D	HC	Biopsy	ST-349
10-1-R6	B2	HC	Biopsy	ST-131
11-1-Ti6	D	HC	Biopsy	ST-69
12-1-Ti12	B2	HC	Biopsy, AIEC	ST-127
12-2-Ti13	B2	HC	Biopsy, non-AIEC	ST-127
13-1-TC6	B2	CD	Biopsy	ST-420
13-3-TC6	B2	CD	Biopsy	ST-420
14-1-R13	D	HC	Biopsy	ST-3056
15-5-C3	B2	HC	Biopsy	ST-6273
16-2-Si18	A	HC	Biopsy	ST-216
16-3-C13	A	HC	Biopsy	ST-216
17-1-R20	F	CD	Biopsy	ST-59
17-2-Si14	F	CD	Biopsy	ST-59
18-1-DC10	B1	HC	Biopsy	ST-906
18-3-Ti5	B2	HC	non-AIEC	ST-73
19-1-TC4	B2	CD	Biopsy	ST-14
19-1-Ti7	B2	CD	Biopsy	ST-14
19-2-TC2	B2	CD	Biopsy	ST-14
19-2-Ti6	B2	CD	Biopsy	ST-14
20-2-TC13	B1	CD	Biopsy, AIEC	ST-1304
20-3-TC12	B1	CD	Biopsy, non-AIEC	ST-1304
20-5-R7	B2	CD	Biopsy	ST-73

21-1-TC7	B2	UC	Biopsy	ST-73
21-2-C8	B2	UC	Biopsy	ST-73
22-1-TiA17	D	CD	Biopsy	ST-59
23-1-TC4	B2	HC	Biopsy	ST-73
23-2-Ti8	B2	HC	Biopsy	ST-73
23-4-R13	B2	HC	Biopsy	ST-73
24-1-Si1	B1	HC	Biopsy	ST-58
24-4-R6	B2	HC	Biopsy	ST-28
25-1-ADC2	D	CD	Biopsy	ST-132
25-2-NDC7	D	CD	Biopsy	ST-362
26-1-C13	D	CD	Biopsy	ST-973
26-2-C12	F	CD	Biopsy	ST-1674
26-AU-glybx	F	CD	Biopsy, Aphthous ulcer	ST-5801
27-1-R1	B2	UC	Biopsy	ST-95
29-1-Si3	D	UC	Biopsy	ST-362
29-2-Si4	B2	UC	Biopsy	ST-589
29-3-Si10	B1	UC	Biopsy	ST-1049
29-4-Si8	D	UC	Biopsy	ST-362
3-1-AU1	A	HC	Biopsy	ST-1316
30-1-R8	F	UC	Biopsy	ST-648
32-1-Si18	B2	UC	Biopsy	ST-420
32-2-R12	B2	UC	Biopsy	ST-131
33-1-Ti5	B2	CD	Biopsy, AIEC	ST-131
33-3-R3	B1	CD	Biopsy	ST-1727
35-9-Si14	B2	CD	Biopsy	ST-95
35-MA-AUDC	B2	CD	Biopsy	ST-73
36-1-TC9	B2	CD	Biopsy	ST-429
36-1-Ti13	B2	CD	Biopsy, non-AIEC	ST-429
37-2-AC20	B1	HC	Biopsy	ST-6171
39-1-AC10	B2	CD	Biopsy	ST-95
39-2-Ti18	B2	CD	Biopsy, non-AIEC	ST-95
4-1-C3	D	CD	Biopsy	ST-69
40-1-R9	B2	CD	Biopsy	ST-95
40-3-R7	B2	CD	Biopsy	ST-95
41-1-Ti9	D	CD	Biopsy	ST-720
42-1-R2	A	CD	Biopsy	ST-48
43-1-R3	B1	CD	Biopsy	ST-2773
43-2-R17	B1	CD	Biopsy	ST-1642

44-1-Ti4	B1	CD	Biopsy	ST-58
44-4-Ti15	B1	CD	Biopsy	ST-58
45-1-Ti1	B2	CD	Biopsy	ST-95
45-2-HU10	A	CD	Biopsy	ST-398
45-3-DC2	B2	CD	Biopsy	ST-144
46-1-Ti2	B2	UC	Biopsy, non-AIEC	ST-95
46-2-Ti11	D	UC	Biopsy	ST-69
47-1-TC4	B2	CD	Biopsy	ST-110
47-3-R4	B1	CD	Biopsy	ST-169
47-4-R1	D	CD	Biopsy	ST-1567
48-1-R18	B2	CD	Biopsy	ST-6327
48-5-R19	B2	CD	Biopsy	ST-6327
5-1-TC12	B2	UC	Biopsy	ST-95
50-4-ACUA1	B1	CD	Biopsy	ST-6169
50-6-TCUA1	A	CD	Biopsy	ST-6174
50-7-TCUA2	B1	CD	Biopsy	ST-154
51-1-Ti1	B1	UC	Biopsy	ST-6333
51-2-R1	D	UC	Biopsy	ST-108
52-1-Ti3	B2	HC	Biopsy, AIEC	ST-95
52-2-Ti10	B2	HC	Biopsy, non-AIEC	ST-550
53-1-AC6	B1	CD	Biopsy	ST-58
53-1-AUC1	B1	CD	Biopsy	ST-58
54-1-TC4	B2	CD	Biopsy	ST-1919
54-1-Ti6	B2	CD	Biopsy, non-AIEC	ST-1919
55-1-AU4	B2	CD	Biopsy, non-AIEC	ST-131
55-1-TC9	B2	CD	Biopsy	ST-131
55-1-Ti19	B2	CD	Biopsy, non-AIEC	ST-131
56-1-TC1	D	HC	Biopsy	ST-6170
56-2-AC5	B2	HC	Biopsy	ST-6170
56-3-AC1	D	HC	Biopsy	ST-6170
56-4-R1	D	HC	Biopsy	ST-6170
57-2-DC4	A	CD	Biopsy	ST-409
57-3-Ti5	B2	CD	Biopsy, AIEC	ST-537
58-2-AC1	B2	UC	Biopsy	ST-95
58-3-Ti7	A	UC	Biopsy	ST-607
59-1-AC3	A	UC	Biopsy	ST-635
6-1-TC16	B2	UC	Biopsy	ST-95
60-1-Ti1	B2	UC	Biopsy, AIEC	ST-80

61-1-Ti1	B2	UC	Biopsy, AIEC	ST-95
62-1-Ti3	B2	UC	Biopsy	ST-12
62-2-Ti6	B2	UC	Biopsy, non-AIEC	ST-95
62-4-Si7	B1	UC	Biopsy	ST-223
63-1-Ti1	D	UC	Biopsy, non-AIEC	ST-963
63-2-TC7	B2	UC	Biopsy	ST-80
65-1-AC2	D	CD	Biopsy	ST-1177
65-2-AC1	A	CD	Biopsy	ST-10
65-3-TC1	D	CD	Biopsy	ST-1177
66-1-AC1	D	HC	Biopsy	ST-938
67-1-Ti13	D	CD	Biopsy	ST-362
67-3-Ti10	D	CD	Biopsy	ST-132
67-4-Ti7	A	CD	Biopsy	ST-216
68-2-Ti3	B1	UC	Biopsy	ST-517
68-3-TC19	D	UC	Biopsy	ST-362
69-1-AU1	B2	CD	Biopsy, Aphthous ulcer, non-AIEC	ST-569
69-1-TC1	B2	CD	Biopsy	ST-569
7-3-R4	D	CD	Biopsy	ST-349
70-1-AC3	D	CD	Biopsy	ST-69
70-2-Ti12	B2	CD	Biopsy, non-AIEC	ST-569
70-3-Ti8	D	CD	Biopsy	ST-362
70-4-DC20	B2	CD	Biopsy	ST-95
71-1-AC1	F	HC	Biopsy	ST-6334
71-2-Ti5	D	HC	Biopsy	ST-362
72-2-AC9	B2	UC	Biopsy	ST-131
72-4-Ti9	D	UC	Biopsy	ST-132
72-5-Ti12	D	UC	Biopsy	ST-362
72-6-Ti12	B2	UC	Biopsy, non-AIEC	ST-569
73-1-Si1	B1	UC	Biopsy	ST-2005
8-1-C8	D	CD	Biopsy	ST-3300
8-2-Ti16	A	CD	Biopsy	ST-10
8-3-DC15	B1	CD	Biopsy	ST-270
9-5-R1	B2	HC	Biopsy	ST-131
CD34LN	B2	CD	Bowel resected sample, LN, non-AIEC	ST-95
CD34NM	B2	CD	Bowel resected sample, NM	ST-95
CD57IM	B1	CD	Bowel resected sample, IM	ST-3759
CD57LN	B1	CD	Lymph node, Non-AIEC	ST-3759

CD57NM	B1	CD	Bowel resected sample, NM	ST-3759
CD59LN	A	CD	Bowel resected sample, LN, AIEC	ST-2606
CD59LN3	A	CD	Bowel resected sample, LN	ST-409
CD59NM	B1	CD	Bowel resected sampleNM	ST 3759
CD62LN	B2	CD	Bowel resected sample, LN, AIEC	ST-95
CD70IM	B1	CD	Bowel resected sample, IM	ST2541
CD70LN1	B1	CD	Bowel resected sample, LN, AIEC	ST2541
CD70LN2	B1	CD	Bowel resected sample, LN	ST2541
CD70LN3	B1	CD	Bowel resected sample, LN	ST2541
CD70NM	B1	CD	Bowel resected sample, NM	ST2541
CD77LN	Clade3	CD	Bowel resected sample, LN, non-AIEC	unknown
CD78LN20	C	CD	Bowel resected sample, LN	ST-23
CD78LN21	C	CD	Bowel resected sample, LN	ST-23
CD78LN5	B1	CD	Bowel resected sample, LN, non-AIEC	ST-767
CD78LN7	C	CD	Bowel resected sample, LN, non-AIEC	ST-23
CD78NM	C	CD	Bowel resected sample, NM	ST- 23
H223	B2	HC	AIEC	ST-141
H305	B2	HC	AIEC	ST-95
H504	B2	HC	AIEC	ST-95
LF82	B2	CD	AIEC	ST-131
232193	B2	CD	Stool sample from paediatric CD	73
233110	D	CD	Stool sample from paediatric CD	349
234122	B2	CD	Stool sample from paediatric CD	538
235134	B2	CD	Stool sample from paediatric CD	73
236146	B1	CD	Stool sample from paediatric CD	58
237158	B2	CD	Stool sample from paediatric CD	73
238170	B2	CD	Stool sample from paediatric CD	141
239182	B2	CD	Stool sample from paediatric CD	127
240194	D	CD	Stool sample from paediatric CD	449
241111	B2	CD	Stool sample from paediatric CD	95
242123	B2	CD	Stool sample from paediatric CD	95
243135	B1	CD	Stool sample from paediatric CD	529
244147	B1	CD	Stool sample from paediatric CD	529
245159	C	CD	Stool sample from paediatric CD	410
246171	B2	CD	Stool sample from paediatric CD	95
247183	B2	CD	Stool sample from paediatric CD	1677
248195	B2	CD	Stool sample from paediatric CD	428

249112	D	CD	Stool sample from paediatric CD	362
250124	B2	CD	Stool sample from paediatric CD	95
251136	B2	CD	Stool sample from paediatric CD	95
252148	B2	CD	Stool sample from paediatric CD	131
253160	B2	CD	Stool sample from paediatric CD	131
254172	B2	CD	Stool sample from paediatric CD	73

AIEC: Adherent invasive *E. coli*, non-AIEC: non-Adherent invasive *E. coli*, AU: Aphthous ulcer, LN: Lymph node, and NM: Normal mucosa, IM: Involved mucosa

Supplemental Table 3.2. Significantly associated genes in adult CD compared with healthy controls (HC)

Gene	**Annotation**	**No.in adult CD**	**No. in (HC)**	**Odds ratio**	**Naive p**
lpfA_1	Putative major fimbrial subunit LpfA	15	0	inf	0.004
lpfB	Putative fimbrial chaperone LpfB	15	0	inf	0.004
lpfC'	Putative outer membrane usher protein LpfC'	15	0	inf	0.004
group_6236	Hypothetical protein/ DUF2732 domain-containing protein	14	0	inf	0.008
group_4352	Hypothetical protein/ chromosome	14	0	inf	0.008
yfjQ_5	Hypothetical protein/ plasmid associated	13	0	inf	0.008
group_2455	Lysozyme RrrD/ putative membrane-associated lysozyme	12	0	inf	0.015
group_195	Outer membrane protein X/ Conserved hypothetical protein, putative transposase hypothetical protein	12	0	inf	0.015
group_4853	CRISPR-associated endoribonuclease Cas2	12	0	inf	0.015
yjbB	Hypothetical protein	65	24	inf	0.02
ecnA	Hypothetical protein	65	24	inf	0.023
group_5263	Hypothetical protein	11	0	inf	0.029
group_2041	Hypothetical protein	11	0	inf	0.029
group_1084	Hypothetical protein/ plasmid associated	10	0	inf	0.030
group_3838	Hypothetical protein	10	0	inf	0.03
ydjO_2	Hypothetical protein	10	0	inf	0.03
group_14972	Tyrosine recombinase XerD	10	0	inf	0.03
group_2394	Hypothetical protein	10	0	inf	0.03
group_15014	5-carboxymethyl-2-hydroxymuconate Delta-isomerase	10	0	inf	0.03
group_79	Protein RhsC	10	0	inf	0.03
group_10622	Hypothetical protein	10	0	inf	0.03

group_939	Putative protein YjgL	10	0	inf	0.03
group_8891	Hypothetical protein	10	0	11.025	0.03
group_2367	Hypothetical protein/ chromosome	63	20	10.7	0.002
pac	Penicillin G acylase	19	1	9.20	0.005
orgA	Hypothetical protein	17	1	9.20	0.01
yigG	Inner membrane protein YigG	17	1	9.20	0.01
group_5728	PTS system 2-O-alpha-mannosyl-D-glycerate-specific EIIABC component	17	1	8.48	0.01
yjgL_1	putative protein YjgL	16	1	8.48	0.01
group_26766	tRNA-Ala(tgc)	16	1	7.86	0.01
yjbE	Hypothetical protein /chromosome associated gene	59	15	7.8	0.0002
group_9158	Hypothetical protein	15	1	7.8	0.032
insC-1_1	Hypothetical protein	15	1	7.8	0.032
group_2979	Outer membrane usher protein HtrE	15	1	7.8	0.032
group_3951	Hypothetical TraR	15	1	7.31	0.032
ybfD	Hypothetical protein	24	2	7.15	0.004
mdaB	Modulator of drug activity B	63	22	6.84	0.021
ccdB	Toxin CcdB	23	2	6.39	0.005
traJ	Protein TraJ	22	2	5.90	0.008
hokA_3	Protein HokA	62	21	5.90	0.017
leuC	3-isopropylmalate dehydratase subunit	62	21	5.55	0.017
group_13502	Hypothetical protein	20	2	5.55	0.016
srp54	Signal recognition particle 54 kDa protein	20	2	5.55	0.016
group_13501	Hypothetical protein	20	2	5.55	0.016
yqeL	Putative protein YqeL	20	2	5.25	0.016
yhaB	Hypothetical protein	39	6	5.24	0.001
group_8719	Hypothetical protein/ ATPase/	31	4	5.16	0.004
stbB	Protein StbB	19	2	5.16	0.02
parM	Plasmid segregation protein ParM	19	2	5.16	0.02
group_26434	tRNA-Asp(gtc)/ chromosome	19	2	4.78	0.02
group_10602	hypothetical protein	18	2	4.78	0.04
spaS	Surface presentation of antigens protein SpaS	18	2	4.78	0.04
hilA	Transcriptional regulator HilA	18	2	4.69	0.04
selB	Selenocysteine-specific elongation factor	62	22	4.69	0.04
ygjK	Glucosidase YgjK	62	22	4.63	0.04
group_4922	Hypothetical protein	29	4	4.38	0.008
group_8465	Hypothetical protein/ chromosome	23	3	4.09	0.02
group_10737	Hypothetical protein/ chromosome	22	3	3.81	0.03
group_3938	Hypothetical protein/ plasmid associated	21	3	3.81	0.03

group_1328	Outer membrane usher protein FimD	21	3	3.81	0.03
ybgD_1	Putative fimbrial-like protein YbgD	21	3	3.81	0.03
group_5172	Aliphatic sulfonates import ATP-binding protein SsuB	21	3	3.60	0.03
group_1803	Putative autotransporter YfaL	33	6	3.59	0.01
group_13908	HTH-type transcriptional regulator PgrR	25	4	3.59	0.02
kefF_2	Glutathione-regulated potassium-efflux system ancillary protein KefF	25	4	3.59	0.02
group_2923	Polyol:NADP oxidoreductase	25	4	3.56	0.02
flp	Protein flp	39	8	3.36	0.01
yjgL	Putative protein YjgL	24	4	3.36	0.04
group_3717	Methylmalonyl-CoA mutase	24	4	3.33	0.04
group_2724	Sensor protein TorS	35	7	3.32	0.02
melB	Melibiose carrier protein	28	5	3.19	0.03
group_4921	Hypothetical protein	31	6	3	0.03
ldrA_1	Small toxic polypeptide LdrA/ small toxic polypeptide	30	6	2.94	0.03
fimA_2	Fimbrial subunit type 1	36	8	2.94	0.03
yidI	Inner membrane protein YidI	36	8	2.94	0.03
yccE	Hypothetical protein	33	7	2.90	0.03
siaM	Sialic acid TRAP transporter large permease protein SiaM	41	10	2.81	0.03
bshB1	N-acetyl-alpha-D-glucosaminyl L-malate deacetylase 1	38	9	2.81	0.03
yciG	Hypothetical protein	47	13	2.77	0.03
ybiU	Putative protein YbiU	35	8	2.65	0.04
yjiJ	Hypothetical protein	50	15	2.65	0.04
yjiE	HTH-type transcriptional regulator YjiE	50	15	2.65	0.04
dlsT	Putative serine transporter	42	11	2.65	0.04
group_10267	HTH-type transcriptional regulator GadX	42	11	2.65	0.04
malS	Periplasmic alpha-amylase	42	11	2.65	0.04
ycaM	Inner membrane transporter YcaM	42	11	2.65	0.04
yeaX	Carnitine monooxygenase reductase subunit	42	11	2.65	0.04
group_6030	Hypothetical protein	42	11	2.65	0.04
ompD	Outer membrane porin protein OmpD/ Outer membrane porin protein from phage origin/ porin activity and ion transport	42	11	0	0.04

Supplemental Table 3.3. Significantly associated genes in adult CD and paediatric CD

Gene	**Annotation**	**No in adult CD**	**No in paediatric CD**	**Naive p**	**Benjamini H p**

group_10035	Hypothetical protein/ rRNA-16S ribosomal RNA	38	0	0.0000002	0.003
hyfG	Hydrogenase-4 component G	38	4	0.0007	0.25
tabA	Toxin-antitoxin biofilm protein TabA	44	7	0.0029	0.29
nupX	Putative nucleoside permease NupX	44	8	0.0074	0.34
bdcA	Cyclic-di-GMP-binding biofilm dispersal mediator protein	43	7	0.0037	0.29
yjgN	Inner membrane protein YjgN	43	7	0.0037	0.29
puuB	Gamma-glutamylputrescine oxidoreductase	43	7	0.0037	0.29
pspG	Phage shock protein G	43	7	0.0037	0.29
tisB	Small toxic protein TisB	43	7	0.0037	0.29
arsB	Arsenical pump membrane protein	43	7	0.0037	0.29
lacZ	Beta-galactosidase	43	8	0.0134	0.47
cbpA	Curved DNA-binding protein	43	8	0.0134	0.47
katE	Catalase HPII	42	7	0.0068	0.32
ompD	Outer membrane porin protein OmpD	42	7	0.0068	0.32
scpB	Methylmalonyl-CoA decarboxylase	42	7	0.0068	0.32
hyfR	DNA-binding transcriptional activator HyfR	42	7	0.0068	0.32
flhA	Flagellar biosynthesis protein FlhA	41	4	0.0002	0.16
mdtO	Multidrug resistance protein MdtO	41	6	0.0032	0.29
sfmA	putative fimbrial-like protein SfmA	41	7	0.0083	0.37
hicB	Antitoxin HicB	39	6	0.0072	0.34
yfcS_2	Putative fimbrial chaperone YfcS	13	18	0.000001	0.01
group_26390	tRNA-Asn(gtt)	27	22	0.00002	0.02
tufA	Elongation factor Tu 1	15	18	0.00002	0.02
gadB	Glutamate decarboxylase beta	30	22	0.00001	0.03
gspA	General stress protein A	20	19	0.00001	0.03
xerD	Tyrosine recombinase XerD	27	18	0.0033	0.29
ada	Bifunctional transcriptional activator/DNA repair enzyme Ada	38	23	0.0001	0.07
tsaR	HTH-type transcriptional regulator TsaR	29	21	0.0001	0.07
klcA	Antirestriction protein KlcA	14	16	0.0001	0.07
vgrG1	Actin cross-linking toxin VgrG1	13	15	0.0002	0.14
torZ	Trimethylamine-N-oxide reductase 2	24	19	0.0002	0.16
xerC	Tyrosine recombinase XerC	41	23	0.0002	0.18
dgcT	putative diguanylate cyclase DgcT	33	21	0.0004	0.23
cbtA	Cytoskeleton-binding toxin CbtA	25	19	0.0005	0.23

hipB	Antitoxin HipB	13	14	0.0005	0.23
nadR	Trifunctional NAD	13	14	0.0005	0.23
icd	Isocitrate dehydrogenase [NADP]	23	18	0.0005	0.23
tibA	putative autotransporter	26	19	0.0005	0.23
yejO	putative autotransporter	18	16	0.0009	0.25
csgD	CsgBAC operon transcriptional regulatory protein	45	23	0.0012	0.25
ybgD	putative fimbrial-like protein YbgD	26	18	0.0031	0.29
papC	Outer membrane usher protein PapC	26	18	0.0031	0.29
vgrG1_1	Actin cross-linking toxin VgrG1	14	13	0.0033	0.29
xerD	Tyrosine recombinase XerD	27	18	0.0033	0.29
dctD	C4-dicarboxylate transport transcriptional regulatory protein DctD	22	16	0.0037	0.29
pspG	Phage shock protein G	22	16	0.0037	0.29
kgtP	Alpha-ketoglutarate permease	25	17	0.0040	0.29
gspE	Putative type II secretion system protein E	28	18	0.0068	0.32
yfcV	Putative fimbrial-like protein YfcV	24	16	0.0083	0.37

Supplemental Table 3.4. Significantly associated genes in paediatric CD compared with healthy controls (HC)

Gene	**Annotation**	**No in paediatric-CD**	**No in HC**	**Naive p**
group_26390	tRNA-Asn(gtt)/ tRNA-Asn	22	7	0.0000004
group_26377	tRNA-Lys(ttt)	23	20	0.011
group_26641	tRNA-Leu(Schulz et al.)	22	19	0.028
yfjQ_5	Hypothetical protein/ DUF945 domain-containing protein/ plasmid associated gene	14	2	0.00006
finO	Fertility inhibition protein/ plasmid associated gene/ conjugal transfer fertility inhibition protein FinO	16	5	0.0005
group_1108	Hypothetical protein/ replication protein/ plasmid associated	16	5	0.0005
traI	Multifunctional conjugation protein TraI/ plasmid associated gene	17	6	0.0005
group_12474	Hypothetical protein/ conjugal transfer pilus assembly protein TraE/plasmid associated	18	8	0.0008
traJ	Protein TraJ/ plasmid associated	10	2	0.006
srp54	Signal recognition particle 54 kDa protein	10	2	0.006

group_13509	Hypothetical protein/ plasmid associated	10	2	0.006
stbB	Protein StbB/ plasmid_associated	10	2	0.006
traM	Relaxosome protein TraM/ plasmid associated	9	2	0.014
traV	Protein TraV/ plasmid associated	9	2	0.014
ccdB	Toxin CcdB/ plasmid associated	12	5	0.017
ccdA_2	Antitoxin CcdA/ plasmid associated	12	5	0.017
group_5024	General stress protein A	19	10	0.002
ydhV	Putative oxidoreductase YdhV/metabolism	23	18	0.002
gadB	Glutamate decarboxylase beta	22	15	0.001
group_4555	Phosphotriesterase homology protein	11	2	0.003
icd_2	Isocitrate dehydrogenase [NADP]	18	10	0.005
pdeG	Putative cyclic di-GMP phosphodiesterase PdeG	23	20	0.011
gadB_1	Glutamate decarboxylase beta	13	6	0.020
pgtB	Phosphoglycerate transport system sensor protein PgtB	9	2	0.014
pgtC	Phosphoglycerate transport regulatory protein PgtC	9	2	0.014
hapE	Glutamate synthase [NADPH] small chain	7	1	0.017
ntdC	Glucose-6-phosphate 3-dehydrogenase	7	1	0.017
csgD	CsgBAC operon transcriptional regulatory protein	23	18	0.002
tufA	Elongation factor Tu 1	18	7	0.001
flhC	Flagellar transcriptional regulator FlhC	23	20	0.011
yehA	Putative fimbrial-like protein YehA	22	17	0.006
yfcS_2	Putative fimbrial chaperone YfcS	18	12	0.021
papG	Fimbrial adhesin PapG	7	1	0.017
vgrG1_3	Actin cross-linking toxin VgrG1	15	7	0.010
group_2576	Cytoskeleton-binding toxin CbtA	7	1	0.017
group_268	Cytoskeleton bundling-enhancing antitoxin CbeA	7	1	0.017
hns_2	DNA-binding protein H-NS	13	6	0.020

Supplemental Table 3.5. List of genes significantly associated with *E. coli* in LN-CD compared with healthy controls (HC)

Gene	**Annotation**	**No. in LN**	**No. in HC**	**Naive p**
group_7279	Hypothetical protein	7	4	0.001
group_3641	Outer membrane usher protein HtrE/ fimbrial assembly protein similar to papC	5	1	0.0017
group_26114	tRNA-Asn(gtt)	4	0	0.0021
group_3863	K( )/H( ) antiporter NhaP2	7	5	0.0024
group_7102	Hypothetical protein	7	5	0.0024

group_7785	Catalase-peroxidase	7	6	0.0049
intS_2	Prophage integrase IntS	7	6	0.0049
yqeL	Putative protein YqeL	5	2	0.0055
group_8543	hypothetical protein/ATPase	6	4	0.0062
group_13809	HTH-type transcriptional regulator PgrR	6	4	0.0062
kefF_2	Glutathione-regulated potassium-efflux system ancillary protein KefF	6	4	0.0062
group_2823	Polyol:NADP oxidoreductase	6	4	0.0062
group_1161	Putative fimbrial-like protein YadM	6	4	0.0062
sopA	E3 ubiquitin-protein ligase SopA /very very important fot bacterial survival	6	4	0.0062
group_14340	Putative fimbrial chaperone YadV	6	4	0.0062
yadN	Putative fimbrial-like protein YadN	6	4	0.0062
frlA	Putative fructoselysine/psicoselysine transporter FrlA	6	4	0.0062
group_6116	Hypothetical protein	6	4	0.0062
pstS1	Phosphate-binding protein PstS 1	6	4	0.0062
pac	Penicillin G acylase	4	1	0.0093
group_9087	PTS system mannitol-specific EIICB component	4	1	0.0093
group_5181	CRISPR-associated endoribonuclease Cas2	3	0	0.0117
group_1055	Hypothetical protein	3	0	0.0117
group_6557	PTS system fructose-specific EIIB'BC component	3	0	0.0117
group_10525	Putative inhibitor of glucose uptake transporter SgrT	6	5	0.0123
setA	Sugar efflux transporter A	6	5	0.0123
group_21862	Putative zinc-binding alcohol dehydrogenase	6	5	0.0123
group_3771	Serine/threonine-protein phosphatase 2	6	5	0.0123
yraK	Putative fimbrial-like protein YraK	6	5	0.0123
ybgD_1	Putative fimbrial-like protein YbgD	5	3	0.013
group_7386	Outer membrane usher protein PapC	6	5	0.012
elfG_2	putative fimbrial-like protein ElfG	5	3	0.013
group_2559	Prophage integrase IntA/ AlpA family phage regulatory protein	3	0	0.011
frlR	Putative fructoselysine utilization operon transcriptional repressor	6	5	0.012
frlC	Fructoselysine 3-epimerase	6	5	0.012
pezT	Toxin PezT	4	2	0.02
vapB	Antitoxin VapB	5	4	0.02
casE	CRISPR system Cascade subunit CasE	6	6	0.03
ygbT	CRISPR-associated endonuclease Cas1	6	6	0.03
casA	CRISPR system Cascade subunit CasA	6	6	0.03
casC	CRISPR system Cascade subunit CasC	6	6	0.03

hilA	Transcriptional regulator HilA	4	2	0.02
thiS	Sulfur carrier protein ThiS	9	15	0.01
rpnA_4	Recombination-promoting nuclease RpnA	8	12	0.02

Supplemental Table 3.6. List of genes significantly associated with *E. coli* isolates in AU-CD compared with HC

Gene	Annotation	Number in AU	Number in HC	Naive p
ydaT	Hypothetical protein	6	2	0.0001
group_10646	DNA-binding and transcription factor activity	5	2	0.001
prtR	HTH-type transcriptional regulator PrtR	5	2	0.001
group_12496	Killing protein KilR	5	2	0.001
group_15336	Hypothetical protein	3	0	0.005
group_15450	DNA-binding and transcription factor activity	3	0	0.005
group_16142	Relaxosome protein TraM	3	0	0.005
group_7578	tRNA-Arg(acg)	3	0	0.005
group_26390	tRNA-Asn(gtt)	6	7	0.007
group_1711	Hypothetical protein	5	4	0.007
traJ	Protein TraJ	4	2	0.009
group_5929	Hypothetical protein	4	2	0.009
yjgL_1	Putative protein YjgL	3	1	0.02
betT_2	Glycine betaine transporter	3	1	0.02
aspA_1	Aspartate ammonia-lyase	3	1	0.006
traI	Multifunctional conjugation protein TraI	5	6	0.006
yhcR	Endonuclease YhcR	5	6	0.006
yeeJ_2	Hypothetical protein	5	6	0.006
traA	Pilin	5	6	0.006
group_26502	tRNA-Ala(tgc)	6	9	0.009
group_6366	Protein StbB	3	1	0.02
group_1890	DNA adenine methyltransferase YhdJ	3	1	0.021
group_997	Replication protien	3	1	0.02

Supplemental Table 3.7. List of genes significantly associated with AIEC compared to non-AIEC

Gene	Annotation	Number in AIEC (15)	Number in Non-AIEC (19)	Naive p
group_5695	Type-1 fimbrial protein, A chain	6	0	0.003
era_3	GTPase Era	6	0	0.003

group_15152	Prophage integrase IntA	6	0	0.003
group_5337	Hypothetical protein	6	0	0.003
group_10960	Hypothetical protein	6	0	0.003
group_10974	Hypothetical protein	6	0	0.003
group_10975	Hypothetical protein	6	0	0.003
group_15175	Hypothetical protein	6	0	0.003
esiB	Secretory immunoglobulin A-binding protein EsiB	6	0	0.003
fimD_3	Outer membrane usher protein FimD	6	0	0.003
group_15225	Chaperone protein FocC	6	0	0.003
group_15228	P fimbrial regulatory protein KS71A	6	0	0.003
group_15229	Hypothetical protein	6	0	0.003
tnpA_1	Transposase for transposon Tn5	6	0	0.003
group_7036	Putative metal chaperone YciC	6	0	0.003
ydiV_2	Anti-FlhC(Polito et al.)FlhD(4) factor YdiV	6	0	0.003
sfaG	S-fimbrial protein subunit SfaG	6	0	0.003
lpfA_4	putative major fimbrial subunit LpfA	6	0	0.003
papB_2	Major pilus subunit operon regulatory protein PapB	6	0	0.003
group_6219	Hypothetical protein	6	0	0.003
group_13316	Major exported protein	12	6	0.007
group_3423	Hypothetical protein	8	2	0.009
group_1761	Elongation factor Tu 2	5	0	0.01
rsmI_2	Ribosomal RNA small subunit methyltransferase I	5	0	0.01
group_9129	S-fimbrial protein subunit SfaH	5	0	0.01
group_3122	Outer membrane protein	7	1	0.01
group_10460	Hypothetical protein	7	1	0.01
group_3422	Putative nuclease YbcO	9	3	0.01

Supplemental Table 3.8. COGS associated with adult CD compared to paediatric CD based on fixed effect model in Pyseer

Gene variant	**AF**	**Annotation/function**	**No in adult CD (65)**	**No in paediatric CD (23)**	**lrt p-value**
group_4994	0.51	Flagellar biosynthesis protein FlhA	41	4	0.0001
yjiC	0.58	Hypothetical protein	45	6	0.0003
rpnA_2	0.83	Recombination-promoting nuclease RpnA	59	14	0.001
group_2748	0.44	Sensor protein TorS	35	4	0.002

group_6839	0.30	Putative acyltransferase YihG	25	2	0.007
group_8511	0.37	Universal stress protein UspG	27	6	0.0003
group_6207	0.22	DUF2169 domain-containing protein peptidase C39	15	5	0.0004
group_1597	0.21	IS110 family transposase	14	5	0.0012
group_249	0.21	S66 family transposase	16	3	0.001
tonB_2	0.46	TonB family protein	29	12	0.003
vgrG1_2	0.23	Type VI secretion system tip protein VgrG	27	6	0.003
group_8595	0.23	DUF2169 domain-containing protein	15	5	0.003
group_13545	0.23	Hypothetical protein	14	5	0.003
mtnA	0.27	S-methyl-5-thioribose-1-phosphate isomerase	17	7	0.003

Supplemental Table 3.9. COGS associated with adult CD compared to paediatric CD based on mixed effect model in Pyseer

Variant	**Annotation**	**AF**	**Adult CD (65)**	**Paediatric CD (23)**	**lrt p-value**
group_4994	Flagellar biosynthesis protein FlhA	0.51	41	4	0.001
yjiC	Hypothetical protein	0.58	45	6	0.002
rpnA_2	Recombination-promoting nuclease RpnA	0.83	59	14	0.007
group_8511	Universal stress protein UspG	0.37	27	6	0.009
group_10291	Hypothetical protein	0.52	39	7	0.009
group_2748	Sensor protein TorS	0.44	35	4	0.010
hyfG	Hydrogenase-4 component G	0.47	38	4	0.011
feoC	DNA-binding transcriptional regulator	0.59			0.011

Supplemental Table 3.10. COGS significantly associated with paediatric CD compared to adult CD based on fixed effect model

Variant	**Annotation**	**AF**	**No in adult CD (65)**	**No in paediatric CD (23)**	**lrt p-value**
tufA	Elongation factor Tu 1	0.375	15	18	0.0000001
group_7835	Hypothetical protein	0.273	9	15	0.0000002
group_2557	Hypothetical protein	0.239	7	14	0.0000002
group_2391	Hypothetical protein	0.295	11	15	0.000005
yfcS_2	Putative fimbrial chaperone YfcS	0.352	13	18	0.000008

group_5024/ gspA_1	General stress protein A	0.443	20	19	0.000018
group_1543	Uncharacterised protein	0.239	7	14	0.00005
yddA	Inner membrane ABC transporter ATP-binding protein YddA	0.636	37	19	0.0001
group_1299	Type III effector	0.148			0.0002
group_857	Hypothetical protein	0.386	16	18	0.0005
group_6211	Hypothetical protein	0.239	9	12	0.0008
group_10364	Protein FlmC/ type I toxin-antitoxin system Hok family toxin	0.159	5	9	0.0008
group_4547	Hypothetical protein	0.432	19	19	0.0009

Supplemental Table 3.11. COGS significantly associated with paediatric CD compared to adult CD based on mixed effect

Variant	**Annotation**	**AF**	**Adult CD (65)**	**Paediatric CD (23)**	**lrt p-value**
group_7835	Hypothetical protein	0.273	9	15	0.0000001
group_2557	AlpA family transcriptional regulator	0.239	7	14	0.0000002
yfcS_2	Putative fimbrial chaperone YfcS	0.352	13	18	0.0000008
group_1543	Uncharacterised protein	0.239	7	14	0.0000023
group_2391	Hypothetical protein	0.295	11	15	0.000004
tufA	Elongation factor Tu 1	0.375	15	18	0.000004
group_857	Hypothetical protein	0.386	16	18	0.00001
group_5024/ gspA_1	General stress protein A	0.443	20	19	0.00003
group_4547	Hypothetical protein	0.432	19	19	0.00003
group_1650	Hypothetical protein	0.398	17	18	0.00004
klcA_1	Antirestriction protein KlcA	0.341	14	16	0.00007
group_3748	Relaxosome protein TraM	0.102	2	7	0.00015
group_4555	Phosphotriesterase homology protein php_2	0.205	7	11	0.00016
group_8847	Protein TraJ	0.114	3	7	0.00016
group_10364	Protein FlmC/ type I toxin-antitoxin system Hok family toxin	0.159	5	9	0.0002
ssb_2	Single-stranded DNA-binding protein/plasmid	0.193	7	10	0.0003
yeeW_2	Phage head morphogenesis protein	0.261	10	13	0.0004

yeeW_1	Type IV toxin-antitoxin system toxin CbtA	0.511	25	20	0.0004

Supplemental Table 3.12. CD vs healthy control (HC): COGs associated with CD compared with healthy control (HC) based on fixed effect model

Variant	Annotation	AF	No in CD	No in HC	lrt p-value
yjbE	Hypothetical protein	0.774	81	15	0.00001
traI	Conjugation and Pillus assembly protein	0.443	48	6	0.00001
group_2769	Lytic transglycosylase	0.365	40	3	0.00006
finO	Fertility inhibition protein	0.417	45	5	0.0001
group_1624	Phage integrase family protein	0.287	32	2	0.0001
group_1711	Type-F conjugative transfer system pilin acetylase TraX	0.365	40	4	0.0001
focC	Fimbrial chaperone protein FimC	0.826	82	21	0.0001
group_12474	IncF plasmid conjugative transfer pilus assembly protein TraE	0.452	47	8	0.0001
group_4184	Hypothetical protein	0.409	43	6	0.0001
group_26390	tRNA-Asn	0.461	49	7	0.0002
selB	Selenocysteinyl-tRNA-specific translation elongation factor SelB	0.87	85	22	0.0003
yehM	Hypothetical protein	0.765	78	17	0.0003
group_2422	Hypothetical protein	0.835	84	20	0.0004
traJ	PAS domain-containing protein	0.287	32	2	0.0004
dsdX	D-serine transporter DsdX	0.826	82	21	0.0004
yhaB	Hypothetical protein	0.417	46	6	0.0004
group_3260	Regulator with HipB	0.365			0.0005
hokA_3	Type I toxin-antitoxin system toxin HokA	0.861	85	21	0.0005
srp54	AAA family ATPase	0.27	30	2	0.0006
group_13509	Hypothetical protein	0.27	30	2	0.0006
ccdB	Type II toxin-antitoxin system toxin CcdB	0.357	38	5	0.0007
group_1108	Replication protein	0.357	38	5	0.0008
yjhQ	N-acetyltransferase	0.27			0.0008
yeiS	DUF2542 family protein	0.896	87	24	0.0009
fimA_2	Major type 1 subunit fimbrin	0.461	47	8	0.004

Supplemental Table 3.13. COGS associated with CD compared to healthy controls (HC) based on mixed effect model

Variants	Annotation	AF	CD (88)	HC (27)	lrt p-value

yjbE	Hypothetical protein	0.826	81	15	0.0000004
group_2422	Hypothetical protein	0.896	84	20	0.0001
hokA_3	Type I toxin-antitoxin system toxin HokA	0.913	85	21	0.0002
yehM	Hypothetical protein	0.817	78	17	0.0004
yfjQ_5	DUF945 domain-containing protein	0.339			0.0007
group_2769	Lytic transglycosylase	0.374	40	3	0.001
selB	Selenocysteinyl-tRNA-specific translation elongation factor SelB	0.922	85	22	0.001
finO	Fertility inhibition protein	0.435	45	5	0.002
traI	Conjugation and Pillus assembly	0.47	48	6	0.002
traJ	PAS domain-containing protein	0.296	32	2	0.003
group_1624	phage integrase family protein/ Tyrosine recombinase XerD	0.296	32	2	0.003
group_1711	Type-F conjugative transfer system pilin acetylase TraX	0.383	40	4	0.003
ygjK	Alpha-glucosidase	0.913			0.004
yhaB	Hypothetical protein	0.452	46	6	0.005
focC	Phage head morphogenesis protein/ fimbrial chaperone protein FimC	0.887	82	21	0.005
dsdX	D-serine transporter DsdX/ recombinase RecT	0.887	82	21	0.005
srp54	AAA family ATPase/ Signal recognition particle 54 kDa protein	0.278	30	2	0.006

Supplemental Table 3.14. Significantly associated COGs in adult CD isolates compared with HC based on fixed effect model

Variant	**Annotation**	**AF**	**No in adult CD**	**No in HC**	**lrt p-value**
group_1624	Phage integrase family protein	0.25	21	2	0.00006
group_2769	Lytic transglycosylase	0.304	24	3	0.00007
traJ	Transcriptional activator	0.261	22	2	0.0001
yfjQ_5/ yubP	DUF945 domain-containing protein	0.272	23	2	0.0001
srp54	Signal recognition particle 54 kDa protein	0.239	20	2	0.0001
group_13509	CopG family transcriptional regulator	0.239	20	2	0.0001
yjbE	Hypothetical protein	0.793	59	14	0.0001
group_13510	Uncharacterised protein	0.239	20	2	0.0003
stbB	Protein StbB	0.228	19	2	0.0003
yhaB	Hypothetical protein	0.489	39	6	0.0003

selB	Phage head morphogenesis protein L-seryl-tRNA (Sec) selenium transferase	0.902	63	20	0.0004
ccdB	Type II toxin-antitoxin system antitoxin CcdB	0.337	26	4	0.0005
group_1711	Type-F conjugative transfer system pilin acetylase TraX	0.337	27	4	0.0006
group_10588	Hypothetical protein	0.217	18	2	0.0006
umuC_3	Translation error-prone DNA polymerase V subunit UmuC	0.217	18	2	0.0006
yehM	Hypothetical protein	0.804	58	16	0.0006
ykfF_2	DUF905 domain-containing protein	0.207	17	2	0.0007
focC	Phage head morphogenesis protein	0.859	60	19	0.0007
group_2422	VWA domain-containing protein	0.891	63	19	0.0008
finO	Fertility inhibition protein	0.37	29	5	0.0009
group_3941	DUF2726 domain-containing protein	0.261	21	3	0.001
traI	Multifunctional conjugation protein	0.402	31	6	0.001
ccdA_2	Type II toxin-antitoxin system antitoxin CcdA	0.326	25	5	0.001

Supplemental Table 3.15. Significantly associated COGs in adult CD isolates based on mixed effect model

Variant	Annotation	AF	No in adult CD	No in HC	lrt p-value
yjbE	Hypothetical protein	0.793	59	14	0.00001
group_2422	VWA domain-containing protein	0.891	63	19	0.0001
yehM	Hypothetical protein	0.804	58	16	0.0007
yhaB	Hypothetical protein	0.489	39	6	0.001
yfjQ_5	DUF945 domain-containing protein	0.272	23	2	0.001
group_2769	Lytic transglycosylase	0.304	24	3	0.001
traJ	Transcriptional activator	0.261	22	2	0.001
group_1624	Uncharacterised protein	0.25	21	2	0.002
hokA_3	Type I toxin-antitoxin system toxin HokA	0.891	62	20	0.002
srp54	Signal recognition particle 54 kDa protein	0.239	20	2	0.003
group_13509	CopG family transcriptional regulator	0.239	20	2	0.003
group_13510	Uncharacterised protein	0.239	20	2	0.004
selB	Phage head morphogenesis protein	0.902	63	20	0.004

group_1711	Type-F conjugative transfer system pilin acetylase TraX	0.337	27	4	0.004
stbB	Protein StbB/ plasmid stability protein	0.228	19	2	0.005
finO	Fertility inhibition protein	0.37	29	5	0.006
group_8701	DUF3343 domain-containing protein	0.38	31	4	0.006
ygjK	Alpha-glucosidase	0.902	62	21	0.007
umuC_3	Translation error-prone DNA polymerase V subunit UmuC	0.217	18	2	0.007
group_10588	Hypothetical protein	0.217	18	2	0.008
group_1823	AIDA-I family autotransporter YfaL	0.413	33	5	0.008
group_3941	DUF2726 domain-containing protein	0.261	21	3	0.008
traI	Multifunctional conjugation protein	0.402	31	6	0.01

Supplemental Table 3.16. Significantly associated COGs in paediatric CD isolates based on fixed effect model

Variant	**Annotation**	**AF**	**No in Paediatric-CD (23)**	**No in HC (27)**	**lrt p-value**
tufA	Elongation factor Tu 1	0.5	18	7	0.00001
group_7835	Hypothetical protein	0.36	15	3	0.00005
yfjQ_5	DUF945 domain-containing protein (plasmid)	0.32	14	2	0.0003
group_1543	Uncharacterised protein	0.38	14	5	0.0004
group_3296	Type VI secretion system PAAR protein	0.46	15	8	0.0006
group_2769	Lytic transglycosylase (plasmid)	0.36	15	3	0.0007
dgcT	GGDEF domain-containing protein	0.66	21	12	0.0007
group_1108	replication protein (plasmid)	0.42	11	1	0.0009
group_14671	FliC/FljB family flagellin	0.16	6	2	0.001
group_3260	Regulator with HipB	0.48	16	8	0.001
group_14755	Peptidase M35	0.18	7	2	0.001
group_10994	Hypothetical protein	0.18	7	2	0.001
group_4184	Hypothetical protein	0.48	16	5	0.001
group_1624	Phage integrase family protein	0.26	11	2	0.001
finO	Fertility inhibition protein FinO	0.42	16	5	0.001

Supplemental Table 3.17. Significantly associated COGs in Paediatric- CD isolates based on mixed effect model

Variant	Annotation	AF	No in paediatric-CD (23)	No in HC (27)	lrt p-value
yfjQ_5	DUF945 domain-containing protein	0.32	14	2	0.00001
group_2769	lytic transglycosylase (plasmid)	0.36	15	3	0.00002
group_7835	hypothetical protein	0.36	15	3	0.00002
group_4184	hypothetical protein (plasmid)	0.48	16	5	0.00002
tufA	Elongation factor Tu 1	0.5	18	7	0.00010
traI	IncF plasmid conjugative transfer protein TraI (plasmid)	0.46	17	6	0.00012
finO	fertility inhibition protein FinO (plasmid)	0.42	16	5	0.00012
group_1108	replication protein (plasmid)	0.42	11	1	0.00012
dgcT	GGDEF domain-containing protein	0.66	21	12	0.00027
tsaR	transcriptional regulator TdcA	0.66	21	12	0.00027
group_12474	similar to TraE protein of plasmid F	0.52	18	8	0.00035
group_5024	General stress protein A	0.58	19	10	0.00077
group_1624	phage integrase family protein	0.26	11	2	0.00079

Supplemental Table 3.18. Significantly associated COGs in UC isolates compared to HC based on fixed effect model

Variant	Annotation	AF	lrt p-value	no in UC (31)	no in HC (27)
yjbE	hypothetical protein	0.724	0.0005	28	15
yhaB	hypothetical protein	0.345	0.0011	14	6
higA	helix-turn-helix domain-containing protein	0.776	0.00118	27	18
group_1823	AIDA-I family autotransporter YfaL	0.328	0.00118	14	5
rhmD	L-rhamnonate dehydratase	0.776	0.00158	25	20
yehA	putative fimbrial-like adhesin protein	0.759	0.00231	27	17
ccdB	Toxin CcdB	0.345	0.00259	15	5
ccdA_2	Antitoxin CcdA	0.345	0.00259	15	5

Supplemental Table 3.19. Significantly associated COGs in UC isolates compared to HC based on mixed effect model (this has more similarity to the Roary analysis results)

Variant	Annotation	AF	No in UC (31)	No in HC (27)	lrt p-value
yjbE	hypothetical protein	0.724	28	15	0.0007
yfjQ_5	Plasmid associated DUF945 domain-containing protein	0.345	12	2	0.004
group_2769	plasmid associated lytic transglucosylase	0.776	13	3	0.008
finO	Fertility inhibition protein	0.328	16	5	0.008
fliD	Flagellar hook-associated protein 2	0.776	25	14	0.008
traJ	Protein TraJ	0.759	11	2	0.009
group_5332	fimbrial assembly protein	0.345			0.009
ccdB	Toxin CcdB	0.345	15	5	0.016
ccdA_2	Antitoxin CcdA	0.724	15	5	0.016

Supplemental Table 3.20. Significantly associated COGs in lymph node CD isolates compared with HC based on fixed effect model

Variant	Annotation	AF	LN (9)	HC (27)	p-value
group_7945	Catalase-peroxidase HPI, katG	3.61E-01	7	6	0.002
group_6734	Hypothetical protein	3.89E-01	7	7	0.005

Supplemental Table 3.21. Significantly associated COGs in lymph node (LN) CD isolates compared with HC based on mixed effect model

Variant	Annotation	AF	LN (9)	HC (27)	p-value
group_26282	tRNA-Phe (gaa)	3.33E-01	6	6	0.014
group_3725	Fructose-bisphosphatase class II, glpX/yggf	3.89E-01	7	8	0.005
group_7945	Catalase-peroxidase HPI, katG	3.61E-01	7	6	0.002

Supplemental Table 3.22. Significantly associated COGS associated with AU compared with HC based on fixed effect model

Variant	Annotation	AF	AU (7)	HC (27)	lrt p-value
group_10646	Helix-turn-helix domain-containing protein	0.206	5	2	0.0005
prtR	Helix-turn-helix transcriptional regulator	0.206	5	2	0.0005

group_12496/ kilR	Cell division inhibitor protein	0.206	5	2	0.0005
ydaT	Regulator/ bacteriophage regulatory protein CII	0.235	6	2	0.0009
group_26390	tRNA-Asn(gtt)	0.382	6	7	0.001

Supplemental Table 3.23. Significantly associated COGS associated with AU compared with HC based on mixed effect model

Variant	Annotation	AF	AU (7)	HC (27)	lrt p-value
ydaT	Regulator/ bacteriophage regulatory protein CII	0.235	6	2	0.0000004
group_10646	Helix-turn-helix domain-containing protein	0.206	5	2	0.00004
prtR	Helix-turn-helix transcriptional regulator	0.206	5	2	0.00004
group_12496/ kilR	Cell division inhibitor protein	0.206	5	2	0.00004
group_15336	DUF4102 domain-containing protein	0.0882	3	0	0.0001
group_15450	Helix-turn-helix domain protein	0.0882	3	0	0.0001
group_16142	Relaxosome protein TraM	0.0882	3	0	0.0001

Supplemental Table 3.24. COGS associated with AIEC compared with non-AIEC based on fixed effect model

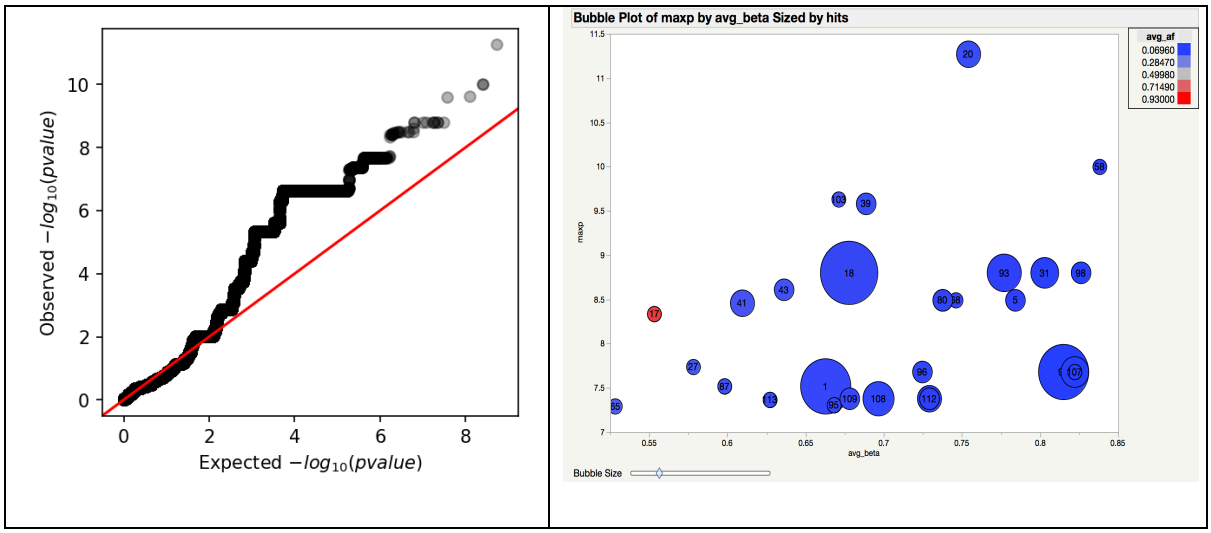
Variant	Annotation	AF	AIEC (15)	Non-AIEC (19)	lrt-p-value
tufB	Elongation factor Tu	0.697			0.00224
higA-2_1	Helix-turn-helix domain-containing protein	0.545	11	7	0.0063
higB-2_1	Phage head morphogenesis protein addiction module toxin RelE	0.545	11	7	0.006
group_2878	Hypothetical protein	0.303	12	7	0.017
group_1777	LysR family transcriptional regulator	0.606	12	8	0.03

Supplemental Table 3.25. Significantly associated COGs in with AIEC compared with non-AIEC based on mixed effect model

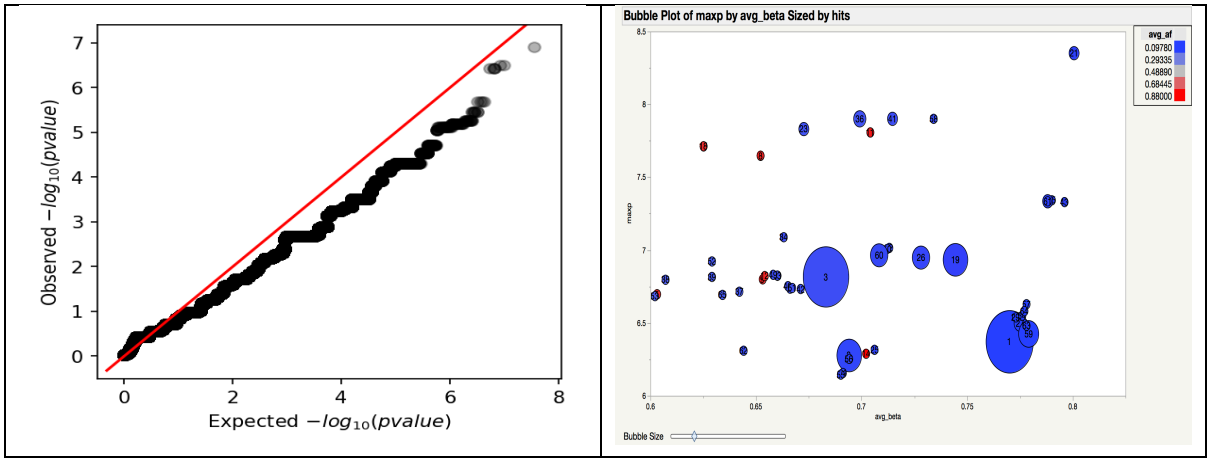
Variant	Annotation	AF	No in AIEC (15)	Non-AIEC (19)	lrt p-value
group_13437	Hypothetical protein	5.45E-01	11	7	0.006

group_2878	Hypothetical protein	3.03E-01	12	7	0.017
ybcO	DUF1364 domain-containing protein	3.64E-01	12	7	0.017
yfaL	AIDA-I family autotransporter YfaL	5.76E-01	12	7	0.017

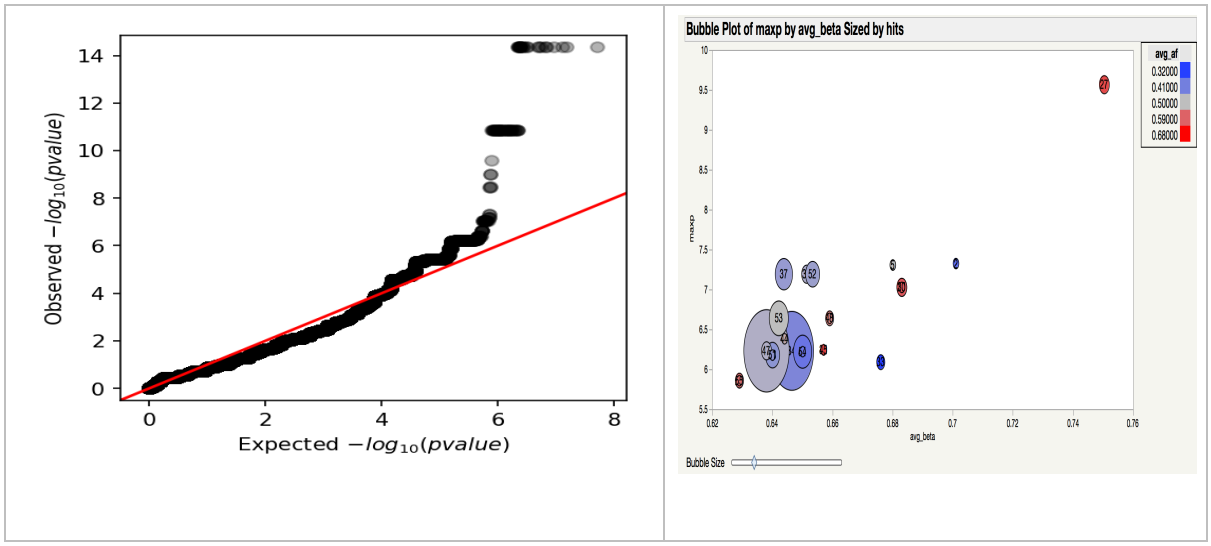
Supplemental Figures-



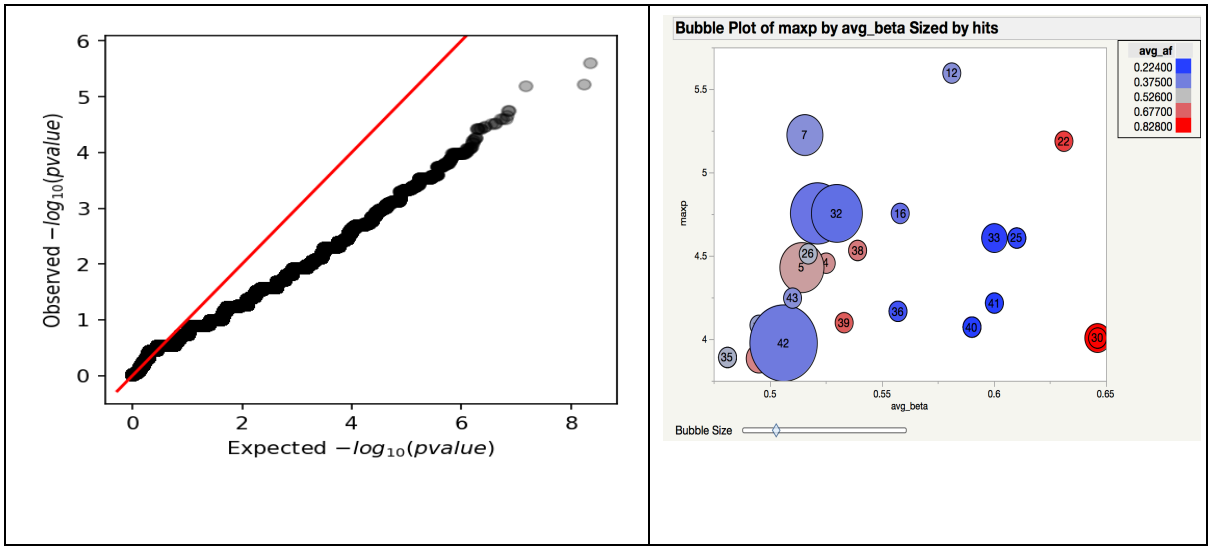
Supplemental Figure 3.1. Significantly associated K-mers with CD compared with HC



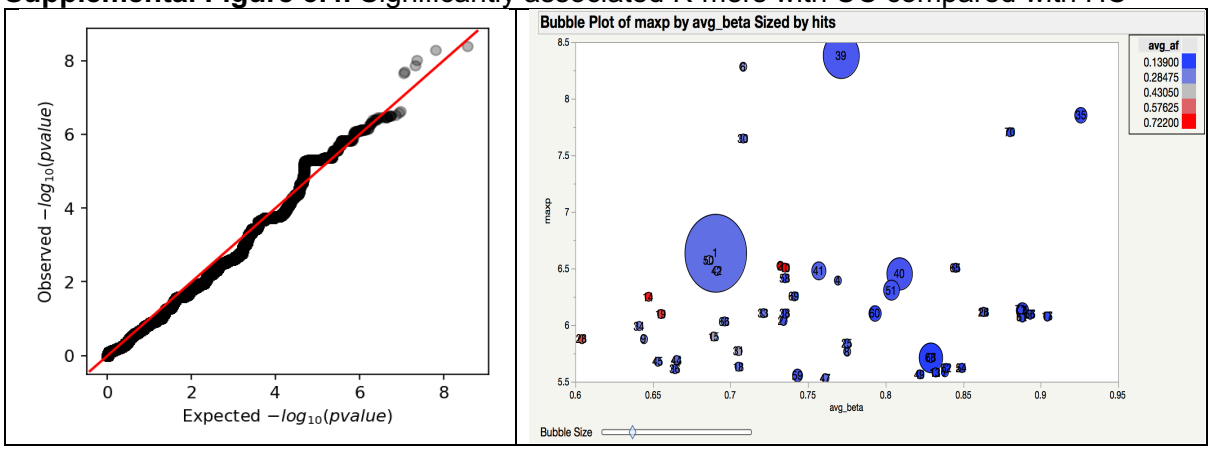
Supplemental Figure 3.2. Significantly associated K-mers with adult CD compared with HC



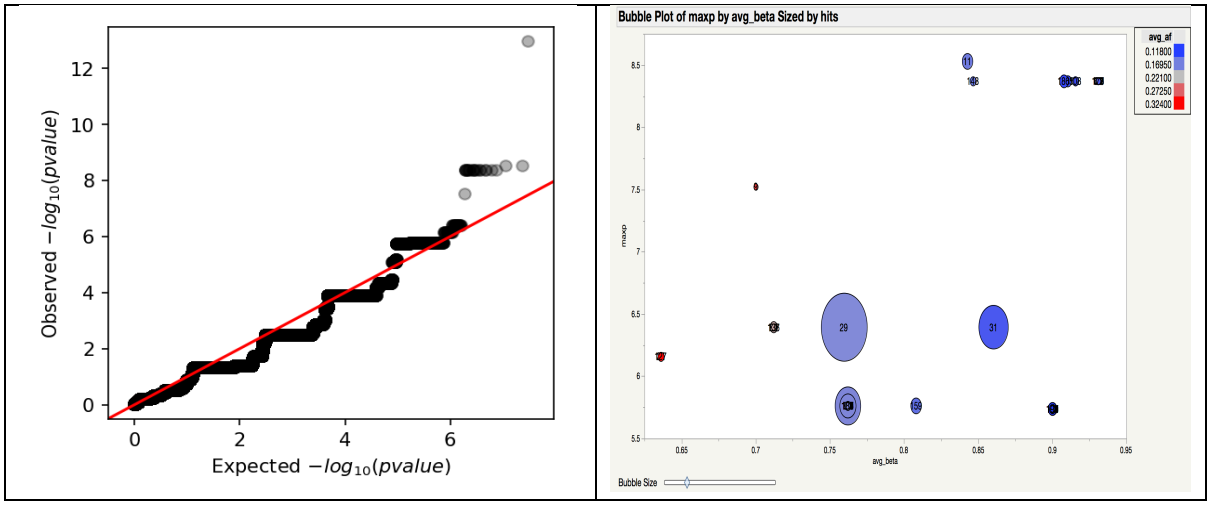
Supplemental Figure 3.3. Significantly associated K-mers with Paediatric CD compared with HC



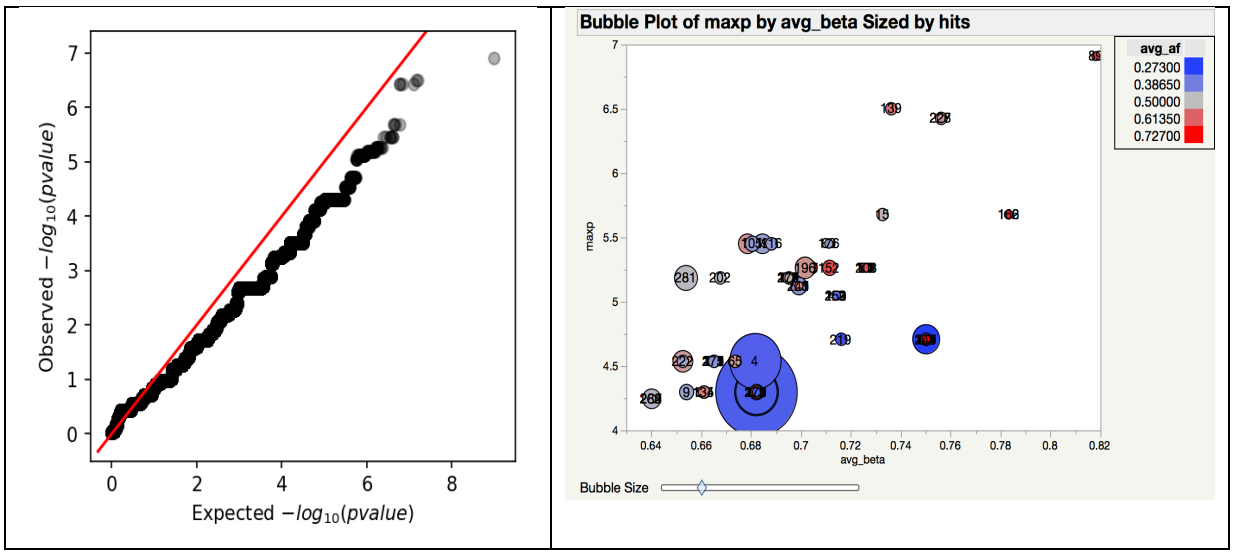
Supplemental Figure 3.4. Significantly associated K-mers with UC compared with HC



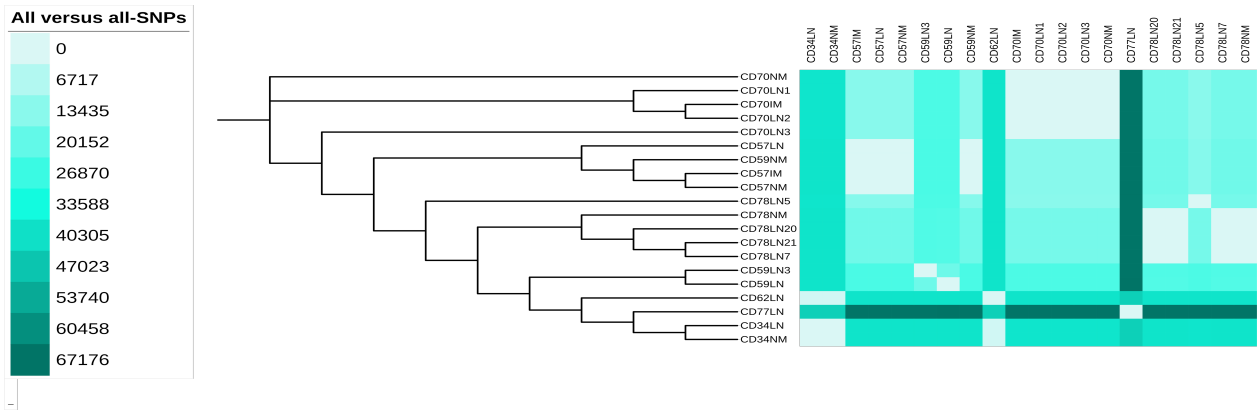
Supplemental Figure 3.5. Significantly associated K-mers with LN-CD compared with HC



Supplemental Figure 3.6. Significantly associated K-mers with AU-CD compared with HC

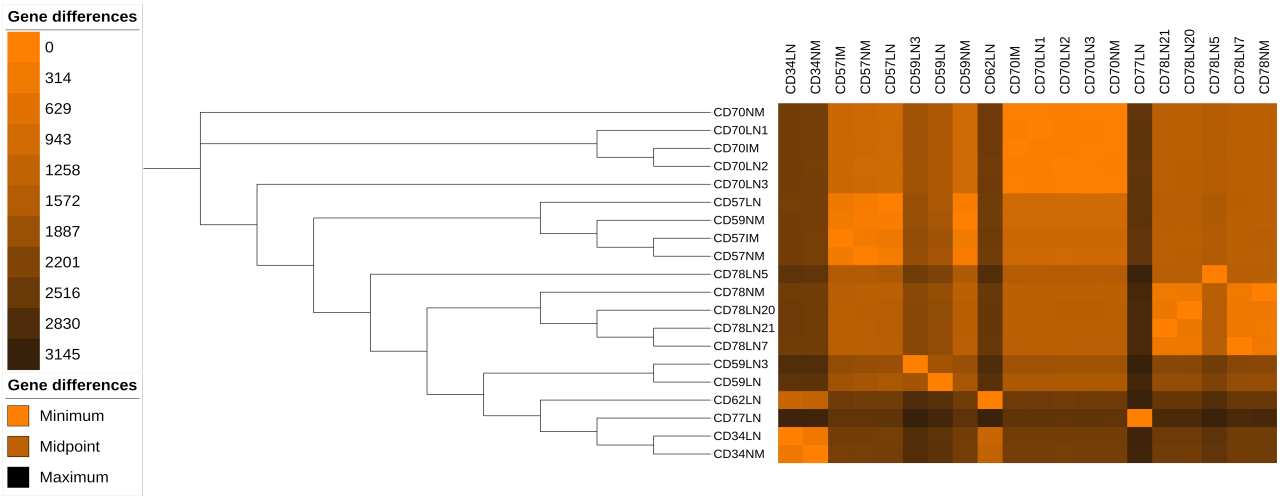


Supplemental Figure 3.7. Significantly associated K-mers with AIEC compared with non-AIEC



Supplemental Figure 3.8. Phylogenetic tree representing the genomic similarity and diversity of *E. coli* isolates from bowel resected samples. Heatmap colours represent the number of

SNPs per nucleotide (all-vs-all method). Lighter colours indicate higher sequence similarity and deeper colours indicate higher diversity



Supplemental Figure 3.9. Phylogenetic tree representing the genomic similarity and diversity of *E. coli* isolates from bowel resected samples based on gene differences. Heatmap colours represent the number of genes varied among the *E. coli* isolates (all-vs-all method). Lighter colours indicating higher similarity and deeper colours indicating higher diversity

Chapter 4: Review on Macrophage- *E. coli* interactions in Crohn's disease

4.1 Introduction

Macrophages are a crucial player of innate immunity. They contact bacteria and release proinflammatory cytokines to eliminate bacteria from the body. Macrophages also play a significant role in removing infected cells from the body through programmed cell death and tissue repair. Accurate knowledge of the macrophage activation process and bacteria that induce this activation could help us develop targeted therapeutics for disease. Our current beli is that interactions between bacteria and host macrophages play a significant role in the pathophysiology of illness associated with chronic inflammation like Crohn's disease (CD).

Several researchers have demonstrated that the number of *E. coli* is increased in CD patients compared with healthy subjects. In addition, the colonization of *E. coli* strains that possess adhesive and invasive properties has been described in the ileum and colon mucosa in patients with CD (Lopez-Siles et al., 2014; Buisson et al., 2016; Palmela et al., 2018; Mirsepasi-Lauridsen et al., 2019; Costa et al., 2020). Our primary focus was to investigate *E. coli* and host macrophage interaction to explore their role in CD. More specifically, in this review chapter, we demonstrate how *E. coli* interact with the host macrophage and escape the host innate immune response and may play a role in developing chronic inflammation in CD.

4.2 Methodology for literature search for this review

A thorough investigation of the literature was carried out to categorise and select all the related articles that studied the interaction of macrophage and *Escherichia coli* in the context of IBD. We performed a systematic search for original articles using PubMed from 2009 to 2021. We considered original articles for this review chapter after an additional manual review based on the relevance according to the topic of

interest. We chose the following conditions for selecting articles manually: 1) articles that demonstrated the origin, development, activation, and the role of macrophage in immunity. 2) articles showed the interaction between macrophage and *E. coli*, highlighting adherent invasive *E. coli* (AIEC). 3) studies that investigated the association of different strains and virulence factors of *E. coli* that enable them to survive within macrophages. 4) Papers that reported macrophage and *E. coli* interaction's immunological pathway associated with IBD and CD pathogenesis.

4.3 Objective of this review

CD is a chronic inflammatory disease of the gastrointestinal tract. It is a complex disease and is thought to be triggered by three major factors: genetic, immunological factors, and environmental. In this PhD research, we investigated CD's immunological association with NLRP3 inflammasome activation when triggered by *E. coli* using bone marrow-derived macrophages.

Macrophage and other crucial immune cells are found in different tissues and submucosa. When they come in contact with the microbes, they are activated. They secrete inflammatory immune substances that trigger an immune response toward the microbes and eliminate them from the body. However, some immunological abnormalities may cause dysfunction of macrophages which make them unable to clear bacteria properly. Eventually, those bacteria can survive inside the cells and cause diseases. We aimed to find out the cause of dysregulated macrophage function that may be triggered by *E. coli* leading to the chronic inflammation in CD. Therefore, it is vital to know the origin, activation, and the role of the macrophage in the immunological process in health and disease.

4.4 The complex biology of macrophage: origin, function, and activation

Understanding the origin of macrophages during early embryonic development is important to know its fundamental role in the innate immune system. Throughout life, macrophages exist in all tissues of the body, regulating and maintaining tissue homeostasis and initiating immune responses to tissue injury or invasive pathogenic bacteria. Ilya Metchnikoff discovered macrophages in the late 19th century. Then, Van Furth suggested that tissue-resident macrophages develop from adult bone marrow-derived blood monocytes. Recent publications have amended this hypothesis of the origin of macrophages by demonstrating that numerous tissue-resident macrophages develop throughout embryonic life and are maintained autonomously by self-renewal in adult life without the contribution of circulating monocytes (Ginhoux et al., 2010; Schulz et al., 2012; Guilliams et al., 2013; Hashimoto et al., 2013; Jakubzick et al., 2013; Yona et al., 2013; Epelman et al., 2014). In mammals, the earliest macrophage progenitors originate in the yolk sac, the initial site of haematopoiesis (McGrath et al., 2015). The theory of about 30 years ago proposed that macrophages originating from the embryonic yolk sac supply the macrophage pool in adults. Throughout embryogenesis progenitor cells migrate to reside inside different tissues of the body. However, these unique progenitor cells are not identified in adult animals.

As described in the Central Illustration by Williams et al., 2018 (Williams et al., 2018), the initial differentiation of macrophages starts at around embryonic day seven (E7.0) as detected in blood islands of the yolk sac in the mouse. By E8.0 to E8.5, erythro-myeloid precursor cells (EMP) that exhibit properties of myeloid cells (myeloid cells include circulating monocytes and tissue-resident macrophages) as well as their capacity to differentiate can be detected in the yolk sac (Frame et al., 2013; Mass et

al., 2016). From E8.5, EMPs gain access to the blood vasculature and migrate to embryonic tissues (Stremmel et al., 2018) including the fetal liver. Between E11 to E17.5, the fetal liver possesses erythro-myeloid progenitor-derived fetal monocytes, which act as the primary source of tissue-resident macrophages before the bone marrow develops (Hoeffel et al., 2015). The present understanding is that these erythro-myeloid progenitor cells generate early tissue-resident macrophages. However, till now, there are debates regarding the origin and differentiation of macrophages from progenitor cells in early embryonic growth.

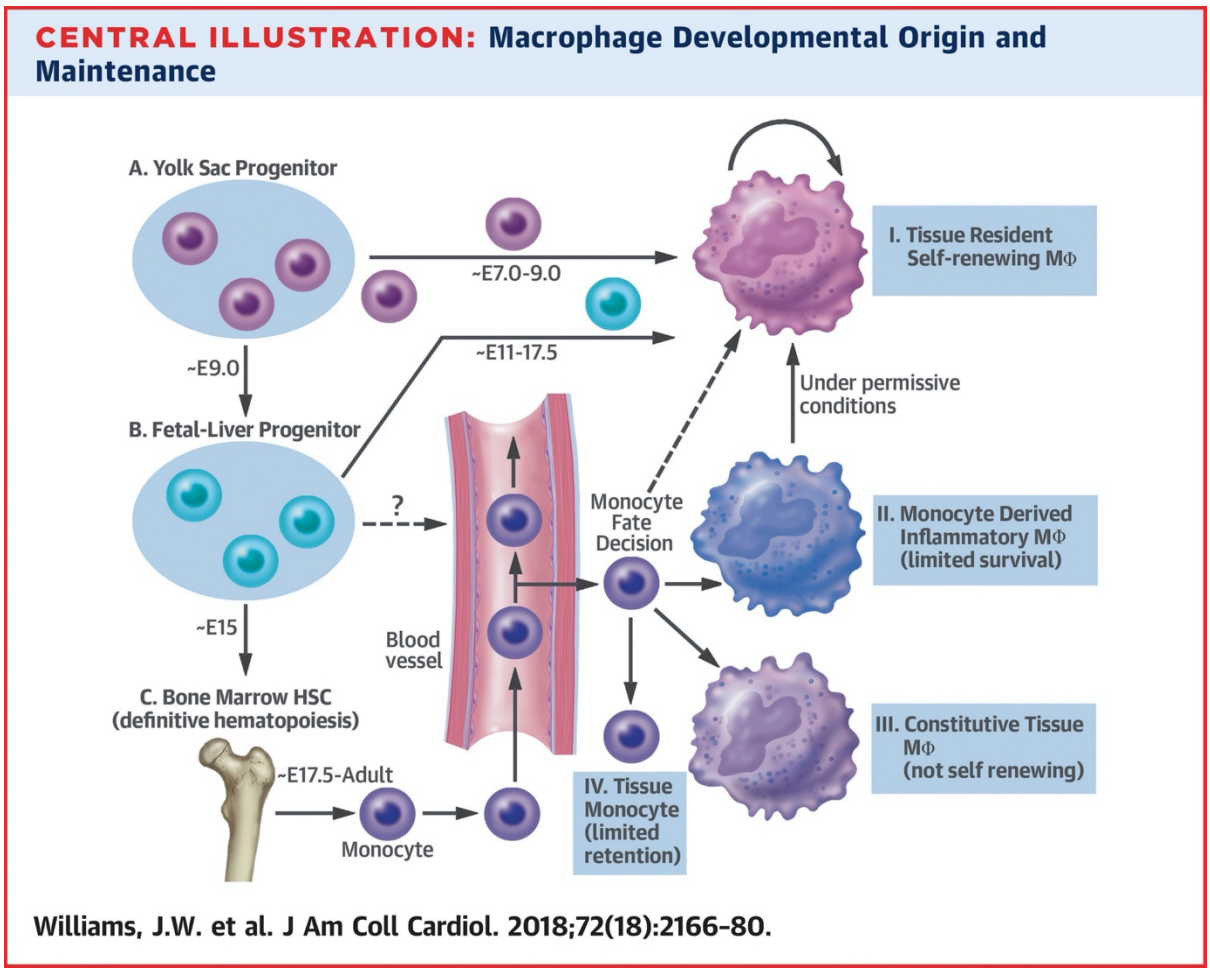


Figure 4.1 Central Illustration: Macrophage Developmental Origin and Maintenance. At different time points macrophages are seeded into different tissues. Throughout embryonic day (E) 7.0 to E9.0, yolk sac precursors give rise to macrophages. Over embryonic day (E) 11 to E17.5, macrophages are seeded from the monocyte-like cells that originate from the fetal liver; and from embryonic day (E) 17.5 till adulthood, fully

developed monocytes in the bone marrow seed macrophages throughout the body. In figure 4.1 Central Illustration, four distinct macrophage populations are displayed. I. tissue-resident macrophages; these are the major population of macrophages in most tissues II. monocyte-derived macrophages originating during inflammatory conditions and tissue injury; III. constitutive tissue macrophages: these macrophages are solely derived from monocytes and cannot self-renew in the tissues; and IV. monocytes transferring through tissues. Here MΦ = macrophage.

4.5 Lineage tracking of tissue-resident macrophages

It is believed that an additional pool of embryonic macrophages strengthens the pool of early macrophages originating from EMPs. It is assumed that macrophages originating from EMP are replenished by macrophages derived from hematopoietic stem cells (HSCs) after these hematopoietic precursor cells become established in adults. However, current studies have challenged this hypothesis as they observe that some tissue resident macrophages arise from a precursor during embryonic development. During adult life, these macrophages can maintain the population by a self-renewal process independently and without the help of HSCs. However, recently, a study proposed that there is an important contribution of HSCs to the macrophage population in different tissues during adulthood (Sheng et al., 2015). The relationship of EMPs to tissue macrophages in adults is still not clear. A current study based on fate-mapping models found that yolk sac macrophages arising from early EMPs do not have a monocytic intermediate. In contrast, transcription factor c-Myb positive late EMPs in the fetal liver generate monocytes that subsequently differentiate into macrophages (Hoeffel et al., 2015). They also stated that before birth, tissue-resident macrophages which have the self-renewal ability during adulthood are seeded (Hoeffel et al., 2015). Although adult macrophages are only supposed to originate from bone marrow precursors and monocyte recruitment, tissue-resident macrophages develop from diverse origins. Based on their location in the body, macrophage populations can

be expanded to a greater or lesser degree by recruitment of blood monocytes, particularly if inflammatory conditions arise.

4.6 Role of macrophages in innate immunity

Numerous microorganisms inhabit our body. Although there is frequent exposure to various microorganisms almost all host immune responses protect us from invasion by those microbes. Our body's immune system consists of two parts: the innate and the adaptive immune systems. Innate immunity is inborn and crucial for maintaining the homeostasis. It operates instantly as the first line of defence against pathogenic microbes when they enter into the body. Neutrophils, macrophages, and dendritic cells are the primary immune cells contributing to innate immunity. They identify then eliminate bacteria from our body through an inflammatory response. Inflammatory reactions are activated when innate immune cell-associated receptors recognise pathogen-associated molecular patterns (PAMPs) using pattern recognition receptors (PRRs).

Similarly, damage-associated molecular patterns (DAMPs) stimulate an immune reaction to damaged tissue (Schenten and Medzhitov, 2011). Toll-like receptors (TLRs) are a type of PRRs associated with the recognition of pathogens. TLR signalling pathway activation results in a signal cascade's activation with a TIR domain-containing adaptor, MyD88. MyD88 is recruited by TLRs, activates mitogen-activated protein (MAP) kinases and transcription factor-like NF-κB resulting in the initiation and expression of inflammatory cytokine genes (Kawai and Akira, 2011). PRRs are not only found on the cell membrane, but also exist inside the cytoplasm, such as NOD-like receptors (NLRs), and RIG-1-like receptors (RLRs). Numerous NLRs form the inflammasome (a multimeric protein complex). Activation of

the inflammasome initiates the conversion and maturation of pro-interleukin IL-1β and IL-18 into the secretory forms, IL-1β and IL-18. The inflammasome also plays a crucial role in eliminating infectious agents and aggravating the inflammatory process by triggering pyroptosis (known as programmed cell death) (Martinon et al., 2009). On the other hand, RLRs identify viral double-stranded RNA and with the help of interferon-β promoter stimulator (IPS-1) activate cellular antiviral signalling. This activation results in the secretion of type I interferons and inflammatory cytokines to remove the virus (Bruns and Horvath, 2014).

Macrophages respond to acute infection, tissue injury and maintain homeostasis in response to the alterations of the body's condition using their regulatory and restorative function. A significant function of macrophages is to engulf foreign substances, including microorganisms and other particulate matter, that enter the body. Moreover, they also maintain tissue homeostasis by eliminating apoptotic cells, recycling nutrients from the tissues and repairing injured tissue (Sieweke and Allen, 2013; Watanabe et al., 2019). Although macrophage activation needs the initial antigen stimulation, their activity is boosted by helper T lymphocytes through cytokines release. Among those cytokines, interferon-gamma (IFN-γ) is the most effective activator. Macrophages also process antigens and present it to lymphocytes through the major histocompatibility complex (MHC) class II molecules on their surface. After the microbes' engulfment, macrophages continue the processing and presentation of the antigen so that T helper cells can recognize those. After that, T lymphocytes secrete cytokines to activate B lymphocytes which subsequently differentiate into plasma cells that produce antibodies against those antigens on the macrophage's surface. Those antibodies form antigen-antibody complex by attaching to the microbes' antigens so that they are phagocytosed more effectively by macrophages.

The latest findings demonstrate that macrophages originate from HSC-derived monocytes and the embryonic yolk sac. The differentiation of tissue macrophages into various phenotypes depends on different microenvironmental signals resulting from tissues' diverse locations. Macrophages are epigenetically oriented to respond to different tissue environment stimuli (Wynn et al., 2013; Lavin et al., 2014). Tissue-resident macrophages and monocyte-derived macrophages play separate roles during injury. Typically, monocyte-derived macrophages exhibit a strong response during inflammation, go through the apoptotic process or gain the capability to self-renew while tissue-resident macrophages either die or via a self-renewal process, repopulate within the tissue. Interestingly, tissue-resident macrophage and monocyte-derived macrophages are phenotypically indistinguishable.

Unlike tissue-resident macrophages in other tissues of the body, the intestinal macrophage pool in the mouse is maintained by ongoing replenishment by Ly6C^{hi} circulating monocyte-derived macrophages (Bain et al., 2014; Bain and Schridde, 2018). Depletion of the macrophage population in the gut quickly causes an impairment of the integrity of the gut epithelial layer that could result in an increased risk of inflammation. A recent study based on transcriptional profiling and macrophage lineage tracing found that macrophages in the gut that were long-lived through the self-renewal process played a crucial role in regulating vascular integrity in the submucosa (De Schepper et al., 2018; Shaw et al., 2018). In the intestine, both tissue-resident and monocyte-derived macrophages take part in clearing pathogenic microbes through developing inflammation (Kim et al., 2011; Weber et al., 2011; Zigmond et al., 2012; Bain et al., 2013b; Grainger et al., 2013; Farro et al., 2017).

4.7 Role of macrophages in inflammation and tissue repair

During tissue injury, the macrophage's role is complex, and includes inflammation activation in response to cellular damage, and healing. During healing, the role of macrophage is to suppress inflammation and start the tissue regeneration process. The macrophage's injury repair process begins with the phagocytosis of tissue debris, microorganisms, and apoptotic cells in the injured site and generating growth-promoting mediators to support the tissue regeneration process (Gordon, 2016; Vannella and Wynn, 2017). Subsequently, after the tissue injury, if the activation of macrophages is interrupted, the initial inflammatory process could be depleted resulting in the inefficient or incomplete resolution of the infection. In some situations, dysregulated macrophage tissue repair processes could be responsible for developing destructive pathological conditions like fibrosis or scarring due to chronic inflammation (Oishi and Manabe, 2018). While the tissue repair process is organized by various immune cells, macrophages have significant interest as a potential therapeutic agent (Getts et al., 2014; Happle et al., 2014) and some researchers have proposed that circulating monocytes and tissue- resident macrophages play distinctive roles in inflammation, and repair of tissue after injury (Gundra et al., 2014; Vannella et al., 2014). Currently, macrophage research's main focus is to explore whether the tissue microenvironment influences the activity of macrophages in different stages of the tissue repair process, or a specific subset of macrophages plays a distinctive role at different time points to control the wound repair process.

4.8 Macrophages in the intestine

After birth, embryonic precursor cells in the intestine are replaced by cells originating from Ly6C blood monocytes (C. Bain et al. 2013b). In both healthy and inflamed intestine these Ly6C monocytes develop into resident macrophages and continuously

supply the macrophage pool. Different subsets of macrophages reside in the intestine, however the stimuli responsible for this differentiation are unknown (Bain et al., 2013b). Studies have found that microbiota also stimulate the differentiation of macrophages (Lavin et al., 2014; Mortha et al., 2014). Also, microbial signals in the intestine regulate the secretion of granulocyte-macrophage colony-stimulating factor (GM-CSF) by innate lymphoid cells (ILCs) and control the number of macrophages in the colon. In addition to dendritic cells, macrophages play a distinct, but complementary role to maintain homeostasis, tissue repair, mucosal layer integrity and immune defence in the gut (Mowat and Bain, 2011b; Gross et al., 2015; Kühl et al., 2015). Any infection or tissue injury leads to the development of inflammatory responses in the gut that recruit monocytes from the circulation to the site of inflammation and trigger their differentiation into proinflammatory macrophages. These newly differentiated macrophages release inflammatory mediators that support immune response progression and wound repair (Cassetta et al., 2011; Mowat and Bain, 2011a). Normally during stable conditions, macrophages in the intestine maintain the homeostatic status of gut mucosa by removing pathogens and their products that cause inflammation. However, activated intestinal macrophages may also trigger polarization of Th1/Th2 and Th17 helper T lymphocytes, and CD4+ T lymphocyte subsets in the inflammation site (Arnold et al., 2015), that may cause the chronic inflammation in a patient with IBD (Weaver et al., 2013).

4.9 Role of macrophages in intestinal injury and repair

Previous reports found that macrophages play an important role in the resolution of tissue after injury. The innate immunity of gut has developed several ways to inhibit inappropriate immune response towards commensal microorganisms. However, the

process through which non-pathogenic bacteria and dysbiotic changes in the gut's bacterial community may result in intestinal inflammation is unknown.

The microbiota-mediated NLRP3 inflammasome induces the release of interleukin-1β following gut injury (Seo et al., 2015). After intestinal epithelial injury, specific bacteria from the Enterobacteriaceae family recruited Ly6C^{hi} CCR2+ blood monocytes that caused the release of IL-1β in the gut. In addition, upon gut injury, hemolysin in bacteria also promoted the activation of NLRP3 inflammasomes in gut macrophages that induced the secretion of IL-1 (Seo et al., 2015). Innate lymphoid cells (ILCs) take part in tissue repair and may also induce pathological conditions like colitis in the gut. Colitis driven by type 3 ILCs depends on granulocyte macrophage-colony stimulating factor (GM-CSF) that recruits and maintains inflammatory monocytes in the gut (Pearson et al., 2016). Their study also suggested that blocking the secretion of GM-CSF or modifying the mobilization and function of ILC3 can resolve intestinal inflammation. On the other hand, Dabritz and colleagues (2015) observed that GM-CSF may have an important protective role against intestinal inflammation through converting inflammatory monocytes to a protective phenotype that helps tissue repair. They demonstrated that GM-CSF also boosted monocytes' ability to adhere and migrate, and primes responses of the monocyte towards secondary microbial stimuli. Another study demonstrated a different pathway that transforms macrophages into a protective phenotype that resolves inflammation in the gut (D'Alessio et al., 2014). In their study they explored the plausible role of lymphangiogenesis in IBD. They found that upregulated lymphangiogenesis and lymphatic function reduces IBD in experimental animal models. They also observed that adenoviral induction of vascular endothelial growth factor-C (VEGF-C) contributed a clear protective response against acute and chronic colitis. VEGF-C supported the translocation of macrophages and

antigen from the inflamed intestinal mucosa to the draining lymph nodes and induced resolution of inflammation. This observation indicates that the VEGF-C pathway could play a potential therapeutic role in inflammatory diseases.

Increased expression of IL-36R promoted mucosal healing in the gut by activating epithelial cells and fibroblasts in the intestine of a mouse model of dextran sodium sulfate (DSS)-induced experimental colitis (Scheibe et al., 2017). The authors detected IL-36γ in the enterocytes and IL-36α in the cytoplasm of CD14+ inflammatory macrophages. They also observed that the secretion of IL-36α and IL-36γ was upregulated in the intestine of patients with active IBD as well as the mouse model. Moreover, blocking IL-36 signalling caused increased sensitivity to DSS and marked impairment of wound healing (Scheibe et al., 2017). As the principal source of IL-36, macrophages and enterocytes together regulate the expression of chemokines and antimicrobial protein, lipocalin 2, along with GM-CSF production that in combination increase intestinal wound repair process in IBD (Scheibe et al., 2017). Esser-von Bieren and colleagues demonstrated that decreased expression of chemokine CXCR2 ligands by macrophages and myofibroblasts in the intestine was associated with impaired wound healing (Esser-von Bieren et al., 2015). They conclude that helminths and antibody-driven macrophage-myofibroblast chemokine signalling is crucial for promoting intestinal mucosa repair (Esser-von Bieren et al., 2015). When the wound resolution process is disturbed, macrophage-driven activation of fibroblasts may also play a pathogenic role. After bone marrow transplantation, a common side effect is organ fibrosis. Yeh and colleagues found that after bone marrow transplantation, fusion between bone marrow-derived monocytic and intestine stromal cells promotes radiation-induced fibrosis development. They suggested that bone marrow-derived

macrophage proliferation promotes activation of local tissue fibroblasts thereby inducing intestine fibrosis (Yeh et al., 2016).

4.10 Macrophages in CD mucosa

Numerous reports have found an increased population of macrophages in the gut of CD patients (Franzè et al., 2013; Thiesen et al., 2014) that express the scavenger receptors CD163 and CD68 (Franzè et al., 2013). Infiltration of the lamina propria by $CD68^+$ and $CD163^+$ macrophages triggers secretion of IL-17 and TNF-α in the inflamed mucosa in CD patients (Franzè et al., 2013). Treatment with the anti-tumour necrosis factor agent, infliximab, resolved intestinal inflammation, a process that was associated with decreased infiltration of CD68+ macrophages in the lamina propria and the downregulation of IL17A (Caprioli et al., 2013). This result indicates the possible association of CD68 and CD163 macrophages with the development and progression of chronic inflammatory activity in CD. CD14+ macrophages in CD patients have also been reported in some studies (Thiesen et al., 2014). They upregulated the secretion of proinflammatory cytokines like IL-23 and TNF-α that may influence the initiation of Th17-driven inflammation in CD patients.

4.11 Genetic variants associated with CD and their effect on macrophage functions

Several genetic variants are associated with CD (Verstockt et al., 2018; Dai et al., 2019). Some of those could have an impact on macrophage function. The first recognized gene locus that increases the risk of CD is *NOD2*. NOD2 plays a major role in the innate immune response as an intracellular sensor to pathogen-associated molecular patterns (PAMPs) (Strober and Watanabe, 2011). NOD2 can identify muramyl dipeptide, a component of the peptidoglycan layer of the bacterial cell wall

(Couturier-Maillard et al., 2013; Knights et al., 2014). In the NOD2 gene, three mutations in the leucine-rich repeat domain are associated with antibacterial pathways linked to the CD pathogenesis (Homer et al., 2010). NOD2 is also associated with recruiting the autophagic protein, ATG16L1, to stimulate autophagy-driven bacterial removal (Ramjeet et al., 2010). Mutations in NOD2 affect the autophagy-derived elimination of intracellular bacteria that may result in the persistence of bacteria leading to CD pathogenesis (Ramjeet et al., 2010). In addition to *NOD2*, many genes linked to autophagy are correlated with CD, such as *ATG16L1* (autophagy-related like 1) (Iida et al., 2018; Larabi et al., 2020), *IRGM (Prescott et al., 2010; Glas et al., 2013; Ajayi et al., 2019; Mehto et al., 2019)*, and *ULK1* (unc-51-like autophagy activating kinase) (Henckaerts et al., 2011), *PTPN2* (protein tyrosine phosphatase nonreceptor type 2) (Glas et al., 2012; Scharl et al., 2012), *LRRK2* (leucine-rich repeat kinase 2) (Franke et al., 2010), *CYLD* (cylindromatosis) (Cleynen et al., 2014), and *XBP-1*(X-box binding protein 1) *(Hoefkens et al., 2013)*. Studies have demonstrated that the initiation of autophagy, processing and presentation of bacterial antigen and bacterial clearance by macrophages is impaired due to the presence of autophagy-associated gene variants (Lapaquette et al., 2010; Lapaquette et al., 2012). Defects in autophagy lead to a failure in killing intracellular microbes and antigen presentation. These defective macrophages cannot effectively activate immune responses against the bacteria, and subsequently fail to secrete higher amounts of IL-1β and IL-18. This inappropriate response may lead to the initiation and progression of chronic inflammation in CD (Nguyen et al., 2013).

4.12 Host macrophage and bacterial interactions in inflammatory bowel disease

In the intestine, macrophages are situated in lamina propria in the subepithelial layer, and in the Peyer's patches (PPs) which form the border between intestinal lymphoid tissue and the luminal microbiome in the gut (Jung et al., 2010). These macrophages are very closely located to luminal microbes (Jung et al., 2010). Macrophages in the intestine regulate homeostasis. They tolerate commensals and trigger an immune response against pathogenic microbes. Bacterial dysbiosis is significantly correlated with inflammation in the gut and IBD (Belkaid and Harrison, 2017; Blander et al., 2017; Ni et al., 2017). An emerging field of research is to identify the resident macrophages' role in the dysbiosis that commonly occurs in IBD patients. It has been hypothesised that an unusual immune response triggered by resident macrophages to the intestinal microbiota may initiate and develop inflammation in the gut of a patient with IBD.

Macrophages can sense microbes and stimulate type 3 innate lymphoid (ILC3) cells to produce GM-CSF (Mortha et al., 2014). In addition, crosstalk between microbiota, macrophages and ILC3 initiate the secretion of IL-22 which is associated with tissue repair in the gut epithelium (Longman et al., 2014). During inflammation intestinal macrophages can differentiate from monocytes in both human and mice (Platt et al., 2010). In CD11c-diphtheria toxin receptor mice (CD11c-DTR), $Ly6C^{hi}$ monocytes can replenish intestinal macrophages in response to diphtheria toxin–driven (Fordtran) depletion of myeloid cells (Rivollier et al., 2012). $Ly6C^{hi}CCR2^{+}$ monocytes may differentiate into all subsets of macrophages in both healthy and inflamed colon (Bain et al., 2013a). In the steady state, through a sequence of processes, CCR2+$Ly6C^{hi}$ blood monocytes differentiate into CX3CR1+ macrophages (Tamoutounour et al., 2012; Bain et al., 2013a; Zigmond and Jung, 2013). Adherent bacteria can interact

with CX3CR1+ mononuclear phagocytes to regulate the immune response. The intestinal microbiota delivers signals to CX3CR1+ mononuclear phagocytes and induces intestinal inflammatory T cell responses (Panea et al., 2015; Kim et al., 2018). Likewise, extensive colitis was found in mice with absent IL-10 signalling associated diet-driven expansion of *B. wadsorthia* in the gut (Devkota et al., 2012).

The importance of CX3CR1+ mononuclear phagocytes in regulating tissue homeostasis through T cell responses towards pathogenic microbes was also observed in Kim and colleagues' study (Kim et al., 2018). CX3CR1+ mononuclear phagocytes downregulate T helper-1 (Th1) cell responses to commensals in the gut and upregulate the responses of regulatory T cells (Treg). However, when pathogenic microbes contact enterocytes, these CX3CR1+ mononuclear phagocytes can stimulate homeostatic regulatory mechanisms by triggering Th1 cell responses with increased inflammatory responses in the gut. (Kim et al., 2018). Depletion of CX3CR1+ cells cause the reduction of Treg cells and increased the number of Th1 and Th17 cells, thereby promoting intestinal pathology. The balance between T effector and T reg cells is important to preserve homeostasis in the gut. This balance is controlled by the CX3CR1+ mononuclear phagocytes (MNP) pathway which may reduce the induction and progression of inflammation in IBD (Shevach, 2009; Kim et al., 2018). The colonization of oral strains of *Klebsiella* spp in the gut causes induction of Th1 in patients with IBD (Atarashi et al., 2017). They also observed a correlation between intestinal colonization by *Klebsiella spp*. with extensive colitis in IL-10 deficiency which was mediated through Th1 immune responses (Atarashi et al., 2017).

4.13 Immune evasion of *E. coli* in IBD

Several pathogens have developed various strategies to evade the innate immune response. In most cases, this immune evasion is mediated through virulence factors like type III secretion system (T3SS) in gram negative pathogens (Baxt et al., 2013). Many pathogens can reprogram host cells by modulating the host cell's defence mechanisms. Moreover, some intracellular pathogens are capable of blocking the maturation of the phagosome and its fusion to lysosomes which enables them to continue their replication and extend the infection (Uribe-Querol and Rosales, 2017). In some cases, pathogens can also escape from the phagosome by avoiding autophagy and spreading the infection from the primarily infected cell to neighbouring cells. Macrophages are the primary phagocytes of immune defences against invaing pathogens (Wynn et al., 2013) and can generate reactive oxygen and nitrogen species within phagosomes to trigger a bactericidal response. Bacteria can protect themselves by several mechanisms, including capsule formation, toxin release, filamentation, and biofilm formation.

E. coli are the best example of many of these mechanisms (Hunstad and Justice, 2010). Immediately after birth, *E. coli* colonizes the gastrointestinal tract and resides as a commensal in the host intestine. However, some studies have shown that avirulent *E. coli* strains in the gut can acquire virulence factors to become pathogenic *E. coli* and cause intestinal disease by horizontal gene transfer through a transfer of plasmids, bacteriophages, pathogenicity islands, and transposons (Govindarajan et al., 2020). Commensal *E. coli* can also evolve into pathogenic *E. coli* (Crossman et al., 2010; Denamur et al., 2010) that are resistant to phagocytosis. Miskinyte et al. (2013) reported that commensal *E. coli* can evolve extremely quickly and develop the ability to survive inside macrophages or escape phagocytosis and

become pathogenic under the selective environmental pressure of certain elements of the host innate immune system (Miskinyte et al., 2013).

AIEC that genetically resemble extra-intestinal pathogenic *E. coli* (Expec) can adhere and invade enterocytes and survive and replicate within phagocytes like macrophages (Miquel et al., 2010; Dogan et al., 2014). IgA-coated AIEC from CD patient with spondyloarthritis stimulated Th17 cell responses and the expression of IL-10 (Viladomiu et al., 2017). These AIEC from patients with Crohn's disease-associated spondyloarthritis caused more severe colitis in IL-10-deficient mice (Viladomiu et al., 2017). They can lead to the activation of immune pathways to reduce intestinal inflammation. Therefore, altering host genetic factors that impair anti-inflammatory signals can make non-pathogenic bacteria into pathogenic (Viladomiu et al., 2017). In addition an increased risk of developing IBD is correlated with the increased use of antibiotics (Belkaid and Hand, 2014). However, Kim et al (2018) (Kim et al., 2018) suggest that developing IBD as a result of exposure to antibiotics can be due to the disruption of microbiota that leads to the loss of anti-inflammatory functions of CX3CR1+ MNPs.

4.14 The interplay between macrophages and *E. coli* in CD

Numerous studies have investigated and demonstrated a close relationship of intestinal macrophages to the pathogenesis of CD. A higher mucosal density of macrophages is predominantly observed in CD patients' intestinal mucosa (Perminow et al., 2009; Kühl et al., 2015; Buisson et al., 2016). CD14+ macrophages secreting IL 23 and TNF have also been found in patients with CD and macrophages are considered a key source of several cytokines associated with its pathogenesis.

Similarly, epithelioid granulomas, a collection of macrophages, are the distinctive pathological features in the intestine of patients with CD (Ye et al., 2015; Rothschild et al., 2017).

However, the actual role of granuloma in CD is not clearly understood. There is a plausible association of *E. coli* particularly adherent-invasive *Escherichia coli* (AIEC) with the development of granulomas in patients with ileal CD. A study observed that the AIEC, LF82, induced the aggregation of macrophages, and subsequently recruited lymphocytes to form granulomas. The presence of epithelioid granulomas is also linked to the aggressive progression of disease based on a six-year follow-up study of CD patients (Johnson et al., 2018). Biopsies from Peyer's patches (a dome like structure composed of a layer of epithelial cells over lymphoid follicles in the gut) frequently contain granulomas. The earliest lesion of ileal CD, the aphthous ulcer that overlies Peyer's patches, originates from activated macrophages (Buisson et al., 2016).

In CD patients, dysbiosis is commonly observed. There is decreased abundance of protective bacteria like *Faecalibacterium prausnitzii* and increased abundance of harmful bacteria with proinflammatory properties such as *Escherichia coli*, more specifically adherent-invasive *E. coli* (AIEC) (Chassaing and Darfeuille-Michaud, 2011; Viladomiu et al., 2017).

AIEC suppress Th17 responses and promote the secretion of IL-10 by inducing CX3CR1+ mononuclear phagocytes (Viladomiu et al., 2017; Kim et al., 2018). The absence of IL-10 signaling stimulated AIEC to intensify intestinal inflammation through the induction of TH17 immune response in IL-10-knockout mice but not in wild-type (Viladomiu et al., 2017).

As an increased number of *Escherichia coli* is observed in the gut mucosa of patients with CD, it has been hypothesised that CD-associated macrophages may have impaired function, specifically towards infectious microbes like *Escherichia coli* (Campos et al., 2011; Sewell et al., 2012; Smith et al., 2015; Vazeille et al., 2015a). A clear understanding of the interaction between macrophages and *E. coli* in CD could help establish new macrophage- targeted therapies.

Table 4.1 Summary of recent studies concerning the intracellular persistence of *E. coli* in IBD

Reference	**Cell types where *E. coli* persisted**	**Type of *E. coli* used in the study**	**Key observations**
Mayorgas et al. 2021 (Mayorgas et al., 2021)	Human Epithelial Organoid 3D Cultures (EpOCs) and Intestine-407 – I407 cell line	CD-associated AIEC strain LF82	They examined the capacity of AIECs to invade epithelial cells in organoid cultures. They also proposed that this model be used for crosstalk studies between AIEC- intestinal epithelial cells
Abdelhalim et al. 2020 (Abdelhalim et al., 2020)	Monolayers of Caco-2 cells and RAW 264.7 macrophage cells	Adherent-invasive *E. coli* (AIEC) strain LF82 and *E. coli* strains from CD patients	In CD patients, 18% *E. coli* strains were not invasive, 82% *E. coli* strains possessed adherent and invasive properties and 59% *E. coli* isolates survived and replicated within macrophages. They also found that virulence genes *chuA* and *ratA* were significantly associated with AIEC compared with non-AIEC.

Leccese *et al.* 2020 (Leccese et al., 2020)	Monocyte-Derived Macrophages (MDM) and Monocyte-Derived Dendritic Cells (MoDC) from blood samples derived from CD patient and HT29 cell line	CD associated AIEC strain LF82	*Lactobacillus* and *Bifidobacterium* strains considerably reduced the adhesion and survival of AIEC LF82 inside HT29 intestinal epithelial cell line and inhibited secretion of IL-8. AIEC-LF82 persisted inside macrophages derived from IBD patients compared to healthy donors. LF82 survived but did not replicate inside monocyte-derived macrophages from healthy donors or patients with UC. This study confirmed the ability of LF82 to replicate inside the CD-derived macrophages. Probiotic strain *B. breve* Bbr8 triggered a significant reduction of the persistence of AIEC followed by a reduction of proinflammatory responses in monocyte-derived dendritic cells in IBD patients.
Ormsby *et al*. 2019 (Ormsby et al., 2019)	RAW264.7 macrophages and THP-1 cells	AIEC (LF82) and *E. coli* isolates from patients with CD	When ethanolamine was added extracellularly at a similar concentration as in the human intestine, AIEC within macrophages were able to access it. Ethanolamine increased the growth and replication of AIEC intracellularly. In addition, the *E. coli* ethanolamine utilization *(eut)* operon was upregulated in children with active CD compared to controls.
Walczuk *et al*. 2019 (Walczuk et al., 2019)	HEp-2 cell line, Caco-2 cell and human monocytic cell line THP-1	Diffusely adherent *E. coli* (DAEC) from patients with CD and UC.	Maximum number of DAEC stains with AfaE3 adhesin were able to adhere and invade Caco-2 cells, and stimulated the secretion of higher amount of TNF-α.
O'Brien *et al.* 2016 (O'Brien et al., 2016)	I-407 epithelial cell line and THP-1 macrophage cell line	B2 phylogroup *E. coli* strains	Among all 41 B2 phylogroup, 79% survived and replicated in macrophages. 11 out of 41 strains from patients with IBD and controls also adhered and invaded epithelial cells and were considered AIEC pathovars. They could not identify any gene or molecular property associated with AIEC.

Vazeille *et al*. 2016 (Vazeille et al., 2016)	Intestinal epithelial cell lines T84, Caco-2, bone marrow-derived macrophage from mice and M-cell monolayers	AIEC	They found a higher prevalence of *E. coli* isolates containing g*ipA* gene in CD patients compared to controls. Deletion of *gipA* gene decreased the translocation of AIEC across M cells and replication within macrophages. Survival of *gipA-negative A*IEC was decreased in medium comprising H2O2 or acidic pH. In mice, deletion *of GipA* also impaired the colonization AIEC in PPs and translocation to mesenteric lymph nodes.
Conte *et al.* 2014 (Conte et al., 2014)	Human Caco-2 cell lines, HEp-2 cell monolayers and macrophage cell line J774	*E. coli* from mucosal biopsies in paediatric CD patients	They identified 36 AIEC-like isolates in CD patients (10%) and five AIEC-like isolates in non-IBD controls (1%). They observed that, genes associated with virulence factors K1, kpsMT II fyuA, and ibeA was significantly associated with AIEC strains isolated from CD patients compared to controls.
Prorok-Hamon *et al*. 2014 (Prorok-Hamon et al., 2014)	Caco2, HEp-2, I-407 cell lines and J774A.1 murine macrophage cell line	Colonic mucosa-associated diffusely adherent afaC+ *E. coli*	*E coli* expressing *afaC* were associated with CD but not with ulcerative colitis compared to controls. *E. coli* possessing *afaC* and *lpfA* (genes associated with M-cell translocation) were predominantly found in CD but not in UC compared to controls. All isolates expressing *dsbA* and *htrA* were associated with intra-macrophage replication. Among those isolates 242 out of 281 expressed *fimH* that encodes type-1 fimbrial adhesin.
Negroni *et al* (Negroni et al., 2012)	HEp2, Caco2, and RAW 264.7 cell lines	AIEC	Two bacterial strains possessed similar adhesion and invasion properties of AIEC strain, LF82. Those strains also upregulated expression of CEACAM6, IL-8 and TNF-α both in vitro and in cultured intestinal mucosa. They also survived within macrophages and destroyed the integrity of intestinal epithelium.
Sepehri *et al*. 2011 (Sepehri et al., 2011)	RAW264.7 macrophages cell line	Mucosal-associated *E. coli*	A significant relationship was observed between *afaE* and IBD, as well as between *afaE* and CD. The survival ability of *E. coli* isolates in macrophages were highest in those isolated from UC patients, followed by CD patients compared with controls. Compared with *E. coli* isolates from controls, *E.*

			coli isolated from patients with IBD were more invasive.
Lapaquette *et al*. 2010 (Lapaquette et al., 2010)	HeLa, Hep-2 and Intestine-407	CD associated *E. coli* strain LF82	Autophagy-associated gene IRGM- and ATG16L1-deficient host cells triggered enhanced intracellular replication of AIEC LF82 bacteria.

4.15 Persistence of *E. coli* in macrophages of patients with CD

Macrophages transport bacteria from the gut mucosa to nearby mesenteric lymph nodes. Bacteria, especially *E. coli*, are more prevalent in mesenteric lymph nodes in CD patients than in healthy subjects (Buisson et al., 2016). O'Brien and colleagues also demonstrated a higher prevalence of *E. coli* in mesenteric lymph nodes from patients with CD compared to controls. In addition, they also observed that the proportion of *E. coli* was higher in ileal CD compared to colonic or ileocolonic CD patients (O'Brien et al., 2014). Transgenic mice that express the carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) receptor in the gut were able to translocate AIEC from the gut mucosa to mesenteric lymph nodes (Chassaing et al., 2011; Vazeille et al., 2016). This translocation is mediated through the M cells that overly Peyer's patches and deliver these bacteria to underlying macrophages. Macrophages cannot kill the internalized AIEC and translocate them to the mesenteric lymph nodes (Chassaing et al., 2011). Macrophages that possess *E. coli* were observed in 71% of patients with CD compared with none in a control group (Elliott et al., 2015).

4.16 Handling of AIEC by macrophages and the role of virulence factors

A study comparing the internalization of bacteria in monocyte-derived macrophages between patients with CD and controls revealed that the ability of AIEC to be taken up

by macrophages is higher than non-pathogenic *E. coli*. This interaction between AIEC and macrophages suggests that AIEC may possess certain molecular properties or unique factors that support their uptake by macrophages (Vazeille et al., 2015b). Strategies to skip degradation by phagocytes include blocking the fusion of endosomes with lysosomes to form phagolysosomes, resisting the toxic molecules of phagolysosomes, and getting entry to a particular niche that helps them to replicate intracellularly. It was also found that non-pathogenic *E. coli* were easily killed by macrophages, while AIEC survived and replicated within monocyte-derived macrophages from patients with CD, UC and controls (Elliott et al., 2015; Vazeille et al., 2015b). Intramacrophage replication of AIEC has been seen in monocyte-derived macrophages in CD compared with that of controls which indicates that macrophages in CD patients fail to control the replication of AIEC (Vazeille et al., 2015b). In general, all intracellular bacteria are targeted by the autophagy pathway mediated through lysosomal degradation within the macrophage leading to their elimination (Lapaquette et al., 2012; Paulus and Xavier, 2015). However, in patients with CD, as outlined above, polymorphisms in some autophagy gene such as ATG16L1, IRGM or NOD2 are associated with intra-macrophagic replication of AIEC leading to an increased release of the proinflammatory cytokines, IL-6 and TNF-α (Lapaquette et al., 2012).

Inducing the autophagy pathway reduced the replication of AIEC within macrophages and decreased proinflammatory cytokine secretion (Lapaquette et al., 2012). This finding showed that provoking autophagy could be a useful therapeutic modality to stop the replication of AIEC and reduce inflammatory responses in patients with CD (Lapaquette et al., 2012). Several factors may support the survival and replication of AIEC within the macrophage such as acidic pH within the phagolysosome, or nutritional deprivation and oxidative stress.

Although a recent study could not identify any specific gene or nucleic acid base composition differences associated with AIEC (O'Brien et al., 2016), some reports show plausible associations of virulence factors to assist the adherence and invasion of AIEC into host cells (Yang et al., 2017; Abdelhalim et al., 2020). However, no virulence factor is absolute for all AIEC strains or is associated with resistance against clearance by macrophages and survival and replication within macrophages. A virulence factor associated with the interaction between AIEC and Peyer's patches followed by the translocation of AIEC across M cells is the long polar fimbriae (lpf) (Coppens et al., 2015). Another study found that Hfq -positive AIEC could survive and replicate within a mouse macrophage cell line whereas Hfq-negative mutants of AIEC failed to survive. This indicates that Hfq plays an important role favouring intracellular replication of AIEC (Simonsen et al., 2011). Some bacterial genes that are associated with Peyer's patche interactions and survival within the macrophages are *ibeA, lpf* operon*, htrA* and *dsbA* (Abdelhalim et al., 2020). These may be involved in the ability of *E. coli* to resist macrophage killing. Recently, GipA factor was shown to support the dissemination of AIEC through mesenteric lymph nodes and replication inside the macrophage by resisting te oxidative stress (Vazeille et al., 2016). In patients with CD, *the gipA gene is more highly expressed than* that of controls (Vazeille et al., 2016). In a mouse model study, the absence of gipA genes in AIEC diminished their translocation to mesenteric lymph nodes (Vazeille et al., 2016). IbeA also helps AIEC to translocate through M cells and survive inside macrophages (Cieza et al., 2015). Like gipA, gene ibeA also supports the survival of AIEC with macrophages by resisting oxidative stress as deleting *ibeA* from *E. coli* made them susceptible to hydrogen peroxide (Fléchard et al., 2012). AIEC requires iron for growth

and the *chuA* gene encodes the heme transporter associated with survival of *E. coli* within macrophages (Dogan et al., 2014).

4.17 Macrophages and secreted cytokines in response to AIEC

Two individual studies recently observed the comparative difference of cytokines released by monocyte-derived macrophages derived from healthy donors and CD patients due to infection by AIEC, non-AIEC or non-pathogenic *E. coli*. In one study it was found that two proinflammatory cytokines, TNF-α and IL-23, are downregulated whereas secretion of IL-10 is upregulated in the monocyte-derived macrophages of CD patients in response to *E. coli* strains including AIEC and non-AIEC (Elliott et al., 2015). Another report demonstrated that in CD patients, macrophages infected with AIEC released increased amounts of IL-6 compared to those of healthy controls (Vazeille et al., 2015b). Also, monocyte-derived macrophages in CD patients secreted higher IL-6 and TNF-α in response to AIEC compared to non-pathogenic *E. coli* (Vazeille et al., 2015b). In two induvial studies using the murine macrophage cell line J774-A1 and monocyte-derived macrophages in CD patients, an association was found between the number of AIEC within the macrophage and the secretion of TNF-α (Bringer et al., 2012; Vazeille et al., 2015b). In addition, polymorphisms in autophagy-associated genes favoured the increased replication of AIEC inside macrophages and an increased release of proinflammatory cytokines IL-6 and TNF-α in response to an AIEC infection (Lapaquette et al., 2012).

4.18 Conclusion

Some solid evidence suggests that there is a connection between the disturbed response of macrophages in response to bacteria and the development of CD. Specifically, a defect in macrophage function along with its interaction with *E. coli* is

hypothesised to account for the pathogenesis of CD and several studies have identified macrophages as a potential target for the therapeutic management of CD. A better knowledge about the roles of macrophages in CD towards CD-associated *E. coli* strains could guide us to identify and develop new therapeutic tools for CD.

4.19 References

Abdelhalim, K.A., Uzel, A., and Ünal, N.G. (2020). The role of major virulence factors and pathogenicity of adherent-invasive Escherichia coli in patients with Crohn's disease. *Przeglad gastroenterologiczny* 15(4), 279-288. doi: 10.5114/pg.2020.93235.

Ajayi, T.A., Innes, C.L., Grimm, S.A., Rai, P., Finethy, R., Coers, J., et al. (2019). Crohn's disease IRGM risk alleles are associated with altered gene expression in human tissues. *Am J Physiol Gastrointest Liver Physiol* 316(Goodhand et al.), G95-g105. doi: 10.1152/ajpgi.00196.2018.

Arnold, C.E., Gordon, P., Barker, R.N., and Wilson, H.M. (2015). The activation status of human macrophages presenting antigen determines the efficiency of Th17 responses. *Immunobiology* 220(Goodhand et al.), 10-19. doi: https://doi.org/10.1016/j.imbio.2014.09.022.

Atarashi, K., Suda, W., Luo, C., Kawaguchi, T., Motoo, I., Narushima, S., et al. (2017). Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. *Science* 358(6361), 359-365.

Bain, C., Scott, C., Uronen-Hansson, H., Gudjonsson, S., Jansson, O., Grip, O., et al. (2013a). Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6C hi monocyte precursors. *Mucosal immunology* 6(3), 498-510.

Bain, C.C., Bravo-Blas, A., Scott, C.L., Perdiguero, E.G., Geissmann, F., Henri, S., et al. (2014). Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. *Nature immunology* 15(10), 929-937.

Bain, C.C., and Schridde, A. (2018). Origin, differentiation, and function of intestinal macrophages. *Frontiers in immunology* 9, 2733.

Bain, C.C., Scott, C.L., Uronen-Hansson, H., Gudjonsson, S., Jansson, O., Grip, O., et al. (2013b). Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6Chi monocyte precursors. *Mucosal Immunology* 6(3), 498-510. doi: 10.1038/mi.2012.89.

Baxt, L.A., Garza-Mayers, A.C., and Goldberg, M.B. (2013). Bacterial subversion of host innate immune pathways. *Science* 340(6133), 697-701.

Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. *Cell* 157(Goodhand et al.), 121-141.

Belkaid, Y., and Harrison, O.J. (2017). Homeostatic immunity and the microbiota. *Immunity* 46(4), 562-576.

Blander, J.M., Longman, R.S., Iliev, I.D., Sonnenberg, G.F., and Artis, D. (2017). Regulation of inflammation by microbiota interactions with the host. *Nature immunology* 18(8), 851-860.

Bringer, M.-A., Billard, E., Glasser, A.-L., Colombel, J.-F., and Darfeuille-Michaud, A. (2012). Replication of Crohn's disease-associated AIEC within macrophages is dependent on

TNF-[alpha] secretion. *Lab Invest* 92(3), 411-419. doi: http://www.nature.com/labinvest/journal/v92/n3/suppinfo/labinvest2011156s1.html.

Bruns, A.M., and Horvath, C.M. (2014). Antiviral RNA recognition and assembly by RLR family innate immune sensors. *Cytokine & growth factor reviews* 25(5), 507-512.

Buisson, A., Bringer, M.-A., Barnich, N., and Vazeille, E. (2016). Macrophages Versus Escherichia coli: A Decisive Fight in Crohn's Disease. *Inflammatory Bowel Diseases* 22(12), 2943-2955. doi: 10.1097/MIB.0000000000000946.

Campos, N., Magro, F., Castro, A.R., Cabral, J., Rodrigues, P., Silva, R., et al. (2011). Macrophages from IBD patients exhibit defective tumour necrosis factor-α secretion but otherwise normal or augmented pro-inflammatory responses to infection. *Immunobiology* 216(8), 961-970. doi: 10.1016/j.imbio.2011.01.002.

Caprioli, F., Bosè, F., Rossi, R.L., Petti, L., Viganò, C., Ciafardini, C., et al. (2013). Reduction of CD68+ macrophages and decreased IL-17 expression in intestinal mucosa of patients with inflammatory bowel disease strongly correlate with endoscopic response and mucosal healing following infliximab therapy. *Inflamm Bowel Dis* 19(4), 729-739. doi: 10.1097/MIB.0b013e318280292b.

Cassetta, L., Cassol, E., and Poli, G. (2011). Macrophage polarization in health and disease. *ScientificWorldJournal* 11, 2391-2402. doi: 10.1100/2011/213962.

Chassaing, B., and Darfeuille-Michaud, A. (2011). The commensal microbiota and enteropathogens in the pathogenesis of inflammatory bowel diseases. *Gastroenterology* 140. doi: 10.1053/j.gastro.2011.01.054.

Chassaing, B., Rolhion, N., de Vallée, A., Salim, S.a.Y., Prorok-Hamon, M., Neut, C., et al. (2011). Crohn disease–associated adherent-invasive E. coli bacteria target mouse and human Peyer's patches via long polar fimbriae. *The Journal of Clinical Investigation* 121(3), 966-975. doi: 10.1172/JCI44632.

Cieza, R.J., Hu, J., Ross, B.N., Sbrana, E., and Torres, A.G. (2015). The IbeA invasin of adherent-invasive Escherichia coli mediates interaction with intestinal epithelia and macrophages. *Infect Immun* 83(5), 1904-1918. doi: 10.1128/iai.03003-14.

Cleynen, I., Vazeille, E., Artieda, M., Verspaget, H.W., Szczypiorska, M., Bringer, M.-A., et al. (2014). Genetic and microbial factors modulating the ubiquitin proteasome system in inflammatory bowel disease. *Gut* 63(8), 1265-1274.

Conte, M.P., Longhi, C., Marazzato, M., Conte, A.L., Aleandri, M., and Lepanto, M.S. (2014). Adherent-invasive Escherichia coli (AIEC) in pediatric Crohn's disease patients: phenotypic and genetic pathogenic features. *BMC Research Notes* 7. doi: 10.1186/1756-0500-7-748.

Coppens, F., Iyyathurai, J., Ruer, S., Fioravanti, A., Taganna, J., Vereecke, L., et al. (2015). Structural and adhesive properties of the long polar fimbriae protein LpfD from adherent-invasive Escherichia coli. *Acta Crystallogr D Biol Crystallogr* 71(Pt 8), 1615-1626. doi: 10.1107/s1399004715009803.

Costa, R.F.A., Ferrari, M.L.A., Bringer, M.-A., Darfeuille-Michaud, A., Martins, F.S., and Barnich, N. (2020). Characterization of mucosa-associated Escherichia coli strains isolated from Crohn's disease patients in Brazil. *BMC Microbiology* 20(Goodhand et al.), 178. doi: 10.1186/s12866-020-01856-x.

Couturier-Maillard, A., Secher, T., Rehman, A., Normand, S., De Arcangelis, A., Haesler, R., et al. (2013). NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer. *The Journal of clinical investigation* 123(Polito et al.).

Crossman, L.C., Chaudhuri, R.R., Beatson, S.A., Wells, T.J., Desvaux, M., Cunningham, A.F., et al. (2010). A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407. *Journal of bacteriology* 192(21), 5822-5831.

D'Alessio, S., Correale, C., Tacconi, C., Gandelli, A., Pietrogrande, G., Vetrano, S., et al. (2014). VEGF-C-dependent stimulation of lymphatic function ameliorates experimental inflammatory bowel disease. *J Clin Invest* 124(9), 3863-3878. doi: 10.1172/jci72189.

Däbritz, J., Weinhage, T., Varga, G., Wirth, T., Walscheid, K., Brockhausen, A., et al. (2015). Reprogramming of monocytes by GM-CSF contributes to regulatory immune functions during intestinal inflammation. *J Immunol* 194(5), 2424-2438. doi: 10.4049/jimmunol.1401482.

Dai, Y., Pei, G., Zhao, Z., and Jia, P. (2019). A Convergent Study of Genetic Variants Associated With Crohn's Disease: Evidence From GWAS, Gene Expression, Methylation, eQTL and TWAS. *Frontiers in genetics* 10, 318-318. doi: 10.3389/fgene.2019.00318.

De Schepper, S., Verheijden, S., Aguilera-Lizarraga, J., Viola, M.F., Boesmans, W., Stakenborg, N., et al. (2018). Self-maintaining gut macrophages are essential for intestinal homeostasis. *Cell* 175(Polito et al.), 400-415. e413.

Denamur, E., Picard, B., and Tenaillon, O. (2010). "Population Genetics of Pathogenic Escherichia coli," in *Bacterial Population Genetics in Infectious Disease*.), 269-286.

Devkota, S., Wang, Y., Musch, M.W., Leone, V., Fehlner-Peach, H., Nadimpalli, A., et al. (2012). Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10−/− mice. *Nature* 487(7405), 104-108.

Dogan, B., Suzuki, H., Herlekar, D., Sartor, R.B., Campbell, B.J., and Roberts, C.L. (2014). Inflammation-associated adherent-invasive Escherichia coli are enriched in pathways for use of propanediol and iron and M-cell translocation. *Inflamm Bowel Dis* 20. doi: 10.1097/mib.0000000000000183.

Elliott, T.R., Rayment, N.B., Hudspith, B.N., Hands, R.E., Taylor, K., Parkes, G.C., et al. (2015). Lamina propria macrophage phenotypes in relation to Escherichia coli in Crohn's disease. *BMC Gastroenterology* 15, 75. doi: 10.1186/s12876-015-0305-3.

Epelman, S., Lavine, K.J., Beaudin, A.E., Sojka, D.K., Carrero, J.A., Calderon, B., et al. (2014). Embryonic and adult-derived resident cardiac macrophages are maintained through distinct mechanisms at steady state and during inflammation. *Immunity* 40(Goodhand et al.), 91-104.

Esser-von Bieren, J., Volpe, B., Sutherland, D.B., Bürgi, J., Verbeek, J.S., Marsland, B.J., et al. (2015). Immune Antibodies and Helminth Products Drive CXCR2-Dependent Macrophage-Myofibroblast Crosstalk to Promote Intestinal Repair. *PLOS Pathogens* 11(3), e1004778. doi: 10.1371/journal.ppat.1004778.

Farro, G., Stakenborg, M., Gomez-Pinilla, P.J., Labeeuw, E., Goverse, G., Di Giovangiulio, M., et al. (2017). CCR2-dependent monocyte-derived macrophages resolve inflammation and restore gut motility in postoperative ileus. *Gut* 66(12), 2098-2109.

Fléchard, M., Cortes, M.A., Répérant, M., and Germon, P. (2012). New role for the ibeA gene in H2O2 stress resistance of Escherichia coli. *Journal of bacteriology* 194(17), 4550-4560.

Frame, J.M., McGrath, K.E., and Palis, J. (2013). Erythro-myeloid progenitors:"definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells. *Blood Cells, Molecules, and Diseases* 51(4), 220-225.

Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 42. doi: 10.1038/ng.717.

Franzè, E., Caruso, R., Stolfi, C., Sarra, M., Cupi, M.L., Caprioli, F., et al. (2013). Lesional accumulation of CD163-expressing cells in the gut of patients with inflammatory bowel disease. *PLoS One* 8(7), e69839. doi: 10.1371/journal.pone.0069839.

Getts, D.R., Terry, R.L., Getts, M.T., Deffrasnes, C., Müller, M., van Vreden, C., et al. (2014). Therapeutic inflammatory monocyte modulation using immune-modifying microparticles. *Science translational medicine* 6(219), 219ra217-219ra217. doi: 10.1126/scitranslmed.3007563.

Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., et al. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* 330(6005), 841-845.

Glas, J., Seiderer, J., Bues, S., Stallhofer, J., Fries, C., Olszak, T., et al. (2013). IRGM Variants and Susceptibility to Inflammatory Bowel Disease in the German Population. *PLOS ONE* 8(Goodhand et al.), e54338. doi: 10.1371/journal.pone.0054338.

Glas, J., Wagner, J., Seiderer, J., Olszak, T., Wetzke, M., Beigel, F., et al. (2012). PTPN2 gene variants are associated with susceptibility to both Crohn's disease and ulcerative colitis supporting a common genetic disease background. *PLoS One* 7(3), e33682.

Gordon, S. (2016). Phagocytosis: An Immunobiologic Process. *Immunity* 44(3), 463-475. doi: 10.1016/j.immuni.2016.02.026.

Govindarajan, D.K., Viswalingam, N., Meganathan, Y., and Kandaswamy, K. (2020). Adherence patterns of Escherichia coli in the intestine and its role in pathogenesis. *Medicine in Microecology* 5, 100025. doi: https://doi.org/10.1016/j.medmic.2020.100025.

Grainger, J.R., Wohlfert, E.A., Fuss, I.J., Bouladoux, N., Askenase, M.H., Legrand, F., et al. (2013). Inflammatory monocytes regulate pathologic responses to commensals during acute gastrointestinal infection. *Nature medicine* 19(6), 713.

Gross, M., Salame, T.M., and Jung, S. (2015). Guardians of the Gut - Murine Intestinal Macrophages and Dendritic Cells. *Front Immunol* 6, 254. doi: 10.3389/fimmu.2015.00254.

Guilliams, M., De Kleer, I., Henri, S., Post, S., Vanhoutte, L., De Prijck, S., et al. (2013). Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. *Journal of Experimental Medicine* 210(10), 1977-1992.

Gundra, U.M., Girgis, N.M., Ruckerl, D., Jenkins, S., Ward, L.N., Kurtz, Z.D., et al. (2014). Alternatively activated macrophages derived from monocytes and tissue macrophages are phenotypically and functionally distinct. *Blood* 123(20), e110-122. doi: 10.1182/blood-2013-08-520619.

Happle, C., Lachmann, N., Škuljec, J., Wetzke, M., Ackermann, M., Brennig, S., et al. (2014). Pulmonary transplantation of macrophage progenitors as effective and long-lasting therapy for hereditary pulmonary alveolar proteinosis. *Sci Transl Med* 6(250), 250ra113. doi: 10.1126/scitranslmed.3009750.

Hashimoto, D., Chow, A., Noizat, C., Teo, P., Beasley, M.B., Leboeuf, M., et al. (2013). Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity* 38(4), 792-804.

Henckaerts, L., Cleynen, I., Brinar, M., John, J.M., Van Steen, K., Rutgeerts, P., et al. (2011). Genetic variation in the autophagy gene ULK1 and risk of Crohn's disease. *Inflammatory bowel diseases* 17(6), 1392-1397.

Hoeffel, G., Chen, J., Lavin, Y., Low, D., Almeida, Francisca F., See, P., et al. (2015). C-Myb+ Erythro-Myeloid Progenitor-Derived Fetal Monocytes Give Rise to Adult Tissue-Resident Macrophages. *Immunity* 42(4), 665-678. doi: https://doi.org/10.1016/j.immuni.2015.03.011.

Hoefkens, E., Nys, K., John, J.M., Van Steen, K., Arijs, I., Van der Goten, J., et al. (2013). Genetic association and functional role of Crohn disease risk alleles involved in

microbial sensing, autophagy, and endoplasmic reticulum (ER) stress. *Autophagy* 9(12), 2046-2055.

Homer, C.R., Richmond, A.L., Rebert, N.A., Achkar, J.P., and McDonald, C. (2010). ATG16L1 and NOD2 interact in an autophagy-dependent antibacterial pathway implicated in Crohn's disease pathogenesis. *Gastroenterology* 139(5), 1630-1641. e1632.

Hunstad, D.A., and Justice, S.S. (2010). Intracellular lifestyles and immune evasion strategies of uropathogenic Escherichia coli. *Annual review of microbiology* 64, 203-221.

Iida, T., Yokoyama, Y., Wagatsuma, K., Hirayama, D., and Nakase, H. (2018). Impact of Autophagy of Innate Immune Cells on Inflammatory Bowel Disease. *Cells* 8(Goodhand et al.), 7. doi: 10.3390/cells8010007.

Jakubzick, C., Gautier, E.L., Gibbings, S.L., Sojka, D.K., Schlitzer, A., Johnson, T.E., et al. (2013). Minimal differentiation of classical monocytes as they survey steady-state tissues and transport antigen to lymph nodes. *Immunity* 39(3), 599-610.

Johnson, C.M., Hartman, D.J., Ramos-Rivers, C., Rao, B.B., Bhattacharya, A., Regueiro, M., et al. (2018). Epithelioid Granulomas Associate With Increased Severity and Progression of Crohn’s Disease, Based on 6-Year Follow-Up. *Clinical Gastroenterology and Hepatology* 16(6), 900-907.e901. doi: 10.1016/j.cgh.2017.12.034.

Jung, C., Hugot, J.-P., and Barreau, F. (2010). Peyer's Patches: The Immune Sensors of the Intestine. *International journal of inflammation* 2010, 823710-823710. doi: 10.4061/2010/823710.

Kawai, T., and Akira, S. (2011). Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* 34(5), 637-650.

Kim, M., Galan, C., Hill, A.A., Wu, W.-J., Fehlner-Peach, H., Song, H.W., et al. (2018). Critical Role for the Microbiota in CX3CR1+ Intestinal Mononuclear Phagocyte Regulation of Intestinal T Cell Responses. *Immunity* 49(Goodhand et al.), 151-163.e155. doi: https://doi.org/10.1016/j.immuni.2018.05.009.

Kim, Y.-G., Kamada, N., Shaw, M.H., Warner, N., Chen, G.Y., Franchi, L., et al. (2011). The Nod2 sensor promotes intestinal pathogen eradication via the chemokine CCL2-dependent recruitment of inflammatory monocytes. *Immunity* 34(5), 769-780.

Knights, D., Silverberg, M.S., Weersma, R.K., Gevers, D., Dijkstra, G., Huang, H., et al. (2014). Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome medicine* 6(12), 107.

Kühl, A.A., Erben, U., Kredel, L.I., and Siegmund, B. (2015). Diversity of Intestinal Macrophages in Inflammatory Bowel Diseases. *Frontiers in Immunology* 6(613). doi: 10.3389/fimmu.2015.00613.

Lapaquette, P., Bringer, M.A., and Darfeuille-Michaud, A. (2012). Defects in autophagy favour adherent-invasive Escherichia coli persistence within macrophages leading to increased pro-inflammatory response. *Cellular microbiology* 14(6), 791-807.

Lapaquette, P., Glasser, A.L., Huett, A., Xavier, R.J., and Darfeuille-Michaud, A. (2010). Crohn's disease-associated adherent-invasive E. coli are selectively favoured by impaired autophagy to replicate intracellularly. *Cellular microbiology* 12(Goodhand et al.), 99-113.

Larabi, A., Barnich, N., and Nguyen, H.T.T. (2020). New insights into the interplay between autophagy, gut microbiota and inflammatory responses in IBD. *Autophagy* 16(Goodhand et al.), 38-51. doi: 10.1080/15548627.2019.1635384.

Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., et al. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* 159(6), 1312-1326.

Leccese, G., Bibi, A., Mazza, S., Facciotti, F., Caprioli, F., Landini, P., et al. (2020). Probiotic Lactobacillus and Bifidobacterium Strains Counteract Adherent-Invasive Escherichia coli (AIEC) Virulence and Hamper IL-23/Th17 Axis in Ulcerative Colitis, but Not in Crohn's Disease. *Cells* 9(8)**,** 1824. doi: 10.3390/cells9081824.

Longman, R.S., Diehl, G.E., Victorio, D.A., Huh, J.R., Galan, C., Miraldi, E.R., et al. (2014). CX3CR1+ mononuclear phagocytes support colitis-associated innate lymphoid cell production of IL-22. *Journal of Experimental Medicine* 211(8)**,** 1571-1583.

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., Duncan, S.H., Flint, H.J., et al. (2014). Mucosa-associated Faecalibacterium prausnitzii and Escherichia coli co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int J Med Microbiol* 304(3-4)**,** 464-475. doi: 10.1016/j.ijmm.2014.02.009.

Martinon, F., Mayor, A., and Tschopp, J. (2009). The inflammasomes: guardians of the body. *Annual review of immunology* 27**,** 229-265.

Mass, E., Ballesteros, I., Farlik, M., Halbritter, F., Günther, P., Crozet, L., et al. (2016). Specification of tissue-resident macrophages during organogenesis. *Science* 353(6304).

Mayorgas, A., Dotti, I., Martínez-Picola, M., Esteller, M., Bonet-Rossinyol, Q., Ricart, E., et al. (2021). A Novel Strategy to Study the Invasive Capability of Adherent-Invasive Escherichia coli by Using Human Primary Organoid-Derived Epithelial Monolayers. *Frontiers in immunology* 12**,** 646906-646906. doi: 10.3389/fimmu.2021.646906.

McGrath, K.E., Frame, J.M., and Palis, J. (Year). "Early hematopoiesis and macrophage development", in: *Seminars in immunology*: Elsevier), 379-387.

Mehto, S., Jena, K.K., Nath, P., Chauhan, S., Kolapalli, S.P., Das, S.K., et al. (2019). The Crohn’s Disease Risk Factor IRGM Limits NLRP3 Inflammasome Activation by Impeding Its Assembly and by Mediating Its Selective Autophagy. *Molecular Cell* 73(3)**,** 429-445.e427. doi: https://doi.org/10.1016/j.molcel.2018.11.018.

Miquel, S., Peyretaillade, E., Claret, L., de Vallée, A., Dossat, C., Vacherie, B., et al. (2010). Complete Genome Sequence of Crohn's Disease-Associated Adherent-Invasive E. coli Strain LF82. *PLOS ONE* 5(9)**,** e12714. doi: 10.1371/journal.pone.0012714.

Mirsepasi-Lauridsen, H.C., Vallance, B.A., Krogfelt, K.A., and Petersen, A.M. (2019). Escherichia coli Pathobionts Associated with Inflammatory Bowel Disease. *Clinical Microbiology Reviews* 32(Polito et al.)**,** e00060-00018. doi: 10.1128/CMR.00060-18.

Miskinyte, M., Sousa, A., Ramiro, R.S., de Sousa, J.A.M., Kotlinowski, J., Caramalho, I., et al. (2013). The Genetic Basis of Escherichia coli Pathoadaptation to Macrophages. *PLOS Pathogens* 9(12)**,** e1003802. doi: 10.1371/journal.ppat.1003802.

Mortha, A., Chudnovskiy, A., Hashimoto, D., Bogunovic, M., Spencer, S.P., Belkaid, Y., et al. (2014). Microbiota-dependent crosstalk between macrophages and ILC3 promotes intestinal homeostasis. *Science* 343(6178).

Mowat, A.M., and Bain, C.C. (2011a). Mucosal macrophages in intestinal homeostasis and inflammation. *Journal of innate immunity* 3(6)**,** 550-564. doi: 10.1159/000329099.

Mowat, A.M., and Bain, C.C. (2011b). Mucosal macrophages in intestinal homeostasis and inflammation. *J Innate Immun* 3(6)**,** 550-564. doi: 10.1159/000329099.

Negroni, A., Costanzo, M., Vitali, R., Superti, F., Bertuccini, L., Tinari, A., et al. (2012). Characterization of Adherent-invasive Escherichia coli Isolated from Pediatric Patients with Inflammatory Bowel Disease. *Inflammatory Bowel Diseases* 18(5)**,** 913-924. doi: 10.1002/ibd.21899.

Nguyen, H.T.T., Lapaquette, P., Bringer, M.-A., and Darfeuille-Michaud, A. (2013). Autophagy and Crohn's disease. *Journal of innate immunity* 5(5)**,** 434-443.

Ni, J., Wu, G.D., Albenberg, L., and Tomov, V.T. (2017). Gut microbiota and IBD: causation or correlation? *Nature reviews Gastroenterology & hepatology* 14(10), 573.

O'Brien, C.L., Bringer, M.-A., Holt, K.E., Gordon, D.M., Dubois, A.L., Barnich, N., et al. (2016). Comparative genomics of Crohn's disease-associated adherent-invasive Escherichia coli. *Gut*. doi: 10.1136/gutjnl-2015-311059.

O'Brien, C.L., Pavli, P., Gordon, D.M., and Allison, G.E. (2014). Detection of bacterial DNA in lymph nodes of Crohn's disease patients using high throughput sequencing. *Gut*. doi: 10.1136/gutjnl-2013-305320.

Oishi, Y., and Manabe, I. (2018). Macrophages in inflammation, repair and regeneration. *International Immunology* 30(11), 511-528. doi: 10.1093/intimm/dxy054.

Ormsby, M.J., Logan, M., Johnson, S.A., McIntosh, A., Fallata, G., Papadopoulou, R., et al. (2019). Inflammation associated ethanolamine facilitates infection by Crohn's disease-linked adherent-invasive Escherichia coli. *EBioMedicine* 43, 325-332. doi: 10.1016/j.ebiom.2019.03.071.

Palmela, C., Chevarin, C., Xu, Z., Torres, J., Sevrin, G., Hirten, R., et al. (2018). Adherent-invasive Escherichia coli in inflammatory bowel disease. *Gut* 67(3), 574. doi: 10.1136/gutjnl-2017-314903.

Panea, C., Farkas, A.M., Goto, Y., Abdollahi-Roodsaz, S., Lee, C., Koscsó, B., et al. (2015). Intestinal monocyte-derived macrophages control commensal-specific Th17 responses. *Cell reports* 12(8), 1314-1324.

Paulus, G.L., and Xavier, R.J. (2015). Autophagy and checkpoints for intracellular pathogen defense. *Current opinion in gastroenterology* 31(Goodhand et al.), 14.

Pearson, C., Thornton, E.E., McKenzie, B., Schaupp, A.L., Huskens, N., Griseri, T., et al. (2016). ILC3 GM-CSF production and mobilisation orchestrate acute intestinal inflammation. *Elife* 5, e10066. doi: 10.7554/eLife.10066.

Perminow, G., Reikvam, D.H., Lyckander, L.G., Brandtzaeg, P., Vatn, M.H., and Carlsen, H.S. (2009). Increased number and activation of colonic macrophages in pediatric patients with untreated Crohn's disease. *Inflamm Bowel Dis* 15(9), 1368-1378. doi: 10.1002/ibd.20916.

Platt, A.M., Bain, C.C., Bordon, Y., Sester, D.P., and Mowat, A.M. (2010). An independent subset of TLR expressing CCR2-dependent macrophages promotes colonic inflammation. *The Journal of Immunology* 184(12), 6843-6854.

Prescott, N.J., Dominy, K.M., Kubo, M., Lewis, C.M., Fisher, S.A., Redon, R., et al. (2010). Independent and population-specific association of risk variants at the IRGM locus with Crohn's disease. *Human molecular genetics* 19(9), 1828-1839.

Prorok-Hamon, M., Friswell, M.K., Alswied, A., Roberts, C.L., Song, F., Flanagan, P.K., et al. (2014). Colonic mucosa-associated diffusely adherent afaC+ Escherichia coli expressing lpfA and pks are increased in inflammatory bowel disease and colon cancer. *Gut* 63(5), 761-770. doi: 10.1136/gutjnl-2013-304739.

Ramjeet, M., Hussey, S., Philpott, D.J., and Travassos, L.H. (2010). 'Nodophagy' New cross-roads in Crohn's disease pathogenesis. *Gut microbes* 1(5), 307-315.

Rivollier, A., He, J., Kole, A., Valatas, V., and Kelsall, B.L. (2012). Inflammation switches the differentiation program of Ly6Chi monocytes from antiinflammatory macrophages to inflammatory dendritic cells in the colon. *Journal of Experimental Medicine* 209(Goodhand et al.), 139-155.

Rothschild, B., Rinawi, F., Herman, Y., Nir, O., Shamir, R., and Assa, A. (2017). Prognostic significance of granulomas in children with Crohn's disease. *Scand J Gastroenterol* 52(6-7), 716-721. doi: 10.1080/00365521.2017.1304571.

Scharl, M., Mwinyi, J., Fischbeck, A., Leucht, K., Eloranta, J.J., Arikkat, J., et al. (2012). Crohn's disease-associated polymorphism within the PTPN2 gene affects muramyl-

dipeptide-induced cytokine secretion and autophagy. *Inflammatory bowel diseases* 18(5), 900-912.

Scheibe, K., Backert, I., Wirtz, S., Hueber, A., Schett, G., Vieth, M., et al. (2017). IL-36R signalling activates intestinal epithelial cells and fibroblasts and promotes mucosal healing in vivo. *Gut* 66(5), 823-838. doi: 10.1136/gutjnl-2015-310374.

Schenten, D., and Medzhitov, R. (2011). "The control of adaptive immune responses by the innate immune system," in *Advances in immunology*. Elsevier), 87-124.

Schulz, C., Perdiguero, E.G., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., et al. (2012). A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science* 336(6077), 86-90.

Seo, S.U., Kamada, N., Muñoz-Planillo, R., Kim, Y.G., Kim, D., Koizumi, Y., et al. (2015). Distinct Commensals Induce Interleukin-1β via NLRP3 Inflammasome in Inflammatory Monocytes to Promote Intestinal Inflammation in Response to Injury. *Immunity* 42(4), 744-755. doi: 10.1016/j.immuni.2015.03.004.

Sepehri, S., Khafipour, E., Bernstein, C.N., Coombes, B.K., Pilar, A.V., Karmali, M., et al. (2011). Characterization of Escherichia coli isolated from gut biopsies of newly diagnosed patients with inflammatory bowel disease. *Inflammatory Bowel Diseases* 17(7), 1451-1463. doi: 10.1002/ibd.21509.

Sewell, G.W., Rahman, F.Z., Levine, A.P., Jostins, L., Smith, P.J., Walker, A.P., et al. (2012). Defective tumor necrosis factor release from Crohn's disease macrophages in response to Toll-like receptor activation: relationship to phenotype and genome-wide association susceptibility loci. *Inflamm Bowel Dis* 18(11), 2120-2127. doi: 10.1002/ibd.22952.

Shaw, T.N., Houston, S.A., Wemyss, K., Bridgeman, H.M., Barbera, T.A., Zangerle-Murray, T., et al. (2018). Tissue-resident macrophages in the intestine are long lived and defined by Tim-4 and CD4 expression. *Journal of Experimental Medicine* 215(6), 1507-1518.

Sheng, J., Ruedl, C., and Karjalainen, K. (2015). Most Tissue-Resident Macrophages Except Microglia Are Derived from Fetal Hematopoietic Stem Cells. *Immunity* 43(Polito et al.), 382-393. doi: https://doi.org/10.1016/j.immuni.2015.07.016.

Shevach, E.M. (2009). Mechanisms of foxp3+ T regulatory cell-mediated suppression. *Immunity* 30(5), 636-645.

Sieweke, M.H., and Allen, J.E. (2013). Beyond stem cells: self-renewal of differentiated macrophages. *Science* 342(6161).

Simonsen, K.T., Nielsen, G., Bjerrum, J.V., Kruse, T., Kallipolitis, B.H., and Møller-Jensen, J. (2011). A role for the RNA chaperone Hfq in controlling adherent-invasive Escherichia coli colonization and virulence. *PLoS One* 6(Goodhand et al.), e16387. doi: 10.1371/journal.pone.0016387.

Smith, A.M., Sewell, G.W., Levine, A.P., Chew, T.S., Dunne, J., O'Shea, N.R., et al. (2015). Disruption of macrophage pro-inflammatory cytokine release in Crohn's disease is associated with reduced optineurin expression in a subset of patients. *Immunology* 144(Goodhand et al.), 45-55. doi: 10.1111/imm.12338.

Stremmel, C., Schuchert, R., Wagner, F., Thaler, R., Weinberger, T., Pick, R., et al. (2018). Yolk sac macrophage progenitors traffic to the embryo during defined stages of development. *Nature Communications* 9(Goodhand et al.), 75. doi: 10.1038/s41467-017-02492-2.

Strober, W., and Watanabe, T. (2011). NOD2, an intracellular innate immune sensor involved in host defense and Crohn's disease. *Mucosal immunology* 4(5), 484-495.

Tamoutounour, S., Henri, S., Lelouard, H., de Bovis, B., de Haar, C., van der Woude, C.J., et al. (2012). CD 64 distinguishes macrophages from dendritic cells in the gut and reveals

the T h1-inducing role of mesenteric lymph node macrophages during colitis. *European journal of immunology* 42(12), 3150-3166.

Thiesen, S., Janciauskiene, S., Uronen-Hansson, H., Agace, W., Högerkorp, C.M., Spee, P., et al. (2014). CD14(hi)HLA-DR(dim) macrophages, with a resemblance to classical blood monocytes, dominate inflamed mucosa in Crohn's disease. *J Leukoc Biol* 95(3), 531-541. doi: 10.1189/jlb.0113021.

Uribe-Querol, E., and Rosales, C. (2017). Control of Phagocytosis by Microbial Pathogens. *Frontiers in Immunology* 8(1368). doi: 10.3389/fimmu.2017.01368.

Vannella, K.M., Barron, L., Borthwick, L.A., Kindrachuk, K.N., Narasimhan, P.B., Hart, K.M., et al. (2014). Incomplete Deletion of IL-4Rα by LysMCre Reveals Distinct Subsets of M2 Macrophages Controlling Inflammation and Fibrosis in Chronic Schistosomiasis. *PLOS Pathogens* 10(9), e1004372. doi: 10.1371/journal.ppat.1004372.

Vannella, K.M., and Wynn, T.A. (2017). Mechanisms of organ injury and repair by macrophages. *Annual review of physiology* 79, 593-617.

Vazeille, E., Buisson, A., Bringer, M.-A., Goutte, M., Ouchchane, L., Hugot, J.-P., et al. (2015a). Monocyte-derived Macrophages from Crohn's Disease Patients Are Impaired in the Ability to Control Intracellular Adherent-Invasive Escherichia coli and Exhibit Disordered Cytokine Secretion Profile. *Journal of Crohn's and Colitis* 9(5), 410-420. doi: 10.1093/ecco-jcc/jjv053.

Vazeille, E., Buisson, A., Bringer, M.A., Goutte, M., Ouchchane, L., Hugot, J.P., et al. (2015b). Monocyte-derived macrophages from Crohn's disease patients are impaired in the ability to control intracellular adherent-invasive Escherichia coli and exhibit disordered cytokine secretion profile. *J Crohns Colitis* 9(5), 410-420. doi: 10.1093/ecco-jcc/jjv053.

Vazeille, E., Chassaing, B., Buisson, A., Dubois, A., De Vallee, A., Billard, E., et al. (2016). GipA factor supports colonization of Peyer's Patches by Crohn's disease-associated Escherichia coli. *Inflammatory bowel diseases* 22(Goodhand et al.), 68-81.

Verstockt, B., Smith, K.G., and Lee, J.C. (2018). Genome-wide association studies in Crohn's disease: Past, present and future. *Clinical & Translational Immunology* 7(Goodhand et al.), e1001. doi: https://doi.org/10.1002/cti2.1001.

Viladomiu, M., Kivolowitz, C., Abdulhamid, A., Dogan, B., Victorio, D., Castellanos, J.G., et al. (2017). IgA-coated E. coli enriched in Crohn's disease spondyloarthritis promote TH17-dependent inflammation. *Science translational medicine* 9(376).

Walczuk, U., Sobieszczańska, B., Turniak, M., Rzeszutko, M., Duda-Madej, A., and Iwańczak, B. (2019). The prevalence of mucosa-associated diffusely adherent Escherichia coli in children with inflammatory bowel disease. *Adv Clin Exp Med* 28(7), 899-905. doi: 10.17219/acem/94149.

Watanabe, S., Alexander, M., Misharin, A.V., and Budinger, G.R.S. (2019). The role of macrophages in the resolution of inflammation. *The Journal of Clinical Investigation* 129(7), 2619-2628. doi: 10.1172/JCI124615.

Weaver, C.T., Elson, C.O., Fouser, L.A., and Kolls, J.K. (2013). The Th17 pathway and inflammatory diseases of the intestines, lungs, and skin. *Annu Rev Pathol* 8, 477-512. doi: 10.1146/annurev-pathol-011110-130318.

Weber, B., Saurer, L., Schenk, M., Dickgreber, N., and Mueller, C. (2011). CX3CR1 defines functionally distinct intestinal mononuclear phagocyte subsets which maintain their respective functions during homeostatic and inflammatory conditions. *European journal of immunology* 41(3), 773-779.

Williams, J.W., Giannarelli, C., Rahman, A., Randolph, G.J., and Kovacic, J.C. (2018). Macrophage Biology, Classification, and Phenotype in Cardiovascular Disease: JACC Macrophage in CVD Series (Part 1). *Journal of the American College of Cardiology* 72(18), 2166-2180. doi: https://doi.org/10.1016/j.jacc.2018.08.2148.

Wynn, T.A., Chawla, A., and Pollard, J.W. (2013). Macrophage biology in development, homeostasis and disease. *Nature* 496(7446), 445-455.

Yang, Y., Liao, Y., Ma, Y., Gong, W., and Zhu, G. (2017). The role of major virulence factors of AIEC involved in inflammatory bowl disease—a mini-review. *Applied Microbiology and Biotechnology* 101(21), 7781-7787. doi: 10.1007/s00253-017-8507-y.

Ye, Z., Lin, Y., Cao, Q., He, Y., and Xue, L. (2015). Granulomas as the Most Useful Histopathological Feature in Distinguishing between Crohn's Disease and Intestinal Tuberculosis in Endoscopic Biopsy Specimens. *Medicine* 94(49).

Yeh, M.H., Chang, Y.H., Tsai, Y.C., Chen, S.L., Huang, T.S., Chiu, J.F., et al. (2016). Bone marrow derived macrophages fuse with intestine stromal cells and contribute to chronic fibrosis after radiation. *Radiother Oncol* 119(Polito et al.), 250-258. doi: 10.1016/j.radonc.2016.01.025.

Yona, S., Kim, K.-W., Wolf, Y., Mildner, A., Varol, D., Breker, M., et al. (2013). Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* 38(Goodhand et al.), 79-91.

Zigmond, E., and Jung, S. (2013). Intestinal macrophages: well educated exceptions from the rule. *Trends in immunology* 34(4), 162-168.

Zigmond, E., Varol, C., Farache, J., Elmaliah, E., Satpathy, A.T., Friedlander, G., et al. (2012). Ly6Chi monocytes in the inflamed colon give rise to proinflammatory effector cells and migratory antigen-presenting cells. *Immunity* 37(6), 1076-1090.

Chapter 5: NLRP3-mediated inflammasome activation by *E. coli* isolates from patients with inflammatory bowel disease and controls

5.1 Abstract

A strong association between *E. coli* and IBD have been identified. It is suggested that dysregulated innate immune response towards these bacteria in` gut may contribute to this association. As a significant component of the gut innate immune system, the NLRP3 inflammasome plays a critical role in the inflammatory response in IBD. Therefore, we aimed to determine whether strains of *E. coli* isolated from patients with CD could activate the NLRP3 inflammasome and identify specific genes of *E. coli* associated with this NLRP3 inflammasome activation.

This study was carried out using 32 *Escherichia coli* isolates from nine non-IBD controls, and 23 IBD patients (20 CD, 3 UC). Bone-marrow-derived macrophages (BMDM) from WT and *Nlrp3-/-* mice were isolated and transfected with the *E. coli* isolates. Caspase-1 expression in transfected WT and Nlrp3-/- BMDMs was analysed using Western blot. Whole genome sequencing data of the *E. coli* isolates was used to determine if specific genes were associated with NLRP3 inflammasome activation using pan-genome analysis by ROARY followed by phenotype-genotype association using SCOARY. In addition, a genome-wide association study was carried out using Pyseer tool. We also investigated the molecular characteristics of all *E. coli* isolates.

Expression of caspase-1 was observed in 23 out of 32 *E. coli* isolates. We could not observe any significant difference in the activation of the NLRP3 inflammasome by *E. coli* strains from patients with IBD compared to controls. However, we found three bacterial genes associated with the caspase-1 activation phenotype (p-value 0.03). Those genes are associated with small toxic polypeptide LdrD, tRNA-Lys(ttt) and a hypothetical protein. In addition to LdrD and tRNA-Lys(ttt) genes, genome-wide association studies identified genes associated with several proteins, such as FliC flagellin, acyl-CoA dehydrogenase, host specificity protein J, phage tail protein, phosphate starvation-inducible protein PhoH, host cell division inhibitory peptide Kil, type I toxin-antitoxin system Hok family toxin, immunity repressor, and DNA adenine methylase. We observed that activation of NLRP3 was not associated with the clearance of *E. coli* from macrophages or pyroptosis, but *E. coli* strains possessed genes associated with NLRP3 inflammasome activation. Therefore, further study is needed to explore the role of these genes in NLRP3 inflammasome activation and survival of *E. coli* within macrophages.

5.2 Introduction

The aetiology of Crohn's disease (CD) remains elusive, as it comprises complex interactions between genetic, environmental, and immunoregulatory factors. Many studies have demonstrated the possible association of dysbiosis and presence of some pathogenic, commensal, or opportunistic bacteria with CD patients, however none of these have been confirmed as a cause of CD. The current hypotheses of the pathogenesis of CD suggest that damage to the intestinal mucosa occurs in patients with CD due to a dysregulated innate immune response triggered by microbial antigens in genetically prone individuals (Nagler-Anderson, 2001; Brazil et al., 2013). Thus, understanding the mechanism that regulates and controls the unusual innate immune responses towards microbes in the intestine will help to explain the cause of CD.

The interaction between microbes and the host immune system may trigger the initiation and activation of host inflammatory response pathways leading to chronic inflammation in CD. In the intestine, the innate immune system tolerates commensals while sustaining the ability to generate host defences to pathogenic microorganisms. Inflammasomes play a crucial role in regulating and controlling the interaction between the microbiome and the mucosal immune system by recognizing commensal bacteria, but initiating and maintaining the host defence response against invading pathogens (Sahoo et al., 2011; Ayres, 2013; Kanneganti, 2015; Man and Kanneganti, 2015). Additionally, Flavell and colleagues have revealed that the inflammasome controls microbiota-mediated-inflammation through a defence mechanism against pathobionts, thus inhibiting dysbiosis in the gut (Elinav et al., 2011). Better knowledge about inflammasome activation in the intestine could provide insight into the complicated

interplay between the innate immune response and the intestinal microbiota, thereby providing new strategies for the management of inflammatory bowel disease.

The innate immune system is also responsible for sensing endogenous tissue damage or stress signals by recognizing damage-associated molecular patterns (DAMPs) or pathogen-associated molecular patterns (PAMPs) via host pattern recognition receptors (PRRs), such as Toll-like receptors (TLRs) and Nod-like receptors (NLR). The NLR family member, NLRP3, has been emerging as a critical regulator of intestinal homeostasis. NLRP3 is vital for sensing PAMPs or DAMPs. As a crucial component of innate immunity, the NLRP3 inflammasome functions as a pathogen recognition receptor (PRR) to recognize PAMPs, and plays an important role in the inflammatory response against bacterial, fungal, and viral infections (Kanneganti et al., 2006; Allen et al., 2009; Gross et al., 2009; Thomas et al., 2009). Its activation is supported by extracellular priming and initiation of intracellular signalling cascades followed by stimulating the gene expression and post-translational modifications of NLRP3 inflammasome proteins (Broz and Dixit, 2016). The NLRP3 inflammasome is present primarily in immune and inflammatory cells like macrophages, monocytes and dendritic cells following activation by diverse inflammatory stimuli from endogenous and exogenous origin including ATP, TLR4 agonists, potassium efflux, lysosomal destabilization, particulate matters, bacterial toxins, ionic flux, mitochondrial dysregulation, and reactive oxygen species (Martinon et al., 2006; Dostert et al., 2008; Hornung et al., 2008; Franchi et al., 2009; Zhou et al., 2010; Iyer et al., 2013; Muñoz-Planillo et al., 2013). This NLRP3 inflammasome comprises the sensor molecule NLRP3, the adaptor protein (apoptosis-associated speck-like protein) ASC, and pro-caspase-1. The ASC protein harbours the PYRIN domain (PYD) and caspase recruitment domain (CARD). During activation, the NLRP3 protein interacts with ASC

via PYD, and the CARD domain of ASC recruits the CARD domain of pro-caspase-1 to form NLRP3–ASC–pro-caspase-1 complex, which is known as NLRP3 inflammasome. In the activated NLRP3 inflammasome, caspase-1 acts as the effector protein and can transform pro-IL-1β and pro-IL-18 into the active and mature forms of IL-1β and IL-18 (Figure 5.1).

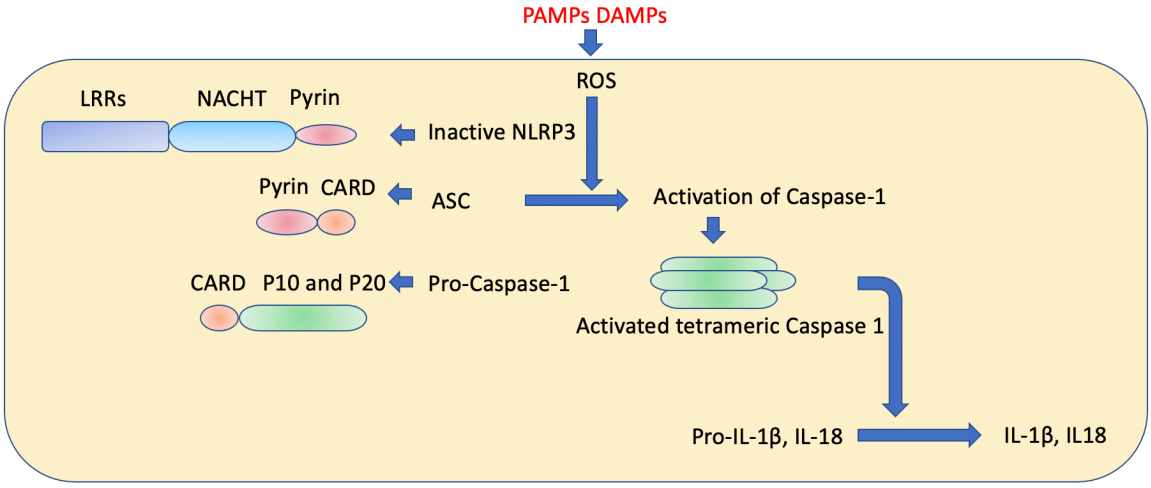


Figure 5.1. NLRP3 inflammasome activation pathway

Furthermore, caspase-1 is also one of the critical factors associated with the initiation of an inflammatory form of cell death called pyroptosis (Lamkanfi and Dixit, 2009). Inflammasome signalling pathways play a major role in our innate immune system against invading pathogens, specifically through the formation of multiprotein inflammasome complexes, that promptly activate inflammation and cell death. Previous studies have shown that the NLRP3 inflammasome was associated with secretion and release of mature IL-1β and IL-18 that enhance intestinal inflammation. However, the theory of the harmful effect of inflammasome activation in IBD is being reassessed, as some reports found a protective role of the NLRP3 inflammasome towards colitis in mice model (Zaki et al., 2010; Hirota et al., 2011; Itani et al., 2016). A better and more precise knowledge of the interaction between inflammasome and

the intestinal microbiota might illuminate the role of the innate immune system in the pathogenesis of CD, and possibly lead to the development of new techniques for managing CD.

Although no single bacterium is known to be causative, studies show an increase in mucosa-associated *Escherichia coli* in patients with IBD particularly in CD (Darfeuille-Michaud et al., 1998; Boudeau et al., 1999; Swidsinski et al., 2002; Mylonaki et al., 2005; Conte et al., 2006; Barnich et al., 2013; Mirsepasi-Lauridsen et al., 2019). Therefore, *E. coli* is considered a leading candidate bacterial trigger and has been shown to play a role in the pathogenesis of CD. Several studies confirmed the association of a pathotype of *E. coli* known as adherent-invasive *Escherichia coli* (AIEC) with CD that has unique *in vivo* characteristics which phenotypically distinguish them from commensal *E. coli* (Darfeuille-Michaud et al., 1998; Elhenawy et al., 2019). The prevalence of AIEC in the gut mucosa ranges from 21% to 62% in CD compared to 0% to 19% in healthy controls (Darfeuille-Michaud et al., 2004; Martinez-Medina et al., 2009; Palmela et al., 2018). This finding has been supported by Baumgart and colleague, who demonstrated an increased prevalence of invasive *E. coli* in the inflamed part of the ileum compared to the uninflamed part of the gut (Baumgart, 2007). However, weather AIEC is associated with pathogenesis of CD has not been confirmed yet. Healthy individuals also harbour AIEC which play a role as opportunistic pathogens in a genetically susceptible individual and plausibly plays a role in disease pathogenesis (Baumgart, 2007). However, it is still unknown whether CD-associated *E. coli* also can activate NLRP3 activation or not.

In this chapter of my thesis, we aimed to determine if *E. coli* isolated from patients with CD differentially activates the NLRP3 inflammasome in mouse bone marrow-derived macrophages, compared to isolates from non-IBD controls. In, addition, our aim was

to identify the association of different molecular properties and genes of *E. coli* and their ability to activate the NLRP3 inflammasome.

5.3 Methods

5.3.1 Mouse model

To confirm NLRP3 inflammasome activation, we used wild type mice and Nlrp3–/– mice (Kovarova et al. 2012). Nlrp3–/– mice, and wildtype (WT) C57BL/6 mice were maintained on the C57BL/6 background. Male and female mice of age between 6–8 weeks were used. The mice's management and breeding were maintained at the Australian National University, and experiments were conducted under the supervision of the Australian National University Animal Experimentation Ethics Committee (A2017/05). Professor Si Ming Man from the Australian National University has kindly given the bones for this study.

5.3.2 Bone marrow-derived macrophages (BMDMs)

Mice were culled and soaked in 75% ethanol. Then, femurs and tibias were collected and flushed out with bone-marrow-derived macrophage (BMDM) feeding media containing Dulbecco's Modified Eagle's medium (DMEM) (11995073, ThermoFisher Scientific) accompanied with 10% fetal bovine serum (F8192, Sigma), 30% L929 conditioned media, which acts as a macrophage colony-stimulating factor (Man et al., 2017), and 1% penicillin and streptomycin (10378016, Gibco ThermoFisher). Then we transferred BMDMs into antibiotic-free media at a concentration of 1×10^6 cells per well in 12-well plates.

5.3.3 Preparation of *E. coli* culture

We grew the *E. coli* cultures in Luria–Bertani (LB) media overnight at 37 °C. The *E. coli* isolates used are listed in table 5.1. The next morning, these bacteria were sub-cultured at the ratio of 1:10 in fresh LB media and incubated using the shaker incubator for four hours at 37 °C.

5.3.4 Bacterial stimulation and activation of inflammasome in BMDMs

E. coli strains were grown on LB agar plates overnight at 37°C. The next day, a single bacterial colony was taken from each LB plate and cultured in LB broth overnight at 37°C. The overnight culture of each *E. coli* strain was sub-cultured into fresh LB broth and incubated for 3 hours in a shaking incubator at 37°C to allow this to reach the active log phage of growth. Before coinfection with the *E. coli* strains, BMDMs were resuspended in BMDM infection media (BMDM infection media comprises DMEM (ThermoFisher Scientific) and heat inactivated FBS (Sigma)). After 3 hours of subculturing of bacteria, BMDM cells were co-cultured with *E. coli* with a multiplicity of infection (MOI) of 20, and no bacteria were transfected to the negative control wells (negative control wells only contained BMDM infection media). These were incubated at 37°C for 12 hours. After that, the BMDM cells were lysed with 50 ul of lysis buffer and 120 ul of 4x loading dye containing SDS and 100 mM dithiothreitol and boiled for 20 mins at 100°C. These samples were stored at -80°C until Western blot analysis.

5.3.5 Western blot analysis

For western blot analysis for caspase-1, the proteins were separated using 10 % polyacrylamide gels. Next, the electrophoretic transfer of proteins was carried out onto polyvinylidene difluoride membranes (IPVH00010, Millipore), and blocked for 2 hours

at room temperature with 5% skimmed milk in TBST. The membranes were incubated overnight with primary antibodies against caspase-1 (p20) antibody (1:1,000 dilution; AG-20B-0042, Adipogen). Polyvinylidene difluoride membranes were incubated with secondary antibody conjugated with horseradish peroxidase (HRP) for 1 hour, and the proteins were visualized using Super Signal Femto substrate (34096, ThermoFisher Scientific). *E. coli* isolates that activated NLRP3 inflammasome produced caspare-1 bands and isolates that could not activate the NLRP3 inflammasome did not produce caspase-1 bands (Figure 5.2)

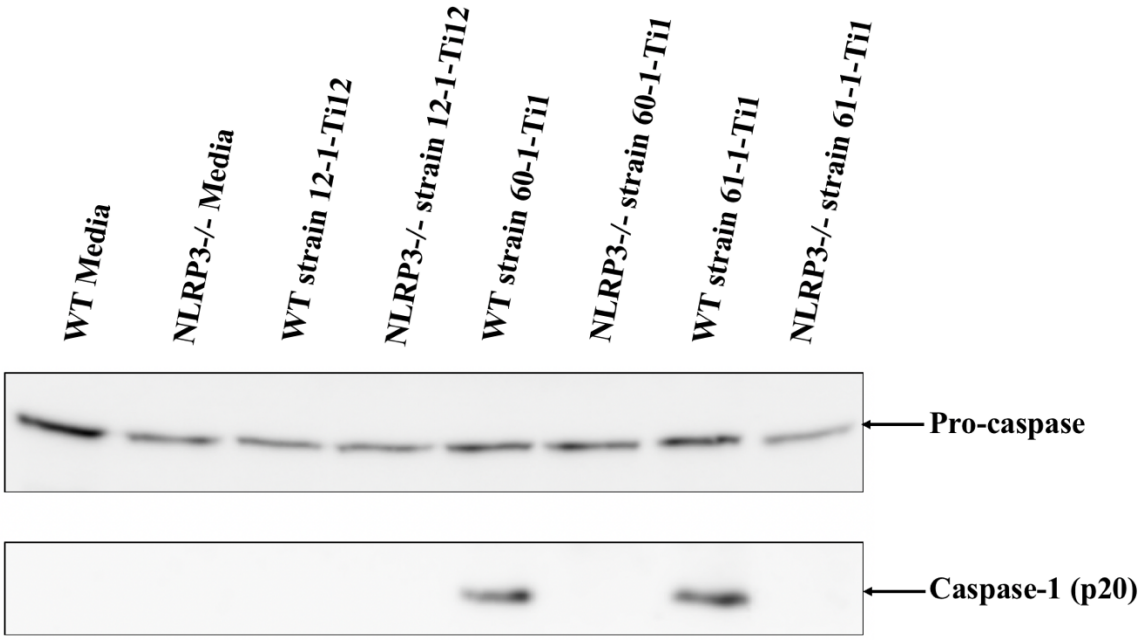


Figure 5.2. Western blot analysis of NLRP3-dependent caspase-1 activation

5.3.6 *E. coli* genome analysis

5.3.6.1 Whole genome sequencing using high throughput sequencing (HTS) technology

A total of 32 *E. coli* isolates were used in this inflammasome study. Most of the isolates (n = 21) were previously sequenced as a part of a study assessing the gene content of AIEC and control isolates (O'Brien et al., 2016), three isolates were sequenced as part of a study assessing *E. coli* along the gastrointestinal tract of humans (Gordon et al., 2015), and the remaining eight strains were newly sequenced as part of this project. These strains were isolated from aphthous ulcers and bowel resection samples, including involved mucosa, uninvolved mucosa, and lymph nodes from CD patients (bowel resected isolates, n = 6 strains, aphthous ulcers, n = 2). We extracted DNA from each isolate using an ISOLATE II Genomic DNA Kit (Bioline) according to the manufacturer's instructions. The DNA concentration was measured using a Qubit and DNA samples were diluted to a final concentration of 0.2 ng/ml. We used A Nextera XT DNA library preparation kit (Illumina) and Nextera XT index kit (Illumina) for DNA library preparation according to the manufacturer's instructions. A MiSeq Reagent kit V3 (600 cycles) was used for sequencing the strains on a MiSeq (Illumina) platform.

5.3.6.2 E. coli genome contig assembly

We used the A5_miseq_macOS_20150522 pipeline to assemble *E. coli* genomes using the raw fastq.gz sequence files according to Coil et al (2015) (Coil et al., 2015)

5.3.6.3 WGS data analysis for molecular characterization of the *E. coli* isolates

The Centre for Genomic Epidemiology (CGE) website (http://www.genomicepidemiology.org/) was used to characterize the strains using the MLST 1.8 (MultiLocus Sequence Typing) (Larsen et al., 2012), ResFinder 3.0 (Zankari et al., 2012) , SerotypeFinder 1.1 (Joensen et al., 2015), PlasmidFinder 1.3 (Carattoli et al., 2014), and VirulanceFinder 1.5 (Joensen et al., 2014) tools. BAGEL4, a web-based bacteriocin genome mining tool (de Jong et al., 2006) was used to identify the antimicrobial peptides, called bacteriocins, produced by bacteria (http://bagel.molgenrug.nl/). To identify prophage sequences within the *E. coli* genome, we used the web server, PHASTER (http://phaster.ca/) (Arndt et al., 2016). The CapsuleFinder tool (Rendueles et al., 2017) was used to determine the capsule type.

5.3.6.4 Phylogenetic Analysis

Parsnp (Treangen et al., 2014) was used for phylogenetic analysis of *E. coli* isolates from patients with IBD and healthy controls. Parsnp also uses maximal unique matches (MUMs) to identify similar genomes for multiple alignment. For visualising the phylogenetic tree of all *E. coli* genomes, along with the different molecular properties (disease status, gender, serotypes, phylogroups, antibiotic resistance, capsule types, bacteriocins and bacteriophage) we used a web-based tool, iTOL, which is accessible online http://itol.embl.de/login.cgi (Letunic and Bork, 2016).

5.3.6.5 Genome annotation and pan-genome studies for detection of genes associated with inflammasome activation

The Prokka rapid prokaryotic genome annotation tool version 1.12 (Seemann, 2014) was used for annotating genomic DNA sequences of *E. coli* (n = 32). The input file was the contig file of each isolate. Roary (v.3.11.2) was used for the pan-genome analysis of all *E. coli* genomes (Page et al., 2015). Annotated assembly outputs from PROKKA (.gff files) were used as the input file for Roary- a pan-genome pipeline. Then, the Scoary tool was used for genotype-phenotype association study, which determines whether or not a particular gene in the pan-genome is associated with a particular treatment or disease state of the host (Brynildsrud et al., 2016). Scoary provides p values for each trait for its phenotype association, a significant association is considered present if $p = <0.05$

5.3.6.6 Bacterial genome-wide association study (GWAS) using Pyseer

We also identified the association of genotypes of *E. coli* with the phenotypes (based on inflammasome activation) based on genome-wide association study using Pyseer tool. Pyseer is based on linear models (fixed or mixed effects) to estimate the contribution of genetic variation in a bacterial population or phenotype of interest, while considering the population structure, such as phylogenetic differences. Pyseer also uses k-mers (k-mers are a specific number of nucleotide sequences that allows it to generate a unique sequence from reads that represent genomic regions with different small sequences) data to identify gene variants that are significantly associated with a particular bacterial population or phenotype (Lees et al., 2018).

5.4 Results

A total of 32 *E. coli* strains were used for inflammasome, molecular properties and whole genome analysis (Table 5.1).

5.4.1 Inflammasome study results

It is important to note that some strains are clone pairs (or triplets) isolated from the same individual (patient 12, 20, 54, 55 and CD57) that differ by 100-200 genes that could be acquired due to within-host mutations. Another point of interest is shown within the table, which highlights strains from the same individual but from different tissue sites, namely an aphthous ulcer (AU) and the terminal ileum (Ti) in patient 55, from transverse colon (TC) and the terminal ileum (Ti) in patient 54 and three different tissues from lymph node, surrounding normal mucosa and involved mucosa in patient CD57 (bowel resected samples). We found that 23 *E. coli* stains activated the NLRP3 inflammasome, and nine did not (Table 5.1). Among CD isolates, 14 out of 20 (70%) activated NLRP3 inflammasome and the remaining six did not. On the other hand, all three (100%) UC isolates activated the NLRP3 inflammasome. In healthy control-associated *E. coli* isolates, six out of nine (67%) activated the NLRP3 inflammasome and the remaining three did not. We could not observe any significant difference in the activation of NLRP3 inflammasome by *E. coli* strains from patients with IBD compared to controls. Based on the invasive phenotype, ten out of 14 (71%) AIEC strains activated NLRP3 inflammasome whereas nine out of 12 (75%) non-AIEC strains activated the NLRP3 inflammasome. Here we also did not find significant difference in the activation of the NLRP3 inflammasome based on the adherent and invasive phenotypes of *E. coli* strains.

Table 5.1. List of *E. coli* strains in the inflammasome study along with their molecular properties, phenotypes, and disease status of the patients from where they were obtained

Patient Number	*E. coli* strains	Disease status	Phenotype	Phylogroup	ST	Inflammasome activation
12	12-1-Ti12	HC	AIEC	B2	127	Activated
12	12-2-Ti13	HC	Non-AIEC	B2	127	Not activated
20	20-2-TC13	CD	AIEC	B1	1304	Activated
20	20-3-TC12	CD	Non-AIEC	B1	1304	Activated
26	26-AU-glybx	CD	N/A	F	5801	Activated
33	33-1-Ti5	CD	AIEC	B2	131	Not activated
36	36-1-Ti13	CD	Non-AIEC	B2	429	Activated
53	53-1-AUC1	CD	N/A	B1	58	Not activated
54	54-1-TC4	CD	N/A	B2	1919	Activated
54	54-1-Ti6	CD	Non-AIEC	B2	1919	Activated
55	55-1-AU4	CD	Non-AIEC	B2	131	Not activated
55	55-1-Ti19	CD	Non-AIEC	B2	131	Activated
57	57-3-Ti5	CD	AIEC	B2	537	Activated
60	60-1-Ti1	UC	AIEC	B2	80	Activated
61	61-1-Ti1	UC	AIEC	B2	95	Activated
63	63-1-Ti1	UC	AIEC	D	963	Activated
CD34	CD34LN	CD	Non-AIEC	B2	95	Activated
CD57	CD57IM	CD	N/A	B1	3759	Activated
CD57	CD57LN	CD	Non-AIEC	B1	3759	Not activated
CD57	CD57NM	CD	N/A	B1	3759	Activated
CD59	CD59LN	CD	AIEC	A	2606	Activated
CD59	CD59LN3	CD	N/A	A	409	Not activated
CD62	CD62LN	CD	AIEC	B2	95	Not activated
CD70	CD70LN1	CD	AIEC	B1	2541	Activated
H223	H223	HC	AIEC	B2	141	Not activated
H305	H305	HC	AIEC	B2	95	Not activated
H413	H413	HC	Non-AIEC	B2	95	Activated
H461	H461	HC	Non-AIEC	B2	95	Activated
H504	H504	HC	AIEC	B2	95	Activated
H588	H588	HC	Non-AIEC	B2	126	Activated
H660	H660	HC	Non-AIEC	B2	95	Activated
LF82	LF82	CD	AIEC	B2	131	Activated

AIEC – adherent invasive *E. coli;* CD – Crohn's disease; HC – healthy controls; UC – ulcerative colitis

5.4.2 Association of sequence types (STs) and phylogroup of *E. coli* with inflammasome activation

We did not find any significant association of any sequence type and phylogroup of *E. coli* with inflammasome activation. We have also noticed that, in patient 12 from healthy control group, *E. coli* clones (ST-127, B2 phylogroup) with same sequence type but different phenotype (one is AIEC and another non-AIEC and only differ by 150-200 genes) gave different result in inflammasome activation (one activated and the other did not). Similarly, in patient CD57 *E. coli* strains with same ST type (ST-3759 and B1 phylogroup) did not give the same result (out of three strains only one activated and two could not activate NLRP3 inflammasome).

5.4.3 Association of molecular properties and genes of *E. coli* with inflammasome activation

No virulence factor, antibiotic resistance gene or plasmid of *E. coli* strains were significantly associated with inflammasome activation. Similarly, there were no significant differences in the presence of bacteriocins or bacteriophages between *E. coli* isolates that activated NLRP3 inflammasomes and those that did not activate NLRP3 inflammasomes (Supplemental Table 5.1 – 5.5). Based on the pan-genome analysis, we investigated the association of *E. coli* genes with inflammasome activation. We found that three bacterial genes were associated with the caspase-1 activation phenotype. Those genes are associated with small toxic polypeptide LdrD (ldrD_3) and tRNA-Lys(ttt) (group_11905) and hypothetical protein (group_953) (Table 5.2).

Table 5.2. Gene associated with inflammasome activation based on the association study between genotypes and phenotypes by Roary

Gene	Annotation	Present in inflammasome positive	Present in inflammasome negative	Naive p*
ldrD_3	Small toxic polypeptide LdrD	10	0	0.03
group_953	hypothetical protein	9	0	0.03
group_11905	tRNA-Lys(ttt)	18	3	0.03

*The naive p- value indicates the p-value without further adjustment or correction.

5.4.4 Association of *E. coli* genes with inflammasome activation

5.4.4.1 Fixed effect model of Pyseer analysis

In fixed effect model of Pyseer analysis we identified some genes in *E. coli* significantly associated with inflammasome activation (Table 5.3). Based on the lrt p-value, the small toxic polypeptide LdrD (ldrD_3) is found to be significantly associated with inflammasome activation corroborating the findings by pangenome analysis by Roary followed by Scoary analysis. Other genes in this analysis are associated with FliC flagellin (fliC group_1241 and group_722), acyl-CoA dehydrogenase (caiA_1), host specificity protein J (group_7), phage tail protein (group_220), UDP-4-amino-4-deoxy-L-arabinose—oxoglutarate aminotransferase (group_1527), phosphate starvation-inducible protein PhoH (phoH_1), host cell division inhibitory peptide Kil (group_6659), type I toxin-antitoxin system Hok family toxin (hokC_2), and enterobacterial Ail/Lom family protein (ompX_2).

Table 5.3. Genes associated with inflammasome activation based on fixed effect model of Pyseer

Variant	Function of the protein	Inflammasome positive (23)	Inflammasome negative (9)	AF	lrt-p value
group_1241	FliC flagellin	10	0	0.312	0.004
fliC	FliC flagellin	10	1	0.344	0.004

group_722	FliC family flagellin	11	1	0.375	0.004
caiA_1	acyl-CoA dehydrogenase	15	2	0.531	0.008
ldrD_3	Small toxic polypeptide LdrD	10	0	0.312	0.01
group_7	Host specificity protein J	7	1	0.25	0.01
group_220	phage tail protein	7	1	0.25	0.01
group_1527	UDP-4-amino-4-deoxy-L-arabinose—oxoglutarate aminotransferase	7	1	0.25	0.03
phoH_1	Phosphate starvation-inducible protein PhoH	14	4	0.562	0.04
group_6659	Host cell division inhibitory peptide Kil	11	3	0.438	0.04
hokC_2	Type I toxin-antitoxin system Hok family toxin	11	1	0.375	0.04
ldrD_2	Type I toxin-antitoxin system toxin Ldr family protein	7	1	0.25	0.04
ompX_2	Enterobacterial Ail/Lom family protein	6	2	0.25	0.04

AF is the allele frequency, and lrt-p-value indicates the p-value of association, adjusted for population structure.

5.4.4.2 Mixed effect model of Pyseer analysis

We identified genes in *E. coli* significantly associated with inflammasome activation based on the mixed effect model of genome-wide association study (GWAS) by Pyseer (Table 5.4). Gene associated with the small toxic polypeptide LdrD (ldrD_3) was significantly associated with inflammasome activation corroborated by the finding by pan-genome analysis by Roary and Scoary as well the fixed effect models of Pyseer analysis. The other genes in the list that are significantly associated with inflammasome activation are tRNA-Arg (group_11886), immunity repressor (group_5261), DNA adenine methylase (dam_1), FliC family flagellin (group_722), FliC flagellin (fliC), tRNA-Lys(ttt) (group_11905), type I toxin-antitoxin system Hok family toxin (hokC_2), two-component-system connector protein YcgZ (ycgZ), and hypothetical proteins (group_5262 and group_1651).

Table 5.4. Genes associated with inflammasome activation based on mixed effect model of Pyseer

Variant	Functions	Inflammasome positive (23)	Inflammasome negative (9)	lrt-pvalue
group_11886	tRNA-Arg	8	0	0.01
ldrD_3	Small toxic polypeptide LdrD	10	0	0.01
group_5261	immunity repressor	4	0	0.02
group_5262	hypothetical protein	4	0	0.02
dam_1	DNA adenine methylase	6	0	0.02
group_1651	Uncharacterised protein	5	0	0.02
group_722	FliC family flagellin	11	1	0.02
group_11905	tRNA-Lys(ttt)	18	3	0.03
fliC	FliC flagellin	10	1	0.03
hokC_2	type I toxin-antitoxin system Hok family toxin	11	1	0.03 0.04
ycgZ	two-component-system connector protein YcgZ	23	7	0.04

5.4.5 Association of k-mers in the *E. coli* with inflammasome activation based on k-mers analysis

Based on k-mer association by Pyseer, we could not find any strong hit or association of any particular k-mer/gene with inflammasome activation Table 5.5.

Table 5.5. K-mers associated with inflammasome activation based on mixed effect model of k-mer analysis

Gene	Hits	Max p	Avg. MAF	Avg. beta	Function
cds-WP_001407222.1	1	8.26	0.21	0.93	Hypothetical protein
cds-WP_032218570.1	1	8.48	0.28	0.83	YtfJ family protein
cds-WP_065226978.1	1	8.26	0.21	0.93	Serine protease
cds-WP_000574091.1	1	8.47	0.21	0.90	GNAT family N-acetyltransferase
cds-WP_000589068.1	1	8.35	0.28	0.8	SoxR-reducing system protein RseC
cds-WP_032314775.1	1	8.49	0.21	0.93	Type III secretion system LEE ATPase EscN

cds-WP_071999079.1	1	8.45	0.21	0.93	fatty acyl-AMP ligase
cds-WP_000691216.1	1	8.49	0.21	0.93	Glycosyltransferase family 4 protein
cds-WP_000150543.1	1	8.45	0.219	0.938	PTS mannose transporter subunit IIAB
cds-WP_001308747.1	1	8.49	0.219	0.938	O8 family O-antigen ABC transporter permease subunit Wzm
cds-WP_001117838.1	1	8.26	0.219	0.939	RnfH family protein
cds-WP_000533643.1	2	8.49	0.219	0.938	Tyrosine-type recombinase
cds-WP_001138064.1	1	8.95	0.219	0.95	Tn3-like element TnAs3 family transposase
cds-WP_000409203.1	1	8.28	0.219	0.956	Ancillary SecYEG translocon subunit YfgM

5.4.6 Phylogenetic analysis

In the phylogenetic tree of *E. coli* isolates, we saw that inflammasome positive *E. coli* isolates did not form a cluster and were scattered among the inflammasome negative isolates. However, it looks like the *E. coli* isolates were clustered in the phylogenetic tree according to the phylogroups of *E. coli*. The phylogenetic tree is presented along with some molecular properties of *E. coli* and presence and absence of significantly associated gene linked with inflammasome activation in Figure 5.2.

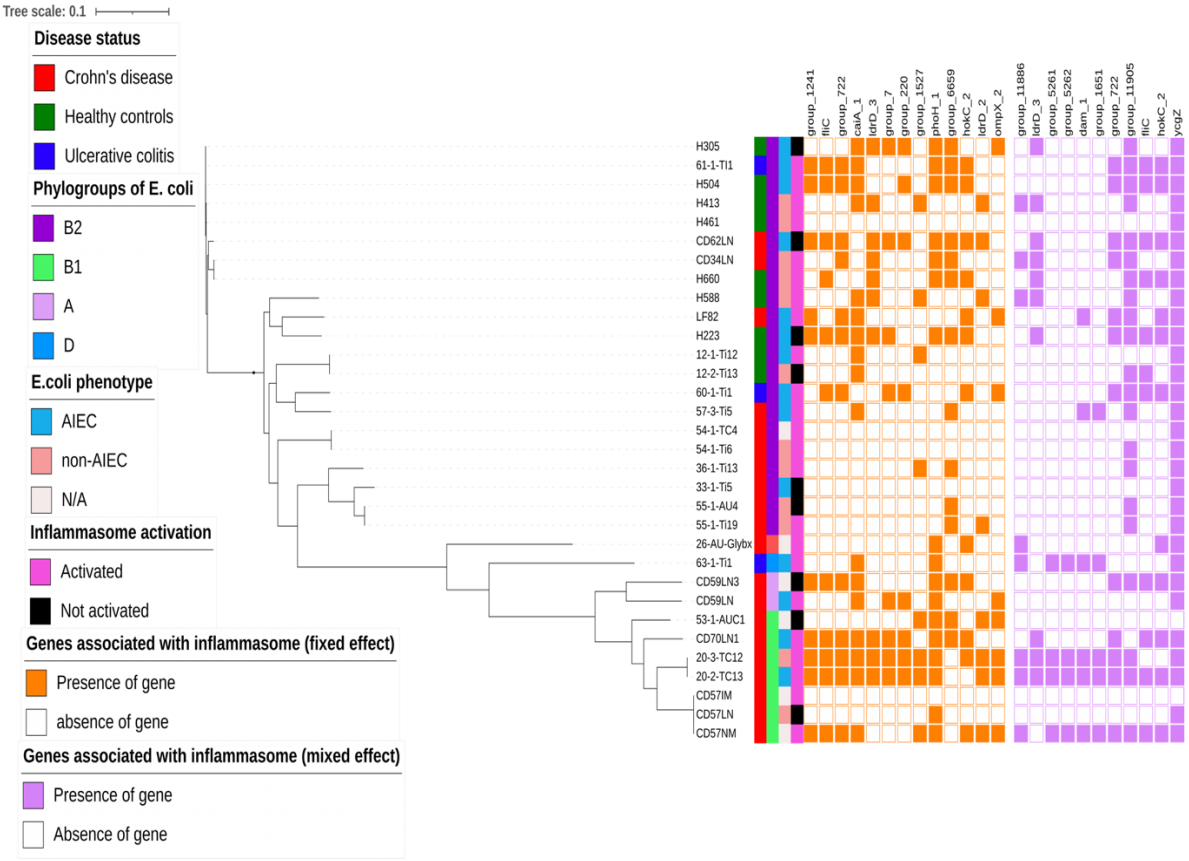


Figure 5.2. Phylogenetic tree of *E. coli* isolates in NLRP3 inflammasome study. Phylogenetic tree of 32 *E. coli* was created with Parsnp tool using maximal unique matches (MUMs). This tree also represents the disease status of the patients, phylogroups and phenotypes of *E. coli* isolates and if those isolates activated inflammasome or not along the right side of the tree. This phylogenetic tree also shows the presence and absence of significantly associated genes linked to inflammasome activation based on fixed effect and mixed effect of genome-wide association study (GWAS).

5.5 Discussion

It is well established that the NLRP3 inflammasome activates caspase-1 through different biochemical processes to provide immunity against infection or stress (Kayagaki et al., 2015; Shi et al., 2015). NLRP3 inflammasome was also found to be associated with many inflammatory conditions like cardiac disease (Sandanger et al., 2013), nephritis (Zhao et al., 2015), Crohn's disease (Villani et al., 2009), macular degeneration (Tseng et al., 2013) and glaucoma (Chi et al., 2014). As *E. coli* was

considered a leading bacterial trigger in CD it is important to explore if *E. coli* plays a role in the pathogenesis of CD through the activation of caspase -1 via the NLRP3-inflammasome pathway.

Some studies have reported the activation of *NLRP3 and NLRP4* inflammasome during uropathogenic *Escherichia coli* (UPEC)-mediated UTI (Verma et al., 2019). Some inflammatory regulators including NLRP3, Caspase-1, and some pro-inflammatory cytokines (IL-6, IL-8, IFN-γ, TNF-α, and MCP-1) were up-regulated in the UTI patients infected in response to UPEC (Verma et al., 2019). NLR proteins of caspase-1-activating inflammasomes function as receptors for flagellin and type III secretion system of bacteria (Zhao and Shao, 2016). Activation of the NLRP3 inflammasome is mediated through the type-1 fimbriae-dependent colonization of UPEC of urinary bladder epithelial cells (Demirel et al., 2018). They also reported that both type-1 fimbriae and α-hemolysin associated with UPEC could modify the activity of the NLRP3 inflammasome (Demirel et al., 2018). So, this indicates that virulence factors in *E. coli* could be associated with NLRP3 inflammasome activation and disease progression. However, to best of our knowledge, only one study (De la Fuente et al., 2014b) tried to identify the association of some of the previously reported virulence genes of *E. coli* from patients with CD, UC and other intestinal pathologies with NLRP3 inflammasome activation (De la Fuente et al., 2014a). But they did not explore all the gene molecular properties of a *E. coli* genome to identify if any particular gene of molecular property associated with NLRP3 inflammasome activation in patients with CD. Therefore, we sought to identify whether NLRP3 inflammasomes are specifically activated by *E. coli* strains isolated from IBD particularly CD patients and genes or molecular propertied associated with NLRP3 inflammasomes. This study was conducted by co-culturing macrophages from Nlrp3-/- and WT mice with *E. coli*

strains from non-IBD and IBD patients. We detected the Caspase-1 protein expression in macrophages through western blot analysis. Understanding how *E. coli* in IBD patients controls or regulates the gut immune system could help us to identify novel therapeutic approaches to block or modify the virulence properties of *E. coli*. In a previous study done by Lazaridis *et al.* peripheral blood mononuclear cells (PBMCs) were collected from patients with CD, ulcerative colitis (UC) patients and healthy controls and stimulated with bacterial lipopolysaccharides (LPS) in the absence or presence of monosodium urate crystals (Lazaridis et al., 2017). They found that NLRP3 inflammasome was activated in PBMCs from patients with CD. Among ulcerative colitis patients, they detected the NLRP3 inflammasome activation is only in patients with long-standing disease. They have also found NLRP3 inflammasome activation in 15 out of 25 (60%) in patients with CD compared to 6 out of 21 (28.6%) in control groups ($p = 0.042$). However, they did not observe any significant difference in NLRP3 activation between UC patients and controls (11/20 or 55% vs 6/21 or 28.6%, $p = 0.12$) (Lazaridis et al., 2017). In our study, we could not observe any significant difference in the activation of the NLRP3 inflammasome by *E. Coli* strains from patients with CD and controls. In most of the cases, *E. coli* isolates from the same patient were closely genetically related, whereas those from different individuals exhibited wide genetic diversity. The only difference between the study by Lazaridis *et al* and our study is that instead of PBMCs from IBD patients we used bone-marrow derived macrophage (BMDM) from wild and Nlrp3–/– mice. We also used live *E. coli* strains from patients with IBD and controls instead of LPS and monosodium urate crystals. This could explain why our results id different from them. We did not find any molecular property of *E. coli* that was associated with inflammasome activation. However, both in the Roary (association of genotype of *E. coli* and inflammasome

activation) and GWAS analysis we found a number of genes in *E. coli* significantly associated with inflammasome activation. Among these genes the most important gene that found in all analyses (Roary analysis, fixed effect of Pyseer and mixed effect of Pyseer) was ldrD_3, that encodes small toxic polypeptide LdrD. It is a toxic component of a type I toxin-antitoxin (TA) system. Expression of toxin protein is harmful to the cell, and it is neutralized by co-expressed antidote of toxin called antitoxin. Overexpression of this gene induces a stress-response and may inhibit ATP synthesis due to its insertion into the cell inner membrane leading to the permanent cessation of all vital functions. This causes rapid cell death and nucleoid condensation of the host cell (Kawano et al., 2002). In addition to ldrD_3 gene, we also observed that the genes (group_722 and fliC) coding for protein FliC family flagellin, were significantly associated with inflammasome activation based on both fixed effect model and mixed effect model of Pyseer. Previous studies found that mutants of *L. pneumophila* that are deficient in the expression of flagellin evade macrophage innate immunity and fail to induce caspase-1-mediated macrophage death. Previous studies suggest that flagellin is required for caspase-1-mediated pyroptosis of macrophages infected with intracellular pathogens (Molofsky et al., 2006; Ren et al., 2006). However, LPS and flagellin are common bacterial PAMPs that are also present in non-pathogenic strains of *E. coli* whereas the type III secretion system and different types of toxins are unique to pathogens. Detection of these distinctive PAMP-like toxins allows the innate immune response to distinguish pathogenic bacteria from non-pathogenic ones which is essential for maintaining immune tolerance towards commensals. In our study, we saw that activation of NLRP3 was not associated with the clearance of *E. coli* from macrophages.

Interestingly, our study showed that the lymph node clone strain from patient 57 did not activate the NLRP3 inflammasome. This *E. coli* strain did not harbour many of the NLRP3 activation associated genes. However, it possesses the ycgZ gene involved in the cellular response to an acidic environment and PhoH, involved in protecting DNA, membranes and stress resistance, and may therefore have enabled it to survive in macrophages within lymph nodes. The unaffected mucosa and involved mucosa strains from the same patient had many more genes associated with NLRP3 activation. Therefore, the lymph node clone strain likely did not activate the inflammasome due to the loss of these genes. Similarly, in the patient 12 clone pair (one AIEC and another non-AIEC), AIEC activated the NLRP3 inflammasome and non-AIEC did not. This result could also be due to the loss of these genes in the non-AIEC clone.

5.6 Conclusion

In this study, activation of NLRP3 was not associated with the clearance of *E. coli* from macrophages or pyroptosis. However, we observed some *E. coli* genes were associated with NLRP3 inflammasome activation. Future study is needed to explore the role of these genes in NLRP3 inflammasome activation and survival of *E. coli* inside macrophages.

5.7 References

Allen, I.C., Scull, M.A., Moore, C.B., Holl, E.K., McElvania-TeKippe, E., Taxman, D.J., et al. (2009). The NLRP3 inflammasome mediates in vivo innate immunity to influenza A virus through recognition of viral RNA. *Immunity* 30(4)**,** 556-565.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44(W1)**,** W16-21. Doi: 10.1093/nar/gkw387.

Ayres, Janelle S. (2013). Inflammasome-Microbiota Interplay in Host Physiologies. *Cell Host & Microbe* 14(5), 491-497. Doi: https://doi.org/10.1016/j.chom.2013.10.013.

Barnich, N., Denizot, J., and Darfeuille-Michaud, A. (2013). E. coli-mediated gut inflammation in genetically predisposed Crohn's disease patients. *Pathologie Biologie* 61(5), e65-e69. Doi: https://doi.org/10.1016/j.patbio.2010.01.004.

Baumgart, M., Dogan, B., Rishniw, M., Weitzman, G., Bosworth, B., and Yantiss, R. (2007). Culture independent analysis of ilealmucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. The ISME Journal 1. doi: 10.1038/ismej.2007.52.

Boudeau, J., Glasser, A.-L., Masseret, E., Joly, B., and Darfeuille-Michaud, A. (1999). Invasive Ability of an Escherichia coliStrain Isolated from the Ileal Mucosa of a Patient with Crohn's Disease. *Infection and immunity* 67(9), 4499-4509.

Brazil, J.C., Louis, N.A., and Parkos, C.A. (2013). The role of polymorphonuclear leukocyte trafficking in the perpetuation of inflammation during inflammatory bowel disease. *Inflammatory bowel diseases* 19(7), 1556-1565.

Broz, P., and Dixit, V.M. (2016). Inflammasomes: mechanism of assembly, regulation and signalling. *Nature Reviews Immunology* 16(7), 407-420.

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biology* 17(Goodhand et al.), 238. Doi: 10.1186/s13059-016-1108-8.

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial Agents and Chemotherapy* 58(7), 3895-3903. Doi: 10.1128/AAC.02412-14.

Chi, W., Li, F., Chen, H., Wang, Y., Zhu, Y., Yang, X., et al. (2014). Caspase-8 promotes NLRP1/NLRP3 inflammasome activation and IL-1β production in acute glaucoma. *Proceedings of the National Academy of Sciences* 111(30), 11181-11186.

Coil, D., Jospin, G., and Darling, A.E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31(4), 587-589. Doi: 10.1093/bioinformatics/btu661.

Conte, M.P., Schippa, S., Zamboni, I., Penta, M., Chiarini, F., and Seganti, L. (2006). Gut-associated bacterial microbiota in paediatric patients with inflammatory bowel disease. *Gut* 55. Doi: 10.1136/gut.2005.078824.

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

Darfeuille-Michaud, A., Neut, C., Barnich, N., Lederman, E., Di Martino, P., Desreumaux, P., et al. (1998). Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease. *Gastroenterology* 115(6), 1405-1413.

De Jong, A., van Hijum, S.A.F.T., Bijlsma, J.J.E., Kok, J., and Kuipers, O.P. (2006). BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Research* 34(Web Server issue), W273-W279. Doi: 10.1093/nar/gkl237.

De la Fuente, M., Franchi, L., Araya, D., Díaz-Jiménez, D., Olivares, M., Álvarez-Lobos, M., et al. (2014a). Escherichia coli isolates from inflammatory bowel diseases patients survive in macrophages and activate NLRP3 inflammasome. *International Journal of Medical Microbiology* 304(3), 384-392. Doi: https://doi.org/10.1016/j.ijmm.2014.01.002.

De la Fuente, M., Franchi, L., Araya, D., Diaz-Jimenez, D., Olivares, M., Alvarez-Lobos, M., et al. (2014b). Escherichia coli isolates from inflammatory bowel diseases patients

survive in macrophages and activate NLRP3 inflammasome. *Int J Med Microbiol* 304(3-4), 384-392. Doi: 10.1016/j.ijmm.2014.01.002.

Demirel, I., Persson, A., Brauner, A., Särndahl, E., Kruse, R., and Persson, K. (2018). Activation of the NLRP3 Inflammasome Pathway by Uropathogenic Escherichia coli Is Virulence Factor-Dependent and Influences Colonization of Bladder Epithelial Cells. *Frontiers in cellular and infection microbiology* 8, 81-81. Doi: 10.3389/fcimb.2018.00081.

Dostert, C., Pétrilli, V., Van Bruggen, R., Steele, C., Mossman, B.T., and Tschopp, J. (2008). Innate immune activation through Nalp3 inflammasome sensing of asbestos and silica. *Science* 320(5876), 674-677.

Elhenawy, W., Tsai, C. N., & Coombes, B. K. (2019). Host-Specific Adaptive Diversification of Crohn's Disease-Associated Adherent-Invasive Escherichia coli. *Cell host & microbe* 25(2), 301–312. Doi: https://doi.org/10.1016/j.chom.2018.12.010

Elinav, E., Strowig, T., Kau, Andrew L., Henao-Mejia, J., Thaiss, Christoph A., Booth, Carmen J., et al. (2011). NLRP6 Inflammasome Regulates Colonic Microbial Ecology and Risk for Colitis. *Cell* 145(5), 745-757. Doi: https://doi.org/10.1016/j.cell.2011.04.022.

Franchi, L., Warner, N., Viani, K., and Nuñez, G. (2009). Function of Nod-like receptors in microbial recognition and host defense. *Immunological reviews* 227(Goodhand et al.), 106-128.

Gordon, D.M., O'brien, C.L., and Pavli, P. (2015). Escherichia coli diversity in the lower intestinal tract of humans. *Environmental microbiology reports* 7(4), 642-648.

Gross, O., Poeck, H., Bscheider, M., Dostert, C., Hannesschläger, N., Endres, S., et al. (2009). Syk kinase signalling couples to the Nlrp3 inflammasome for anti-fungal host defence. *Nature* 459(7245), 433-436.

Hirota, S.A., Ng, J., Lueng, A., Khajah, M., Parhar, K., Li, Y., et al. (2011). NLRP3 inflammasome plays a key role in the regulation of intestinal homeostasis. *Inflammatory bowel diseases* 17(6), 1359-1372.

Hornung, V., Bauernfeind, F., Halle, A., Samstad, E.O., Kono, H., Rock, K.L., et al. (2008). Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nature immunology* 9(8), 847.

Itani, S., Watanabe, T., Nadatani, Y., Sugimura, N., Shimada, S., Takeda, S., et al. (2016). NLRP3 inflammasome has a protective effect against oxazolone-induced colitis: a possible role in ulcerative colitis. *Scientific Reports* 6(Goodhand et al.), 39075. Doi: 10.1038/srep39075.

Iyer, S.S., He, Q., Janczy, J.R., Elliott, E.I., Zhong, Z., Olivier, A.K., et al. (2013). Mitochondrial cardiolipin is required for Nlrp3 inflammasome activation. *Immunity* 39(Polito et al.), 311-323.

Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol* 52(5), 1501-1510. Doi: 10.1128/jcm.03617-13.

Joensen, K.G., Tetzschner, A.M., Iguchi, A., Aarestrup, F.M., and Scheutz, F. (2015). Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 53(8), 2410-2426. Doi: 10.1128/jcm.00008-15.

Kanneganti, T.-D. (2015). The inflammasome: firing up innate immunity. *Immunological reviews* 265(Goodhand et al.), 1.

Kanneganti, T.D., Ozoren, N., Body-Malapel, M., Amer, A., Park, J.H., Franchi, L., et al. (2006). Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3. *Nature* 440(7081), 233-236. Doi: 10.1038/nature04517.

Kawano, M., Oshima, T., Kasai, H., and Mori, H. (2002). Molecular characterization of long direct repeat (LDR) sequences expressing a stable mRNA encoding for a 35-amino-acid cell-killing peptide and a cis-encoded small antisense RNA in Escherichia coli. *Molecular Microbiology* 45(Polito et al.)**,** 333-349. Doi: https://doi.org/10.1046/j.1365-2958.2002.03042.x.

Kayagaki, N., Stowe, I.B., Lee, B.L., O'Rourke, K., Anderson, K., Warming, S., et al. (2015). Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. *Nature* 526(7575)**,** 666-671.

Lamkanfi, M., and Dixit, V.M. (2009). The inflammasomes. *PloS Pathog* 5(12)**,** e1000510.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 50(4)**,** 1355-1361. Doi: 10.1128/jcm.06094-11.

Lazaridis, L.D., Pistiki, A., Giamarellos-Bourboulis, E.J., Georgitsi, M., Damoraki, G., Polymeros, D., et al. (2017). Activation of NLRP3 Inflammasome in Inflammatory Bowel Disease: Differences Between Crohn's Disease and Ulcerative Colitis. *Dig Dis Sci* 62(9)**,** 2348-2356. Doi: 10.1007/s10620-017-4609-8.

Lees, J.A., Galardini, M., Bentley, S.D., Weiser, J.N., and Corander, J. (2018). Pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* 34(24)**,** 4310-4312. Doi: 10.1093/bioinformatics/bty539.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44(W1)**,** W242-W245. Doi: 10.1093/nar/gkw290.

Man, S.M., and Kanneganti, T.-D. (2015). Regulation of inflammasome activation. *Immunological reviews* 265(Goodhand et al.)**,** 6-21. Doi: 10.1111/imr.12296.

Man, S.M., Karki, R., Briard, B., Burton, A., Gingras, S., Pelletier, S., et al. (2017). Differential roles of caspase-1 and caspase-11 in infection and inflammation. *Scientific Reports* 7(Goodhand et al.)**,** 45126. Doi: 10.1038/srep45126.

Mann, E.A., and Saeed, S.A. (2012). Gastrointestinal infection as a trigger for inflammatory bowel disease. *Current opinion in gastroenterology* 28(Goodhand et al.)**,** 24-29.

Martinez-Medina, M.X., Aldeguer, M., Lopez-Siles, F., González-Huix, C., López-Oliu, G., and Dahbi, G. (2009). Molecular diversity of Escherichia coli in the human gut: new ecological evidence supporting the role of adherent-invasive E. coli (AIEC) in Crohn's disease. *Inflamm Bowel Dis* 15. Doi: 10.1002/ibd.20860.

Martinon, F., Pétrilli, V., Mayor, A., Tardivel, A., and Tschopp, J. (2006). Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* 440(7081)**,** 237-241.

Mirsepasi-Lauridsen, H.C., Vallance, B.A., Krogfelt, K.A., and Petersen, A.M. (2019). Escherichia coli Pathobionts Associated with Inflammatory Bowel Disease. *Clinical microbiology reviews* 32(Polito et al.)**,** e00060-00018. Doi: 10.1128/CMR.00060-18.

Molofsky, A.B., Byrne, B.G., Whitfield, N.N., Madigan, C.A., Fuse, E.T., Tateda, K., et al. (2006). Cytosolic recognition of flagellin by mouse macrophages restricts Legionella pneumophila infection. *The Journal of experimental medicine* 203(4)**,** 1093-1104.

Muñoz-Planillo, R., Kuffa, P., Martínez-Colón, G., Smith, B.L., Rajendiran, T.M., and Núñez, G. (2013). K+ efflux is the common trigger of NLRP3 inflammasome activation by bacterial toxins and particulate matter. *Immunity* 38(6)**,** 1142-1153.

Mylonaki, M., Rayment, N.B., Rampton, D.S., Hudspith, B.N., and Brostoff, J. (2005). Molecular characterization of rectal mucosa-associated bacterial flora in inflammatory bowel disease. *Inflammatory bowel diseases* 11(5)**,** 481-487.

Nagler-Anderson, C. (2001). Man the barrier! Strategic defences in the intestinal mucosa. *Nature Reviews Immunology* 1(Goodhand et al.)**,** 59-67. Doi: 10.1038/35095573.

Nataro, J.P., and Kaper, J.B. (1998). Diarrheagenic Escherichia coli. *Clin Microbiol Rev* 11(Goodhand et al.), 142-201.

O'Brien, C.L., Bringer, M.-A., Holt, K.E., Gordon, D.M., Dubois, A.L., Barnich, N., et al. (2016). Comparative genomics of Crohn's disease-associated adherent-invasive Escherichia coli. *Gut*. Doi: 10.1136/gutjnl-2015-311059.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22), 3691-3693. Doi: 10.1093/bioinformatics/btv421.

Palmela, C., Chevarin, C., Xu, Z., Torres, J., Sevrin, G., Hirten, R., et al. (2018). Adherent-invasive Escherichia coli in inflammatory bowel disease. *Gut* 67(3), 574. Doi: 10.1136/gutjnl-2017-314903.

Ren, T., Zamboni, D.S., Roy, C.R., Dietrich, W.F., and Vance, R.E. (2006). Flagellin-deficient Legionella mutants evade caspase-1-and Naip5-mediated macrophage immunity. *PloS Pathog* 2(3), e18.

Rendueles, O., Garcia-Garcera, M., Neron, B., Touchon, M., and Rocha, E.P.C. (2017). Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. *PloS Pathog* 13(7), e1006525. Doi: 10.1371/journal.ppat.1006525.

Sahoo, M., Ceballos-Olvera, I., del Barrio, L., and Re, F. (2011). Role of the inflammasome, IL-1β, and IL-18 in bacterial infections. *The Scientific World Journal* 11, 2037.

Sandanger, Ø., Ranheim, T., Vinge, L.E., Bliksøen, M., Alfsnes, K., Finsen, A.V., et al. (2013). The NLRP3 inflammasome is up-regulated in cardiac fibroblasts and mediates myocardial ischaemia–reperfusion injury. *Cardiovascular research* 99(Goodhand et al.), 164-174.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30. Doi: 10.1093/bioinformatics/btu153.

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., et al. (2015). Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature* 526(7575), 660-665.

Swidsinski, A., Ladhoff, A., Pernthaler, A., Swidsinski, S., Loening–Baucke, V., Ortner, M., et al. (2002). Mucosal flora in inflammatory bowel disease. *Gastroenterology* 122(Goodhand et al.), 44-54.

Thomas, P.G., Dash, P., Aldridge Jr, J.R., Ellebedy, A.H., Reynolds, C., Funk, A.J., et al. (2009). The intracellular sensor NLRP3 mediates key innate and healing responses to influenza A virus via the regulation of caspase-1. *Immunity* 30(4), 566-575.

Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biology* 15(11), 524. Doi: 10.1186/s13059-014-0524-x.

Tseng, W.A., Thein, T., Kinnunen, K., Lashkari, K., Gregory, M.S., D'Amore, P.A., et al. (2013). NLRP3 inflammasome activation in retinal pigment epithelial cells by lysosomal destabilization: implications for age-related macular degeneration. *Investigative ophthalmology & visual science* 54(Goodhand et al.), 110-120.

Verma, V., Gupta, S., Kumar, P., Yadav, S., Dhanda, R.S., Gaind, R., et al. (2019). Involvement of NLRP3 and NLRC4 Inflammasome in Uropathogenic E. coli Mediated Urinary Tract Infections. *Front Microbiol* 10, 2020. Doi: 10.3389/fmicb.2019.02020.

Villani, A.-C., Lemire, M., Fortin, G., Louis, E., Silverberg, M.S., Collette, C., et al. (2009). Common variants in the NLRP3 region contribute to Crohn's disease susceptibility. *Nature genetics* 41(Goodhand et al.), 71.

Zaki, M.H., Boyd, K.L., Vogel, P., Kastan, M.B., Lamkanfi, M., and Kanneganti, T.-D. (2010). The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. *Immunity* 32(3)**,** 379-391.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy* 67(11)**,** 2640-2644. Doi: 10.1093/jac/dks261.

Zhao, J., Wang, H., Huang, Y., Zhang, H., Wang, S., Gaskin, F., et al. (2015). Lupus nephritis: glycogen synthase kinase 3β promotion of renal damage through activation of the NLRP3 inflammasome in lupus-prone mice. *Arthritis & Rheumatology* 67(4)**,** 1036-1044.

Zhao, Y., and Shao, F. (2016). Diverse mechanisms for inflammasome sensing of cytosolic bacteria and bacterial virulence. *Curr Opin Microbiol* 29**,** 37-42. Doi: 10.1016/j.mib.2015.10.003.

Zhou, R., Tardivel, A., Thorens, B., Choi, I., and Tschopp, J. (2010). Thioredoxin-interacting protein links oxidative stress to inflammasome activation. *Nature immunology* 11(Polito et al.)**,** 136-140.

5.8 Supplemental materials

Supplemental Table 5.1. List of virulence genes in *E. coli* isolates used in inflammasome study

Virulence factor	NLRP3 activated (23)	NLRP3 not activated (9)
afaD	2	1
air	2	0
cea	3	2
celb	1	0
chuA	17	6
cia	4	1
cib	1	0
clbB	8	4
cma	0	1
cnf1	4	2
cvaC	2	1
eilA	2	0
etsC	2	2
focC	2	1
focCsfaE	3	1

fyuA	17	6
gad	11	5
hlyF	1	1
hra	8	2
ibeA	10	2
iha	2	2
ireA	6	1
iroN	5	3
irp2	17	7
iss	13	7
iucC	4	3
iutA	4	3
kpsE	15	6
kpsM_K15	1	0
kpsMII	1	0
kpsMII_K1	8	3
kpsMII_K23	1	0
kpsMII_K5	3	3
lpfA	7	2
mchB	3	1
mchC	2	1
mchF	4	2
mcmA	2	1
neuC	8	3
ompT	16	7
papA_F11	2	1
papA_F43	4	3
papA_F48	1	1
papC	6	4
pic	2	1
sat	2	2
senB	8	4
sfaD	5	2
sfaS	3	1
sitA	15	8
tcpC	3	2
terC	22	9
traT	12	5

usp	13	6
vat	14	4
yfcV	16	5

Supplemental Table 5.2. List of antibiotic resistance genes associated with *E. coli* isolates in inflammasome study

Antibiotic resistant	**Antibiotic resistance gene**	**Inflammasome activation positive (23)**	**Inflammasome activation negative (9)**
Aminoglycoside resistance	aadA5	0	1
Aminoglycoside resistance	aph (3'')-Ib	0	3
Aminoglycoside resistance	aph (6)-Id	0	3
Beta-lactam resistance	blaTEM-1B	1	4
Beta-lactam resistance	blaTEM-1C (blaTEM-1C_FJ560503)	2	0
Phenicol resistance	catA1	0	1
Trimethoprim resistance	dfrA17	0	1
Trimethoprim resistance	dfrA5	0	1
Trimethoprim resistance	dfrA7	0	1
Macrolide resistance	mph (A)	0	1
No resistance	No resistance	20	4
Quinolones and fluoroquinolones	344are p.I529L (Quinolones and fluoroquinolones)	0	1
Sulphonamide resistance	sul1	0	2
Sulphonamide resistance	sul2	0	3
Tetracycline resistance	tet (A)	1	2

Supplemental Table 5.3. List of plasmids associated with *E. coli* isolates in inflammasome study

Plasmid	**Inflammasome positive (23)**	**Inflammasome negative (9)**
Col (MG828)	0	1
Col156	8	4
IncB/O/K/Z	1	0
IncFIA	1	0

IncFIB(AP001918)	13	6
IncFIB(K)	1	0
IncFIB(pB171)	2	0
IncFIB(pLF82)	1	0
IncFIC(FII)	1	0
IncFII	3	2
IncFII(29)	5	3
IncFII(pCRY)	0	1
IncFII(pHN7A8)	1	0
IncHI1A	0	1
IncHI1B(R27)	0	1
IncI1	4	3
IncI2	1	0
IncN	1	0
IncQ1	0	2
IncX4	2	0
IncY	0	1
no plasmid	3	0
pSL483	2	0

Supplemental Table 5.4. List of bacteriocins associated with *E. coli* isolates in this inflammasome study

Bactreriocins	**Inflammasome activated (23)**	**Inflammasome not activated (9)**
Colicin	10	4
Colicin_E1	2	0
Microcin_PDI	2	2
Colicin_E1_protein	0	1
Colicin_E7	1	0
Colicin_E9	3	1
Colicin_Ib	0	1
Colicin_Ib	2	0
Colicin_protein	2	1
Colicin-E1	1	0
Colicin-M	1	0
MccI47_mchS2	2	0
Microcin_H47_(MccH47)	2	0
Microcin_M	0	1

Microcin_V_(Colicin_V)	1	0
Carocin_D	6	3
Linocin_M18Bacteriocin	0	1
Bottromycin	19	8
Microcin	6	2

Supplemental Table 5.5. List of prophages associated with *E. coli* isolates in this inflammasome study

Most Common Phage	**Inflammasome activation**	**Inflammasome not activated**
PHAGE_Entero_lambda_NC_001416	13	6
PHAGE_Entero_Sf101_NC_027398	11	4
PHAGE_Entero_mEp460_NC_019716	7	3
PHAGE_Entero_P4_NC_001609	7	1
PHAGE_Shigel_SfII_NC_021857	7	2
PHAGE_Entero_DE3_NC_042057	6	0
PHAGE_Entero_fiAA91_ss_NC_022750	5	1
PHAGE_Salmon_Fels_2_NC_010463	5	1
PHAGE_Shigel_Sf6_NC_005344	5	1
PHAGE_Burkho_BcepMu_NC_005882	4	4
PHAGE_Burkho_phiE255_NC_009237	4	1
PHAGE_Entero_cdtI_NC_009514	4	0
PHAGE_Entero_HK630_NC_019723	4	0
PHAGE_Entero_N15_NC_001901	4	1
PHAGE_Entero_P88_NC_026014	4	2
PHAGE_Escher_RCS47_NC_042128	4	0
PHAGE_Entero_BP_4795_NC_004813	3	3
PHAGE_Entero_HK629_NC_019711	3	0
PHAGE_Entero_SfI_NC_027339	3	2
PHAGE_Escher_phiV10_NC_007804	3	0

Chapter 6: General Discussion

6.1 General discussion

The microbiome of the gastrointestinal tract is an essential constituent of our body and is linked to health and disease. It is now well recognised that diverse and complex microbial communities in the gastrointestinal tract play a crucial role in establishing and maintaining health and contribute to disease pathogenesis (Lavelle and Sokol, 2020). However, the underlying mechanism of the gastrointestinal microbiome in disease development has not been fully elucidated. At first, to understand the effect of microbiota on health and disease, our primary focus should be diverted from the individual pathogens and towards the microbial community. Based on the microbiome studies, once the desired microbiota composition and function in a healthy gut is determined, we can decide which component, when disturbed, may play a role in disease pathogenesis, such as CD. Sometimes, inter-individual variability and high microbiota complexity make it challenging to define the desired microbiome. A better understanding and characterisation of the diverse microbial composition in the healthy gut could improve management and therapeutic treatment in diseases, like CD.

6.1.1 Characterisation of microbial communities in different disease stages and tissues from Crohn's disease patients and non-IBD controls

The actual cause of CD is still unknown. However, many studies suggest an interplay among genetic, immunological, and environmental factors, including microbes, may play a role in the pathogenesis of CD. Based on previous studies, three theories have been proposed:

1. Bacterial dysbiosis may result in an increase in the harmful bacteria and decrease the beneficial bacteria.

2. The presence of an unidentified pathogen in a genetically susceptible host may play a role; and

3. Putative bacterial triggers such as *Escherichia coli*

This study demonstrated that all these theories could be admissible to CD development. Previous findings show that in CD patients, OTU richness and species diversity are significantly decreased compared with that of controls. Our findings were consistent with that of previous reports, as decreased alpha diversity in patients with CD was consistently observed (Kowalska-Duplaga et al., 2019; Clooney et al., 2021).

Based on disease activity, in different tissue types in CD and controls, species richness was significantly reduced in aphthous ulcers (AU) early-active-CD, inactive small bowel CD, active small bowel CD, inactive colon CD, and active colon CD compared with that of the normal biopsies in controls. However, this result was not consistent with a recent study where no significant differences in alpha diversity was reported between patients with active CD and inactive CD (Clooney et al., 2021). Their results might be different from us because of their different sampling approach. They used stool samples, whereas we used mucosal biopsies. When we compared mucosal microbial communities in active terminal ileum with inactive rectum, we observed that alpha diversity based on observed OTUs was significantly reduced in the active terminal ileum but not in inactive rectum, when compared with the normal biopsies from controls. This result illustrates that decreased microbiota richness is found only in the inflamed site but not in the uninflamed site in patients with CD, therefore reduced alpha diversity in CD is not diffuse but focal. A recent study also demonstrated that patients with active CD had lower species diversity compared to those CD patients with inactive disease and healthy controls (Mirsepasi-Lauridsen et al., 2018). This

result suggests that inflammation is required for reduced alpha diversity in patients with CD.

Beta diversity (differences in microbial community composition between individuals) differed significantly between CD and controls. Microbiome composition in different disease stages in CD also differed significantly compared with controls. Significant differences in microbial communities have been also observed in these studies, such that aphthous ulcers (AU)-early active CD, inactive small bowel CD, active small bowel CD, inactive colon CD, and active colon CD compared with healthy controls. Our findings were consistent with that of previous research, in which a decrease of beta diversity in CD patients was observed (Galazzo et al., 2019; Kowalska-Duplaga et al., 2019).

We could not identify common bacteria present in all the patients with CD but totally absent in healthy controls. However, we found that at the genus level, some bacteria were significantly increased in CD compared to controls such as *Ruminococcus gravus* group ($p = 0.003$), *Clostridium innocuum* group ($p = 0.02$), *Escherichia_Shigella* ($p = 0.01$), *Flavonifractor* ($p = 0.01$), *Klebsiella* ($p = 0.01$), and *Ralstonia* ($p = 0.01$). Previous studies also reported that *Ruminococcus gravus* was increased in CD (Png et al., 2010; Willing et al., 2010; Joossens et al., 2011; Hall et al., 2017; Nishino et al., 2018). Recently a study characterized an inflammatory polysaccharide produced by *Ruminococcus gravus*, transient blooms of which correspond to the disease flares in patients with CD and suggest that this could be a plausible molecular mechanism of inflammation in CD (Henke et al., 2019). The significant association of *Clostridium innocuum* in our study is supported by a recent report that found mesenteric adipose tissue (MAT) from creeping fat in CD patients

was associated with *Clostridium innocuum* (Dickson, 2020). Creeping fat is an extraintestinal manifestation in CD, in which MAT migrates to sites of inflammation.

In addition, *Flavonifractor* was also overrepresented in CD in our study. A recent study also identified *Flavonifractor plautii* as a pathobiont in paediatric inflammatory bowel diseases using IgG as a marker (Armstrong et al., 2019). In addition, a few studies reported the anti-inflammatory effects of flavonoids (Vezza et al., 2016) that is degraded by *F. plautii* in the human gut (Braune and Blaut, 2016; Musumeci et al., 2020) and is also associated with colorectal cancer patients (Gupta et al., 2019). Ju and collaborators also found that the flavonol reduced gut inflammation and changed the intestinal microbiome in a mouse model with experimental colitis through modulating gut macrophage function (Ju et al., 2018). The study also suggested the therapeutic potential of flavonoids in patients with IBD (Salaritabar et al., 2017).

Likewise, our study observed an increased abundance of *Klebsiella* in CD that corroborates Rishid and colleagues' findings, as their findings suggest that *Klebsiella* might play a significant role in initiating and perpetuating intestinal inflammation in patients with CD (Rashid et al., 2013). High dietary intake of starch-rich food also enhanced the growth of gut microbes, particularly *Klebsiella*. Increased growth of *Klebsiella* boosted the production of antibodies against *Klebsiella* and autoantibodies to the self-antigens present in collagen, resulting in inflammation in the gut of CD patients (Rashid et al., 2009). Recently, a study demonstrated that ectopic colonization in the colon by oral *Klebsiella pneumoniae* in the dysbiotic gut strongly induced T helper 1 (TH1) cells to stimulate severe inflammation in the gut (Atarashi et al., 2017).

Similar to our results, the increased abundance of *Ralstonia* has also been demonstrated by a recent study where OTUs belonging to genus *Ralstonia* increased in phagocytes in the mucosa in CD patients (Dheer et al., 2020). However, as the previous studies did not exclusively identify these bacteria in only CD patients, the exact role of *Ruminococcus gravus* group, *Clostridum innocuum* group, *Escherichia_Shigella*, *Flavonifractor*, *Klebsiella* and *Ralstonia* remains to be defined.

The most significant finding of this study is identifying the association of genera *Escherichia_Shigella* with CD. The abundance of *Escherichia-shigella* was also significantly increased in different disease stages of CD (active-colon, active-small bowel, and inactive-colon) compared to normal controls. In many studies, a remarkable increase in Gammaproteobacteria that include genera belonging to the Enterobacteriaceae family have consistently been reported (Gevers et al., 2014; Pascal et al., 2017). In addition, a study observed that Proteobacteria was significantly enriched in the phagocyte-associated microbiota of both CD and UC patients when compared with gut mucosa-associated microbiota (Dheer et al., 2020). Over the last few decades, the bacterium that has become the centre of all attention, in relation to CD aetiology, is *Escherichia coli*. Winter and colleagues hypothesized that by-products of reactive oxygen and nitrogen species produced during inflammation supported the overgrowth of various *E. coli* strains, which outcompete mutant strains in the large intestine of mice (Winter et al., 2013).

In this study, we also observed the decreased abundance of helpful bacteria commonly associated with the healthy non-IBD gut. For example, we demonstrated that the abundance of *Faecalibacterium* was significantly reduced in active and inactive colon CD and active-small bowel CD patients compared with healthy individuals. This bacterium has been found to play a significant role in promoting gut

homeostasis by restoring the intestinal barrier structure (Xu et al., 2020). The role of *F. prausnitzii* in maintaining intestinal health through secreting anti-inflammatory substances like butyrate, and microbial anti-inflammatory molecule (MAM) has also been reported in previous studies (Sokol et al., 2008; Louis and Flint, 2009; Lopez-Siles et al., 2012; Breyner et al., 2017). This bacterium also played an anti-inflammatory as well as a protective role in mice with TNBS-induced colitis (Sokol et al., 2008; Martín et al., 2014). We also demonstrated a significant depletion of genera belonging to *Clostridiales* in CD microbiomes compared to controls, particularly *Ruminococcus_torques_group*, *Roseburia*, *Erysipelotrichaceae_UCG_003*, *Dorea*, *Eubacterium_hallii_group*, *Lachnospiraceae_NK4A136_group*, *Lachnospiraceae_UCG_004*, *Agathobacter*, *Phascolarctobacterium and Fusicatenibacter.* Our findings suggest that dysbiotic gut microbiota could contribute to CD pathogenesis through dysregulating the host's gut homeostasis. However, the gut microbial imbalances in CD at the level of different phyla is variable across several studies (Frank et al., 2007; Nishida et al., 2018; Franzosa et al., 2019; Vester-Andersen et al., 2019). This imbalance could be due to the various approaches in microbiome studies, sampling techniques or heterogeneity of disease-associated microbiota. Instead of faecal samples, we used gut mucosal biopsies obtained during colonoscopy, and bowel resected mucosal samples in the late stage of disease. Although most of the previous studies used faecal samples for microbiome analysis due to high availability and stability, it has been reported that mucosa-associated microbiota may play a critical role in CD pathogenesis rather than luminal microbiota (Willing et al., 2009).

Studies have identified a significant difference in microbiome composition between faecal and mucosal samples in the colon (Morgan et al., 2012; Gevers et al., 2014). Faecal samples represent microbes derived from any part of the intestine and

are not ideal for studying the difference in microbiome composition between inflamed and adjacent non-inflamed sites. Moreover, they are not suitable samples for exploring the microbiome of a particular gut location like ileum. However, studies commonly use faecal samples due to the higher costs and invasive nature of colonoscopy to collect biopsy samples. Therefore, a strength of our study was the use of mucosal samples instead of faecal material.

6.1.2 The plausible association of *Escherichia coli* in CD

Many bacterial pathogens have been implicated in CD. However, the results have been conflicting. Among all, *E. coli* has received the greatest attention for its association with CD. Results of many studies suggested that *E. coli* are more prevalent in CD compared with healthy controls. However, there was no clear indication whether increased numbers of *E. coli* is the cause or result of inflammation in CD. Studies have identified viable *E. coli* within the lymph nodes of patients with CD, which could provide continuous stimulation of the immune system and thereby perpetuate inflammation in CD.

Our whole genome sequences data analysis results of the *E. coli* isolates revealed that the CD strains were phylogenetically diverse and possess interpatient heterogeneity based on different molecular properties. Moreover, no common genotype of *E. coli* was identified in all the CD isolates, nor do they share a common origin. Therefore, our results suggest that these CD-associated phenotypes could have arisen by horizontal gene transfer possible from other bacteria.

Based on the bacteriocin analysis, we found that the bacteriocin, bottromycin, was significantly associated with CD compared to healthy controls. In CD, bottromycin bacteriocin may help bacteria to survive in the inflamed intestine by way of its

bactericidal activity, possibly by outcompeting other bacteria. We observed that Group IV_e capsule and bottromycin bacteriocin were associated with lymph node CD isolates, and bottromycin bacteriocin, ABC capsule, and *E. coli* pUTI89 plasmid were significantly associated with paediatric CD when compared to HC isolates. The factor that may have driven this result in paediatric CD could be the different geographic locations of control patients compared to paediatric patients. In addition, this bacteriocin may also enable *E. coli* to survive within the lymph nodes because of its bactericidal activity, possibly enabling it to compete with other bacteria. Other molecular properties like capsule and plasmid presumably provide *E. coli* with an increased ability to survive and play a pathogenic role in CD patients.

Pan-genome analysis and genome wide association studies also found that CD isolates were significantly associated with genes linked to adhesion, the toxin-antitoxin system, plasmid partitioning, and conjugation compared with HC. In addition, *E. coli* isolates in lymph node CD were more likely to be associated with genes involved in adhesion, phosphate binding and catalase-peroxidase activity compared to HC. Catalase-peroxidase gene (KatG) functions to protects bacteria against the toxic reactive oxygen species (ROS) including hydrogen peroxide and organic peroxides to support the survival of bacteria within host macrophages by preventing the oxidative burst by phagocytes. CD-associated *E. coli* survival within the macrophage is associated with their resistance to superoxide (Tawfik et al., 2019). Under normal circumstances, activated macrophages generate reactive oxygen and nitrogen species as a defence mechanism against microorganisms during infection.

In our study, catalase-peroxidase activity may allow *E. coli* to survive within lymph nodes, subsequently provide a continuous stimulus to the host immune system. However, the number of *E. coli* isolates from lymph node and paediatric CD patients

in this study was small (13 lymph node, 33 paediatric CD, and 23 healthy controls) and functional significance of this association requires further investigation.

6.1.3 NLRP3 mediated inflammasome activation by IBD and non-IBD associated *E. coli* isolates

Inflammatory bowel disease is the chronic and relapsing inflammation of the gastrointestinal tract, and it is hypothesized that, this chronic inflammation is a result of dysregulated innate immune response in the gut towards intestinal microflora or pathogens (Goyette et al., 2007). NLRP3, a member of NLRs family has emerged as a critical regulatory factor of intestinal immune response and gut homeostasis. This immune mediator can detect the microbial signals and promote the secretion of proinflammatory cytokines interleukin 1β (IL-1β) and IL-18, that has been associated with IBD (Mao et al., 2018). Previous reports have suggested a pathogenic role of IL-1β and IL-18 to the intestinal inflammation. Early studies also found that many virulence factors in *E. coli* could be associated with NLRP3 inflammasome activation as well as disease progression. The role of NLRP3 inflammasome in the pathogenesis of UTI by uropathogenic *Escherichia coli* (UPEC) was extensively studied (Nagamatsu et al., 2015; Demirel et al., 2018; Verma et al., 2019). However, the genes and molecular properties of *E. coli* from a patient with CD has not been properly investigated. As in our microbiome study we also found that *E. coli* is significantly associated with CD and we also demonstrated that some genes and molecular properties of *E. coli* was associated with CD, we wanted to check if these genes and molecular properties of *E. coli* play any role in inflammasome activation in patients with IBD. In this study, we could not find any significant difference in the activation of the NLRP3 inflammasome by *E. coli* isolates between CD and controls. Our present study did not find any molecular property of *E. coli* associated with inflammasome

activation. However, based on our pan-genome analysis and genome-wide association study, we demonstrated that some genes of *E. coli* are associated with NLRP3 inflammasome activation. Among these genes, the most important ldrD_3 gene encodes small toxic polypeptide LdrD, a toxic component of a type I toxin-antitoxin (TA) system. This toxin protein causes rapid cell death and nucleoid condensation of the host organism. This toxin is regulated by the sRNA antitoxin, RdlD of these toxin-antitoxins (TA) system. This type I toxin-antitoxins system exists on the *E. coli* chromosome, but their biological functions are not clearly known.

However, lymph node clones from CD patient 57 did not activate the NLRP3 inflammasome and did not harbour many of the NLRP3 activation-associated genes. However, it carries the ycgZ gene involved in the cellular response to an acidic environment and PhoH, which provides protection of DNA and stress resistance, enabling them to survive within the macrophage. This result is consistent with our lymph node comparison, where lymph node isolates were compared to healthy control isolates. We found genes associated with lymph node strains that may help them to better survive in a phagolysosome. The unaffected mucosa and involved mucosa-associated isolates from the same patient had many more genes associated with NLRP3 activation. Therefore, that the lymph node clone strain did not activate the inflammasome is likely due to the loss of these genes in this clone pair, and may allow it to evade host responses upon translocation.

In addition, genes (group_722 and fliC) coding for protein FliC family flagellin were significantly associated with NLRP3 inflammasome. Previous reports demonstrated that flagellin is required for caspase-1-mediated pyroptosis of macrophages infected with intracellular pathogens (Molofsky et al., 2006; Ren et al., 2006). The flagellin of *Salmonella enterica serovar typhimurium* activated NLRP3 inflammasomes in human

macrophages (Gram et al., 2021). That study also showed that flagellin from *S. typhimurium* induced pyroptotic cell death of primary human monocyte-derived macrophages (Gram et al., 2021). However, in our study, we saw that activation of NLRP3 is not associated with the clearing of *E. coli* from macrophages, nor pyroptotic cell death of mouse bone marrow-derived macrophages. Instead, *E. coli* increased in number within the macrophages. This result indicates that *E. coli* in patients with CD may have acquired some genes, such as those described above in the comparison of lymph node and mucosal isolates, that confer increased survival inside macrophages, thereby possible providing a continuous stimulus to the immune system and thereby release of pro-inflammatory signals leading to chronic inflammation in CD. By understanding these pathways of how macrophages respond to *E. coli* and fail to eradicate them from the gut, we may explain the underlying mechanism of CD pathogenesis. Future studies should deeply explore *E. coli* macrophage interactions in CD, which may lead to the development of effective therapies for CD.

6.2 References

Armstrong, H., Alipour, M., Valcheva, R., Bording-Jorgensen, M., Jovel, J., Zaidi, D., et al. (2019). Host immunoglobulin G selectively identifies pathobionts in pediatric inflammatory bowel diseases. *Microbiome* 7(Goodhand et al.), 1. doi: 10.1186/s40168-018-0604-3.

Atarashi, K., Suda, W., Luo, C., Kawaguchi, T., Motoo, I., Narushima, S., et al. (2017). Ectopic colonization of oral bacteria in the intestine drives T(H)1 cell induction and inflammation. *Science (New York, N.Y.)* 358(6361), 359-365. doi: 10.1126/science.aan4526.

Braune, A., and Blaut, M. (2016). Bacterial species involved in the conversion of dietary flavonoids in the human gut. *Gut microbes* 7(3), 216-234.

Breyner, N.M., Michon, C., de Sousa, C.S., Vilas Boas, P.B., Chain, F., Azevedo, V.A., et al. (2017). Microbial anti-inflammatory molecule (MAM) from Faecalibacterium prausnitzii shows a protective effect on DNBS and DSS-induced colitis model in mice through inhibition of NF-κB pathway. *Frontiers in microbiology* 8, 114.

Clooney, A.G., Eckenberger, J., Laserna-Mendieta, E., Sexton, K.A., Bernstein, M.T., Vagianos, K., et al. (2021). Ranking microbiome variance in inflammatory bowel disease: a large longitudinal intercontinental study. *Gut* 70(3), 499. doi: 10.1136/gutjnl-2020-321106.

Demirel, I., Persson, A., Brauner, A., Särndahl, E., Kruse, R., and Persson, K. (2018). Activation of the NLRP3 Inflammasome Pathway by Uropathogenic Escherichia coli Is Virulence Factor-Dependent and Influences Colonization of Bladder Epithelial Cells. *Frontiers in cellular and infection microbiology* 8**,** 81-81. doi: 10.3389/fcimb.2018.00081.

Dheer, R., Davies, J.M., Quintero, M.A., Damas, O.M., Deshpande, A.R., Kerman, D.H., et al. (2020). Microbial Signatures and Innate Immune Gene Expression in Lamina Propria Phagocytes of Inflammatory Bowel Disease Patients. *Cellular and Molecular Gastroenterology and Hepatology* 9(3)**,** 387-402. doi: https://doi.org/10.1016/j.jcmgh.2019.10.013.

Dickson, I. (2020). Creeping fat in Crohn's disease explained. *Nature Reviews Gastroenterology & Hepatology* 17(12)**,** 713-713. doi: 10.1038/s41575-020-00379-0.

Frank, D.N., St. Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences* 104(34)**,** 13780-13785. doi: 10.1073/pnas.0706625104.

Franzosa, E.A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H.J., Reinker, S., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature microbiology* 4(Polito et al.)**,** 293-305.

Galazzo, G., Tedjo, D.I., Wintjens, D.S.J., Savelkoul, P.H.M., Masclee, A.A.M., Bodelier, A.G.L., et al. (2019). Faecal Microbiota Dynamics and their Relation to Disease Course in Crohn's Disease. *Journal of Crohn's and Colitis* 13(10)**,** 1273-1282. doi: 10.1093/ecco-jcc/jjz049.

Gevers, D., Kugathasan, S., Denson, L.A., Vazquez-Baeza, Y., Treuren, W., and Ren, B. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15. doi: 10.1016/j.chom.2014.02.005.

Goyette, P., Labbe, C., Trinh, T.T., Xavier, R.J., and Rioux, J.D. (2007). Molecular pathogenesis of inflammatory bowel disease: genotypes, phenotypes and personalized medicine. *Ann Med* 39(3)**,** 177-199. doi: 10.1080/07853890701197615.

Gram, A.M., Wright, J.A., Pickering, R.J., Lam, N.L., Booty, L.M., Webster, S.J., et al. (2021). Salmonella Flagellin Activates NAIP/NLRC4 and Canonical NLRP3 Inflammasomes in Human Macrophages. *The Journal of Immunology* 206(3)**,** 631-640. doi: 10.4049/jimmunol.2000382.

Gupta, A., Dhakan, D.B., Maji, A., Saxena, R., P.K., V.P., Mahajan, S., et al. (2019). Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. *mSystems* 4(6)**,** e00438-00419. doi: doi:10.1128/mSystems.00438-19.

Hall, A.B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., et al. (2017). A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome medicine* 9(Goodhand et al.)**,** 1-12.

Henke, M.T., Kenny, D.J., Cassilly, C.D., Vlamakis, H., Xavier, R.J., and Clardy, J. (2019). Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proceedings of the National Academy of Sciences* 116(26)**,** 12672. doi: 10.1073/pnas.1904099116.

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., et al. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60(5)**,** 631-637. doi: 10.1136/gut.2010.223263.

Ju, S., Ge, Y., Li, P., Tian, X., Wang, H., Zheng, X., et al. (2018). Dietary quercetin ameliorates experimental colitis in mouse by remodeling the function of colonic macrophages via

a heme oxygenase-1-dependent pathway. *Cell Cycle* 17(Goodhand et al.), 53-63. doi: 10.1080/15384101.2017.1387701.

Kowalska-Duplaga, K., Gosiewski, T., Kapusta, P., Sroka-Oleksiak, A., Wędrychowicz, A., Pieczarkowski, S., et al. (2019). Differences in the intestinal microbiome of healthy children and patients with newly diagnosed Crohn's disease. *Scientific reports* 9(Goodhand et al.), 18880-18880. doi: 10.1038/s41598-019-55290-9.

Lavelle, A., and Sokol, H. (2020). Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol* 17(4), 223-237. doi: 10.1038/s41575-019-0258-z.

Lopez-Siles, M., Khan, T.M., Duncan, S.H., Harmsen, H.J.M., Garcia-Gil, L.J., and Flint, H.J. (2012). Cultured Representatives of Two Major Phylogroups of Human Colonic Faecalibacterium prausnitzii Can Utilize Pectin, Uronic Acids, and Host-Derived Substrates for Growth. *Applied and Environmental Microbiology* 78(Polito et al.), 420-428. doi: doi:10.1128/AEM.06858-11.

Louis, P., and Flint, H.J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiology Letters* 294(Goodhand et al.), 1-8. doi: 10.1111/j.1574-6968.2009.01514.x.

Mao, L., Kitani, A., Strober, W., and Fuss, I.J. (2018). The Role of NLRP3 and IL-1β in the Pathogenesis of Inflammatory Bowel Disease. *Frontiers in Immunology* 9. doi: 10.3389/fimmu.2018.02566.

Martín, R., Chain, F., Miquel, S., Lu, J., Gratadoux, J.-J., Sokol, H., et al. (2014). The commensal bacterium Faecalibacterium prausnitzii is protective in DNBS-induced chronic moderate and severe colitis models. *Inflammatory bowel diseases* 20(3), 417-430.

Mirsepasi-Lauridsen, H.C., Vrankx, K., Engberg, J., Friis-Møller, A., Brynskov, J., Nordgaard-Lassen, I., et al. (2018). Disease-Specific Enteric Microbiome Dysbiosis in Inflammatory Bowel Disease. *Frontiers in Medicine* 5(304). doi: 10.3389/fmed.2018.00304.

Molofsky, A.B., Byrne, B.G., Whitfield, N.N., Madigan, C.A., Fuse, E.T., Tateda, K., et al. (2006). Cytosolic recognition of flagellin by mouse macrophages restricts Legionella pneumophila infection. *The Journal of experimental medicine* 203(4), 1093-1104.

Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome biology* 13(9), R79.

Musumeci, L., Maugeri, A., Cirmi, S., Lombardo, G.E., Russo, C., Gangemi, S., et al. (2020). Citrus fruits and their flavonoids in inflammatory bowel disease: An overview. *Natural product research* 34(Goodhand et al.), 122-136.

Nagamatsu, K., Hannan, T.J., Guest, R.L., Kostakioti, M., Hadjifrangiskou, M., Binkley, J., et al. (2015). Dysregulation of Escherichia coli alpha-hemolysin expression alters the course of acute and persistent urinary tract infection. *Proc Natl Acad Sci U S A* 112(8), E871-880. doi: 10.1073/pnas.1500374112.

Nishida, A., Inoue, R., Inatomi, O., Bamba, S., Naito, Y., and Andoh, A. (2018). Gut microbiota in the pathogenesis of inflammatory bowel disease. *Clinical journal of gastroenterology* 11(Goodhand et al.), 1-10.

Nishino, K., Nishida, A., Inoue, R., Kawada, Y., Ohno, M., Sakai, S., et al. (2018). Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *Journal of gastroenterology* 53(Goodhand et al.), 95-106.

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., et al. (2017). A microbial signature for Crohn's disease. *Gut* 66(5), 813-822.

Png, C.W., Lindén, S.K., Gilshenan, K.S., Zoetendal, E.G., McSweeney, C.S., Sly, L.I., et al. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augmentin vitroutilization of mucin by other bacteria. *Official journal of the American College of Gastroenterology| ACG* 105(11)**,** 2420-2428.

Rashid, T., Ebringer, A., Tiwana, H., and Fielder, M. (2009). Role of Klebsiella and collagens in Crohn's disease: a new prospect in the use of low-starch diet. *European Journal of Gastroenterology & Hepatology* 21(8).

Rashid, T., Ebringer, A., and Wilson, C. (2013). The role of Klebsiella in Crohn's disease with a potential for the use of antimicrobial measures. *International journal of rheumatology* 2013**,** 610393-610393. doi: 10.1155/2013/610393.

Ren, T., Zamboni, D.S., Roy, C.R., Dietrich, W.F., and Vance, R.E. (2006). Flagellin-deficient Legionella mutants evade caspase-1-and Naip5-mediated macrophage immunity. *PLoS Pathog* 2(3)**,** e18.

Salaritabar, A., Darvishi, B., Hadjiakhoondi, F., Manayi, A., Sureda, A., Nabavi, S.F., et al. (2017). Therapeutic potential of flavonoids in inflammatory bowel disease: A comprehensive review. *World journal of gastroenterology* 23(28)**,** 5097.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L.G., Gratadoux, J.-J., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences* 105(43)**,** 16731-16736.

Tawfik, A., Knight, P., Duckworth, C.A., Pritchard, D.M., Rhodes, J.M., and Campbell, B.J. (2019). Replication of Crohn's Disease Mucosal E. coli Isolates inside Macrophages Correlates with Resistance to Superoxide and Is Dependent on Macrophage NF-kappa B Activation. *Pathogens* 8(Polito et al.)**,** 74.

Verma, V., Gupta, S., Kumar, P., Yadav, S., Dhanda, R.S., Gaind, R., et al. (2019). Involvement of NLRP3 and NLRC4 Inflammasome in Uropathogenic E. coli Mediated Urinary Tract Infections. *Front Microbiol* 10**,** 2020. doi: 10.3389/fmicb.2019.02020.

Vester-Andersen, M., Mirsepasi-Lauridsen, H., Prosberg, M., Mortensen, C., Träger, C., Skovsen, K., et al. (2019). Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. *Scientific reports* 9(Goodhand et al.)**,** 1-10.

Vezza, T., Rodríguez-Nogales, A., Algieri, F., Utrilla, M.P., Rodriguez-Cabezas, M.E., and Galvez, J. (2016). Flavonoids in Inflammatory Bowel Disease: A Review. *Nutrients* 8(4)**,** 211.

Willing, B., Halfvarson, J., Dicksved, J., Rosenquist, M., Jarnerot, G., Engstrand, L., et al. (2009). Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis* 15(5)**,** 653-660. doi: 10.1002/ibd.20783.

Willing, B.P., Dicksved, J., Halfvarson, J., Andersson, A.F., Lucio, M., Zheng, Z., et al. (2010). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 139(6)**,** 1844-1854. e1841.

Winter, S.E., Winter, M.G., Xavier, M.N., Thiennimitr, P., Poon, V., Keestra, A.M., et al. (2013). Host-derived nitrate boosts growth of E. coli in the inflamed gut. *science* 339(6120)**,** 708-711.

Xu, J., Liang, R., Zhang, W., Tian, K., Li, J., Chen, X., et al. (2020). Faecalibacterium prausnitzii-derived microbial anti-inflammatory molecule regulates intestinal integrity in diabetes mellitus mice via modulating tight junction protein expression. *Journal of diabetes* 12(3)**,** 224-236. doi: 10.1111/1753-0407.12986.