META-ANALYSIS OF GUT MICROBIAL ASSOCIATIONS WITH MURINE ULCERATIVE COLITIS

Salma El-Sahhar

A thesis submitted for
the degree of Doctor of Philosophy

School of Life Sciences
University of Essex

July 2023

Abstract

There has been a growing interest in understanding the microbiome and the role it plays in disease and maintaining homeostasis since the launch of the Human Microbiome Project. While some members of the gut flora have proven to be beneficial to the host other commensals have been linked to all sorts of disease. Inflammatory bowel diseases (IBD) are one of the most investigated group of diseases within the realm of intestinal microbial communities. Pathogenesis of ulcerative colitis (UC), one of the two main forms of IBD, is highly dependent and influenced by gut microbial composition. Despite the amount of substantial research conducted on the microbiome within the context of IBD, the role played by microbes in this group of diseases remains poorly understood. Some microbial associations have been defined and are consistently found in UC patients. However, there is a lack of consistent microbial associations in colitis mouse models owed to multiple factors, like genetics, diet, and the environment, introducing variability in the structure of the gut microbiome. Variations in bioinformatics pipelines used in microbiome analysis is one of the shortcomings contributing to the absence of reproducible links between with gut microbiome and UC.

Using meta-analysis, this thesis aims to reduce the heterogeneity of microbiome data by standardising computational methods across 13 published datasets. In doing that, first relevant datasets were screened, selected, and pre-processed prior to analysis. A pipeline using the ASV method was optimised and tested before the analysis. Initially, a total of 27 datasets were used for the meta-analysis only to find that the different types of colitis mouse models introduced a strong batch effect making it even more difficult to decipher an already complex community. Based on that, the focus of the meta-analysis shifted to investigating gut microbial patterns associated with colitis in the DSS mouse model, which is one of the most commonly used models in experimental colitis. To untangle microbiota data further, two additional meta-

analyses were done on different groupings of the chosen DSS datasets. Prior to that each dataset was analysed individual to better understand the nature of the data and the inter-population differences encountered in the meta-analysis.

Links between members of the gut microbiota and both colitogenic and colitis states in mice were established in this thesis. Consistent with the current literature, the meta-analysis identified a decrease in the phyla Firmicutes and increase in Bacteroidetes in mice prone to colitis. Peptostreptococaceae species, Erysipelotrichales species, as well as, described pathogens like Helicobacter and *Esherichia*/*Shigella spp* were found predominantly in both colitogenic and colitis conditions in mice. Furthermore, this thesis first reports associative links between the genera *Terrisporobacter* and *Herbinix* and experimental UC.

The last chapter of this thesis focuses on the role of the microbiome in phenotypic modulations in the intestinal epithelial lining elicited by dietary inulin. This work was done in collaboration with the Laboratory of Immunoinflammation in University of Campinas, Brazil. Using flow cytometry, immunostaining, clonogenicity assay, and transcriptomic sequencings they described enhanced proliferative activity of $Lgr5^{+}$ cells in colonic crypts when mice were put on an inulin diet. Additional experiments, including 16S analysis of microbiome data from SPF mice, GF mice with faecal matter transplant (FMT) and gnotobiotic mice showed that the observed phenotype is dependent on gut microbial composition.

Acknowledgments

Firstly, I'd like to express my gratitude to both my supervisor Dr. Patrick Varga-Weisz for granting me the opportunity to embark on this PhD his support and guidance throughout these unchartered territories. Thank you for trusting my abilities in performing my research and constantly challenging me and pushing me out of my comfort zone which allowed me to grow and develop my skill greatly.

I'd also like to thank my other supervisor Dr. Nicolae Radu Zabet for all his support and words of encouragement. Thank you for welcoming me into your lab once I lost all hope post project change. I appreciate all the help you have provided me and your constant reminders to take breaks.

I would also like to thank Dr. Boyd McKew, Professor Alex Dumbrell, and Dr. David Clark for their guidance and valuable advice during my early days of 16S analysis. I am very grateful to Stuart Newman who is always quick to help when issues with the cluster arise. I thank all members of both my labs and members of the Genomics group, especially Liv, Mia and Ellie who've made my PhD experience a lot more pleasant.

A special thanks goes out to my husband, Obi, for showering me with support throughout this PhD. Thank you for all time you spent helping me debug my code, generating my all-time favourite (NOT) heatmaps and keeping me hydrated during my write-up period. Thanks to my friends Kay and Sara for their constant support and check-ins throughout this journey. Thanks to my parents for flying out from Egypt to hold down the fort while I finish writing my thesis and their never-ending support. Finally, thanks to my cat Maple for keeping me company and bringing smiles to my face during all the tough times in this PhD journey.

Impact of COVID-19 Statement

The COVID-19 pandemic has impacted my thesis significantly. The beginning of my program was in October 2019, due to the pandemic lab access was restricted from March 2020 till July 2020. The original project for this thesis was to develop a droplet-based single cell transcriptomic platform for microbiomics. I had to drop this project during my second year due to limited lab access imposed by COVID-19 precautionary measures. I started a new fully computationally based project (the one presented in this thesis) with minimal bioinformatics experience and none in microbiomics, forcing me to retrain and pick up a new skill set in order to complete this thesis.

Declaration

I, Salma El-Sahhar, declare that the data used in this thesis was publicly available and obtained using the accession codes specified in chapter 2. The data analysed in chapter 5 was generated by a collaborating lab in University of Campinas, Brazil.

Table of Contents

Abstract **1**

List of figures **8**

List of tables **10**

Abbreviations and Acronyms **11**

Chapter 1 – Introduction **12**

1.1 Chapter Overview 12

1.2 Introduction 12

1.3 Genetic analysis highlights the role of the gut microbiota in IBD 17

1.4 How do changes in the microbiome influence IBD? 28

1.5 How does the environment contribute to gut microbiome changes? 36

1.6 Can microbial-based treatments be the answer? 43

1.7 Conclusion and Outlook 48

Chapter 2 – Methods **51**

2.1 Chapter Overview 51

2.2 Introduction 51

2.3 Study Selection 56

2.4 Data Retrieval 59

2.5 Metadata Curation 61

2.6 Dataset Processing 63

2.6.1 OTU Method 63

2.6.2 ASV Approach 64

2.7 Taxonomic Profiling and Microbial Community Analysis 67

2.7.1 Microbiome Analysis 67

2.7.2 MMUPHin 67

2.7.3 GitHub Repository ……. 68

Chapter 3 – Results ……... **69**

3.1 Chapter Overview ……. 69

3.2 Study Selection ……. 69

3.3 Dataset Processing ……. 70

3.4 Meta-Analysis on DSS Mouse Model Studies ……. 78

3.4.1 Batch Effect Adjustment and Meta-Analysis on All Datasets …… 78

3.4.2 Meta-Analysis on DSS Datasets …… 78

3.4.3 Meta-Analysis on Different. Groupings of DSS Datasets …..…….....84

3.5 Taxonomic Profiling and Microbial Community Analysis ……..... 90

3.5.1 Surana & Kasper …….……... 90

3.5.2 Tran *et al.* …….…..... 92

3.5.3 Spalinger *et al.* ……..... 95

3.5.4 Roy *et al.* ……..... 98

3.5.5 Song *et al.* …….....…….....101

3.5.6 Nakajima *et al.* ……..... 103

3.5.7 Kazakevych *et al.* ……...……..... 106

3.5.8 Qi *et al.* ……...……..... 109

3.5.9 Mamantopoulos *et al.* ……...……..... 112

3.5.10 Schmidt *et al.* ……...……..... 113

3.5.11 Kimura *et al.* ……...……..... 115

3.5.12 Khan *et al.* ……...……..... 116

3.5.13 Ma *et al.* ……...……..... 118

3.6 Results Summary ……...……...…..... 120

Chapter 4: Discussion **125**

4.1 Thesis Summary and Importance 125

4.2 Gut Microbial Associations with UC in Humans and Mice 131

4.3 Limitations and Challenges 133

4.4 Future Work 135

4.5 Conclusions 135

Chapter 5: Faecal matter transfer in GF shows that phenotypic changes in intestinal epithelial lining modulated by the immune system are microbiota dependent **137**

5.1 Chapter Overview 137

5.2 Manuscript summary 137

5.3 Methods 139

5.3.1 Data Processing 139

5.3.2 Data Analysis and Taxonomic Profiling 139

5.4 Results 140

5.4.1 Data Processing 140

5.4.2 Data Analysis and Taxonomic Profiling 141

5.5 Conclusion 145

References **146**

List of Figures

Figure 1.1 Histogram of the number of studies on the human microbiome 13

Figure 1.2 Illustration summarising factors influencing the microbiome 15

Figure 1.3 Graphic representation of the intestinal epithelial barrier 20

Figure 1.4 Key features linked to a healthy vs IBD microbiome 35

Figure 2.1 Illustration of the 16S rRNA gene 53

Figure 2.2 Schematic of meta-analysis design 58

Figure 3.1 Mouse models employed by studies included in meta-analysis 70

Figure 3.2 ASV vs OTU: taxonomic profiling 72

Figure 3.3 Heatmap of ASV vs OTU comparison 74

Figure 3.4 Ordination plots of all colitis datasets 78

Figure 3.5 Ordination plots of all DSS datasets 81

Figure 3.6 Bar plot of abundant phyla in condition groups 82

Figure 3.7 DESeq2 heatmap of all DSS datasets 83

Figure 3.8 Ordination plots of group A datasets 85

Figure 3.9 Ordination plots of group B datasets 86

Figure 3.10 DESeq2 heatmap of group A datasets 87

Figure 3.11 DESeq2 heatmap of group B datasets 88

Figure 3.12 Ordination plot of Surana & Kasper (2017) 92

Figure 3.13 DESeq2 heatmap of Surana & Kasper (2017) 92

Figure 3.14 Ordination plot of Tran *et al.* (2019) 93

Figure 3.15 DESeq2 heatmap of Tran *et al.* (2019) 94

Figure 3.16 Ordination plots of Spalinger *et al*. (2019) 97

Figure 3.17 DESeq2 heatmap of Spalinger *et al.* (2019) 98

Figure 3.18 Ordination plots of Roy *et al*. (2017) 100

Figure 3.19 DESeq2 heatmap of Roy *et al.* (2017) 101

Figure 3.20 Ordination plot of Song *et al*. (2020) 102

Figure 3.21 DESeq2 heatmap of Song *et al.* (2020) 103

Figure 3.22 Ordination plots of Nakajima *et al*. (2018) 104

Figure 3.23 DESeq2 heatmap of Nakajima *et al*. (2018) 105

Figure 3.24 Ordination plots of Kazakevych *et al.* (2020) 107

Figure 3.25 DESeq2 heatmap of Kazakevych *et al*. (2020) 108

Figure 3.26 Ordination plot of Qi *et al*. (2020) 109

Figure 3.27 DESeq2 heatmap of Qi *et al*. (2020) 111

Figure 3.28 Ordination plot of Mamantopoulos *et al*. (2017) 112

Figure 3.29 DESeq2 heatmap of Mamantopoulos *et al*. (2017) 113

Figure 3.30 Ordination plot of Schmidt *et al*. (2019) 114

Figure 3.31 DESeq2 heatmap of Schmidt *et al*. (2019) 114

Figure 3.32 Ordination plot of Kimura *et al*. (2020) 115

Figure 3.33 DESeq2 heatmap of Kimura *et al*. (2020) 116

Figure 3.34 Ordination plots of Khan *et al*. (2020) 117

Figure 3.35 DESeq2 heatmap of Khan *et al*. (2020) 118

Figure 3.36 Ordination plots of Ma *et al*. (2020) 119

Figure 3.37 DESeq2 heatmap of Ma *et al*. (2020) 120

Figure 5.1 Ordination plots of Corrêa *et al.* (2023) 143

Figure 5.2 Heatmap and LDA barplot of Corrêa *et al.* (2023) 144

Table List

Table 1.1 Literature review glossary 16

Table 2.1 Initial list of studies for meta-analysis inclusion 60

Table 2.2 Arguments/parameters used in data OTU and ASV pipelines 65

Table 3.1 Pipeline optimisation and comparison 71

Table 3.2 Summary of pre-processed datasets 74

Table 3.3 Summary of DADA2 processed datasets 76

Table 3.4 Grouping of studies used in meta-analysis 84

Table 3.5 P values for Wilcoxon Test on groups A and B 84

Table 3.6 Summary of differentially abundant Clostridiales species 122

Table 3.7 Summary of differentially abundant Bacteroidales species 123

Table 3.8 Summary of differentially abundant Erysipelotrichales species and other taxa 124

Table 5.1 Parameter comparison from Corrêa *et al.* (2023) 141

Abbreviations and Acronyms

Abbreviation/ Acronym	Definition
16S data	16S rRNA amplicon sequencing data
ASV	Amplicon sequence variant
CD	Crohn's disease
DADA2	Divisive Amplicon Denoising Algorithm 2
DSS	Dextran sulfate sodium
FDR	False discovery rate
FMT	Faecal matter/microbiota transplant
GF	Germ free
Gnotobiotic	An animal with a defined flora
HDAC	Histone deacetylase
IBD	Inflammatory bowel diseases
IECs	Intestinal epithelial cells
IL	Interleukin
KO	Knockout
MANOVA	Multivariate analysis of variance
MMUPHin	Meta-Analysis Methods with a Uniform Pipeline for Heterogeneity in microbiome studies
OTU	Operational taxonomic unit
PCoA	Principal coordinate analysis
PERMANOVA	Permutational multivariate analysis of variance
Relative abundance	Evenness of distribution of species among other taxa in a community (percentage)
RNAseq	mRNA bulk sequencing
SCFA	Short-chain fatty acids
scRNAseq	single cell mRNA sequencing
SPF	Specific pathogen free
Th17	Pro-inflammatory cells that secrete IL17, IL-21, and IL22 cytokines
UC	Ulcerative colitis
WT	Wildtype

Chapter 1: Literature Review

1.1 Chapter Overview

The gut microbiome is known to play an essential role in maintaining homeostasis in the human body. Studies have shown that shifts in gut microbial composition can lead to disease. Inflammatory bowel diseases (IBD) are a group of diseases that are highly associated with the microbiome. Ulcerative colitis (UC) is one of the two main forms of IBD and is the focus of this thesis. The aetiology of UC remains elusive but is highly dependent on and influenced by gut microbes. Environmental cues, particularly dietary and lifestyle changes, seem to play a huge role in development of the disease. In this chapter, a comprehensive literature review on links between the gut microbiota and IBD is presented as a book chapter (titled "The gut microbiome in health and disease: Inflammatory bowel disease") published in *Advances in Ecological Research* on the 27th of October 2022. This chapter provides an in-depth description of the effects of genetics and the environment on the gut microbiota and IBD. It also provides an extensive background for the work covered throughout this thesis.

1.2 Introduction

When Japan's long-standing prime minister Shinzo Abe resigned in 2020 because of his ulcerative colitis, it put the debilitating nature of inflammatory bowel diseases in the international press. What is largely not appreciated is that these diseases seem to reflect an aberrant reaction to what is normally an innocuous gut microbiome. Despite their bad reputation in relation to health, microbes are important for maintaining a state of mental and physical health and wellness. Indeed, there are at least as many microbes inhabiting our bodies,

primarily in the colon, as there are human cells (Gilbert and Stephens, 2018; Sender *et al*., 2016).

Bacteria, viruses, fungi and archaea reside in communities throughout the body, and are collectively known as the human microbiome. Yet, when most studies mention the microbiome, they are usually referring to bacterial communities rather than the whole assembly of microbial organisms. The term "microbiome" was coined by the Nobel laureate, Joshua Lederberg, in 2001 (Prescott, 2017). Research into the topic dates back to the 19th century (Farré-Maduell and Casals-Pascual, 2019). However, technical limitations in culturing (often anaerobe) constituents of the human microbiomes were the main reason why this field was not broadly explored. Recent advancements in high throughput DNA sequencing technologies and decreasing sequencing costs have been followed by a surge of investigations into the microbiome (figure 1.1).

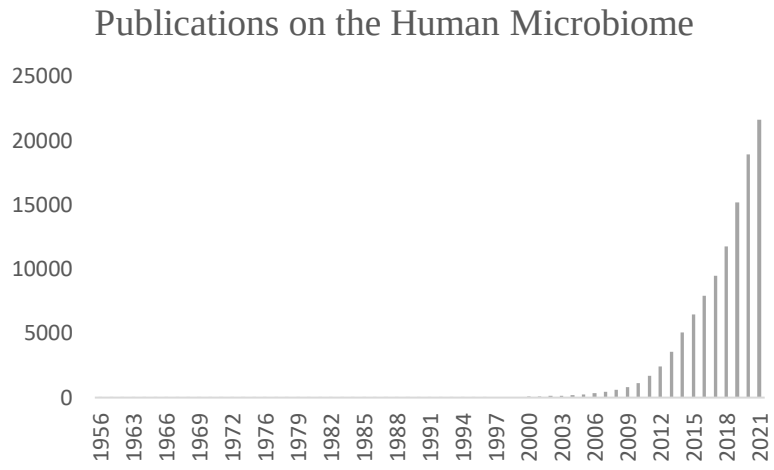


Figure 1.1: The number of studies on the human microbiome is on the rise since the boom in next generation sequencing in the early 2000s, generated by a basic search on PubMed (El-Sahhar & Varga-Weisz, 2022)**..**

Over the last decade, there has been an influx of new findings about the microbiome expedited by the Human Microbiome Project (HMP) (figure 1.1). The HMP is a group of interdisciplinary

projects launched throughout the US, Europe and Asia with an aim to discover new biomarkers associated with health and disease (Turnbaugh *et al.*, 2007). The first phase of the HMP laid the groundwork for successive studies on the interplay between the human host and microbes. It focused on characterising the structural and functional properties of microbial populations inhabiting five key body areas from a cohort of healthy, Western adults (Huttenhower *et al.*, 2012). 5,177 microbial taxonomic profiles were identified using 16S ribosomal RNA sequencing while whole genome shotgun/metagenomics sequencing elucidated microbial functional properties (Methé *et al.*, 2012). The results showed that the make-up and diversity of the microbial populations were differed markedly at each body site. Oral and gut communities were the most diverse compared to the populations present at other sites. However, microbial populations found in saliva among subjects were more or less similar (Huttenhower *et al.*, 2012). Vaginal sites accommodated the simplest of communities, low in diversity. In fact, high microbial diversity in vaginal tissues has been correlated with vaginosis (Kang *et al.*, 2015; Gilbert and Stephens, 2018). The skin microbiome was shown to be unique for each person. Overall, an individual's microbiome is stable over time and has a relatively unique makeup that contributes to the metabolic pathways essential for human health (Caporaso *et al.*, 2011). Fluctuations in the biodiversity and abundance of certain taxa in the microbiome are constantly occurring with environmental changes at different temporal rates depending on the individual (Flores *et al.*, 2014). There are many factors that influence microbial composition, particularly in the gut, which is the most studied community and is the focus of this chapter. Genetics, sex, age, the immune system, birth mode, diet, lifestyle, geography and exposure to antibiotics and other drugs have all been shown to play a significant role in shaping the host’s gut microbial communities (figure 1.2).

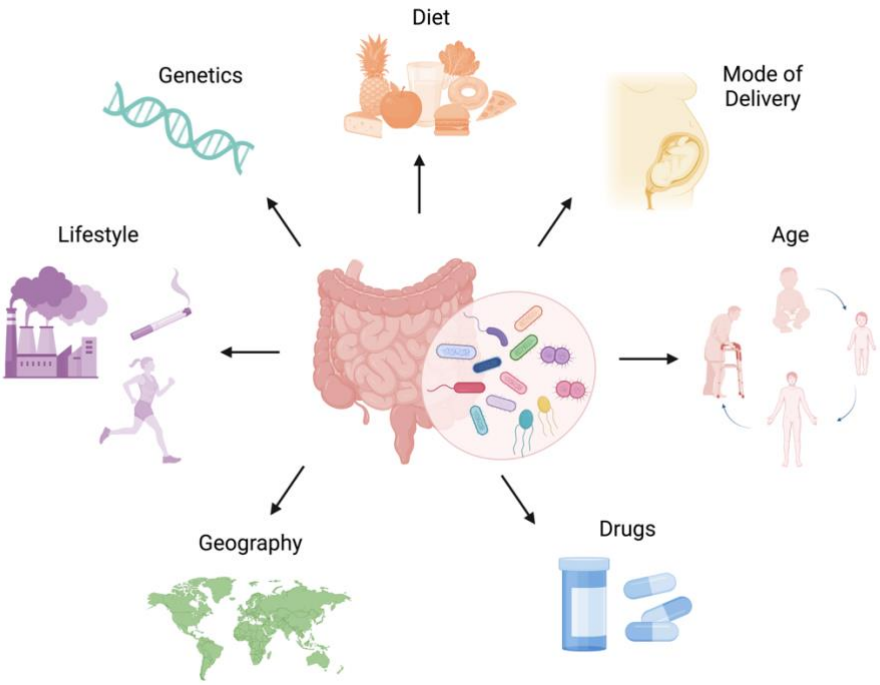


Figure 1.2: A summary graphic illustrating factors that influence the gut microbiome (El-Sahhar & Varga-Weisz, 2022)..

The second phase of the HMP explored the interplay between the host and their microbiota over time in representative microbiome-associated conditions, one of which is inflammatory bowel disease (IBD) (Proctor *et al.*, 2019). IBD are a group of disorders that are thought to be both affected by and in turn affect the gut microbiome (Manichanh *et al.*, 2006; Frank *et al.*, 2007). These disorders involve an altered immune response towards gut microbiota in genetically predisposed individuals, under the influence of certain environmental factors, such as diet, antibiotic exposure, lifestyle (Piovani *et al.*, 2019). Whether **dysbiosis** is a causative or risk factor in developing IBD is yet to be clearly understood. The multidimensionality of this condition makes it very challenging to draw causal links between the gut microbiota and IBD, despite it being the most widely studied disease in relation to the microbiome. Extensive research examining the associations between the microbiome and IBD will be discussed

thoroughly throughout this chapter. We include a Glossary (table 1.1) for the explanation of some of the concepts and technical terms in bold.

Table 1.1: Glossary

Term	Definition
Dysbiosis	An imbalance of the microbiome causing compositional and functional changes in its constituents
Intestinal Mucosa	The inner lining of the intestinal tract composed of mucus, intestinal epithelium, lamina propria and microbiota
Mucin	Glycoproteins that are the major components of the mucosa, which lubricate and protect the gastrointestinal tract
Lamina propria	Thin layer of loose connective tissue that forms part of the intestinal mucosa
Crypts	Intestinal glands in the colon
Granuloma	Aggregation of macrophages forming in response to chronic inflammation
Paneth cells	Secretory cells residing in the base of intestinal crypts
Colectomy	A surgical procedure that involves a segmental resection of the colon
Genome-wide association studies (GWAS)	Studies based on genetic variation, especially single nucleotide polymorphisms (SNPs), exploring the link between genotype and phenotype using many DNA samples from large populations. These are compared, e.g., using microarrays build around SNPs.
Pattern recognition receptors (PRR)	A class of receptors that detect specific molecular structures on the surface of microbes.
Toll-like receptor (TLR)	A class of receptors present on sentinel cells that play a role in the innate immune system by detecting microbial derived molecules
Nod-like receptor (NLR)	PRRs that detect pathogens and regulate innate immune responses
RIG-I-like receptors	Cytosolic PRRs that detect viruses
Autophagy	A cellular process that involves the engulfment of organelles and macromolecules by vesicles for degradation
Nuclear factor-κB (NF-κB)	A transcription factor involved in cell survival and cytokine production
Immunoglobulin A (IgA)	An antibody that plays a role in mucosal immunity
T-helper cells 17	A subset of pro-inflammatory T-helper cells that produce interleukin 17 (IL-17)
Transforming growth factor beta (TGF-β)	A pleiotropic cytokine that regulates immune responses in the gut, among other things
Cytokine	A small signalling protein involved in regulating immune responses
16S rRNA gene	A gene encoding for the prokaryotic small subunit 16S rRNA commonly used in phylogenetic studies
18S rRNA gene	A gene encoding for the eukaryotic small subunit 18S rRNA commonly used in phylogenetic studies
Internal Transcribed Spacer (ITS)	A region of non-coding DNA situated between the small-subunit ribosomal RNA (rRNA) and the large-subunit rRNA
Fluorescent in situ hybridization	A cytogenetic technique that targets chromosomal locations using fluorescent probes detected by a fluorescent microscope
Taxonomy Nomenclature	As this chapter predominantly refers to published literature that uses the old nomenclature, we use it here throughout, noting that the new nomenclature is not universally accepted. The new nomenclature can be found here https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.005056
Bacteriocin	Toxic proteins produced by specific species of bacteria to inhibit the growth of other related species
Regulatory T cells (Treg cells)	A specialised subpopulation of T cells that maintain homeostasis by suppressing immune responses
Histone deacetylase (HDAC)	An enzyme that plays a role in modulating chromatin architecture by removing acetyl groups from lysine residues in histone tails
Hypoxia-inducible factor 1-alpha (HIF1-α)	A subunit of transcription factor HIF1

Adherent-Invasive E. coli (AIEC)	Strains of *E. coli* that adhere to IECs and colonise the mucosa
Dendritic cells	Antigen presenting cells of the immune system
Occludin	A plasma membrane protein present in tight junctions
Ileal pouch anastomosis (IPAA)	A surgical procedure done post colectomies to restore gastrointestinal continuity
Perianal disease	A type of disease that occurs in CD patients that causes inflammation around the anus
Luminal disease	A type of disease that occurs in CD patients that causes inflammation of the intestinal lining and/or wall

1.3 Genetic analysis highlights the role of the gut microbiota in IBD

Digestive diseases, particularly Inflammatory Bowel Diseases (IBD), are one of the most explored areas within microbiome research (iHMP, 2014). IBD is one of the most prevalent gastrointestinal diseases globally, with rising incidence rates in newly industrialised countries in Asia, Africa and South America (Guan, 2019; Freeman *et al.*, 2021; Ng *et al*., 2021) . Europe and North America have reported the highest prevalence of IBD (Ng *et al*., 2021). However, the mortality rates associated with the disease are low (Freeman *et al.*, 2021). It is usually diagnosed in patients within the 20-40 years age bracket (Guan, 2019). The common symptoms include diarrhoea, bloody stools, abdominal pain, weight loss, obstruction of the gastrointestinal tract (GIT) and, in some cases, extra-gastrointestinal manifestations like rashes, arthritis and eye complications (Frank *et al.*, 2007; Ghosh and Premchand, 2015). IBD is characterised by chronic inflammation of the **intestinal mucosa** mediated by the immune system and driven by genetics and also environmental factors, including diet, smoking and socio-economic status (Ni *et al*., 2017). Crohn's disease (CD) and Ulcerative Colitis (UC) are the two main forms of IBD, about 5% of IBD cases remain unclassified due to unclear symptoms. UC is distinguished by mucosal inflammation starting in the rectum and extending proximally into the colon, while CD involves areas of transmural inflammation surrounded by normal mucosa that can span the entire gastrointestinal tract (Ni *et al.*, 2017). These traits are clearly observed in histopathology procedures, where biopsies from patients with UC show

inflammation restricted to the mucosa and submucosa with **mucin** depletion and increased neutrophil migration to the **lamina propria** and **crypts** forming micro-abscesses (Xavier and Podolsky, 2007; Guan, 2019). CD, on the other hand, is characterised by different histological features, including **granulomas** infiltrated by aggregations of macrophages and other inflammatory cells, as well as **Paneth cell** deficiency (Xavier and Podolsky, 2007). These histopathological hallmarks together with the symptoms presented by patients are used in determining the severity of the disease which impacts the course of treatment. Topical medication is prescribed to mild cases of UC, while mild-to-moderate cases require oral anti-inflammatories progressing to immunomodulatory drugs for severe cases (Ghosh and Premchand, 2015). Steroids are given to subdue flare-ups post remission periods. Long and frequent UC flares pose a risk of bowel cancer. About 25% of UC patients require **colectomy** surgery due the risk of cancer development or poor response to medical treatment (Ghosh and Premchand, 2015). Meanwhile, 80% of CD cases undergo some sort of surgery be it as minor as an abscess drainage or as complex as a bowel resection. However, the standard treatment for CD is immunomodulatory drugs and biological therapy for more severe cases (Ghosh and Premchand, 2015).

There is an underlying genetic component to the pathogenesis of IBD. **Genome-wide association studies (GWAS)** have been conducted to pinpoint IBD genetic risk loci (Jostins *et al*., 2012; Liu *et al*., 2015; Huang *et al*., 2017). Over 200 IBD-associated loci have been identified (Jostins *et al.*, 2012; Liu *et al.*, 2015; Graham and Xavier, 2020). Only a fraction of those markers has been validated and shown to have a causal role on development of IBD. A majority of the genes implicated in the pathology of the disease are involved in pathways that regulate mucosal integrity (barrier function), immunity and host-microbe interactions, thus

highlighting the role of the microbiota in the disease (Graham and Xavier, 2020). Below we discuss some of the key genes that have been linked to IBD.

The intestinal mucosal barrier plays a critical role in maintaining homeostasis and disease progression (Turner, 2009) (figure 1.3) by segregating intestinal epithelial cells (IECs) from luminal contents, like macromolecules, digestive enzymes and microbes – both commensals and pathogens (Okumura and Takeda, 2018). It acts as a physical barrier as well as a chemical one that signals to the immune system when needed. The mucosal lining is comprised of an undulating monolayer of specialised epithelial cells such as enterocytes, goblet cells and enteroendocrine cells and their precursors, organised into crypts and villi (Peterson and Artis, 2014). The crypts are comprised of intestinal stem cells (ISCs), transit amplifying cells as well as **Paneth cells** (in the small intestine), while villi consist of terminally differentiated cells, like absorptive enterocytes and goblet cells. This epithelial layer is physically protected via a mucus layer and the glycocalyx (Maloy and Powrie, 2011). The latter is a network of glycolipids and glycoproteins covering the villi, particularly absorptive IECs, preventing the passing of bacteria or unwanted molecules. A viscous gel-like fluid enriched in Mucin2 (MUC2), a glycoprotein secreted by goblet cells, forms the mucus layer. The large intestine has a double mucus layer for added protection, since it holds the largest microbial community in the body (Peterson and Artis, 2014) (figure 1.3, right side). The inner mucus layer is anchored to the intestinal epithelium and is usually impermeable, due to polymerisation and stratification of MUC2, blocking the entry of bacteria. Whereas the looser outer mucus layer is more porous thereby harbouring some bacterial species (Okumura and Takeda, 2018).

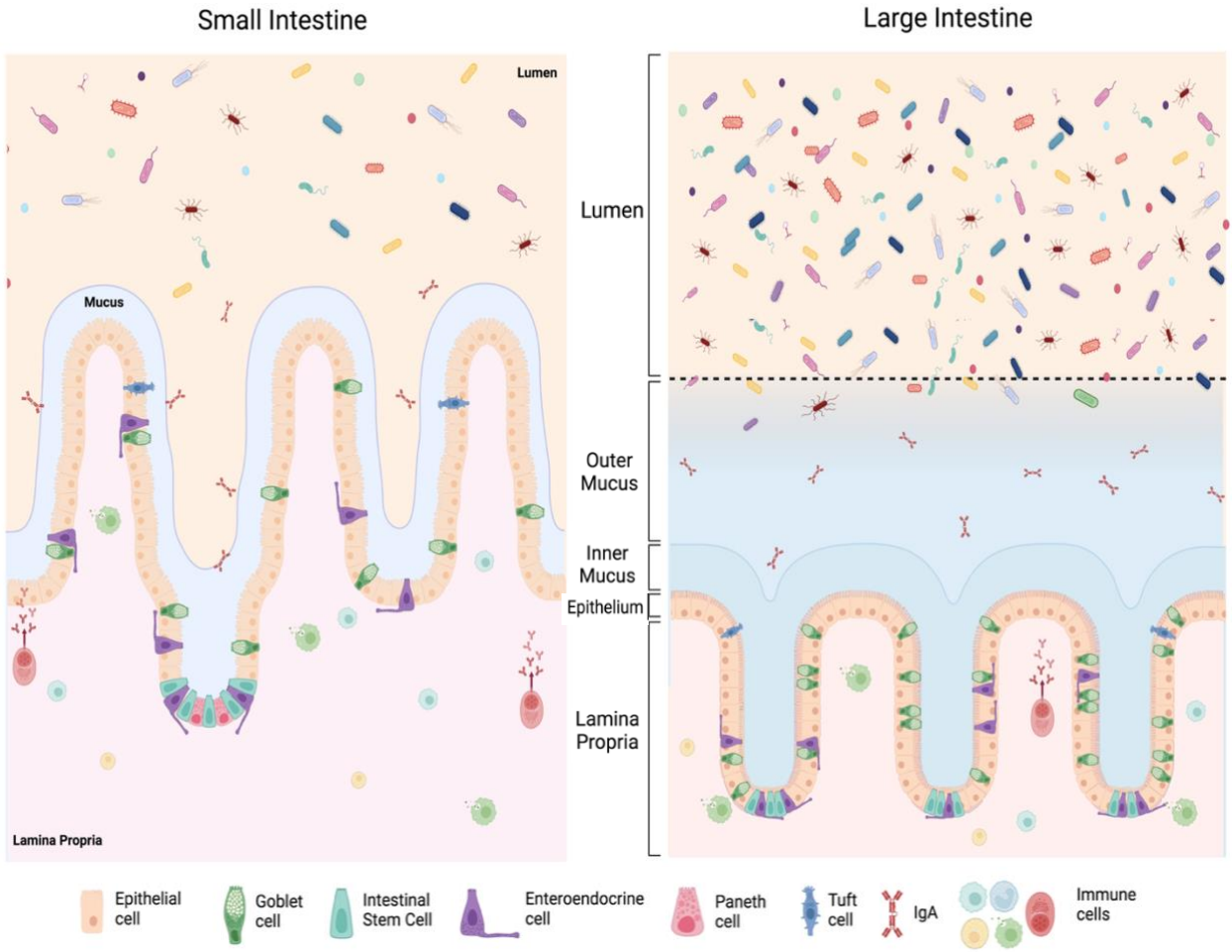


Figure 1.3: Intestinal epithelial barrier. The intestinal mucosa acts both as a chemical and physical barrier that protects IECs from harmful molecules and microbes present in the intestinal lumen. The physical barrier is comprised of a single mucus layer in the small intestine while a double layer of a firm inner mucus layer and a looser outer mucus layer. The porous outer mucus layer allows for some commensal habitation. AMPs, immunoglobulins, cytokines and other signalling molecules secreted by IECs and immune cells of the mucosa make up the chemical barrier (El-Sahhar & Varga-Weisz, 2022).

Patients with UC tend to have a disjointed and diminished mucus layer, especially during flare ups, allowing for direct contact between microbes and the epithelium (Kang *et al*., 2022). A decrease in MUC2 secretion accompanied by reduced numbers of goblet cells account for this phenotype which is reversed upon remission (Sun *et al.*, 2016). In contrast, CD patients tend to have a thicker mucus layer compared to healthy individuals, however, the properties of the mucus seem to be compromised (Sun *et al.*, 2016). GWA studies have linked IBD pathogenesis to mutations in several *MUC* genes, including *MUC2*, *MUC3* and *MUC19* (Kyo *et al*., 1999; Rivas *et al.*, 2011; McCole, 2014; Visschedijk *et al.*, 2016). A study done on two Dutch and one German cohorts has drawn a link between *MUC2* and UC susceptibility, although this association was lacking in the German cohort (Visschedijk *et al.*, 2016). A correlation between rare *MUC3A* alleles and both UC and CD was examined by Kyo *et al.* (2001). They hypothesised that conformational changes in mucin proteins encoded by these rare alleles could lead to a decrease in glycosylation – a mechanism which gives mucins their protective properties – making the glycoproteins more susceptible to bacterial proteases and degradation, thus affecting barrier stability (Kyo *et al*., 1999). Despite GWAS linking *MUC19* single nucleotide polymorphisms (SNPs) to IBD, a study in mice reported the absence of *MUC19* mRNA and protein in the intestines and their presence in the oral cavity (Das *et al.*, 2010) which questions the relevance of *MUC19* in the pathology of IBD. Further investigations into the role of *MUC* genes in IBD are required in determining the nature of the association. Despite this, the most commonly used experimental mouse model used to study colitis, the dextran sodium sulphate (DSS) model, is known to exhibit swift changes in the colon mucus layer, rendering it permeable to bacteria, upon DSS administration, bringing about mucosal inflammation (Johansson *et al.*, 2010). Therefore, it can be concluded that mucins play a part in IBD, given their integral role in maintaining the physical mucosal barrier, but their exact role in the pathogenesis of the disease should be investigated further.

The *Cadherin 1* (*CDH1*) gene encodes for the transmembrane epithelial protein, E-cadherin, essential for cell adhesion and maintaining tight junctions. It is present in the *IBD1* locus, along with a few other IBD susceptibility genes, and is known to be associated with CD and UC (Barrett *et al.*, 2009; Muise *et al.*, 2009). Mutations in *CDH1* can lead to increased mucosal permeability, allowing for the infiltration of luminal microbes through the intestinal wall, with the exact mechanisms still unknown (Younis *et al.*, 2020). SNPs in *CDH1* in *in vitro* models result in a cytoplasmic accumulation of the protein rather than localisation to the plasma membrane – its expected location (Muise *et al.*, 2009). This observation is shared in CD patients carrying the same SNPs, with increased cytoplasmic E-cadherin staining. Deficient E-cadherin compromises the integrity of the epithelial barrier increasing the risk of inflammation.

Hepatocyte nuclear factor 4 α (*HNF4A*) is another gene involved in the maintenance of the mucosal barrier and associated with IBD. As a DNA binding transcription factor, it regulates the expression of genes involved in epithelial cell proliferation and differentiation as well as tight junction formation. GWA studies in the Netherlands and the UK have shown a strong link between variations in *HNF4A* and UC (Barrett *et al.*, 2009; Van Sommeren *et al.*, 2011). Moreover, a study investigating the role of intestinal *HNF4A* in mice revealed that selective deletion of the gene in IECs increased mucosal permeability and caused severe DSS-induced colitis in mice (Ahn *et al.*, 2008). *HNF4A* seems to play a protective role against UC by maintaining mucosal barrier function and architecture.

Aside from acting as a physical barrier to foreign bodies and toxic materials, the mucosa and its constituents also provide a chemical barrier. IECs have pathogen recognition and processing mechanisms in addition to secretory defences that keep infiltrators at bay. They express an

array of **pattern recognition receptors (PRRs)**, such as **toll-like receptors (TLRs)**, **Nod-like receptors (NLRs)** and **RIG-I-like receptors**, that screen the environment for foreign microbes and, in turn, orchestrate immune cell responses in the mucosa (Peterson and Artis, 2014). Mucins are one of the secretory defences produced by goblet cells, forming the first line of defence in the intestine. Goblet cells also regulate the physical mucosal barrier via secretion of trefoil factor 3 (TFF3) and resistin-like molecule-β (RELMβ). TTF3 is a protein involved in epithelial repair, IEC migration, anti-apoptotic pathways and mucin crosslinking – adding structural integrity to the mucus layer (Peterson and Artis, 2014). In addition to promoting MUC2 secretion, RELMβ plays a role in regulating immune responses during inflammation particularly of macrophages and T-cells (Artis *et al.*, 2004; Nair *et al.*, 2009). Intestinal homeostasis is fortified further by the secretion of antimicrobial proteins (AMPs) by several IECs. For instance, Paneth cells release several AMPs, including defensins, cathelicidins, and lysozymes, that disrupt cell walls and membrane surfaces in both Gram-positive and Gram-negative bacteria via peptidoglycan breakage and pore-formation, respectively (Peterson and Artis, 2014). REGIIIγ is another enterocyte and Paneth cell-derived AMP which is essential in maintaining mucosal homeostasis by spatially segregating members of the microbiota from the small intestinal lining (Vaishnava *et al.*, 2011). IECs are not only sentinels but also mediators, transporting immunoglobulins released by secretory B-cells from the lamina propria across the epithelial barrier in response to pathogens (Peterson and Artis, 2014).

Autophagy and the unfolded protein response (UPR) are processes that regulate immune responses and integral in regulating secretory IECs. The invasion of microbes into the epithelium is limited by autophagy and interruptions in this process can lead to spontaneous inflammation (Peterson and Artis, 2014). UPRs promotes protein folding and translocation in the endoplasmic reticulum (ER) as well as regulation of protein synthesis, arresting of the cell

cycle and induction of apoptosis. Regulation of these mechanisms prevent ER stress in secretory cells averting inflammation (Kaser *et al.*, 2008). The range of pathogen recognition and processing mechanisms described reinforces the intestinal barrier. Any defects in these operations leaves the mucosa exposed to inflammation and can result in health complications, like in IBD.

Nucleotide-binding oligomerization domain 2 (*NOD2*) is the most studied susceptibility gene in IBD and is commonly mutated in CD patients of European descent with a prevalence of ~33% (Guan, 2019). Certain polymorphisms in *NOD2* are linked to poor prognosis for CD patients, however, not all carriers of these develop CD. The *NOD2* gene encodes for a cytosolic receptor expressed in IECs, particularly Paneth cells, and antigen presenting cells (APCs), like macrophages and dendritic cells (Guan, 2019). The function of NOD2 may vary depending on cell-type, but antibacterial immunity seems to be the primary function of NOD2 (Xavier and Podolsky, 2007). Broadly, the leucine rich repeat (LRR) domains in NOD2 recognises the muramyl dipeptide (MDP) component present in bacterial peptidoglycans, which leads to the activation **of nuclear factor-κB (NF-κB)** through a signalling cascade (Xavier and Podolsky, 2007). This, in turn, induces the secretion of either defensins or pro-inflammatory cytokines (like IL-12), depending on the cell type. NOD2 is also known to initiate autophagy by interacting with the protein ATG16L1 through MDP (Guan, 2019). Mutations in *NOD2* can impair pathogen recognition and processing leading to an attenuated immune response, resulting in inflammation. CD patients carrying loss-of-function *NOD2* mutations tend to have lower expression of the AMP α-defensin (Younis *et al.*, 2020). Studies have shown that there is a 10-fold increase in risk for CD with bi-allelic risk carriers versus individuals heterozygous for *NOD2* variants (Hugot *et al.*, 2001, 2007; Ogura *et al.*, 2001).

Genetic variations in the *autophagy-related 16-like 1 protein* (*ATG16L1*) are also strongly associated with CD (Fowler *et al*., 2008; Liu *et al.*, 2015; McGovern, Kugathasan, and Cho, 2015). The product of this gene is an autophagy protein that is widely expressed throughout the intestinal epithelium in several cell types, including APCs, T cells, and B cells (Xavier and Podolsky, 2007). Histology samples from CD patients homozygous for *ATG16L1* T300A substitution show aberrant numbers of Paneth cells with irregular morphologies (Graham and Xavier, 2020). Mutated *ATG16L1* impairs the ability of Paneth cells to properly identify and eliminate pathogens partly due to aberrant TLR signalling (Guan, 2019). This interrupted signalling hinders autophagy in Paneth cells, affecting AMP production, thus allowing bacteria to breach the epithelial barrier which then elicits an immune response bringing about inflammation (Younis *et al.*, 2020). Mice with selective deletion of *ATG16L1* in T cells exhibit spontaneous intestinal inflammation, highlighting that ATG16L1 is essential in maintaining stability in the mucosa (Kabat *et al.*, 2016). The *T300A* genotype is also implicated in ER stress pathways (Graham and Xavier, 2020). Carriers of this variant, both healthy individuals and CD patients, experience polyreactive IgA responses following elevated epithelial ER stress which target both pathogenic and commensal bacteria, potentially leading to dysbiosis (Grootjans *et al.*, 2019; Guan, 2019). IgA plays an important role in gut mucosal immunity by targeting pathogens and promoting tolerance to commensal microbes (Jackson *et al.*, 2021). Thus, ATG16L1 plays a pivotal role in maintaining homeostasis in the gut.

Deficient immune responses are one of the main causes of inflammation. Interleukin 23 (IL-23) is a pro-inflammatory cytokine involved in adaptive immunity and an orchestrator in IBD inflammation, along with its receptor (IL23R). IL-23 is upregulated in the mucosa of both CD and UC patients and it can be protective or a risk factor (Xavier and Podolsky, 2007; Younis *et al.*, 2020). Following bacterial recognition, the cytokine is released by dendritic cells and

macrophages promoting the differentiation of pro-inflammatory **T-helper 17 cells (Th17)** from naïve T cells via **TGF-β** and IL-6 signalling. In addition to inducing differentiation, TGF-β and IL-6 initiate the expression of IL23R on Th17 cell surfaces (McGovern and Powrie, 2007). IL23R is expressed in dendritic cells and macrophages, thus in *IL23R* and *IL-23* variants, this cytokine cascade can trigger a state of unrestrained inflammation through a continuous supply of pro-inflammation molecules like IL-6, IL-12, IL-17, INF-γ, TNF-α and IL-23 (Younis *et al.*, 2020). IL-23 serves as an effective drug target against defective immune responses; risankizumab -rzaa, a monoclonal antibody targeting the p19 subunit of the cytokine has been approved by the FDA to treat cases of moderate to severe CD (D'Haens *et al.*, 2022).

Interleukin 10 (IL-10) is another cytokine with anti-inflammatory properties associated with very early-onset IBD in ages 6 and under (Younis *et al.*, 2020). It is secreted by APCs, lymphocytes, and macrophages to dampen rogue immune responses avoiding development of autoimmune diseases. Genetic variants of the encoding gene cause the downregulation of IL-10 eliciting an aggravated immune response through pro-inflammatory cytokines, like TNF-α and IL-12. Both humans and mice with defects in IL-10 signalling experience severe colitis (Kühn *et al*., 1993; Kotlarz *et al*., 2012). However, *IL10* aberrance is not enough to cause IBD, and its function is tightly linked with the microbiota. Studies show that the microbiota is essential in spontaneous colitis development and immune system activation in the *IL10* mouse model (Sellon *et al.*, 1998; Wirtz and Neurath, 2007). Members of the microbiota induce IL-10 secretion via TLR signalling, downregulating immune response and preventing inflammation (Levast *et al.*, 2015).

Genetic predisposition to IBD has been mapped by GWA studies. However, these studies provide correlations to IBD and do not identify causal genes. What the genetics of IBD reveals

is that defects in bacterial recognition in innate immunity responses is at the core of IBD, highlighting the importance on host-microbiota interactions in the disease. More functional studies need to be carried out to better understand the associations between genetic polymorphisms and IBD. Interestingly, only a small percentage of IBD-associated risk loci carriers develop IBD which would imply that genetics is not the main driver of IBD. The development of IBD is caused by a convoluted mix of genetic risk factors, and alterations to the microbiome and its interactions with the mucosal immune system as well as environmental factors. The development of IBD, particularly UC, seems to be highly dependent on and influenced by gut microbes and their products (Sellon *et al*., 1998; Roy *et al*., 2017). Studies have linked imbalances in the composition of the gut microbiota, dysbiosis, to IBD.

The interdependency between the microflora and occurrence of colitis has been Observed in several IBD mouse models in the studies discussed throughout this section. The DSS model (briefly described in text above) chemically induces colitis through IEC damage and exposure of the lamina propria and submucosal environment to luminal contents (Low *et al.*, 2013). It is one of the most commonly used mouse models to study UC, and its effectiveness depends on several factors, only one of which is microbiota composition. Germ-free (GF) mice were shown to have reduced DSS colonic inflammation when compared to partial microbial depletion using antibiotics (Hernández-chirlaque *et al.*, 2016). Roy *et al.* (2017) and Kazakevych *et al*. (2020) reported that a DSS-mediated colitis response is determined by the composition of an otherwise Specific-Pathogen Free (SPF) microbiome. 2, 4, 6 Trinitrobenzene (TNBS) is a chemical agent that causes mucosal inflammation in mice that resembles CD and is known to reduce gut microbial diversity in mice (Low *et al*., 2013; Kozik *et al*., 2019). Its colitic effect can be alleviated with administration of butyrate producing *Butyricicoccus* in rats (Eeckhaut *et al.*, 2013). Efficacy of immunologically induced colitis models using bacterial agents, like

Adhesion invasive *E. coli* (AIEC) and *Salmonella typhimurium*, is dependent on commensal bacteria (Low *et al.*, 2013). Antibiotics are generally administered in these models to deplete biofilm-forming bacteria, like *Lactobacillus casei*, and other commensals, allowing for better colonisation for the inflammation causing pathological agents. Gene knockout animal models of IBD are generally used to study the immune responses and inflammatory pathways involved in mucosal inflammation (Wirtz and Neurath, 2007). Gut microbes are28ccludingl for development of spontaneous colitis in both $IL2^{-/-}$ and $IL10^{-/-}$ mice (Rakoff-Nahoum, Hao and Medzhitov, 2006). Mice deficient in IL-10 or IL-2 do not develop colitis in germ-free conditions unless transferred to a specific pathogen-free environment (Sellon *et al.*, 1998; Hoshi *et al.*, 2012). However, intestinal inflammation mediated by the microbiota in both animal models are regulated by different signalling pathways (Rakoff-Nahoum, Hao and Medzhitov, 2006; Kiesler *et al.*, 2015). Another method of inducing colitis in mice is through adoptive T cell transfer in immunodeficient mouse models, like $Rag^{-/-}$ or SCID mice (Low *et al.*, 2013). Susceptibility of T cell induced colitis is reliant on gut microbial composition in this model, too (Webb *et al*., 2018). In conclusion, animal models used to study IBD have shown the significance of the microbiome in the aetiology of the disease. Studies have linked dysbiosis to IBD, recent findings on this topic will be discussed in detail throughout the next section of this chapter.

1.4 How do changes in the microbiome influence IBD?

Identifying and quantifying the taxa present within the gut microbiota is crucial in understanding their role in the human body. There are a several established methods that work on the classification of gut microbes like amplicon sequencing, metagenomic sequencing and shotgun sequencing (Young *et al.*, 2021). Amplicon sequencing, particularly of the **16S rRNA gene**, is the most widely used technique in characterising the microbiome despite it lacking

taxonomic resolution at the species/strain level (Johnson *et al.*, 2019). 16S sequencing is paired with other phylogenetic markers, most commonly the 18S rRNA gene and ITS region, are used to identify eukaryotic and fungal members of the microbiome. On the other hand, metagenomic sequencing is better suited for understanding the functional properties of microbes and microbial communities, since the technique samples all genes instead of the single 16S rRNA gene; this allows for both microbial classification as well as highlighting putative metabolic functions (Iyer, 2016). Functional properties of the microbiome can be revealed further by transcriptomic, proteomic and metabolomic analyses. Microbiota community structure is an area much less explored and can be visualised using **fluorescent *in situ* hybridization (FISH)**, which labels bacterial communities in host tissues by utilizing 16S rRNA fluorescent probes (Iyer, 2016). Most sampling of the gut microbiome is non-invasive and is done through stool analysis, which is considered representative of the colonic luminal community. Microbes inhabiting the mucosal layer are sampled via endoscopic tissue biopsies or luminal brushings and other less frequently used techniques, like intestinal fluid aspiration (Tang *et al.*, 2020). When analysing microbiome data potential issues of each sampling approach should be taken into consideration, like the bias introduced in faecal specimens as well as the effect of bowel preparation prior endoscopies on the microbiota and the invasiveness of the procedure. Despite this, using a combination of the currently available techniques, we have gained great insight into the composition, structure, and dynamics of the gut microflora in health and disease.

Studies over the past two decades have shown that, the most prevalent bacteria in the gastrointestinal tract (GIT) are broadly classified into four phyla: Firmicutes, Bacteriodetes, Actinobacteria and Proteobacteria (check glossary for nomenclature) (Hillman *et al.*, 2017; Cresci and Izzo, 2019) while bacteria belonging to the Fusobacteria, Tenericutes, Spirochaetes, Cyanobacteria and TM7 phyla are rarer (Hillman *et al.*, 2017). The diversity and richness of

the microbes varies axially and longitudinally along the digestive tract, changing dramatically throughout the first few years of life, then stabilising and eventually declining with old age (Wilmanski *et al.*, 2021). As summarised in figure 1.2, the architecture of gut microflora communities and their dynamic stability is shaped by both intrinsic and extrinsic factors including, the birthing process, diet, gender, geographical location, pharmaceutical drugs and age (Hillman *et al.*, 2017; Schmidt *et al.*, 2018; Cresci and Izzo, 2019). Seeding of the gut microbiome is believed to start during the birthing process, where the newborn is inoculated with microbes from either the vaginal-anal area or the skin, depending on the mode of delivery (Browne *et al.*, 2022). Feeding style post birth affects early-life colonisation, where breast-fed infants have higher numbers of *Bifidobacterium* species selectively enhanced by oligosaccharides present in breast milk (McBurney *et al.*, 2019). In early years, members of the microbiota are mostly aerobic shifting with time to facultatively anaerobic bacteria and finally to largely anaerobic microbes (Cresci and Bawden, 2015).

Various patterns of microbial colonisation have been linked to numerous health states, whether steady or disease conditions. Despite the extensive research carried out in the field over the last two decades, there is a lack of a clear definition of a “healthy microbiome”. Essentially, health is a dynamic state; therefore it is proposed that the associated microbiota should be diverse and resilient in order to overcome perturbations and return to equilibrium (McBurney *et al.*, 2019). Microbial community composition varies greatly among individuals, it is almost as unique as a thumbprint. About 70% of interindividual microbiota variation is unexplained, while the remaining deviation is due to both genetic and environmental factors (Rothschild *et al*., 2018; Wang *et al*., 2018; McBurney *et al.*, 2019). This evidence is a plausible justification to why a state of equilibrium cannot be defined by a distinct microbial composition. However, there are certain traits and phenotypes that correlate with a diverse gut microbiota, including longer

colonic transit times and stool with soft to firm consistency (Roager *et al.*, 2016; Vandeputte *et al.*, 2016).

Scientists have developed a method of classifying microbiomes of individuals based on their composition, into what they termed as enterotypes in attempt to classify microbial markers with diagnostic potential (Arumugam *et al.*, 2011). Arumugam and colleagues identified 3 main enterotypes that are recognisable by variations in the levels of three main genera. Two of the clusters are always denoted by *Bacteroides* and *Prevotella*, while the third cluster is driven by different groups, *Ruminococcus*, *Blautia*, and unclassified Lachnospiraceae depending on the studied dataset. Further analysis revealed that enterotypes are influenced by long-term dietary habits (Wu *et al.*, 2011). A study has shown that infants have two enterotypes driven by Firmicutes and *Bifidobacterium* dominance which shifts to *Bacteroides* and *Prevotella* as they get older (Xiao *et al*., 2021). However, the concept of enterotypes has been challenged since most microbiome data exhibit a continuous gradient of dominant taxa and not microbial populations with discrete taxa (Knights *et al.*, 2014). Despite all efforts, there is also a lack of valid microbial biomarkers which adds further challenges to the present complexities of defining a "healthy microbiome". Patterns of colonisation associated with steady and disease states are very difficult to identify due to the lack of longitudinal studies and conflicting observations. For instance, *Akkermansia* is often linked to healthier metabolic features in obese individuals but negatively correlated with incidence of multiple sclerosis – suggesting that microbial behaviour is context dependent just as found in the natural environment (Dao *et al.*, 2016; Alzarhani *et al*., 2019; McBurney *et al*., 2019). However, a recent multi-generational study on a large Dutch cohort investigating environmental factors causing microbiome variations highlighted over 1000 associations between disease and bacterial taxa (Gacesa *et al.*, 2022). *Alistipes senegalensi* was one of the microbes identified by Gacesa *et al*. (2022) that is

associated with 43 disease phenotypes, including IBD. Overall, it might be more pertinent to identify a set of "healthy" microbial metabolic functions rather than a "healthy" microbial composition (Bäckhed *et al.*, 2012).

Human metabolic processes are highly reliant on members of the microbiota. Intestinal bacteria are involved in the synthesis of amino acids, enzymes, and vitamins, in addition to nutrient and mineral absorption, and fermentation. For instance, *Bifidobacterium* spp. are known to synthesize vitamin K and B (Nishida *et al.*, 2017). Many plant-derived dietary fibres are indigestible by the human body and can only be fermented by commensal bacteria. Short chain fatty acids (SCFAs, primarily acetate, propionate and butyrate) are by-products of this process and are important in maintaining homeostasis in the gut. Butyrate is the main energy source of IECs and prevents dysbiosis by enhancing IEC uptake of oxygen through β oxidation, maintaining a hypoxic environment for commensal anaerobes, reviewed in Valdes *et al.* (2018). On the other hand, acetate is vital for the growth of other gut flora and is involved in central appetite regulation, cholesterol metabolism and lipogenesis (Frost *et al.*, 2014). Firmicutes such as, *Faecalibacterium prausnitzii*, *Clostridium leptum* as well as *Eubacterium rectale* and *Roseburia* spp of the family Lachnospiraceae, are mainly butyrate producers while members of Bacteroidetes mostly produce acetate and propionate (reviewed in Venegas *et al.*, 2019). *Eubacterium hallii* and *Anaerostipes* spp. are secondary producers of butyrate from lactate and acetate. Members of the genus *Bifidobacterium* produce acetate and lactate, while *Akkermansia* spp. generate propionate and acetate (Venegas *et al.*, 2019). A shift in abundance of these SCFA producers, among other species, alters the metabolic profile of an individual thus disrupting the functional properties of the intestinal ecosystem (Cummings and Macfarlane, 1991; Venegas *et al*., 2019).

Research shows that a less diverse microbiome renders an individual more susceptible to disease (Bäckhed *et al.*, 2012). IBD is one of many diseases associated with dysbiosis and specific commensal microbes without an established causal relationship. Changes in metabolic pathways linked to inflammation and the flourishing of pathobionts could potentially act as microbial stressors, supporting the outgrowth of species associated with dysbiosis. Anaerobic SCFA producers are depleted in the intestinal mucosa and faeces of IBD patients, leaving space for the expansion of populations of pathobionts, like *E. coli* and *Salmonella* (Venegas *et al.*, 2019). Studies have shown that IBD patients have low numbers of gut microbes belonging to the phyla Bacteroidetes and Firmicutes alongside an increased abundance of Proteobacteria, Actinobacteria and Bacilli (Frank *et al.*, 2007; Ni *et al.*, 2017; Venegas *et al.*, 2019). The SCFA producers, *F. prausnitzii* and *R. intestinalis* are particularly reduced in CD cases (Ni *et al.*, 2017; Franzosa *et al.,* 2019; Venegas *et al.*, 2019). *F. prausnitzii* is involved in the production of the anti-inflammatory IL-10 and prevents the release of inflammatory cytokines (IL-12 and INF-γ) thus maintaining mucosal equilibrium (Nishida *et al.*, 2017). Therefore, altering the presence of *F. prausnitzii* can disrupt mucosal stability, bringing about inflammation. Inflammation is strongly tied to the suppression of genes involved In SCFA utilisation and metabolism. Genes encoding enzymes involved with butyrate uptake and oxidation are downregulated in dysbiotic UC patients and this reduction can be reversed with anti-TNF-α antibody treatment (Venegas *et al.*, 2019).

Commensal bacteria maintain gut homeostasis by also mediating intestinal mucosal immunity while others suppress pathobionts directly through nutrient competition and secretion of antimicrobial substances. *Bacillus thuringiensis*, for instance, releases a bacteriocin targeting spore-forming bacteria such as other *Bacillus* species or Clostridia (Nishida *et al.*, 2017). *Candidatus Arthomitus*, also known as segmented filamentous bacteria (SFB), promote the

maturation of the mucosal immune system, IgA secretion, as well as Th17 development (Nishida *et al.*, 2017). SCFA producers can indirectly modulate immune system interactions, such as AMP (REGIIIγ and β-defensins) production, induction of the anti-inflammatory IL-10, and T-cell differentiation (Nishida *et al.*, 2017; Vaga *et al.*, 2020). Butyrate produced by Clostridium clusters IV, XIVa and XVIII enhances the function of colonic regulatory T cells (Treg cells) (Ni *et al.*, 2017; Nishida *et al.*, 2017). Moreover, butyrate has the ability of preventing inflammation by suppressing NF-κB activation, inhibiting proinflammatory cytokines, like TNF-α, INF-γ, and IL-6, through HDAC inhibition and stabilising HIF-α (Ni *et al.*, 2017).

Adherent-Invasive E. coli (AIEC) is another bacterial species linked to IBD that affects the permeability of the intestinal mucosa and is increased in CD patients (Ni *et al.*, 2017; Nishida *et al.*, 2017). IBD has also been associated with an increase in the sulphate-reducing bacteria *Desulfovibrio*, where the hydrogen sulphate produced from that process causes damage to IECs and inflammation (Nishida *et al.*, 2017). Elevated levels of tryptophan metabolites in IBD patients indicates an increase in tryptophan degradation which, can be linked to the abundance of *Bacteroides* species involved in tryptophan metabolism in IBD (Nikolaus *et al.*, 2017; Vaga *et al.*, 2020; Lee and Chang, 2021). Secondary bile acids generated by microbiota has anti-inflammatory properties and plays a role in maintaining mucosal homeostasis. IBD patients exhibit imbalances in bile acid absorption and deficiencies in secondary bile acids which can induce the growth of bile-tolerant pathobionts, like the sulphite-reducing *Bilophila wadsworthia*, as shown in IL10-deficient mice (Devkota *et al.*, 2012; Natividad *et al.*, 2018; Lee and Chang, 2021). A visual summary of some of the key findings related to the gut microbiome and IBD is presented in figure 1.4.

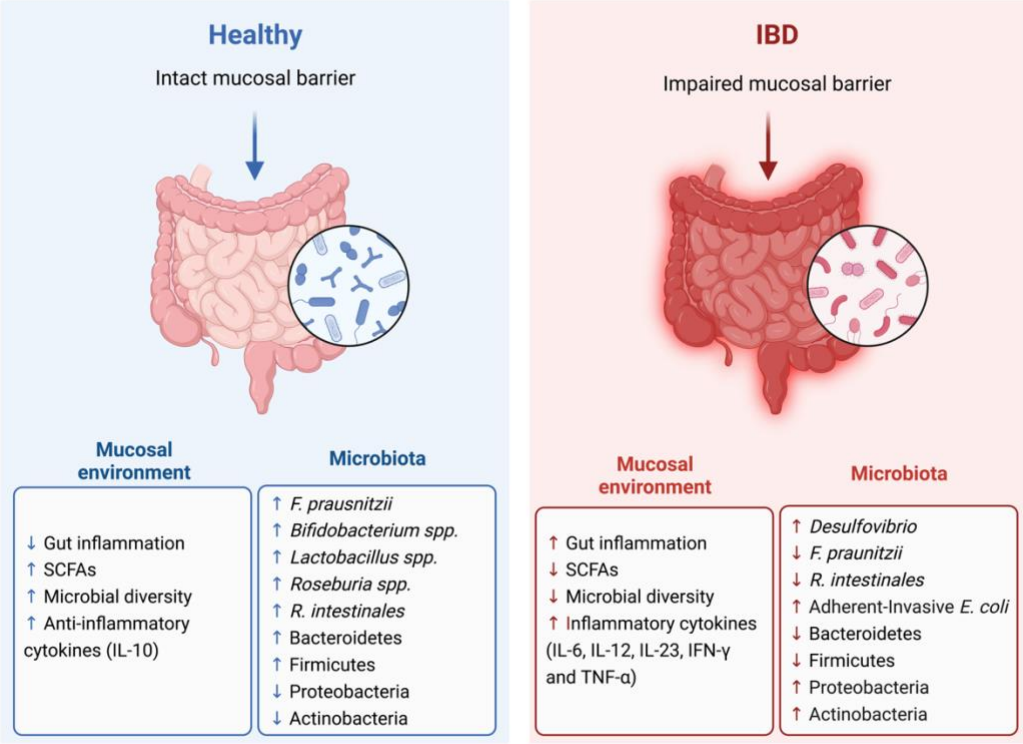


Figure 1.4: Summary of some of the key features linked to a healthy microbiome or a microbiome linked to IBD (right) (El-Sahhar & Varga-Weisz, 2022).

All the findings discussed earlier are correlations with IBD and causality has not been shown. Most microbial associations have not been replicated, some have even been contradicted, implying that they may be context-dependent or heavily influenced by environmental factors. Only 2% of taxa at the genus level are found to be different in IBD when compared to healthy individuals as opposed to 12% of change in metabolic pathways (Ni *et al.*, 2017). A study has shown over 2700 differentially abundant metabolites in faecal samples from IBD patients, of which 244 metabolites are significantly increased (Franzosa *et al.*, 2019). This suggests that perhaps identifying metabolic pathways and metabolite profiles associated with IBD rather than a microbial signature is a better way to understand the role of the microbiota in disease progression.

1.5 How does the environment contribute to gut microbiome changes?

Diet, lifestyle, physical activity, smoking, and exposure to pollution are some of the environmental factors that shape the gut microbiome and, in some cases, are associated with IBD prevalence. Incidence rates of IBD are higher in urban and western environments, particularly in areas that have undergone a period of industrialization over the past few decades (Ng *et al.,* 2013; Benchimol *et al*., 2017). A large-scale population-based study showed that, within 50 years, there has been a 10-fold increase in incidence of IBD in Japan, South Korea, China and India (Ng *et al.*, 2013). Incidence rates are still higher in western populations, however, IBD is an emerging threat in developing countries. The link between urbanization and IBD development could be mediated by multiple lifestyle factors, like diet, exposure to air pollution, hygiene practices, and antimicrobials (Ananthakrishnan *et al*., 2017). There have been changes in microbial patterns associated with urbanization, like an increase in Bacteroidetes, particularly *Alistipes*, the Firmicutes *Blautia*, *Faecalibacterium* and *Ruminococcus*, as well as *Bifidobacterium* and *Bilophila* of the phyla Actinobacteria and Proteobacteria, respectively (Zuo *et al.*, 2018). Generally, there is a shift in microbial composition from gram positive to gram negative species with pathogenic tendencies (Hills *et al.*, 2019). Studies investigating variation within microbiota structure in relation to geography have pointed out that there are deviations across countries as well as between ethnicities within the same country (Prideaux *et al.*, 2013; Cresci and Bawden, 2015). Gut microbial communities of Malawian and Amerindian populations (of the Venezuelan Amazonia) are divergent from those of the US, with less diversity in the latter (Yatsunenko *et al.*, 2012). These differences can be attributed to diet since they have been mapped to metabolic pathways involved in carbohydrate metabolism and vitamin B biosynthesis (Yatsunenko *et al.*, 2012). Malawian and Amerindian diets are carbohydrate rich, high in corn and cassava, while a US diet is rich in

protein (Cresci and Bawden, 2015). Another study comparing microbiotas of African and Western children showed that microbial communities in children from Burkina Faso coevolved with their diets, maximising energy intake from fibrous foods (Filippo *et al*., 2010). There was a higher prevalence of Bacteroidetes, notably *Prevotella* species involved in cellulose and xylan hydrolysis, and less of Firmicutes in children from Burkina Faso, compared to a total absence of these microbes in an Italian cohort (Filippo *et al.*, 2010). The difference in gut microbial composition can be linked back to the African diet, which is high in plant fibre (Zuo *et al.*, 2018).

Diet is one of the environmental factors that has a huge impact on the microbiome. Metagenomic studies on twins have shown that the influence of dietary habits and cohabitation on the microbiome outweighs the effect of genetics on the structure and functionality of the gut microflora *(Rothschild et al.*, 2018; Hills *et al., 2019; Gacesa et al., 2022*). This starts at infancy where breastfeeding selects for a microbiome dominated by Bifidobacterium due to the oligosaccharides present in breastmilk and can reverse infancy dysbiosis from intrapartum antibiotic exposure (Cresci and Bawden, 2015; Azad *et al*., 2016). Development of IBD is negatively correlated with breastfeeding, however, the causal links are not clear (Ananthakrishnan *et al*., 2017). Dietary habits continue to shape the microbiome throughout life with various ramifications. A western diet – one typically high in simple carbohydrates, fats, additives and low in fibre – has a negative effect on microbiome diversity as well as host mucosal barrier and immunity (Levine, *et al.*, 2018; Hills *et al.*, 2019). An abundance of *Bifidobacteria* and a reduction in *Lactobacillus*, *Streptococcus* and *Roseburia* are commonly associated with diets high in simple carbohydrates (Hills *et al.*, 2019). Reduced mucosal barrier function favouring AIEC colonisation and mucin-degrading *Ruminococcus* accompanied by dysbiosis is induced by a western diet in both CD patients and mouse models (Martinez-Medina

et al., 2014; Agus *et al.*, 2016). Histological assays on mice fed a western diet showed depletion of goblet cells followed by a reduction in MUC2 expression and increased intestinal permeability (Martinez-Medina *et al.*, 2014). Moreover, the dysbiosis caused by a western diet decreases production of SCFAs and expression of the SCFA receptor GPR43 modulating the immune response via Treg cell reduction (Agus *et al.*, 2016). Interestingly, biopsies from CD patients also showed reduced expression of the butyrate receptor GPR43 highlighting the role of butyrate in modulating inflammation (Agus *et al.*, 2016). A few studies have linked diets high in fat with intestinal inflammation. Mice put on high fat diets exhibited an increase in pro-inflammatory markers, including TNF-α and INF-γ, decreased Treg cells, recruitment of **dendritic cells** and Th17 cells along with downregulation of tight junction proteins collectively exacerbating intestinal inflammation (Ma, Hua, and Li, 2008; Suzuki and Hara, 2010; Gruber *et al.*, 2013). Furthermore, structure and function of the microbiota were highly impacted by a diet high in fat, presenting reduced numbers of Ruminococcaceae, altered antimicrobial pathways, defective bile metabolism and shifts in nutrient metabolism highlighted by a large set of differentially abundant bacterial proteins and metabolites (Daniel *et al.*, 2014). Risk of UC is significantly reduced with high polyunsaturated dietary fat intake as opposed to unfavourable trans-unsaturated fats (Levine *et al.*, 2018).

High-fibre diets, on the other hand, are linked to a highly diverse microbiome. Increased dietary intake of fibre in $IL10^{-/-}$ mice suppresses gut inflammation which protects against acute colitis through several mechanisms, including increased expression and proper localisation of tight junction proteins which maintain the mucosal barrier as well as a decrease in production of the pro-inflammatory IFN-γ accompanied by an increase in SCFA and Treg production (Bassaganya-Riera *et al.*, 2011; Silveira *et al.*, 2017; Wang *et al.*, 2016). Supplementing the diet of mice with fermentable fibres also increased SCFA production and decreased mucosal

permeability through an increase in expression of **occludin** and other tight junction proteins, reducing inflammation score (Hung and Suzuki, 2016). However, increased fibre intake did not resolve the low levels of SCFAs in UC patients despite an established link between low fibre and the condition, suggesting that not all findings in mouse models are translatable in humans (James *et al.*, 2015; Levine *et al.*, 2018). Regardless, a study revealed a bloom of bile tolerant bacteria, like *Alistipes* and *Bilophila* and Bacteroides (a pattern previously linked to IBD), in the microbiota of individuals following an animal-based diet compared to those on a plant-based diet, which supports the connection between animal protein and IBD (David *et al.*, 2014; Levine *et al.*, 2018). There have been no dietary patterns tied to CD. However, a diet low in fermentable oligosaccharides, disaccharides, monosaccharides, and polyols, known as the FODMAP diet, seems to provide symptom relief for IBD patients with no attenuation of inflammation (Cox *et al.*, 2020). This diet is only recommended for the short-term due to its low levels of fibres and effect on the microbiome, with lower abundances of *Bifidobacterium* species and *F. prausnitzii* (Hills *et al*., 2019; Cox *et al*., 2020).

Processed foods are high in food additives which proved to have harmful effects on both mucosal stability and the gut microflora (Levine *et al.*, 2018; Hills *et al.*, 2019). Carboxymethylcellulose is an emulsifier commonly used in food products and was found to increase intestinal permeability through depletion of the mucus layer and distension of inter-villi spaces, allowing for bacterial invasion in $IL10^{-/-}$ mice (Swidsinski *et al*., 2009; Chassaing *et al*., 2015). Another emulsifier, polysorbate 80, was shown to cause similar deleterious effects on the intestinal epithelium mucosa as well as the microbiome (Chassaing *et al*., 2017). Both emulsifiers alter the microbial composition to a less diverse one, with an increase in mucin-degrading bacteria (like *Ruminococcus gnavus*), Verrucomicrobia and inflammatory Proteobacteria and Enterobacteriaceae species along with a decrease in key Bacteriodaceae

(Chassaing *et al.*, 2017). Maltodextrin is an artificial sweetener linked to gut inflammation through elevated levels of IgA (Miyazato *et al*., 2016). The sweetener has been shown to cause dysbiosis promoting the expansion of AIEC and biofilm formation (Nickerson and Mcdonald, 2012). It is evident that additive-rich processed foods have adverse effects on mucosal immunity and the microbiome. However, not all foods have a solely harmful effect on health and the microbiome, as some may provide context dependent beneficial effects. Administering dietary alcohol in mice promoted colonic hyperpermeability and decreased butyrate and other SCFA levels in stool (Forsyth *et al.*, 2017). Yet, red wine in moderation exhibits prebiotic benefits–- owed to high levels of polyphenols–- with an increase in Proteobacteria, Fusobacteria, Firmicutes and Bacteroidetes species, especially *F. prausnitzii* and *Bifidobacterium* (Queipo-Ortuño *et al*., 2012; Hills *et al*., 2019). These findings have not been replicated with other kinds of alcoholic beverages.

With the significant effect of diet on both the gut microflora and IBD, investigations into IBD and obesity comorbidities has been an area of interest. Generally, obesity is reflected in gut microbiota shifts in abundance of Bacteroidetes and Firmicutes, but no connection can be drawn due to the minimal research done on the co-morbidity of IBD and obesity within the context of the microbiome.

Another major player in microbiome shifts is drug exposure. Intake of drugs, like antibiotics and non-steroidal anti-inflammatory drugs (NSAIDs), result in significant changes in gut microbial composition and are considered a major predisposition for IBD (Dethlefsen *et al.*, 2008; Ananthakrishnan *et al.*, 2012). A 5-day course of the antibiotic ciprofloxacin in humans is enough to reduce microbial diversity and richness of the microbiome within 3-4 days of antibiotic administration (Dethlefsen and Relman, 2011). This effect is mostly reversible within

4 weeks to 6 months with the exception of a few taxa (Dethlefsen *et al.*, 2008). Studies investigating the use of antibiotics in early life and IBD predisposition found a strong link in Western populations, but only with CD and not UC (Ungaro *et al.*, 2014). Another study showed in a cohort of 36 IBD subjects, 21 (58%) had at least 1 dispensation of antibiotics in their first year of life compared to 39% of the controls (Shaw *et al.*, 2010). Furthermore, children receiving antibiotics during their first year of life were almost 3 times as likely to develop IBD compared to those that did not (Shaw *et al.*, 2010). An increase in use of NSAIDs poses a higher risk of developing IBD (Ananthakrishnan *et al*., 2017). The group of drugs is known to cause intestinal mucosal permeability and oxidative stress (Bjarnason *et al.*, 2018). Changes in gut microbial composition have been reported upon NSAID intake, these alterations vary depending on the type of NSAID and tend to favour taxa able to metabolise these drugs (Maseda and Ricciotti, 2020). *Prevotella* spp., *Bacteroides* spp. Ruminococcaceae and *Barnesiella* spp. were increased with the use of aspirin, whereas an enrichment of Acidaminococcaceae and Enterobacteriacceae were associated with ibuprofen and celecoxib users (Rogers and Aronoff, 2016). Another study reported no changes in microbiota upon celecoxib use but a decrease in butyrate production as well as inflammatory markers *in vitro* (Hernandez-sanabria *et al.*, 2020).

Other lifestyle factors, like supplement intake, smoking, and exposure to pollution, have been connected to both the microbiome and IBD, although, how these factors influence IBD are poorly understood (Piovani *et al.*, 2019). There has been a lot of speculation on the role of vitamin D in IBD pathogenesis. However, a population study on 169 participants showed that vitamin D deficiency is not associated with development of the disease (Opstelten *et al.*, 2018). Vitamin D and its receptor play a role in modulating the immune system through the butyrate producing bacteria but there is no causative link between that and pathogenesis of IBD

(Battistini *et al.*, 2021). Smoking is a lifestyle habit with controversial effects on IBD. It is considered a risk factor in CD development and somehow protective against UC, however, this protective effect was only observed in some populations (Piovani *et al.*, 2019). No associations were found between smoking and IBD in Asian and Jewish populations. On the other hand, smoking affects the human microbiome and causes dysbiosis in mice with reduced numbers of Firmicutes, Actinobacteria, *Bifidobacterium*, and *Lactococcus* and an enrichment in Bacteroidetes, *Prevotella*, *Bacteroides*, Proteobacteria and Clostridium, reviewed in (Savin *et al.*, 2018). Functional studies need to be carried out to assess the nature of the association between smoking, IBD and the microbiome. Pollution is another environmental factor connected to socioeconomic status and urbanisation that affects the microbiome and has been linked to IBD. Early exposure to air pollution was associated with an increased risk of IBD in children and young adults (Kaplan *et al.*, 2010). However, this link has been questioned by Opstelten *et al.* (2016). Particulate matter (PM) exposure in mice was shown to increase intestinal permeability via rearrangement of tight junction proteins of the epithelium and promoted inflammation through IL-6 release and NF-κB activation (Mutlu *et al.*, 2011; Salim *et al.*, 2014). A similar pro-inflammatory cytokine response in $IL10^{-/-}$ mice was observed along with increased gut permeability with PM supplemented chow (Kish *et al.*, 2013). The mice experience shifts in Bacteroidetes, Firmicutes and Verrucobacteria leading to altered SCFA production and reduced butyrate levels. Again, these observations are insufficient for assigning solid associations between smoking and IBD, let alone suspect causative effects. The lifestyle factors discussed are clearly connected to microbiome perturbations but have a less causal effect on IBD.

1.6 Can microbial-based treatments be the answer?

As briefly mentioned in the first section of this contribution, IBD is managed through different treatment protocols depending on the condition, severity of the disease and response to treatment. A large number of IBD patients fail to respond to initial treatment and a fraction of those who respond to medication initially become secondary non-responders, especially with biological therapy (Chudy-Onwugaje *et al*., 2019; Lamb *et al*., 2019). Current regimens target the aberrant immune responses involved in IBD and can change during active and quiescent phases of the disease. The first line of treatment for mild to moderate pancolitis in UC patients is administration of oral and/or enema 5-aminosalicylic acid (5-ASA), which is an anti-inflammatory drug that captures damaging free radicals in the bowel (Kuehbacher *et al.*, 2008; Lamb *et al.*, 2019). 5-ASA is also used for maintenance therapy in UC patients. Patients experiencing major side effects or are refractory to 5-ASA are prescribed corticosteroids. Due to their severe side effects, corticosteroids are only used in moderate to severe cases as well as refractory cases of both UC and CD (Chudy-Onwugaje *et al.*, 2019; Lamb *et al.*, 2019). Immunosuppressive or immunomodulator therapies in the form of biological agents, like human anti-TNF-α monoclonal antibodies, are used in the event of failure in response to second line drugs (Kuehbacher *et al.*, 2008; Lamb *et al.*, 2019). In the case of complete resistance to medication and persistence of health complications, a colectomy followed by **ileal pouch anastomosis (IPAA)** is the remaining option (Lamb *et al.*, 2019).

Primary therapy in CD patients with moderately to severely active disease includes systemic corticosteroids (Lamb *et al.*, 2019). However, nutritional intervention through exclusive enteral nutrition (EEN) has proven to be an effective first line of treatment, inducing remission with a success rate of 73% but only in children (Lamb *et al.*, 2019; Eindor-Abarbanel *et al.*, 2021). EEN involves the administration of a nutritionally-whole liquified diet orally or through

nasogastric tubes for 4-8 weeks (Borrelli *et al.*, 2006; Lamb *et al.*, 2019). The mechanism of action for this nutritional therapy remains poorly understood, yet it has been shown to lower gut microbial diversity with reduction in *Bacteroides* species and *Clostridium coccoides* (Maclellan *et al.*, 2017). When compared to children treated with corticosteroids, those on EEN exhibited significant mucosal healing but lower numbers of butyrate-producing bacteria, which seems very contradictory since lower levels of butyrate are generally associated with active disease (Pigneur *et al.*, 2019). Nutritional intervention is not as effective in adults, early use of biological agents is considered with adult CD patients who have poor prognoses (Lamb *et al.*, 2019). Failure of conventional therapy and recurrence of the disease can lead to ileocaecal resection to induce and maintain remission.

Antibiotics are also part of the IBD treatment regimens discussed above and have been largely used in the incidence of pouchitis (infection post IPAA), abdominal abscesses, enteric infections, IBD associated *C. difficile* infection, **perianal disease** and **luminal disease** (Ledder *et al.*, 2014; Eindor-Abarbanel *et al*., 2021). They are also used to modulate gut microflora by targeting bacteria associated with inflammation in CD patients and selecting for the beneficial ones, since patients with CD tend to harbour mucosal infiltrating pathobionts (Eindor-Abarbanel *et al.*, 2021). However, exposure to antibiotics could have a negative effect on the microbiome on its already unstable state. A recent study has shown reduced microbial diversity and enrichment of *Escherichia* species along with a reduced UC activity index in young UC patients post antibiotic treatment (Turner *et al*., 2019). Two other studies reported similar conflicting responses to antibiotics in CD patients, where the microbiota community in both study groups was less diverse but faecal calprotectin (an inflammation marker) was reduced in patients from one of the studies (Levine *et al*., 2018; Sprockett *et al.*, 2019). Interestingly, increased abundance of *Bifidobacterium* and *F. prausnitzii* along with an increase in SCFA

were observed post antibiotic treatment in 4 CD patients (Maccaferri *et al.*, 2010). However, the findings of all the studies combined with a small remainder of studies are not enough to draw a strong conclusion on the effect of antibiotics on IBD and microbiota of IBD patients.

The standard treatment protocol is not always effective and usually comes with serious side effects, thus the search remains for an alternative treatment that is more robust. Since the microbiome is heavily associated with IBD, there has been an interest in microbe-based or microbe-related interventions. Probiotics are live bacteria and fungi that are present in the form of food products or supplements that provide health benefit to the host by restoring balance and function in the gut microbial communities when administered in the right dosage (Martyniak *et al*., 2021). Probiotic strains have been shown to maintain the mucosal barrier function and modulate immune responses (Ng and Hart, 2014). There are countless probiotic preparations on the market, and they are designed in a way so that the microbes can survive transit until they reach the larger bowel (Eindor-Abarbanel *et al.*, 2021). The main microbial strains used in most probiotic supplements include *E. coli* Nissle 1917, E. coli (serotype 06:K5:H1), *Lactobacillus reuteri*, *L. casei* ATCC PTA-3945, *Bifidobacterium* and yeast species *Saccharomyces boulardii* (Eindor-Abarbanel *et al.*, 2021). VSL#3 is one of the most commonly administered and investigated probiotic in IBD, and it is comprised of *Streptococcus thermophilus*, three bifidobacteria strains (*B. longum*, *B. breve*, and *B. infantis*) as well as 4 *Lactobacillus* strains (*L. casei*, *L. acidophilus*, *L. bulgaricus*) (Cheng *et al*., 2020). There have been contradictory observations from studies evaluating the efficacy of probiotics in inducing and maintaining remission in IBD patients; a few patients had positive outcomes (Shadnoush *et al.*, 2015; Yoshimatsu *et al.,* 2015; Palumbo *et al.*, 2016; Tamaki *et al.*, 2016) while others showed no significant effects (Fedorak *et al.*, 2015; Bjarnason *et al.*, 2018; Matsuoka *et al.*, 2018). Meta-analysis studies and reviews also present opposing findings, where two groups

found probiotics to be beneficial in UC but not CD patients (Ganji-Arjenaki and Rafieian-Kopaei, 2018; Zhang *et al*., 2021) while Astó *et al*. (2019) reported a significant effect of probiotics inducing remission but not maintaining it.

The effects of probiotics are short-lived and cease once treatment is stopped, which is probably due to the lack of nutrients or substrates required for the microorganisms to flourish (Tannock *et al.*, 2000). Instead, patients can be prescribed prebiotics, which are compounds (commonly found in foods) that are selectively utilised by members of the microbiota and promote their growth and function, thus providing health benefit to the host (Eindor-Abarbanel *et al.*, 2021). Glucooligosaccharides, fructooligosaccharides, lactulose and inulin are the most widely used prebiotics and are indigestible by humans (Martyniak *et al*., 2021). Generally, they are known to enhance the growth of lactic acid-producing bacteria, like *Lactobacillus* as well as *Bifidobacterium* and *Bacteroides* species, and promote SCFA production (Cummings and Macfarlane, 2002; Martyniak *et al*., 2021). Clinical studies assessing the effect of prebiotics on IBD patients have been contradictory; Benjamin et al. (2011) reported deterioration of CD patients on prebiotics while another group reported no significant improvement or worsening in both UC and CD on prebiotics (Hafer *et al.*, 2007). A few studies have reported improvement in clinical activity of the disease and remission states with different prebiotics (Fernandez-Banares *et al*., 1999; Kanauchi *et al.* , 2003; Hallert *et al*., 1991). These findings counter each other, and presently no significant positive effect can be linked to prebiotics intake and IBD. However, when combined with probiotics, prebiotics some studies have suggested at a beneficial effect (the combination of prebiotics and probiotics are collectively referred to as synbiotics) (Zhang *et al.*, 2021). Usually, synbiotic supplementation comes in the form of *Lactobacillus* GG and/or *Bifidobacteria* coupled with fructooligosaccharides and/or inulin (Martyniak *et al*., 2021). Compared to a placebo group, 13 CD patients put on *B. longum* and

a mix of inulin and fructoligosaccharides showed significant clinical and histopathological improvement, in addition to enrichment of *Bifidobacterium* species and reduction in inflammatory markers. However, this was only observed in the first 3 months of treatment and not the total 6 months (Steed *et al.*, 2010). Chermesh *et al*. (2007) reported no significant improvement in CD patients on synbiotics. Again, results are conflicting with no significant effect highlighted and investigations in UC and synbiotics were equally paradoxical (Martyniak *et al*.,, 2021).

Just like probiotics, faecal matter transplantation (FMT) works through modulation of the recipient's immune response via microbiota compositional changes. FMT involves the delivery of stool derived microbes from a healthy donor to the gastrointestinal tract of a patient (Ni *et al.*, 2017). It is argued that it is a more robust method to treat dysbiosis as it has been shown to increase microbial diversity due to a larger and more varied microbial load during transfer (Khoruts *et al.*, 2010; Hamilton *et al.*, 2013; Bojanova and Bordenstein, 2016). However, the screening process for donors is very complex to avoid infections and potential allergic reactions that can be passed on through the faecal matter (Weingarden and Vaughn, 2017; Bibbò *et al.*, 2020). This procedure has been widely explored to restore microbial composition and successfully used to treat *C. difficile* infections. The success rate of over 90% in recurrent *C. difficile* infection has made FMT a therapeutic of interest (Drekonja *et al.*, 2015). However, the clinical response to FMT is not equally positive in IBD patients. Success rates varied from 25% to 79% in UC patients, while response to FMT in CD and placebo groups was mostly not significant due to small cohort size and differences in procedure protocols (Paramsothy *et al.*, 2017; Costello *et al*., 2019; Sood *et al*., 2019; Yang *et al*., 2020; Sokol *et al.*, 2020; Crothers *et al*., 2021).

In general, more comprehensive investigations into the mentioned treatments should be done as current work is inconsistent and is based on small cohorts. Furthermore, the efficacy of current therapeutics should be linked to its functional effects on the microbiome. A study has shown that two thirds of commonly prescribed drugs are metabolised by at least 1 bacterial strain of the microflora and that each strain can process from 11-95 different drugs (Zimmermann *et al.*, 2019). The authors identified and validated 30 microbiome-associated enzymes able to convert 20 known drugs into 59 candidate metabolites. This could be a reason why many patients are resistant to conventional treatment. Identifying microbes associated with current IBD medication metabolism could help alleviate refractory therapy. With the steady drop of sequencing expenses, personalised treatment for IBD patients could be implemented and potentially solve inefficiencies of current IBD treatment regimens.

1.7 Conclusion and Outlook

The gut microbiota plays a role in maintaining homeostasis and disease pathogenesis. Despite the lack of clear causal relationships between the human microbiome and many health conditions, the evidence for links between commensals and health are intriguing. The gut microbiome is a very complex network built on intricate interactions among microbes and the crosstalk between host and microbes. Current research into the field is largely based on the bacterial component of the microbiome and fails to address the importance of the other members, like fungi, viruses, and archaea. There is very little work done on investigating fungal and viral communities in the gut and even less on archaea. Different bacterial strains can contribute to the same metabolic pathways conferring a health or diseased state. Perhaps there could be an unknown co-dependency between other members of the gut microflora influencing major functional properties involved in immunity or metabolism. Exploring the neglected part of the human microbiome would give us better, more holistic insight into its role in health and disease. The added complexity of environmental variables causing microflora perturbations

makes it very challenging to assign causal relationships. This could explain why many of the findings in mice are not replicated in humans. For instance, the mouse models used to study IBD focus on isolated events or a number of variables in a tightly controlled environment; however, it is very difficult to do the same in a clinical setting due to great variability in environmental and genetic factors.

These difficulties highlight the importance of animal models in identifying microbiome-host association with disease. To correctly identify microbiome links to disease the appropriate model must be chosen. In the case of IBD, the spontaneous CD model SJL/J is less favoured, especially with oxazolone induced inflammation, due to the development of other autoimmune diseases along with CD via Th1 mediated inflammation (Baydi *et al.*, 2021). Furthermore, the sex of the mice used can be as influential as the model used especially when looking into gut microbial composition. It is critical to consider such technical matters when designing a study to avoid bias. Similar considerations should be taken in clinical studies as well, most of the clinical studies investigating the effect environmental factors in IBD have very low statistical power due to the small number of patients in their study groups. More effort should be put into designing larger cohort studies in order to get a more accurate representation of the disease and its risk factors.

In the future, it will be pivotal to perform longitudinal studies in large patient- and control-cohorts, to capture associations and changes in the gut microbiome and linking these to disease or disease progression, paving the way to identify causations rather than just correlations. As the power of DNA sequencing technologies and bioinformatic analysis increase, these will be combined with proteomic, transcriptomic and metabolomic analysis to capture the dynamic nature and function of the human microbiome in greater detail, allowing us to understand the ecology of this system with respect to health and disease.

Genetics are less influential within the context of IBD and the microbiome, whereas environmental factors, especially diet, play a leading role in gut microbial composition and can affect disease susceptibility. Further investigations are required for the identification of optimal manipulations of gut microbes whether prophylactically or in treatment form. As our understanding of the human microbiome increases, we are beginning to unlock its secrets in order to promote health and fight disease.

Chapter 2: Methods

2.1 Chapter Overview:

The aim of this chapter is to describe the experimental design of this thesis and outline the pipelines implemented for both the meta-analyses and individual datasets, with a detailed explanation of the methods used and the justification of their use. Preceding this is a brief background about the thesis topic and the importance of the work.

2.2 Introduction

Ulcerative colitis (UC) is one of the two main forms of inflammatory bowel disease (IBD) and the focus of this thesis. It is characterised by mucosal inflammation starting in the rectum and extending proximally into the colon. Common symptoms include diarrhoea, abdominal pain, weight loss, obstruction of the gastrointestinal tract and extragastrointestinal manifestations in some cases (Frank *et al.*, 2007; Ghosh and Premchand, 2015). Treatment protocol of UC involves anti-inflammatory drugs, corticosteroids, immunomodulatory drugs and surgical intervention depending on the stage and progression of the illness (Ghosh and Premchand, 2015). Microbial interventions like prebiotics, probiotics, antibiotics, and faecal microbiota transplants are also used to alleviate IBD symptoms and restore microbial balance in the gut. A combination of genetic and environmental factors modulating gut microbial composition, triggering a dysregulated immune response is known to cause this disease. The exact pathogenesis of UC is very complex and remains elusive, however environmental cues, particularly dietary and lifestyle changes, seem to play a huge role the development of the disease. As extensively discussed in chapter 1, the occurrence of IBD, particularly UC, is highly dependent on and influenced by gut microbes (Sellon *et al.*, 1998; Roy *et al.*, 2017; Shen *et al.*,

2018). Studies have linked imbalances in the composition of the gut microbiota, known as dysbiosis, to IBD. Changes in microbial diversity reshapes the metabolome, affecting the host in ways we have yet to discover.

In fact, IBD is one of the most studied microbiome-linked diseases. Yet there is an absence of consistent and reproducible microbial associations, particularly in colitis mouse models. Murine colitis models have been expediently used to investigate deeper connections between the gut microbiome and UC. However, experimental design and procedures as well as data computation methods within the field are highly variable and inconsistent, which could be contributing to the lack of unified taxonomic patterns tying the microbiome to UC. Therefore, I hypothesize that standardising the methods used to analyse microbiome data will reduce variability among results producing reproducible taxonomic patterns linked to UC.

16S rRNA amplicon sequencing and metagenomic shotgun sequencing are the two most used tools in dissecting the microbiome. The 16S rRNA gene is used as a biomarker as it ubiquitous among prokaryotes, carrying ample interspecific polymorphisms. It stretches for about 1.5kb and has 9 short hypervariable regions that are used to differentiate between bacterial taxa (Weinstock, 2012; Kang *et al.*, 2015). The sequences of one or more of the variable regions are targeted to identify the composition of the microbial populations. This is demonstrated in figure 2.1 (Shahi, Freedman and Mangalam, 2017). Shotgun sequencing can also detect the functional properties of the taxa as well as classification (Weinstock, 2012). However, this meta-analysis will only employ 16S rRNA datasets since they seem to be more widely available in the given context of our research question.

Bioinformatic pipelines for 16S rRNA sequencing are generally follow two different approaches producing distinct feature types: Operational Taxonomic Units (OTUs) or Amplicon Sequence Variants (ASVs).

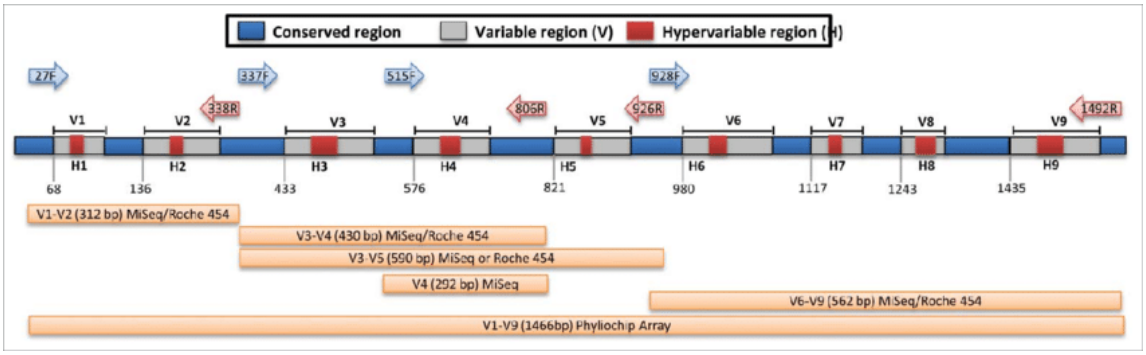


Figure 2.1: Illustration of the 16S rRNA gene, highlighting its conserved, variable, and hypervariable regions as well as multiple primer pairs used for metagenomic sequencing taken from Shahi, Freedman and Mangalam (2017)**.**

OTUs are generated using a clustering approach which combines similar sequences together, usually with a similarity of 97%, to account for sequencing errors/noise. Grouping similar sequences together reduces the misinterpretation of sequencing errors as biological variation. Each OTU is often represented by a representative consensus sequence. This approach is based on the idea that related organisms will have similar target gene sequences while sequencing errors, if present, will not affect the consensus sequences of the OTUs.

There are three main methods of generating OTUs: *de novo* clustering, closed reference clustering and open reference clustering. *De novo* clustering creates OTUs only from observed sequences, without relying on a reference database. This process is time consuming and computationally expensive. Moreover, if data is added or removed from the study clustering must be repeated as it is highly dependent on the sequences identified. Closed reference clustering is a more efficient method for generating OTUs and uses a reference database of target gene sequences from known taxa. Despite this method being both time and computationally efficient it is subject to reference bias. In other words, novel taxa from the

samples will be overlooked. Open-reference clustering was developed to avoid the loss of novel sequences. It does that by using a mix of both *de novo* and closed reference clustering. The nature of the data or samples being used dictates the clustering method to be implemented. Open-reference clustering was used to process the data on hand.

QIIME/QIIME2 and MOTHUR are the commonly used platforms for microbiome bioinformatics, however, they tend to inflate the number of OTUs present in a sample (Prodan *et al.*, 2020). Therefore, another pipeline was adapted for the processing of the sequencing data in this meta-analysis (Dumbrell, Ferguson and Clark, 2016). This pipeline uses freely available packages including **Sickle**, **SPAdes**, **PANDAseq**, **VSEARCH,** and **UCHIME** (Joshi *et al.,* 2011; Edgar *et al.*, 2011; Bankevich *et al.*, 2012; Masella *et al.*, 2012; Nikolenko, Korobeynikov and Alekseyev, 2013; Zhang *et al.*, 2014; Rognes *et al.*, 2016).

Rather than clustering similar sequences into an abstract consensus sequence, the ASV method determines the exact sequences that were read and the number of times they were read. Similar read sequences are then compared, using the error model generated by the algorithm for each sequencing run, to determine the probability that a certain read at a given frequency is an actual read and not a sequencing error. Essentially, this process generates a p-value for every exact sequence detected, where the null hypothesis is equal to that same sequence being error-derived instead of a true sequence. The sequences are then filtered based on a set threshold value for confidence, resulting in a list of "true" sequences with statistical confidence. DADA2 (Divisive Amplicon Denoising Algorithm 2) is a common algorithm used to infer exact ASVs from amplicon sequencing data (Callahan *et al.*, 2016). It is generally run as an R script using its corresponding R package and was used throughout this work.

Meta-analysis of existing studies is an approach that could potentially help generalise and integrate discoveries across different studies. It involves the merging of findings from independent studies to calculate an absolute effect using statistics. Heterogeneity in study approach, sample sizes and findings are some of the issues addressed when conducting a meta-analysis. There is no uniform method for meta-analysis but, there are five common steps or components including: formulating a question, conducting a systematic review, extracting data, standardising, and scoring studies and final estimation of effect (Shorten and Shorten, 2013). Each of the steps involved in this meta-analysis will be addressed throughout this chapter. Meta-analysis of microbial community profiles is a good way to unify the findings from multitude of studies. However, it presents its own set of challenges including strong batch, inter-individual and inter-population differences; in addition to, statistical complications due to zero-inflation and compositionality. Methods for cohort- and batch-effect corrections for other 'omic' data are present but they cannot be directly applied to microbial data. There are several normalisation approaches that can be used except they do not account for batch-effect. Thus, Meta-analysis Methods with a Uniform Pipeline for Heterogeneity in microbiome studies (MMUPHin) was used in the analysis of the pre-processed colitis mouse model datasets (Ma *et al*., 2020).

MMUPHin is an R package designed for joint normalisation, meta-analysis and population structure discovery using microbial community taxonomy and functional profiles. It does so by combining data transformation and linear modelling combinations for 16S data with fixed and random effect modelling. This tool has been tested on a set of 10 published 16S studies of the IBD gut microbiome in humans and identified a gradient in the gut microbiome indicative of increasing dysbiosis in a subset of samples.

2.3 Study Selection

In identifying the microbial associations in murine UC using a meta-analysis approach, a systematic review of a list potential datasets was conducted. This review included primary studies exploring the gut microbiome in UC mouse models using 16S rRNA sequencing. In September 2020, the following PubMed search query was used to generate an unbiased catalogue of studies investigating the gut microbiome in colitis mouse models:

((Microbiome[ALL] OR microbiota[ALL] OR microflora[ALL] OR microbial ecology[ALL] OR gut[ALL] OR fecal[ALL] OR feces[ALL]) AND ("Dextran Sodium Sulphate"[ALL] OR DSS[ALL] OR colitis[ALL]OR ulcerative[ALL] OR bowel disease[ALL] OR "IBD"[ALL] OR colitogenic[ALL] OR Inflammatory Bowel Disease[TIAB]) AND (Mouse[ALL] OR murine[ALL] OR mus musculus[ALL] OR "animal model"[ALL] OR mice[ALL]) AND (16S[ALL] OR metagenome[ALL] OR "ribosomal RNA"[ALL] OR sequencing[ALL] OR metagenomic[ALL])) NOT (review[PT] OR clinical trial[PT]) AND (colitis[TIAB])

The results were filtered to only include papers published between 2015-2020, which yielded 336 potential studies for inclusion in the meta-analysis. The search results were downloaded from NCBI and divided equally among 3 lab members for the first part of the selection process. The papers were first filtered by relevance, where the titles and abstracts were scanned for keywords like “colitis”, “ulcerative colitis”, “microbiome”, and “mouse model”. Only the studies with publicly available 16S rRNA data were chosen. The remainder of these publications was narrowed down further by eliminating studies that used less than 5 mice per cohort. The selection workflow described above resulted in a total of 34 datasets that were

eligible for further analysis.

A rigorous approach to a systematic review is essential in picking datasets for meta-analysis, thus a checkpoint was added to the detailed protocol. Two studies that were known to meet the set criteria were used as controls, to test the accuracy of the study selection pipeline. Despite following a very meticulous selection and elimination method, both control studies were absent from the final list of filtered papers.

Investigations into the failed quality check concluded that NCBI's algorithm searches for the term "16S" and its synonyms, provided in the search query, only within the title and abstract of a paper rather than throughout the whole text. Consequently, the term "16S" was removed from the new search query to access a list of papers that was a better suited for the aim of this meta-analysis study:

((Microbiome[ALL] OR microbiota[ALL] OR microflora[ALL] OR microbial ecology[ALL] OR gut[ALL] OR fecal[ALL] OR feces[ALL]) AND ("Dextran Sodium Sulphate"[ALL] OR DSS[ALL] OR colitis[ALL]OR ulcerative[ALL] OR bowel disease[ALL] OR "IBD"[ALL] OR colitogenic[ALL] OR Inflammatory Bowel Disease[TIAB]) AND (Mouse[ALL] OR murine[ALL] OR mus musculus[ALL] OR "animal model"[ALL] OR mice[ALL])) NOT (review[PT] OR clinical trial[PT]) AND (colitis[TIAB])

The new search query generated a set of 1,855 papers, which was too large to process. Thus, a journal impact factor (IF) filter of 7 was added to the selection criteria to produce a more manageable list of 448 studies. The addition of this filter does not in any way suggest that studies published in journals with an IF of less than 7 are less favorable. In fact, about 75% of the studies produced by the search query were published in journals with an IF of less than 7, with the majority lying between 4 and 7. While applying such a filter is not ideal, it was simply a way to truncate the list of papers for manual screening considering the processing of 1,855 studies would not have been feasible within the timeframe of the project. The filters described above (also shown in figure 2.1B below) were applied during the manual screening I did on the 448 publications. 63 studies were selected post screening and were subjected to an additional filtering step which was the sequencing platform. Filtering for studies with sequences generated by Illumina platforms – the most used across the studies – would enable smoother downstream analysis. Incomplete and improperly annotated datasets were also excluded resulting in a final list of 49 selected studies.

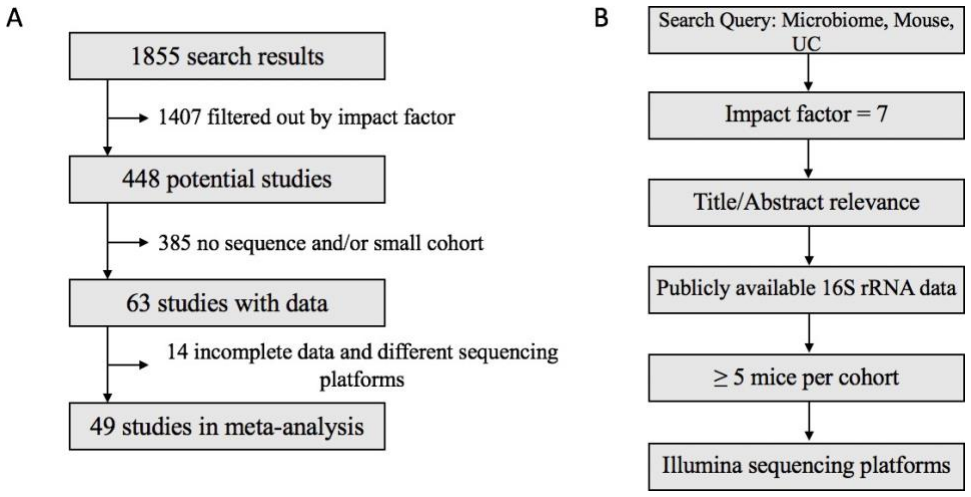


Figure 2.2: Schematic of meta-analysis design highlighting the features used for study selection (B) and the output of each screening step (A). (A) A total of potential 1855 studies were identified by the search query, of which 49 were selected. **(B)** The selection criteria used in the study selection pipeline to refine the search query results.

2.4 Data Retrieval

Sequence data was downloaded directly from SRA, or ENA using the accession numbers listed in table 2.1 onto The University of Essex high-performance computing (HPC) cluster. All datasets were downloaded into their allocated directories and then checked for missing data files; missing data files were downloaded again. Data files from humans or other animals from the selected studies were filtered out. Metagenomic data as well as sequence files from non-Illumina sequencing platforms from the final list of papers were also excluded.

Table 2.1: Initial list of studies for meta-analysis inclusion with their sequence database source and accession numbers.

Deposited Data	Source	Accession Number
Surana *et al*. (2017)	SRA	PRJEB23029
Desai *et al*. (2016)	SRA	PRJNA300261
Llewellyn *et al*. (2018)	SRA	PRJNA403997
Britton *et al*. (2019)	SRA	PRJNA436992
Levy *et al*. (2015)	SRA	PRJEB11300
Zhu *et al*. (2018)	SRA	PRJEB19192
Leonardi *et al*. (2018)	SRA	SRP124782 & SRP124783
Seregin *et al*. (2017)	SRA	SRP076233
Contijoch *et al*. (2019)	SRA	PRJNA413199
Seishima *et al*. (2019)	SRA	PRJNA511382
Miranda *et al*. (2018)	SRA	PRJNA394874
Toubai *et al*. (2019)	SRA	PRJNA491725
Tran *et al*. (2019)	SRA	PRJEB35012
Wheeler *et al*. (2016)	SRA	SRP073269
Gerassy-Vainberg *et al*. (2018)	SRA	PRJNA353641 & PRJNA342751
Sun *et al*. (2018)	SRA	PRJNA351834
Burrello *et al*. (2018)	SRA	PRJNA494680
Viennois *et al*. (2019)	SRA	PRJEB32441
Mamantopoulos *et al*. (2017)	SRA	PRJNA390049
Schulfer *et al*. (2018)	SRA	ERP104982
Nagao-Kitamoto *et al*. (2020)	SRA	PRJNA594915
Ihekweazu *et al*. (2019)	SRA	PRJNA454325
Canesso *et al*. (2018)	SRA	PRJEB22143
Schmidt *et al*. (2019)	SRA	SRP131846
Spalinger *et al*. (2019)	SRA	PRJEB26758
Moschen *et al*. (2016)	SRA	ERP014639
Lemire *et al*. (2017)	SRA	PRJEB22561
Roy *et al*. (2017)	SRA	SRP119278
Sarashina-Kida *et al*. (2017)	SRA	PRJDB5950
Miyoshi *et al*. (2017)	SRA	PRJNA376026
Fachi *et al*. (2019)	SRA	PRJNA486872
Kimura *et al*. (2020)	DDBJ	DRP004759
Reber *et al*. (2016)	SRA	ERP015380
Song *et al*. (2020)	SRA	PRJNA573477
Viladomiu *et al*. (2017)	SRA	PRJNA349809
Nakajima *et al*. (2018)	SRA	GSE115902
Cornick *et al*. (2019)	SRA	PRJNA551025
Powell *et al*. (2015)	SRA	PRJEB6328
Ramakrishnan *et al*. (2019)	SRA	PRJNA531395
Wang *et al*. (2020)	SRA	PRJNA596333
Sun *et al*. (2019)	SRA	PRJNA559351
Khan *et al*. (2020)	SRA	PRJNA616059
Kim *et al*. (2020)	SRA	PRJNA592599
Liu *et al*. (2020)	SRA	PRJNA601328
Ma *et al*. (2020)	SRA	PRJNA574780
Kazakevych *et al*. (2020)	SRA	SRP187112
Qi *et al*. (2020)	SRA	PRJNA577541, PRJNA577385, PRJNA577387 & PRJNA577388

2.5 Metadata Curation

For consistency, sample-specific metadata across the remaining 27 datasets were manually curated. The curated variables were as follows:

1. ***Condition*** consisted of three criteria: "healthy", "colitogenic" and "colitis". "Healthy" referred to a microbiota that did not induce colitis or was not known to induce colitis. A "colitogenic" microbiota was one that induced colitis but was collected and sequenced prior development of colitis, as opposed to a "colitis" sample which was analysed after disease occurrence.
2. ***Treatment*** varied across studies depending on the experimental design. It was generally either a different diet, an immunomodulatory agent, DSS, an antibiotic, or a microbiota transfer of some sort.
3. ***Colitis mouse models*** included DSS, *Il10*$^{-/-}$, *Il6*$^{-/-}$, T cell transfer and modified microbiotas.
4. ***Sex*** of mice was unavailable for most datasets included.
5. ***Length of Treatment*** (in days)
6. ***Intensity of the disease*** in mice, unlike humans, does not have a universal classification and was assigned to three categories (mild, intermediate, and severe) based on my interpretation of each study's observations of the disease. All "healthy" mice under colitogenic treatment (e.g., DSS) were labelled as having "mild" disease, while "intermediate" disease was for those that had mild episodes of colitis. "Severe" was for mice experiencing severe colitis.
7. ***DNA extraction kits*** by Qiagen (QIAmp DNA mini kit, QiaQuick columns, PowerLyzer soil kit), MO BIO (PowerFecal DNA isolation kit, PowerSoil isolation kit), MP Biomedicals (GNOME DNA isolation kit and FastDNA kit for soil), TIANGEN (TIANamp stool kit) and other custom protocols.
8. ***Sequencing platforms*** were either Illumina's MiSeq or HiSeq sequencers.

9. ***Library preparation*** was generally a custom protocol used in each study except for a few which used the Nextera Kit by Ilumina, Pippen Prep by Sage Sciences, an unspecified kit by Thermo Fisher Scientific.
10. ***16S region*** sequenced, when specified, was generally V4 and in a few studies V3-4, V4-5, V1-2.
11. ***Mouse strains*** used in the selected studies were mainly C57BL/6j and the remainders were C3H/HeNCr, BALB/c, GM Swiss Webster and CX3CR1DTR mice.
12. ***Genotypes*** were generally WT except for some *Il10*$^{-/-}$, *Il*18$^{-/-}$, *NLRP6*$^{-/-}$, *NLRP6*$^{-/+}$, *PTPN22* 619W, *PTPN22*$^{-/-}$, *muMY*$^{-/-}$, *Tcrbd*$^{-/-}$, *CD4*$^{-/-}$, *CD8*$^{-/-}$, *Rag2*$^{-/-}$, *Sftpd*$^{-/-}$, *GSDMD*$^{-/-}$, *ASC*$^{-/-}$, *Reg4*$^{-/-}$, BTΔMAFF, BVΔMAFF, *CD11c*.cre$^{+/-}$, and *CD11c*.cre$^{-/-}$.

2.6 Dataset Processing

Two different dataset processing pipelines were tested on a single dataset from a study by Kazakevych *et al*. (2020). In this study, the authors described that the response to DSS is highly dependent on the composition of the microbiome – they realised – while investigating the role of *Smarcad1* in murine DSS-induced colitis.

2.6.1 OTU Method

The first workflow that was tested was adapted from the one detailed by Dumbrell *et al*.(2016), which is based on the OTU clustering method. Pair-ended Illumina sequence read files were downloaded from SRA using fastq-dump and their allocated accession number (SRP187112). The read files were then split and counted to ensure there were no missing data files. Quality trimming was carried out using **Sickle** (Joshi *et al*., 2011). This quality filtering tool uses a "sliding window" method in determining the exact location on the 3' end of the read to trim. Following that, **SPAdes** (Bankevich *et al.*, 2012) was used for error correction and denoising. The **BayesHammer** (Nikolenko, Korobeynikov and Alekseyev, 2013) denoising algorithm is implemented and parallelised within **SPAdes** allowing for faster computation. Pair-end assembly was done using **PANDAseq** (Masella *et al.*, 2012) which, like **SPAdes**, utilises several algorithms. The **PEAR** algorithm (Zhang *et al.*, 2014) within **PANDAseq** was used for merging the quality filtered and error corrected reads. The sequences were then dereplicated and clustered using **VSEARCH** (Rognes *et al.*, 2016). Chimera checks were carried out with **UCHIME** (Edgar *et al.*, 2011). Finally, taxonomic ranks were assigned using the Ribosomal Database Project (RDP) Classifier 2.13 (Wang *et al.*, 2007) with the training set no18. All the programs used were run using their default parameters (table 2). An OTU table and a taxonomy table are the main outputs of this pipeline which were used in downstream analysis described in section 2.6 of this chapter.

2.6.2 ASV Approach

DADA2 (Divisive Amplicon Denoising Algorithm 2) is the most widely used algorithm to infer exact ASVs from amplicon sequencing data (Callahan *et al.*, 2016). It is generally run as an R script using its R package. This package was used for the pre-processing of the dataset (Kazakevych *et al.*, 2020) and generated an ASV table for the downstream analysis to follow. As described in the OTU approach, the data was downloaded from NCBI using the assigned accession number. The sequencing files were split and counted for missing files. Then the DADA2 pipeline was run on this data using R. First, the reads were filtered by quality using the "filterandTrim" function (table 2) while testing multiple filtering parameters to allow for 3 and 4 expected errors per read instead of the default value '2'. Only 25 bases were trimmed from the ends of the reverse reads based on the quality plots. Then, the error rates were estimated from a random set of subsampled reads. This is generally done for each sequencing run since each run tends to have a different error profile. The reads were then deduplicated prior to sample inference. ASV inference is carried out using the quality scores associated with each of the unique sequences. The forward and reverse reads are then merged followed by chimera detection. Taxonomy is then assigned to chimera-removed sequence variants using the same RDP Naïve Bayesian Classifier algorithm mentioned previously. The product of this pipeline is an ASV table (equivalent to an OTU table) and a taxonomy table.

Three DADA2 pipeline flows were run on the Kazakevych *et al.* (2020) sequencing data using different quality score filters. The first filtered reads to a maximum of 2 expected errors per read (which is the default filter), while the second and third flows allowed for 3 and 4 expected errors per read, respectively. All three flows removed the last 25 bases from the reverse reads and none from the forward reads. This cutoff was based off the quality plots generated at the start of the pipeline.

The remainder of the datasets were also processed following the DADA2 pipeline and the steps mentioned above. First, a quality check was run on each dataset using FASTQC. When present, 15 bp of primer sequences were trimmed off the reads. Reads were trimmed according to the quality of plots presented in the FASTQC reports. Several trimming parameters were tested for each of the datasets. Parameters which yielded the highest number of quality reads without compromising pair-ended read alignment and taxonomy assignment were chosen for the pre-processing pipeline for each of the datasets (table 2.2). In some cases, only the forward reads were used due to poor quality of reverse reads, failed read merging and no taxonomy assignment. The datasets that failed to assign taxonomy were eliminated from the list (highlighted in table 3.2).

Table 2.2: Description of arguments used to set important parameters in both data pre-processing pipelines.

Tool/Function	Parameter	Description	Additional Comments
OTU Clustering			
Sickle	pe	Runs the pair-end mode	
	-f, -r	Specifies the forward and reverse read files	
	-o	Specifies the output files and their location	
	-t	Sets the FASTQ quality format	It was set to 'sanger'
SPAdes	--only-error-correction	Performs error correction only	
	-1	Specifies file with forward reads	
	-2	Specifies file with reverse reads	
	-t	Number of threads to be used	8 (default is 16)
	-m	Set memory limit	32Gb
	-f, -r	Specifies the forward and reverse read files	
PANDAseq	-A	Sets the algorithm to use	PEAR was used
	-B	Allows the sequence to be missing a barcode/tag	
	-w	Writes all assembled sequences in FASTA format	
vsearch	--derep_fulllength	Merges identical sequences during dereplication	
	--sizeout	Adds abundance annotations to the output fasta file in dereplication step	

	--minseqlength	Discards sequences shorter than the value specified	Set at 50
	--fasta_width	Sets values for fasta file wrapping	Value = 0 which eliminates wrapping
	--minuniquesize	Discards sequences with a post- dereplication abundance of specified value	Set at 2 (i.e., Discard OTUs with less than 2 reads)
	--relabel	Relabels output files	
	--uchime_denovo	Detects chimeras present in fasta files without a reference	
	--nonchimeras	Specifies the name of the output non- chimeric file in fasta-format	
	--cluster_size	Clusterizes the fasta sequences in specified file sorted by decreasing sequence abundance	
	--id	Clustering identity threshold	set at 0.97 (i.e., Sequences with 97% similarity)
	--centroids	Creates an output file with centroid sequences (i.e., the first sequence of each cluster)	
	--usearch_global	Compares the reference sequences with the merged read files, using global pairwise alignment	
	--db	Reads the reference sequences annotated with taxonomy	
	--threads	Number of computation threads to be used	40

DADA2			
filterAndTrim	tuncLen	Truncates reads after set value.	F=250 (no trimming) R= 225 based on the quality plots
	maxN	Allows a set number of ambiguous bases	Default value = 0
	maxEE	Discards reads with higher than set value of "expected errors" after filtering	Value = 2
	trunQ	Truncates reads that have a quality score less than or equal to set value	Default value = 2
removeBimeraDenovo	method	Sets the method by which chimeras are removed	Default method was used, "consensus". The samples are independently checked for chimeras instead of being pooled. Each sequence is checked for variants.

2.7 Taxonomic Profiling and Microbial Community Analysis

2.7.1 Microbiome Analysis

Raw data in the form of an ASV table, a taxonomy table and a FASTA file were used to, first, generate a phylogenetic tree in R. The sequences were aligned first using the R package DECIPHER (Wright, 2020) before a phylogenetic tree was constructed using ape and phangorn v. 2.11.1 (Paradis & Schliep, 2019; Schliep *et al*., 2017). For easier data management, the ASV table, taxonomy table, FASTA file, phylogenetic tree and sample data file were all merged into a "phyloseq object" using Phyloseq v1.38 (McMurdie & Holmes, 2013). Samples with less than 7000 reads were filtered out. The data was rarefied to even depth with replacement and taxa with less than 5% prevalence across samples were omitted to avoid false positives. Wilcoxon signed-rank tests and PERMANOVA tests were used to test for statistical significance. Ordination of the data points was done using the beta diversity measure Bray-Curtis and visualised in a PCoA plot. DESeq2 v1.32 (Love, Huber & Anders, 2014) was used for differential abundance testing with a p value cut-off of 0.01 and multiple testing correction using the false discovery rate (FDR) method. While DESeq2 was originally developed to test for differential gene expression in RNA sequencing analysis, the tool has been adapted for 16S analysis. The differentially abundant taxa detected by DESeq2 were visualised in the form of a heatmap using the package ComplexHeatmaps (Gu, 2022).

2.7.2 MMUPHin

MMUPHin requires two input files: feature-by-sample matrix and the metadata data frame. The feature-by-sample files is created by merging the taxonomy and feature abundance tables generated by the DADA2 pipeline. Surana and Kasper (2017), Leonardi *et al.* (2018), and Seishima *et al.* (2019) are the three studies used for the testing and optimization of the MMUPHin pipeline. Post metadata curation, the abundance tables for each study was merged

into one table then re-assigned taxonomy to avoid repeated ASVs. The feature abundance and taxonomic tables were merged into the feature-by-sample table and were run on the MMUPHin pipeline along with the metadata file which allowed for joint normalisation as well as batch and study effect correction. Once the pipeline was optimised, the test was repeated multiple times including all 27 datasets, then 14 datasets based on the DSS mouse model and two more follow up runs building on the latter.

For the meta-analysis, all 27 ASV and taxonomy tables were combined into one feature abundance table. Along with the meta-data, the feature abundance table was used as in input for MMUPHin which allowed for joint normalisation as well as batch and study effect correction. This was repeated several times for different combinations of studies using the DSS mouse model. A multivariate analysis of variance (MANOVA) test was run before and after MMUPHin to determine the efficacy of the algorithm in reducing batch effect.

2.7.3 GitHub Repository

All the scripts used in the pipelines described above are available in my GitHub repository using this link:

https://github.com/SalmaES/meta-analysis-microbiome-vs-uc

Chapter 3: Results

3.1 Chapter Overview:

The aim of this chapter is to list and discuss the results of the explained methods in the previous chapter. Firstly, the outcome of the pre-processing steps including study selection, decision behind the chosen pipeline, and effects of used parameters throughout pre-processing of sequencing data is detailed. Then results from taxonomic analyses is discussed, starting with the meta-analysis in 3 different runs as well as the justification behind their formats, followed by the individual datasets.

3.2 Study Selection

As detailed in the methods chapter, a search query designed to generate a list of papers studying gut microbial associations in murine ulcerative colitis (UC) produced a set of 1,855 papers. Since this number of studies is too large to manually process single-handedly, an additional selection criterion of a journal impact factor filter of 7 was added to shortlist datasets for this meta-analysis. This step reduced the results to 448 papers which were then manually screened, using the filters described in the methods chapter. 63 studies were selected post screening. An additional filter for sequencing platforms was implemented for smoother downstream analysis. Incomplete and improperly annotated datasets were excluded resulting in a final list of 47 selected studies.

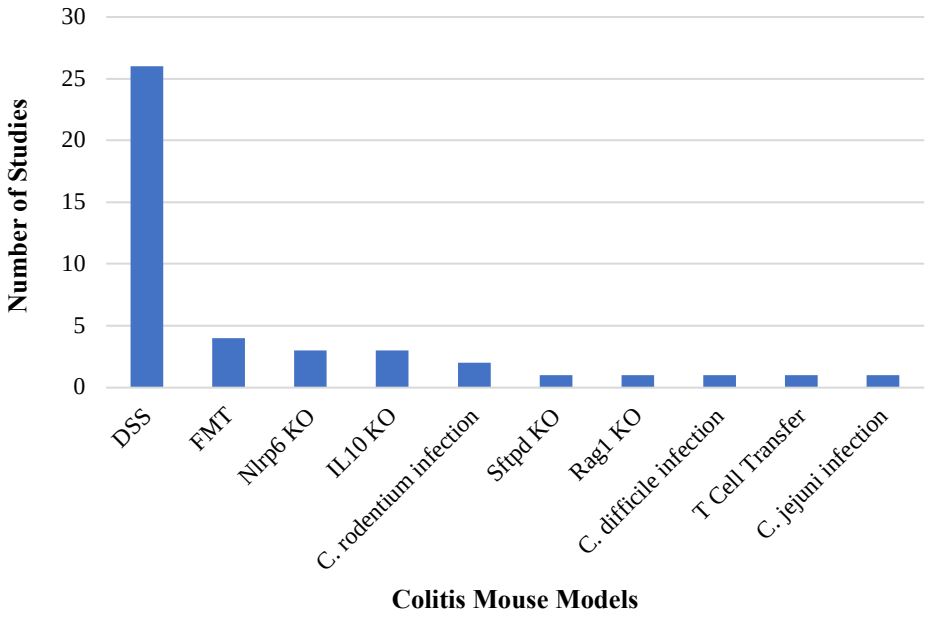


Figure 3.1: Study breakdown by UC mouse model in bar plot format, showing that DSS is the most used method to induce colitis in mice in our dataset consortium.

3.3 Dataset Processing

The study by Kazakevych *et al*. (2020) was used to test the differences between outputs of two 16S rRNA sequencing analysis workflows of different methodologies, the OTU method using a pipeline of multiple algorithms (named Dumbrell pipeline hereafter), and ASV approach using DADA2 v1.20.0 (Callahan *et al*., 2016). This study investigated the role of *Smarcad1* in murine DSS-induced colitis and the authors described how DSS response is highly reliant on the microbiome. The efficiency of each pipeline was evaluated by comparing read tracking information and feature counts (table 3.1). 15% of the sequencing reads were filtered out in the trimming and quality filtering steps using the default DADA2 pipeline while under 1% of the reads were filtered out in the same steps using the OTU method (table 3.1). The filtering parameters were relaxed to allow for more than 2 expected errors per reads to avoid unnecessary loss of sequences. Values for expected errors per read were adjusted to 3 then 4, which only resulted in a 2 percent increase in both the number of reads after quality trimming and retained reads. Given that the increase in read count after relaxing the quality trimming

filters was minimal, the default filtering values were used (table 3.1). Almost half the reads were lost after pair-end assembly in the OTU pipeline. It should be noted that reads are merged further down the ASV workflow, post quality trimming, error correction and deduplication. On the other hand, pair-ended reads are merged right after error correction before deduplication in the OTU pipeline. There is almost a 3-fold increase in chimera detection in the DADA2 method when compared to UCHIME (Edgar *et al.*, 2011). Overall, there was about 17% more reads retained when the DADA2 package was used. Yet, that increase in reads did not translate to feature counts. There were 512 more OTUs generated than there were ASVs.

Feature count is not an ideal measure to compare the efficiency of both pipelines as both features are of different "units" and are generated in two different methods. Therefore, taxonomic profiling was done to determine the biological significance of the discrepancies between both pipelines. It showed no major difference in taxonomic abundance across both pipelines, except for the phylum Bacteroidetes and unclassified taxa (figure 3.2).

Table 3.1: Pipeline optimisation and comparison. Read tracking information and feature output generated by ASV and OTU pipelines. DADA2 (default) used the default parameters during quality filtering. DADA2 (III) allowed for 3 errors per read during quality filtering while DADA2 (IV) permitted 4 errors.

Method	Feature	Total Reads	Reads After Trimming	Merged Reads	Chimeric Sequences	Retained Reads	Feature Count
DADA2 (default)	ASV	4641280	3980618 (85%)	3529290 (76%)	658190 (14.2%)	2871100 (61.7%)	1410
DADA2 (III)	ASV		4075940 (88%)	3962378 (77%)	1716464 (37%)	2924816 (63%)	1409
DADA2 (IV)	ASV		4082246 (88%)	3596230 (77.5%)	1712672 (37%)	2928608 (63%)	1409
OTU Clustering	OTU		4614004 (99.4%)	2294836 (49.9%)	224169 (4.8%)	2070667 (44.6%)	1922

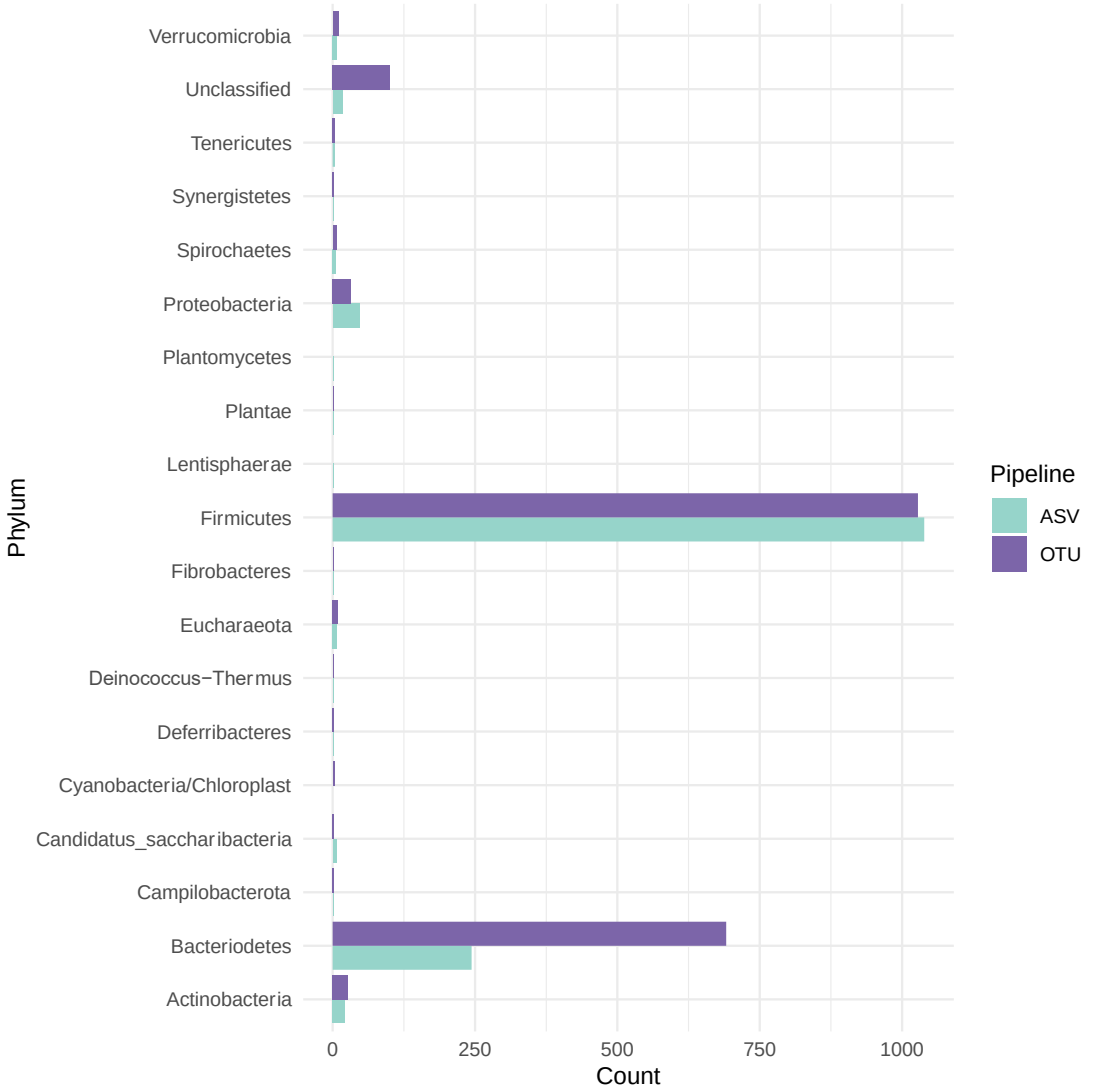


Figure 3.2: Taxonomic profiling of raw feature counts from both ASV and OTU pipelines of the Kazakevych *et al.* (2020) dataset.

This can probably be attributed to the presence of erroneous OTUs as well as less specific taxonomic assignment due to the relaxed nature of the clustering method. Further elaboration of the similarity among outputs of both pipelines is highlighted in figure 3.3. Log10 transformation of relative abundance of the 20 most prevalent taxa is almost identical except for a few taxa between both pipelines (distinguished with an orange circle in figure 3.3). The genus *Mucispirillum* was only present in the output of the ASV pipeline. However, a few genera absent in the OTU pipeline output, like *Lawsonibacter, Flintibacter, Clostridium_XIVa* and *Oscillibacter*, can be tied to less specific

classification of OTUs like “unclassified_Clostridiales” (also known as Eubacteriales) and *Odoribacter* to “unclassified_Bacteroidales”. *Kineothrix* and *Acetifactor* from DADA2 can also be linked to “unclassified_Lachnospiraceae” while *Prevotella* is potentially associated with “unclassified_Prevotellaceae” from the OTU output. Other unclassified taxa from the OTU pipeline, like Ruminococcaceae and Muribaculaceae, could be related to more specific taxa from the ASV pipeline, like the genera *Faecalibaculum*, *Muribaculum* and *Paramuribaculum*. In summary, there is no significant difference between both pipelines. The DADA2 pipeline with its default parameters was chosen for its stringency, specificity, and flexibility in handling more than one dataset, which is ideal for this meta-analysis.

The remaining 47 datasets were processed following the default DADA2 pipeline mentioned above. 8 datasets that failed to assign taxonomy were discarded from downstream analysis (table 3.2). 13 more studies were excluded due to ambiguous metadata, resulting in 27 final datasets for meta-analysis inclusion (table 3.2). These 27 remaining datasets were narrowed down to 13 DSS datasets (elaborated in section 3.4) and processed using DADA2 (table 3.3) for meta-analysis as well as individual taxonomic profiling.

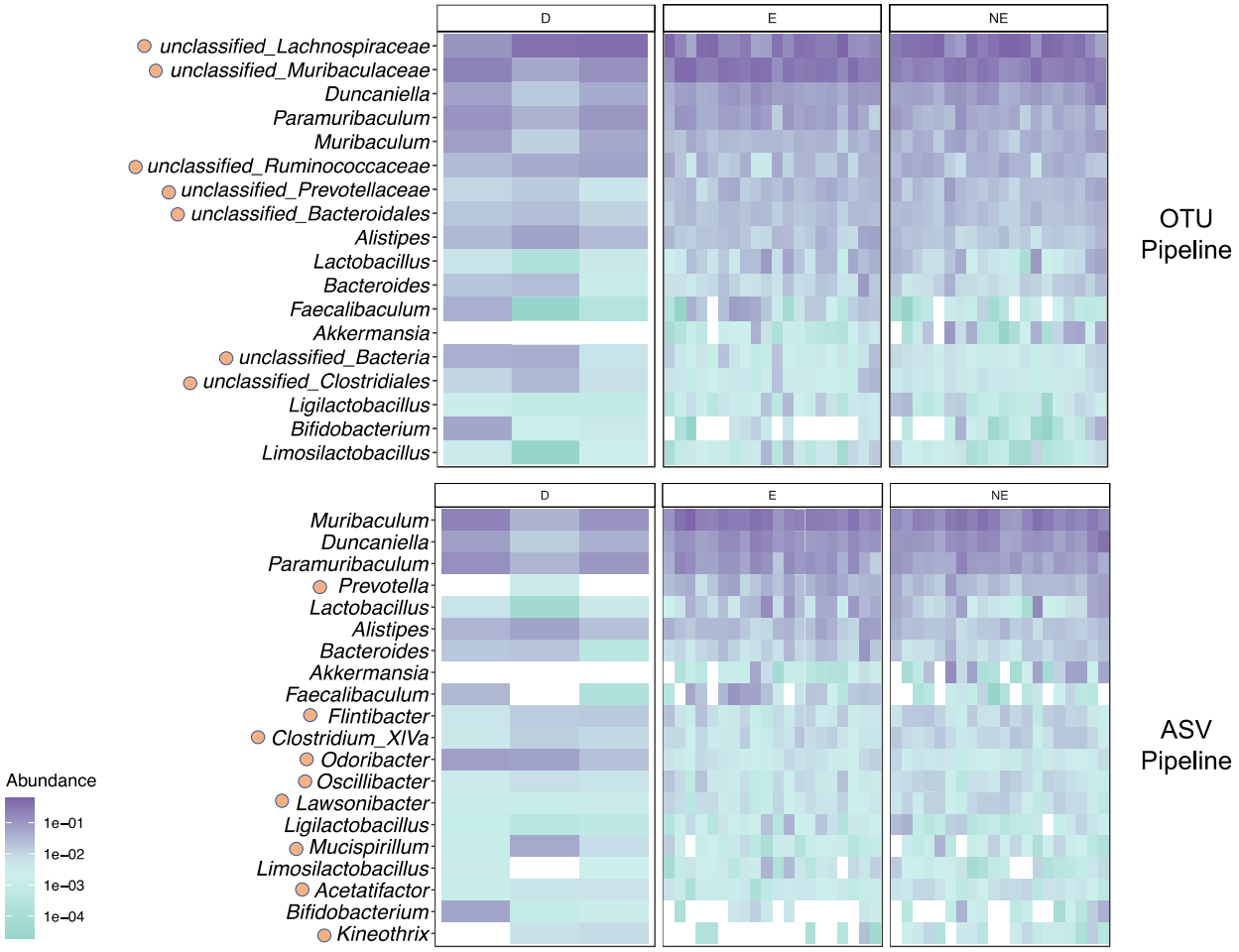


Figure 3.3: Top 20 features (OTU/ASV) from Kazakevych *et al.* (2020) study. Heat map of log 10 transformation of the relative abundance of the 20 most prevalent taxa among the 3 groups (D = donor group, E = enriched microbiota group, NE = non-enriched microbiota group) generated by the Dumbrell (A) and DADA2 (B) pipelines at the genus level, with a prevalence threshold of 10% and abundance threshold of 1%. The orange circle marks taxa that do not match exactly in both pipelines.

Table 3.2: Summary of the pre-processed datasets, trimming parameters, read tracking, feature counts and taxonomy assignment.

Dataset	Library Layout	Trimming Parameters	Total Reads	Retained Reads	Number of ASVs	Assigned Taxonomy
Surana *et al.* (2017)	Single	F=200	4092581	80%	1311	Yes
Desai *et al.* (2016)	Paired	F=225; R=200	32044719	80%	2662	Yes
Llewellyn *et al.* (2018)	Single	F=235	2699674	60%	579	Yes
Britton *et al.* (2019)	Single	F=235	6125271	66%	708	Yes
Levy *et al.* (2015)	Paired	F=250; R=225	38598798	0	N/A	N/A
Zhu *et al.* (2018)	Paired	R=200	954833	92%	737	Yes
Leonardi *et al.* (2018)	Paired	F=175	3630819	58%	1809	Yes
Schroeder *et al.* (2018)	Paired	F=215; R=190	4403799	74%	1491	Yes
Seregin *et al.* (2017)	Paired	F=225; R=150	2939299	79%	1037	Yes

Contijoch *et al*. (2019)	Single	F=235	63634961	68%	7044	Yes
Seishima *et al*. (2019)	Paired	F=225	5715167	79%	1717	Yes
Miranda *et al*. (2018)	Paired	F=225	3106378	95%	1646	No
Toubai *et al*. (2019)	Paired	F=225; R=150	1599481	79%	1356	Yes
Tran *et al*. (2019)	Single	F=250	1296322	93%	1943	Yes
Wheeler *et al*. (2016)	Paired	F=250	5544742	68%	2099	No
Gerassy-Vainberg *et al*. (2018)	Single	F=200	5142274	83%	3093	Yes
Sun *et al*. (2018)	Paired	F=175	8115438	89%	375	Yes
Burrello *et al*. (2018)	Paired	170_170	7693941	62%	1892	Yes
Viennois *et al*. (2019)	Single	F=250	1655719	95%	1389	Yes
Mamantopoulos *et al*. (2017)	Paired	F=200	24730088	88%	10575	Yes
Schulfer *et al*. (2018)	Single	F=135	24988550	94%	1953	Yes
Nagao-Kitamoto *et al*. (2020)	Paired	F=225; R=200	528551	78%	256	Yes
Ihekweazu *et al*. (2019)	Paired	235_165	5964106	71%	1623	Yes
Canesso *et al*. (2018)	Single	F=250	443968	84%	668	Yes
Schmidt *et al*. (2019)	Paired	F=215	4943256	89%	1215	Yes
Spalinger *et al*. (2019)	Paired	F=220; R=220	18037481	72%	6318	Yes
Moschen *et al*. (2016)	Paired	F=235	4069119	70%	2747	Yes
Lemire *et al*. (2017)	Paired	F=250; R=225	14415841	0	N/A	N/A
Roy *et al*. (2017)	Paired	F=225	14595515	73%	8713	Yes
Sarashina-Kida *et al*. (2017)	Paired	F=215	1408356	81%	1981	Yes
Miyoshi *et al*. (2017)	Paired	F=250; R=225	13699727	0	N/A	N/A
Fachi *et al*. (2019)	Paired	F=185	6477281	87%	3012	Yes
Kimura *et al*. (2020)	Paired	F=240; R=190	1772226	71%	996	Yes
Reber *et al*. (2016)	Single	F=100	35605122	77%	1788	Yes
Song et al. (2020)	Paired	F=200; R=150	3618167	82%	446	Yes
Viladomiu *et al*. (2017)	Paired	F=200; R=150	3855419	79%	1765	Yes
Nakajima *et al*. (2018)	Paired	F=175	11704644	83%	2941	Yes
Cornick *et al*. (2019)	Paired	F=216; R=160	3163553	90%	1586	No
Powell *et al*. (2015)	Paired	F=175; R=200	736302	73%	748	Yes
Ramakrishnan *et al*. (2019)	Paired	F=225; R=160	12360516	81%	3228	Yes
Wang *et al*. (2020)	Paired	F=200; R=200	12457760	0	N/A	N/A
Sun *et al*. (2019)	Single	F=180	631274	94%	2665	No
Khan *et al*. (2020)	Paired	F=200	3471149	91%	1675	Yes
Kim *et al*. (2020)	Paired	F=225	1125137	83%	1762	Yes
Liu *et al*. (2020)	Paired	F=210	8275698	84%	927	Yes
Ma *et al*. (2020)	Paired	F=260	705367	87%	971	Yes
Kazakevych *et al*. (2020)	Paired	F=250; R=225	4641280	62%	1410	Yes
Qi *et al*. (2020)	Paired	F=200	3461357	73%	3023	Yes

Table 3.3: Summary of DADA2 processed DSS datasets and meta-analysis of DSS datasets, grouping of DSS datasets which clustered based on "Condition" (group A), and grouping of DSS datasets that either clustered based on "Condition" or other variables (group B).

Study	Number of ASVs	Number of Samples	ASVs > 7000 reads	Samples > 7000 reads	ASVs post Rarefication	Samples post Rarefication	ASVs post Prevalence Filtering	Samples post Prevalence Filtering
Surana & Kasper (2017)	1311	38	1311	38	1231	38	735	38
Tran *et al*. (2019)	1943	56	1943	56	1665	54	665	54
Spalinger *et al*. (2019)	6318	292	6318	287	5788	287	1377	287
Roy *et al*. (2017)	8713	448	8713	443	6950	443	528	443
Song *et al*. (2020)	446	25	446	25	445	25	395	25
Nakajima *et al*. (2018)	2941	85	2941	85	2000	85	643	85
Kazakevych *et al*. (2020)	1410	43	1410	43	1394	43	433	43
Qi *et al*. (2020)	1610	38	1610	38	1588	38	1035	38
Mamantopoulos *et al*. (2017)	10575	22	10575	22	3280	22	1047	22
Schmidt *et al*. (2019)	1215	32	1215	30	931	30	515	30
Kimura *et al*. (2020)	996	12	996	12	977	12	977	12
Khan *et al*. (2020)	1675	49	1675	49	1671	49	1158	49
Ma *et al*. (2020)	971	16	971	16	964	16	964	16
MMUPHin (all DSS)	42881	1156	42881	1142	712	1142	156	1142
MMUPHin (group A)	23278	1023	23278	1011	519	1011	147	1011
MMUPHin (group B)	3367	101	3367	101	353	101	183	101

3.4 Meta-Analysis on DSS Mouse Model Studies

Meta-analysis Methods with a Uniform Pipeline for Heterogeneity in microbiome studies (MMUPHin) is an R package developed by the Huttenhower lab (Ma *et al*., 2020) and was used in this meta-analysis of microbial profiles from murine colitis datasets. The approach implemented by this package addresses the challenges faced when performing meta-analysis (or even regular analysis) on microbiome data such as strong batch effect, massive inter- and intra-population variations as well as zero-inflated compositional values.

3.4.1 Batch Effect Adjustment and Meta-Analysis on All Datasets

A total of 4,396 samples were processed for joint normalisation and batch effect adjustment using the R package MMUPHin. This package accounts for the quantitative challenges faced when analysing microbiome data, like zero inflation, strong batch effects and great variation both within and across samples. The substantial variability in experimental design across the studies contributed to inconsistent patterns across microbial profiles, even after batch effect adjustment. Permutation testing showed that the batch effect from studies accounts for about 44% of microbial profile variability. Only 16% of the batch effect from studies was corrected using MMUPHin. The study effect was too great which made it difficult to draw a link between colitis and gut microbial composition, as demonstrated in the PCoA plots in figure 3.4A. Clustering samples based on colitis model, revealed that gut microbial profiles were highly influenced by the murine colitis model used in each study (figure 3.4B). Thus, in an attempt to reduce complexity of the highly variable data, samples employing one colitis model were selected for further analysis. Since the DSS mouse model is one of the most used models to study UC and around half of the studies in this meta-analysis were based on this model, downstream analysis focuses on those 14 datasets.

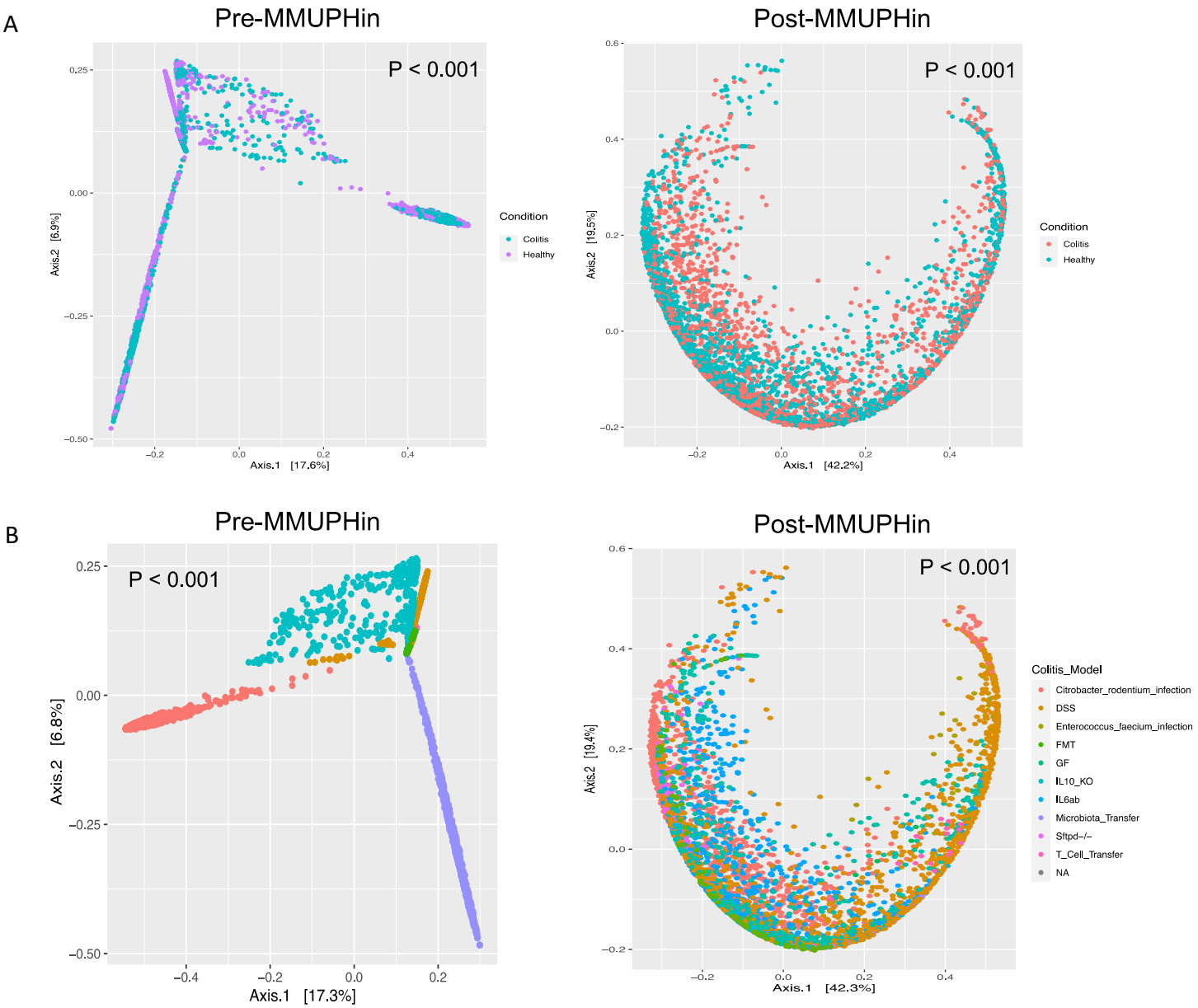


Figure 3.4: Visualisation of beta diversity (Bray-Curtis dissimilarity) across 4,396 microbial profiles from all datasets using PCoA plots coloured based on condition (A) and colitis model (B) before (left) and after batch-correction (right) shows high variability across samples even after adjusting for batch effect using MMUPHin.

3.4.2 Meta-Analysis on DSS Datasets

After changing the "Condition" variable in the metadata to reflect the stage of colitis more accurately (detailed in methods chapter) ("healthy", "colitogenic", "colitis), the study by *Burello et al.* was excluded from the analysis as it did not meet these criteria. Briefly, "healthy" samples were taken from

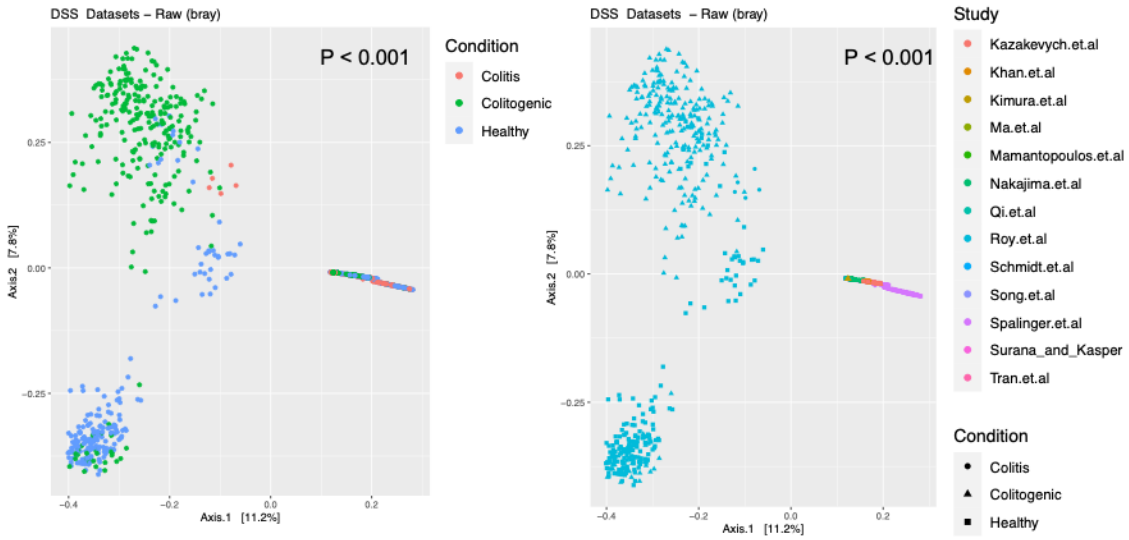
mice that did not experience colitis symptoms, “colitogenic” ones were sampled from mice before developing colitis while “colitis” samples were sequenced from mice experiencing colitis. 1,395 samples were corrected for batch effect and normalised using MMUPHin. Multivariate Analysis of Variance (MANOVA) was used to assess the effect of batch effect adjustment calculated by the adonis function in the R package *vegan* (Okansen *et al*., 2020). The test demonstrated that MMUPHin was able to reduce the batch effect by study variability from 35% to 25%. Despite the large effect of study on microbiome composition, pairwise comparison using the Wilcoxon Rank Sum test indicated significant differences between “colitogenic” and “healthy” microbiotas, as well as “colitogenic” and “colitis” microbiotas, and “healthy” and “colitis” microbiotas ($p < 0.001$). This is demonstrated in the ordination plots from figure 3.5, where samples from each condition mostly aggregate, rather than form separate clusters, using raw data and batch adjusted data with and without outlier samples ($p < 0.001$, PERMANOVA).

Further analysis of the data showed differentially abundant taxa that are associated with UC, whether it being its onset, presence, or absence. DESeq2 was used to identify differentially abundant taxa that are of statistical significance between different condition groups. DESeq2 was chosen over LEfSe, due to the latter’s reports of high false discovery rates (Nearing *et al*., 2022). Members of the order Erysipelotrichales, including *Turicibacter*, *Clostridium cocleatum* (now reclassified as *Thomasclave*lia *cocleata*), and *Faecalibaculum*, were more abundant in microbiotas from colitic mice. Erysipelotrichales species are known to be highly coated with IgA and have been associated with severe colitis in DSS-induced mice (Palm *et al*., 2014). Bacterial genera from Clostridiales (also known as Eubacteriales) and Bacteroidales were more prominent in inflammatory states (colitis and colitogenic conditions), except for *Ruminoccus* and *Acetobacteroides*, respectively, which were more abundant in healthy mice. *Robinsoniella peoriensis* was particularly enriched (21 log2 fold change) in the colitogenic group. This taxon has been previously linked with gut inflammation in IL10-/- mice

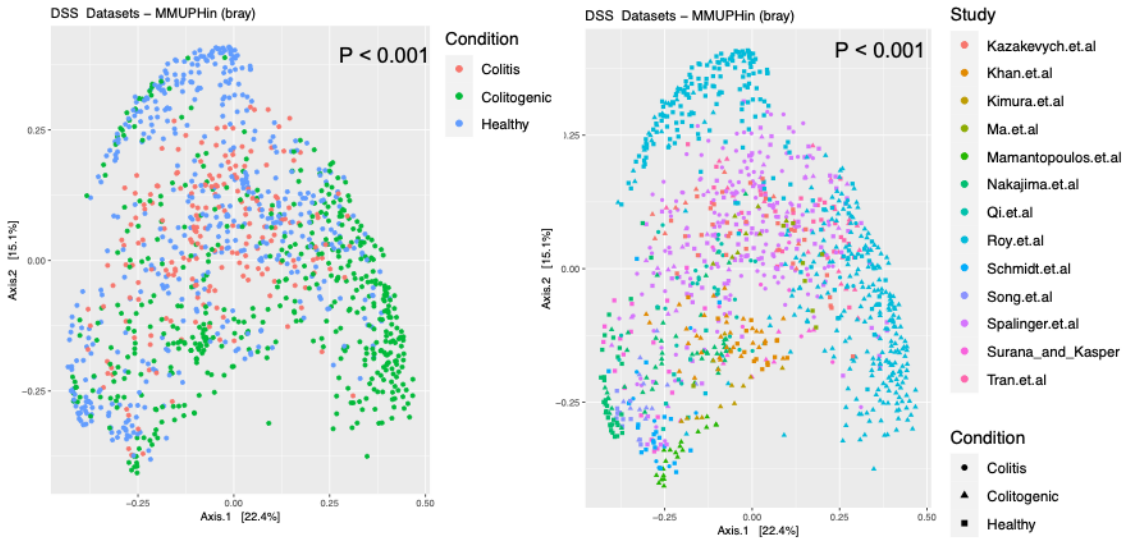
(Wohlgemuth *et al*., 2011) and is known to be a confounder in *C. difficile* infections (Roberts *et al*., 2016). Other taxa, including *Bifidobacteria*, *Rodentibacter*, *Akkermansia*, *Escherichia*/*Shigella* spp., *Helicobacter typhlonius* and *Lactobacillus intestinalis*, were more abundant in non-healthy gut microbiotas and have been previously linked with IBD, inflammation and tumour development (Wang *et al*., 2014; Dingemanse *et al*., 2015; Chai *et al*., 2017; Fornefetti *et al*., 2018; Presti *et al*., 2019; Jochum *et al*., 2022). Healthy mice had higher proportions of the less described *Acetobacteroides* and Lachnospiraceae members that are known for their paradoxical roles in the human gut (Vacca *et al*., 2020). *Eggerthella lenta* was vastly enriched (26 log2 fold change) in healthy mice despite its described role in Th17 activation in murine colitis (Alexander *et al*., 2021). Overall, there was a decrease in Firmicutes and an increase in Bacteroidetes species (figure 3.6).

Condition Study

A



B



C

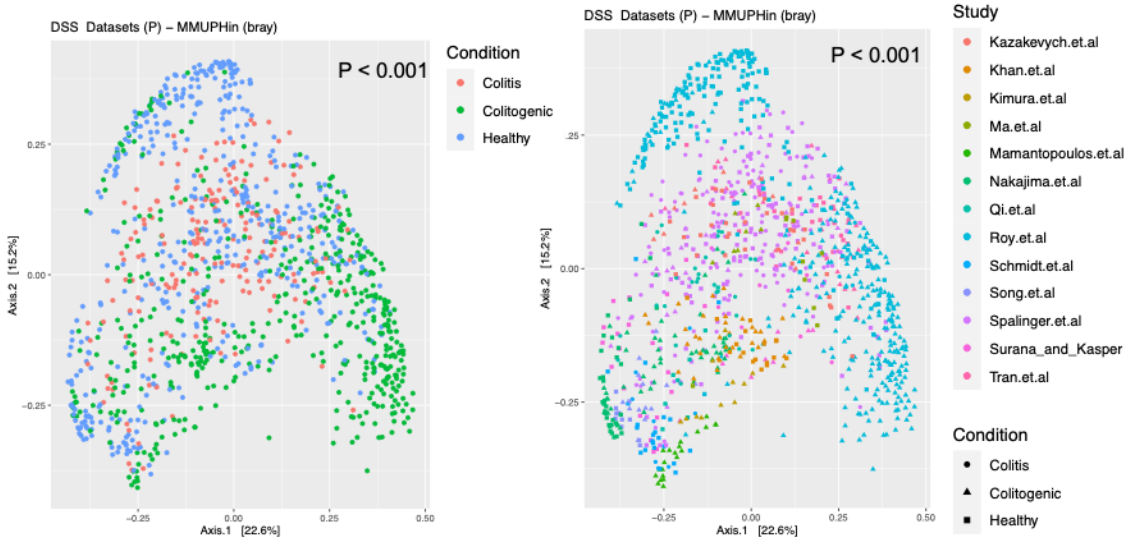


Figure 3.5: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from all DSS datasets using PCoA plots coloured based on condition (left) and study (right) before (A) and after batch-correction with (B) and without (C) samples of less than 7000 reads using MMUPHin shows differences in gut microbial profiles across different condition groups.

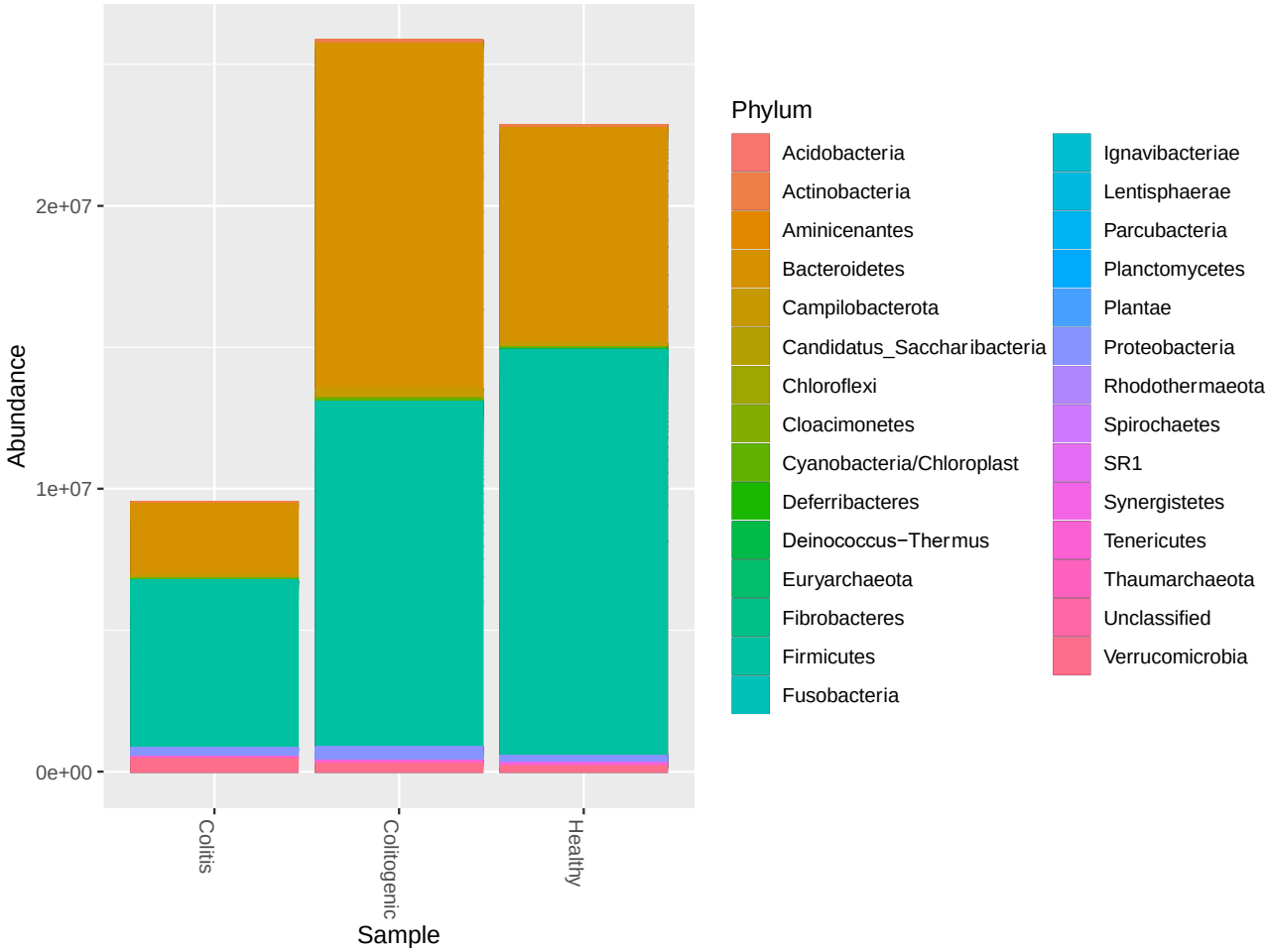


Figure 3.6: Bar plot demonstration of the mean of total abundance of each phylum present in the healthy, colitogenic an colitis groups, showing a reduction in Firmicutes/Bacteroidetes (F/B) ratio in the colitogenic group and an increase in F/B ratio in the colitis group when compared to the healthy group.

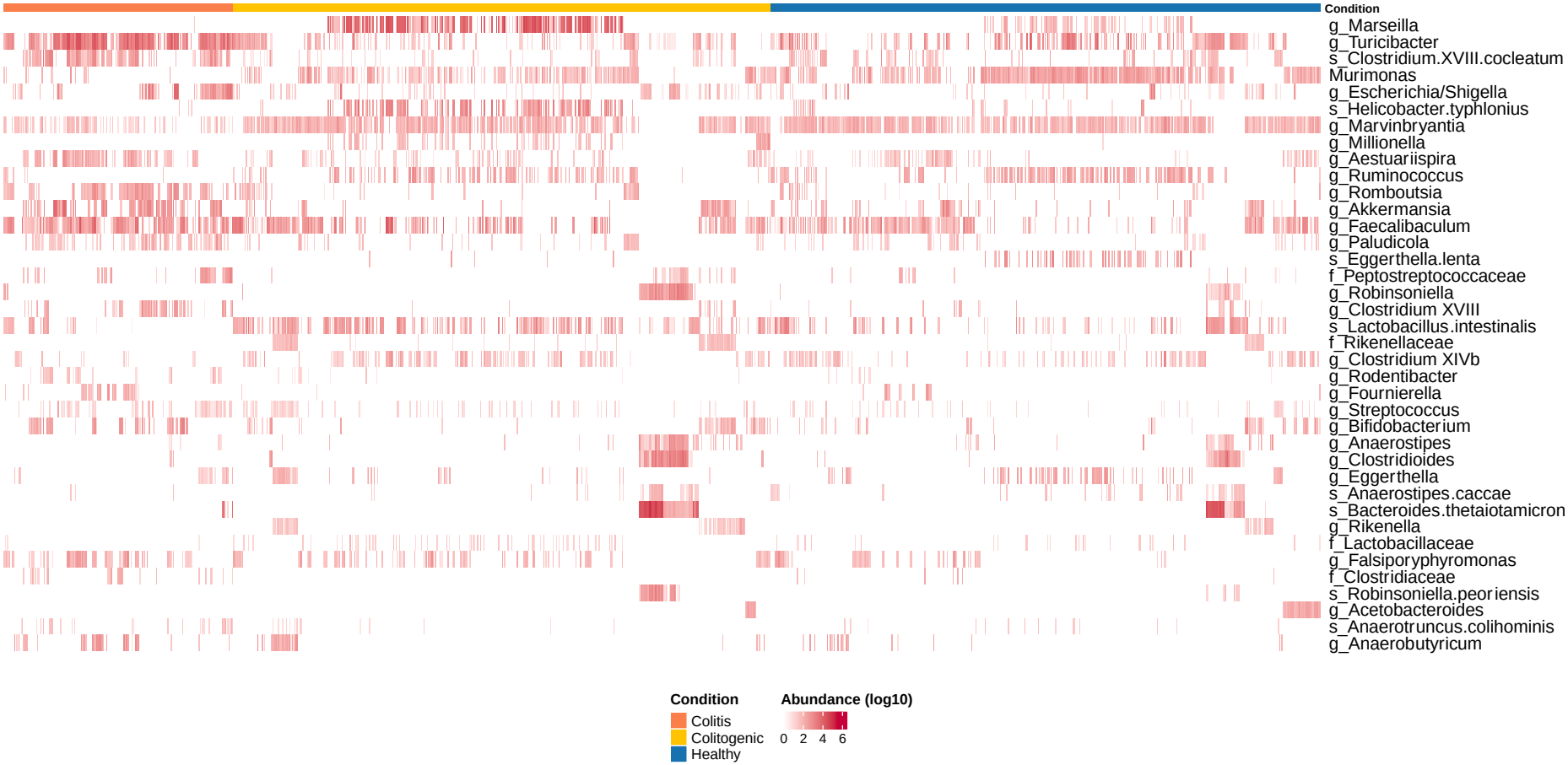


Figure 3.7: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa (at or above a ± 2 log2 fold change) at the lowest taxonomic classification from all samples in the 14 DSS datasets, grouped by condition (blue = healthy, yellow = colitogenic, orange = colitis) and identified by DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.4.3 Meta-Analysis on Different Groupings of DSS Datasets

In attempt to disentangle the datasets and simplify the results further, the meta-analysis was repeated using two groups of studies. The first group (A) included DSS studies from which samples clustered based on condition of the mouse (i.e., "healthy", "colitogenic", "colitis") while the second group (B) included both samples from group A and additional samples which clustered based on other variables like treatment, genotype, and length of treatment (table 3.4). Clustering patterns of samples were identified by analysing each study separately (results in next section). MANOVA testing indicated a 3% and a 4% reduction in batch effect post adjustment in group A and B, respectively. Ordination of the samples from both groups (figures 3.8 & 3.9) using Bray-Curtis Dissimilarity as a distance metric shows that there is a statistically significant condition-based grouping ($p < 0.001$). Furthermore, a pairwise comparison indicated considerable differences between gut microbial communities of healthy, colitogenic and colitic mice, with the exception for healthy vs colitis in group A (table 3.5).

Table 3.4: Grouping of studies used in meta-analysis. Group A included samples which clustered based on condition while group B samples formed clusters using other variables.

Group A	Group B
Surana & Kasper (2017)	Surana & Kasper (2017)
Song *et al*. (2020)	Song *et al*. (2020)
Qi *et al*. (2020)	Qi *et al*. (2020)
	Tran *et al*. (2019)
	Spalinger *et al*. (2019)
	Roy *et al*. (2017)
	Nakajima *et al*. (2018)
	Kazakevych *et al*. (2020)

Table 3.5: P values for Wilcoxon Rank Sum tests on group A and group B of DSS datasets.

Group A		
	Colitis	**Colitogenic**
Colitogenic	5.66E-05	NA
Healthy	0.71655166	0.0001344
Group B		
	Colitis	**Colitogenic**
Colitogenic	1.36E-07	NA
Healthy	2.47E-21	7.73E-07

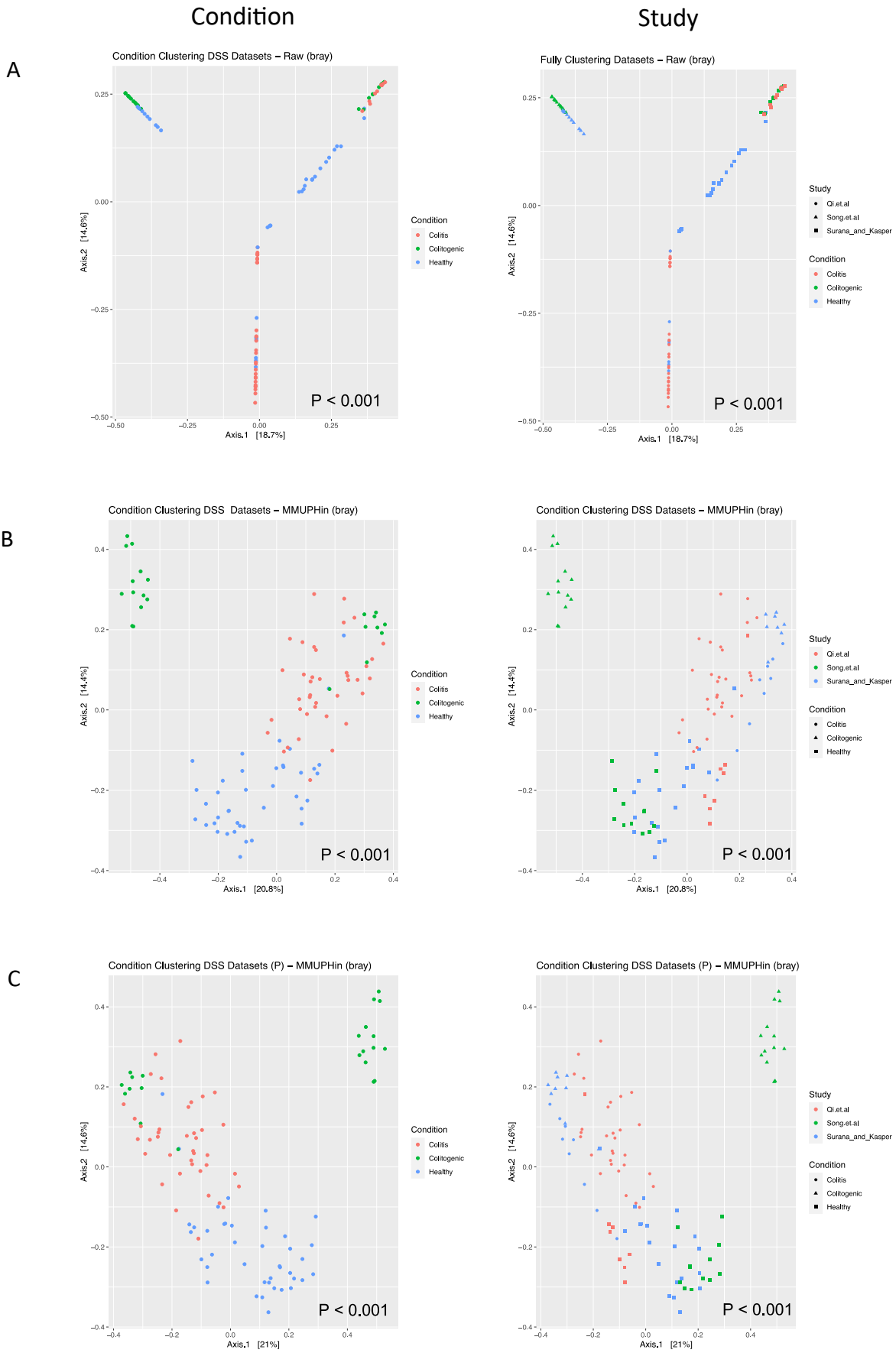


Figure 3.8: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from group A using PCoA plots coloured based on condition (left) and study (right). Repeated for pre (A) and post batch-correction with (B) and without (C) samples of less than 7000 reads using MMUPHin shows differences in gut microbial profiles across different condition groups.

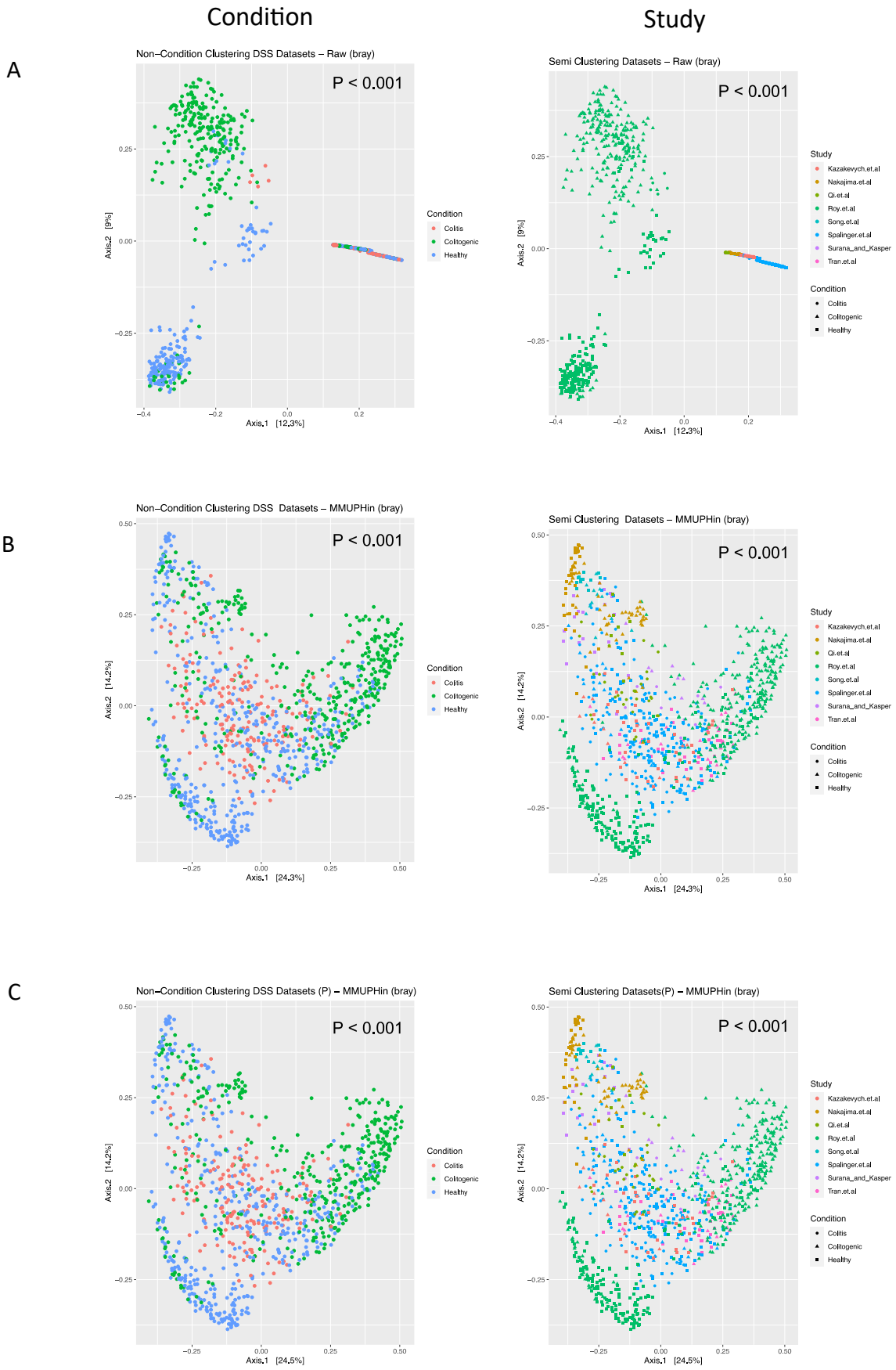


Figure 3.9: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from group B using PCoA plots coloured based on condition (left) and study (right). Repeated for pre (A) and post batch-correction with (B) and without (C) outlier samples of less than 7000 reads using MMUPHin shows differences in gut microbial profiles across different condition groups.

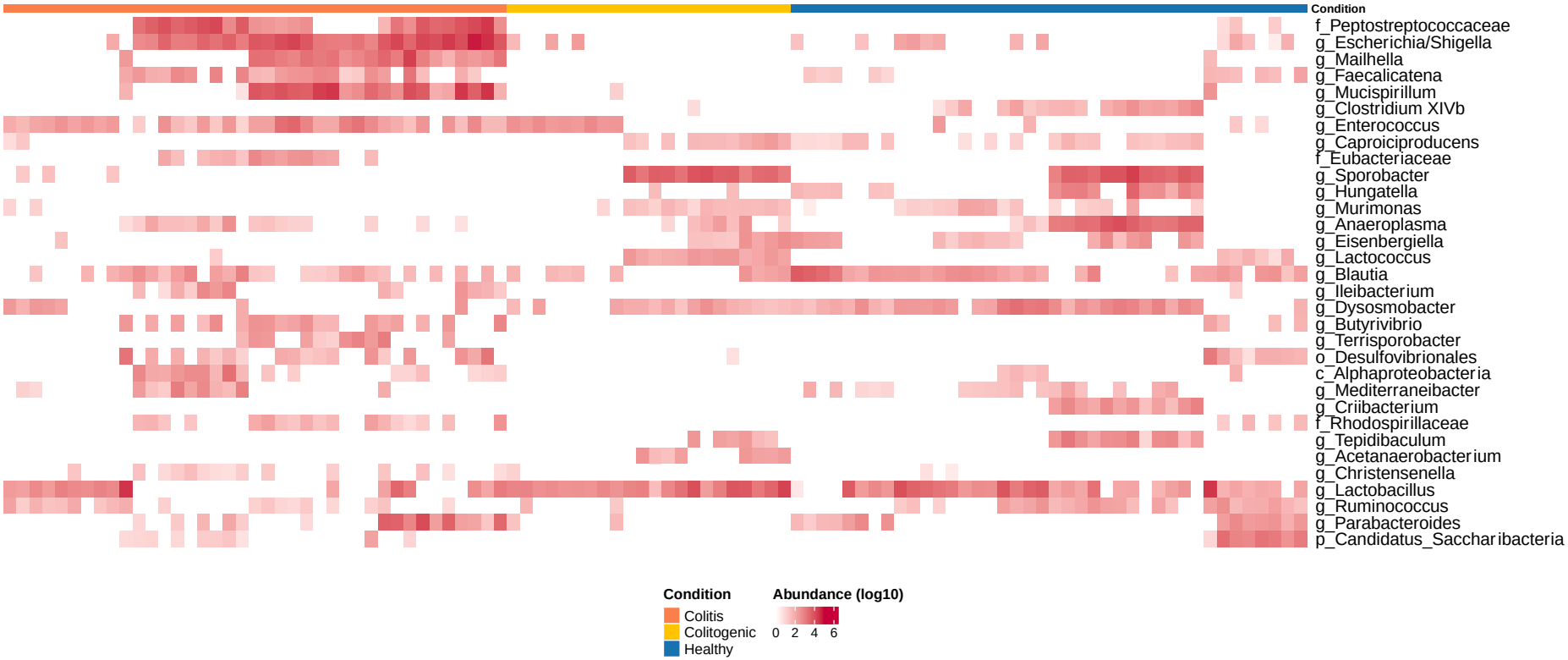


Figure 3.10: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa (at or above a ± 2 log2 fold change) at the lowest taxonomic classification from all group A samples, grouped by condition (blue = healthy, yellow = colitogenic, orange = colitis) and identified by DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

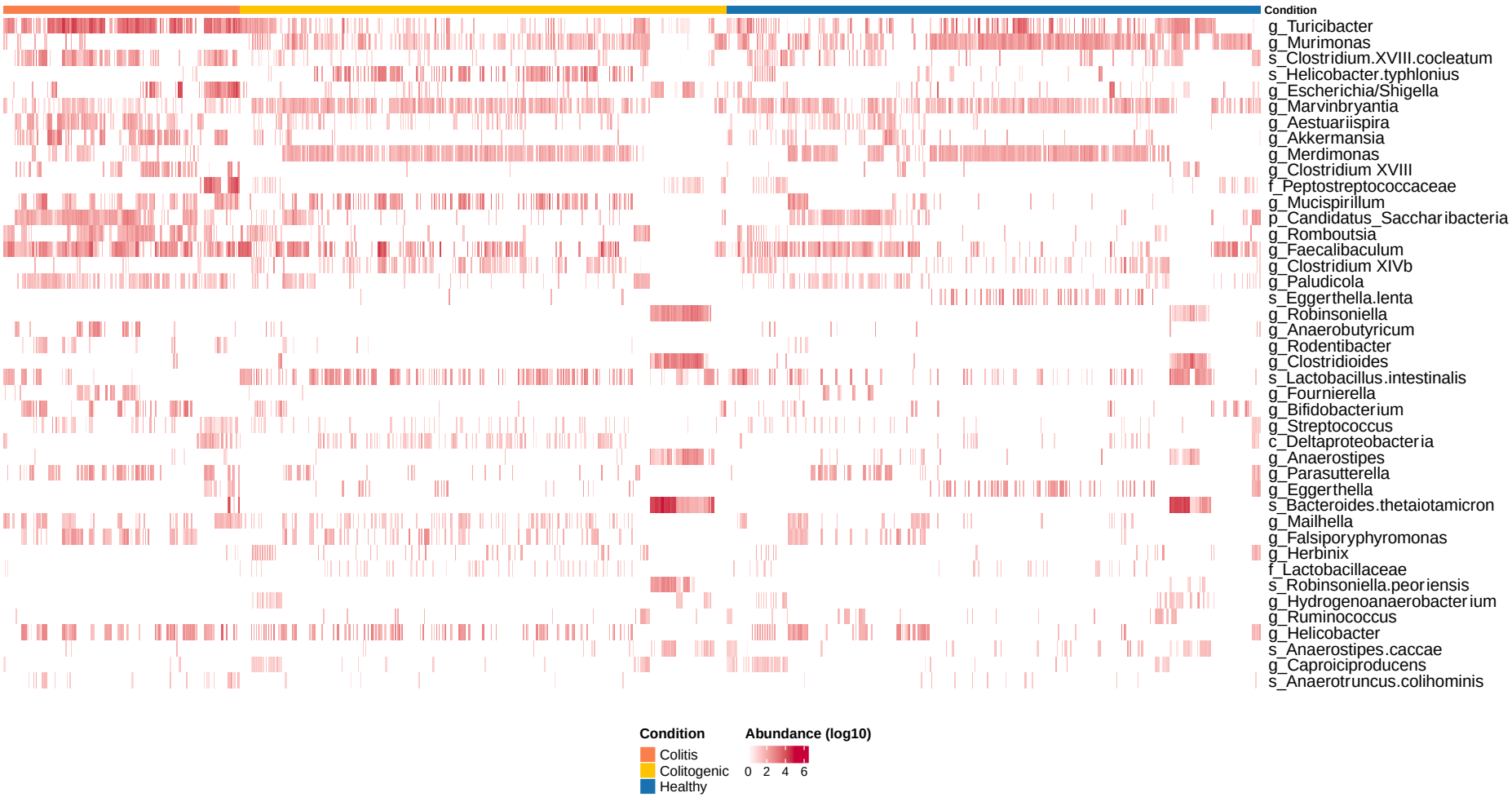


Figure 3.11: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa (at or above a ± 2 log2 fold change) at the lowest taxonomic classification from all group B samples, grouped by condition (blue = healthy, yellow = colitogenic, orange = colitis) and identified by DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

Differential analysis on both groups A and B highlighted numerous taxa prevalent in each of the condition groups. Some differentially abundant taxa were shared across both dataset groups in colitogenic and colitic mice, like Peptrostreptococcaceae, *Escherichia*/*Shigella* spp., *Mailhella* and *Mucispirillum*, which are known to be associated with colitis and even described as pathobionts (Berry *et al*., 2015; Wan *et al*., 2022; Sun *et al*., 2023). Enrichment of the family Peptrostreptococcaceae in non-healthy groups is common with the analysis described in section 3.2 and has been reported in UC patients, increased especially in those with a *C. difficile* infection (Wan *et al*., 2022). This observation coincides with the increased abundance of Clostridium spp. in colitis and colitogenic groups across the 3 different meta-analysis runs. Lactobacillales, Erysipelotrichales, and Bacteroidales orders were also enriched in non-healthy mice as well as Lachnospiraceae in healthy ones, from both groups A and B. Candidatus Saccharibacteria, described in association with murine colitis by Kazakevych *et al*. (2020), was more abundant in mice with colitis from group B. Inversely, this phylum was enriched in healthy mice from group A. This is probably due to differences in abundance of Saccharibacteria in the studies where this taxon was flagged as differentially abundant. Overall, differential abundance analyses showed that the microbial profiles of the 3 condition groups from all DSS datasets and group B were very similar compared to those from group A.

3.5 Taxonomic Profiling and Microbial Community Analysis

This section discusses the results of community (beta diversity) and differential abundance analyses on each of the studies employing the DSS model. These datasets were individually analysed to:

1) identify whether the same differences among microbiotas of mice with different health conditions detected in the meta-analysis were reflected in each dataset.
2) find different ways to group the data, in order to simplify the results and deduce clearer associations with microbial members and health status.

3.5.1 Surana & Kasper

The authors of this study (Surana & Kasper, 2017) investigated the link between microbial composition and phenotypic variations in DSS-induced colitis. They used gnotobiotic mice colonised with either a mouse microbiota (MMb) or a human microbiota (HMb), that were previously generated and characterised in-house, in multiple co-housing experiments to triangulate microbial-phenotypic observations and identify a causative rather than implicative relationship between commensals and disease pathogenesis. Surana & Kasper (2017) demonstrated that the bacterial family Lachnospiraceae protects against colitis-associated death.

Following the same analysis workflow from the previous section, a pairwise comparison using Wilcoxon Rank Sum test showed a significant difference between microbiotas of colitogenic and colitic mice ($p = 0.04$) as well as healthy and colitogenic mice ($p = 0.01$) but no significant distinction between microbiotas from healthy and mice with colitis ($p = 0.7$). Ordination of a beta diversity measure (Bray-Curtis dissimilarity) shows microbiotas from healthy mice clustering away from those of colitic and colitogenic mice. PERMANOVA test confirmed that there is a significant difference in

microbial composition between the groups and that 41% of the variance in distances between samples can be attributed to the condition of the mice. Differential abundance analysis using DESeq2 highlighted associations between bacterial taxa and health condition (figure 3.13). Two ASVs assigned to the genus *Ligilactobacillus* were more enriched in healthy murine guts from this study. Species from this genus, particularly *L. salivarius*, play a role in maintaining the intestinal barrier as well as curbing inflammation via reduction of inflammatory cytokines and increase in anti-inflammatory cytokines in both *in vitro* and *in vivo* models (Yao *et al*., 2021). *Paramuribaculum* was absent from healthy guts and present in the colitis group. While the genus itself and its role in IBD is poorly understood, the family it belongs to Muribaculaceae has been both positively and negatively linked to colitis (Shang *et al*. 2021). Another two less investigated taxa, *Intestinimonas* and *Duncaniella*, were found to be differentially and significantly abundant in both colitogenic and colitis groups. Both have been previously associated with colitis, the former with severe colitis symptoms and the latter as a protectant against DSS-induced injury in mice (Webb *et al*., 2018; Chang *et al*., 2021). Similar to the meta-analysis, two different ASVs assigned to the family Lachnospiraceae have been tied to either the healthy or both disease groups. This probably due to the large number of genera ranked under this family with various functions, some of which can be beneficial and others harmful to the host (Vacca *et al*., 2020). Like in group A, *Enterococcus* is also present in disease groups and depleted in the healthy group, which has been implicated in promoting colonic inflammation (Seishima *et al*., 2019).

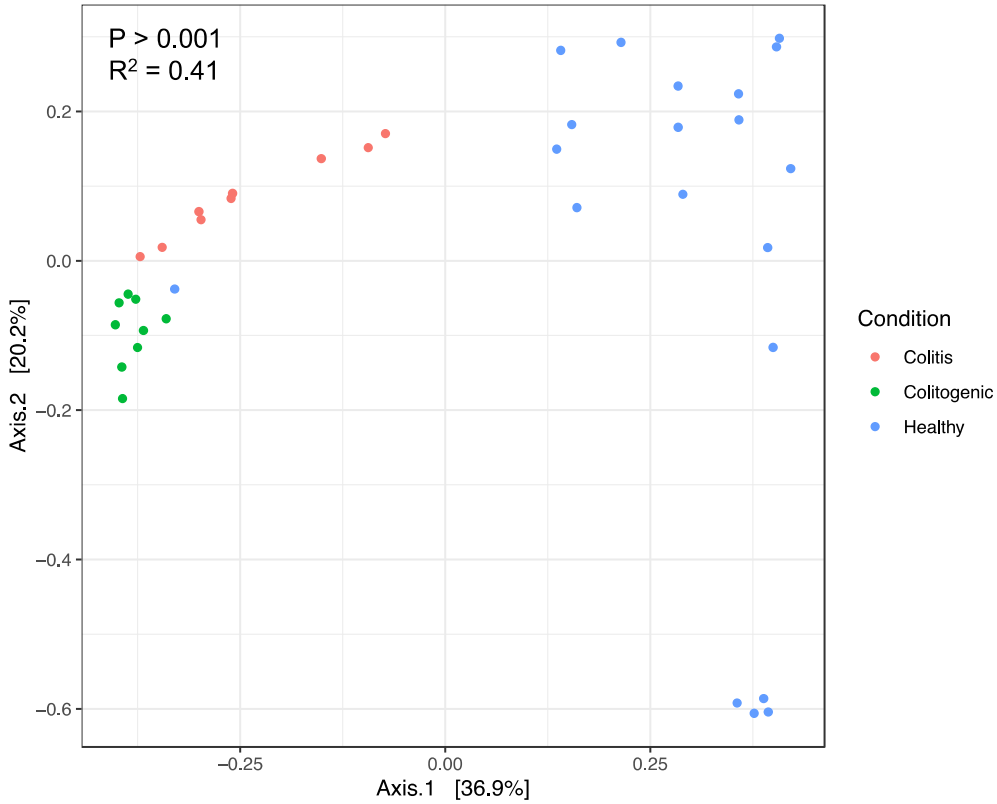


Figure 3.12: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Surana & Kasper (2017) using PCoA plots coloured based on condition (healthy = blue, colitogenic = green, colitis = red) shows differences in microbial populations across condition groups.

Figure 3.13: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa at the lowest taxonomic classification from Surana & Kasper (2017) samples, grouped by condition (blue = healthy, yellow = colitogenic, orange = colitis) and identified by DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.2 Tran *et al.*

This study (Tran *et al.*, 2019) attempted to develop a vaccine against recurrent inflammation of the gut using purified flagellin injections. These injections protected against IL-10 deficiency-induced colitis but not DSS-induced colitis in mice. The authors ran the sequencing experiments of the flagellin-immunised Il10$^{-/-}$ mice pre-DSS exposure at different timepoints during weekly flagellin treatment.

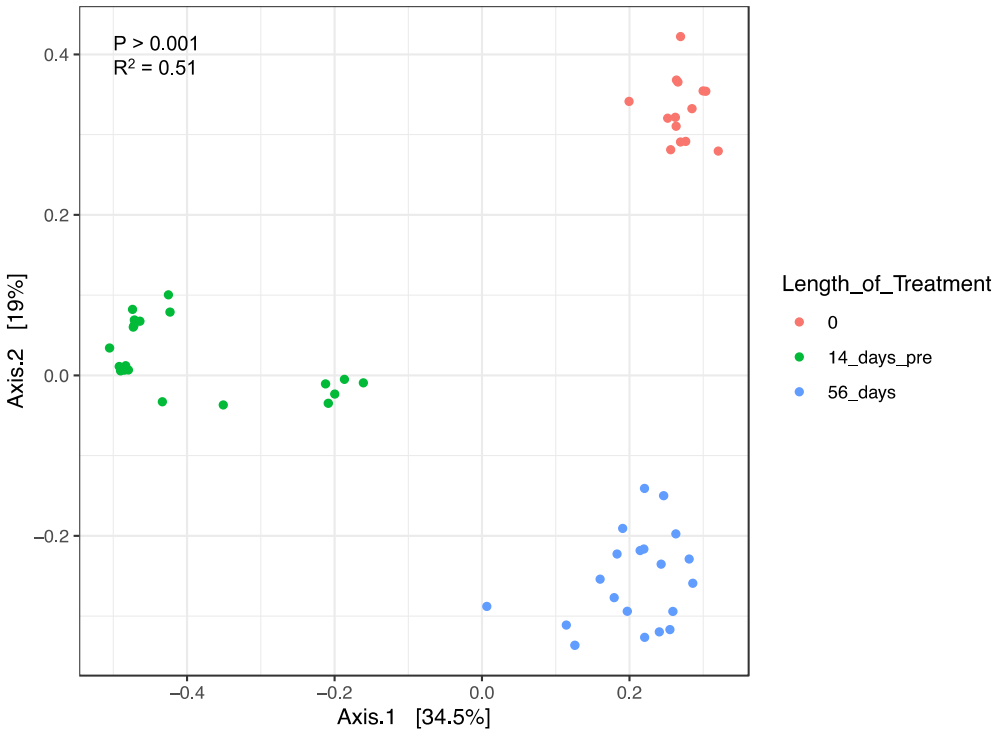


Figure 3.14: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Tran *et al.* (2019) using PCoA plots coloured based on length of treatment (56 days = blue, 14 days = green, 0 days = red) shows differences in microbial composition at the various timepoints.

Based on the established taxonomic pipeline, in this investigation, the condition of the mouse did not play a role in gut microbial composition ($p = 0.6$, Wilcoxon's test) rather the length of treatment (time post flagellin injections) is what shaped the microbiome ($p < 0.001$, Wilcoxon's test). The samples from different timepoints during treatment distinctly cluster as illustrated in figure 3.14. DESeq2 analysis showed a clear-cut distinction between microbiotas at different treatment timepoints when compared to condition (figure 3.15). A pairwise comparison between samples from day 0 and day 56 showed that despite there being significant differences in the taxa present across

both groups, most of them belonged to Bacteroidales or Clostridia. Most of these taxa, like *Alistipes*, *Phocaeicola*, *Clostridium* XIVa, *Duncaniella* and *Limosillactobacillus* have been previously associated with positive colitis outcomes or healthy status (Parker *et al*. 2020; Sun *et al*., 2022; Van den Abbeele *et al*., 2012; Chang *et al*., 2021; Liu *et al*. 2022). The inefficacy of the flagellin immunisation in DSS treated mice could be due to the abundance of *Bacteroides acidifaciens* at the 56-day timepoint. Despite it being a SCFA producer, this species is a mucin degrader and might have amplified the effects of DSS on the gut barrier (Busbee *et al*., 2020).

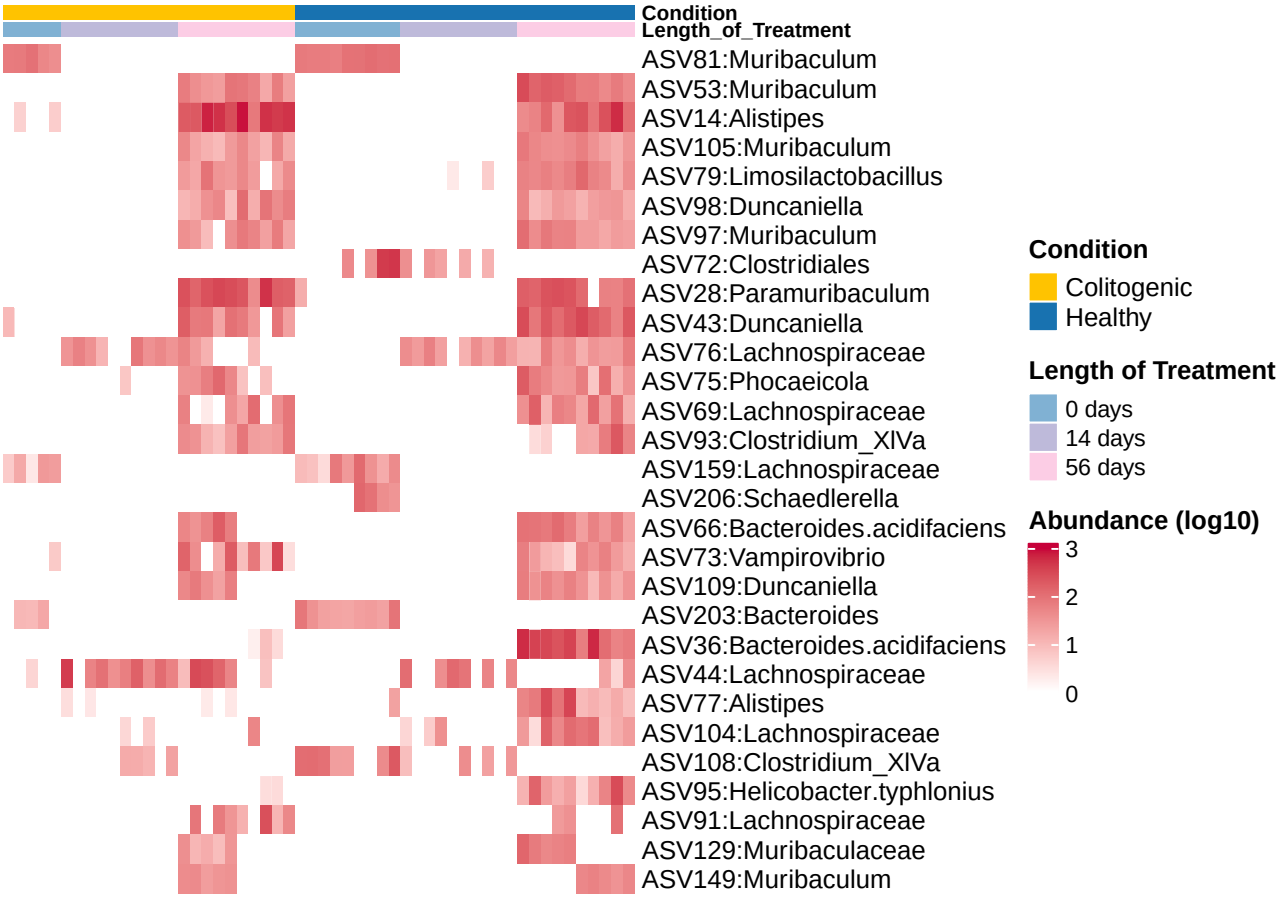


Figure 3.15: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa (at or above a ± 7 log2 fold change) at the lowest taxonomic classification from Tran *et al.* (2019) samples, grouped by condition (blue = healthy, yellow = colitogenic, orange = colitis) and length of treatment (sky blue = 0 days, lilac = 14 days, pink = 56 days). Plotted taxa were identified by a pairwise comparison of samples from timepoints 0 and 56 using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.3 Spalinger *et al*.

Spalinger *et al*. (2019) were investigating the role of protein tyrosine phosphatase non-receptor type 22 (PTPN22) in modulating IBD pathogenesis. By subjecting PTPN22 wildtype (WT), gene deletion (KO), and PTPN22 variant mice to 4 cycles of DSS (with a 10-day break in between each cycle) they found that the PTPN22 variant enhances the severity of colitis through the host's response to changes in intestinal microbial composition.

The structure of the microbiome varied significantly over the course of DSS treatment, while the condition of the mouse seemed to have minimal to no effect as identified by the analysis workflow (figure 3.16). Gut microbial composition at the end of the treatment (day 79) was drastically different than it was at day 0, day 16 and day 26 ($p = 0.0087$, $p = 0.0106$, $p = 0.0431$, respectively, Wilcoxon test). Pairwise comparison testing also showed significant changes in the microbiota at day 18 when compared to those from day 16 and 26 ($p = 0.007$ and $p = 0.0271$, respectively).

A pairwise comparison of the DESeq2 results between the first (day 0) and last (day 79) timepoints showed a 22 log2 fold increase in *Alloprevotella* at day 79, indicating an inactive stage of UC or potential remission. There is a negative correlation between the abundance of this genus with inflammatory markers in mice (Gu *et al*., 2022) and it has also been associated with UC patients in remission post FMT treatment (Fan *et al*., 2019). Microbiotas sequenced at day 0 had an abundance of taxa that are negatively and positively associated with UC. Members of the order Erysipelotrichales, including *Dubosiella*, *Faecalibaculum*, *Turicibacter* and *C. cocleatum*, were significantly enriched at the beginning of the treatment protocol (Li *et al*., 2022). As discussed in the meta-analysis, the abundance of these taxa is negatively associated with UC. Moreover, potentially, pro-inflammatory taxa highlighted in the meta-analysis were also present in high numbers at day 0 and other early stages of the DSS treatment course in this study, like Bacteroidales (*B. acidifaciens*, *Prevotella*) (Iljazovic *et al*., 2021). There were also beneficial genera, like *Alistipes*, *Odoribacter*, *Ligilactobacillus*,

Duncaniella, *Mucispirillum*, *Kineothrix* and *Clostridium* XIVa, present in abundance at the first time point (Hiippala *et al*., 2020; Tao *et al*., 2021). Many of these taxa and/or the higher order they belong to have also been flagged up in the meta-analysis.

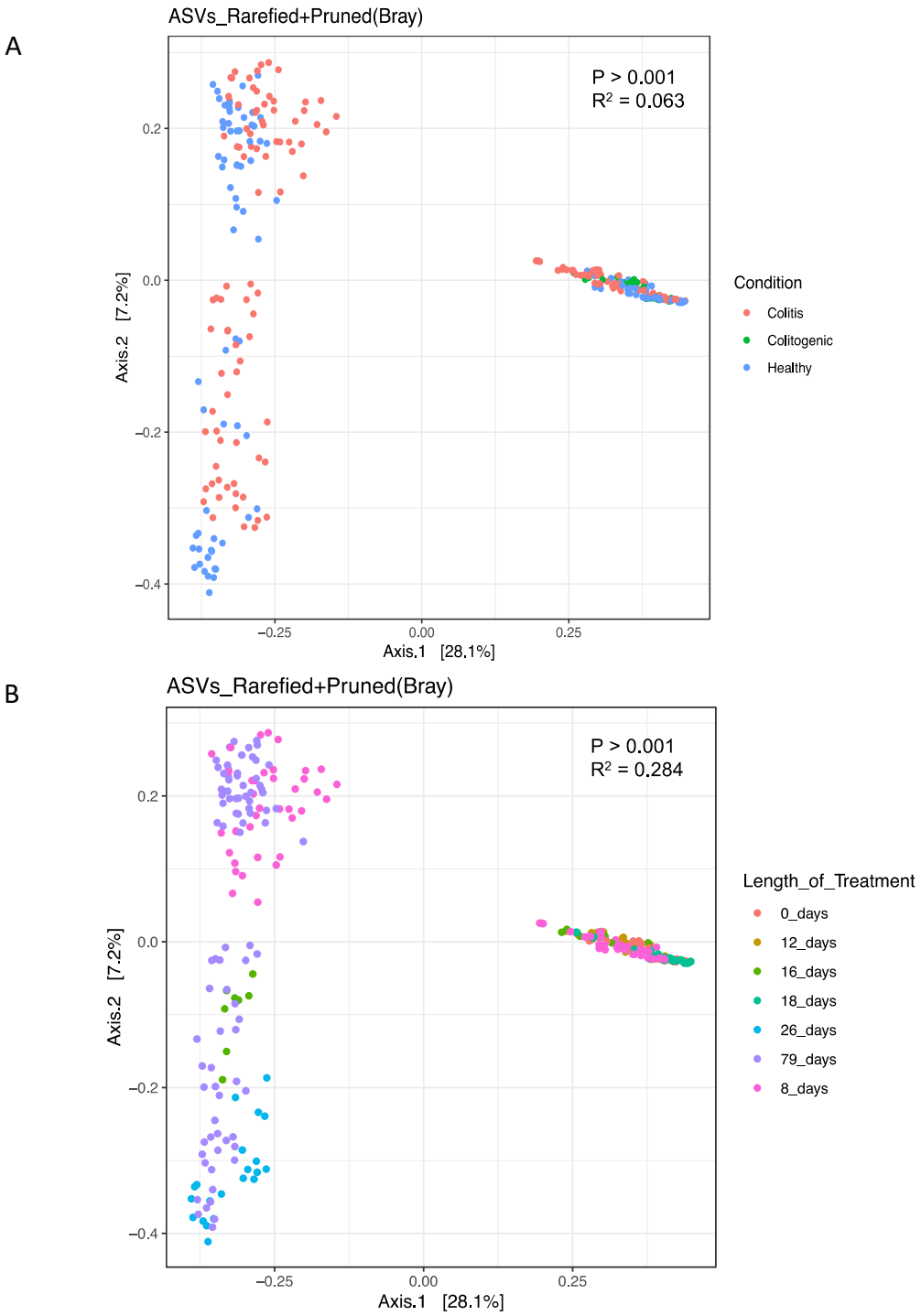


Figure 3.16: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Spalinger *et al.* (2019) using PCoA plots coloured based on (A) condition (healthy = blue, colitogenic = green, colitis = red) and (B) length of treatment shows differences in microbial composition at the various timepoints.

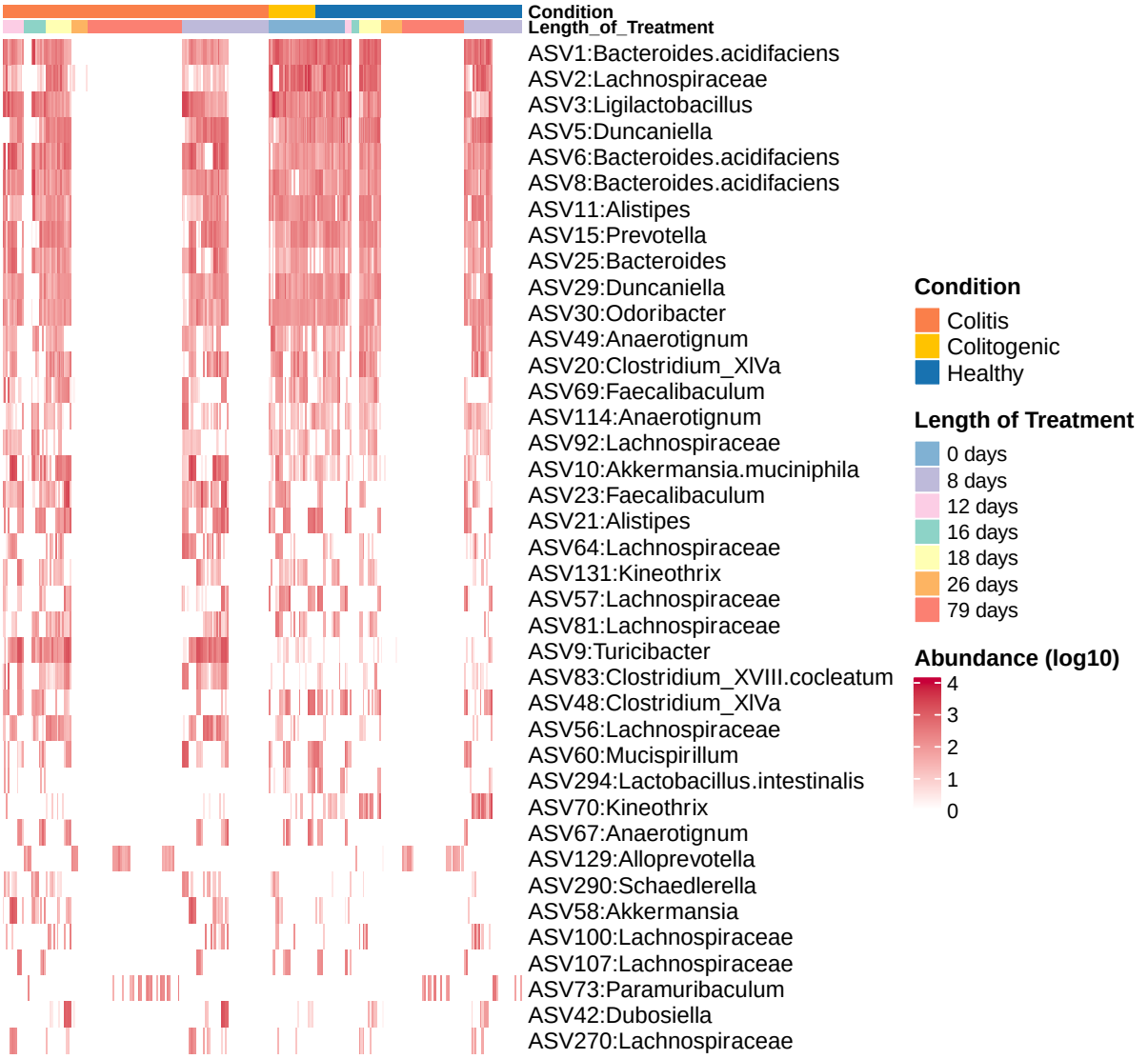


Figure 3.17: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa (at or above a ± 22 log2 fold change) at the lowest taxonomic classification from Spalinger *et al.* (2019) samples, grouped by condition (blue = healthy, yellow = colitogenic, orange = colitis) and length of treatment (sky blue = 0 days, lilac = 8 days, pink = 12 days, green = 16 days, yellow = 18 days, orange = 26 days, red = 79 days). Plotted taxa were identified by a pairwise comparison of samples from timepoints 0 and 79 using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

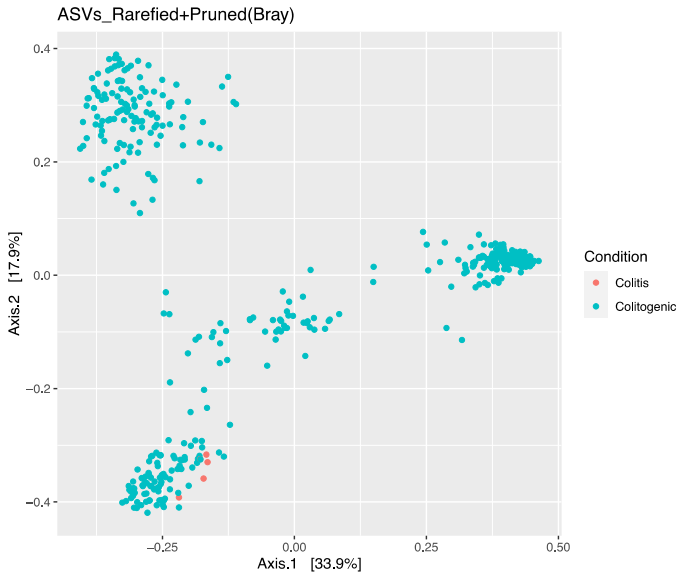
3.5.4 Roy *et al.*

The work by Roy and colleagues (Roy *et al.*, 2017) describes how innate and/or adaptive immunity driven by members of the gut microbiome is implicated in the development and severity of intestinal inflammation. To reach this conclusion they ran cohousing experiments using 6 SPF mouse lines as well as an inhouse Nlrp6$^{-/-}$ dysbiotic mouse line (DysN6) with 16S sequence microbiome analysis

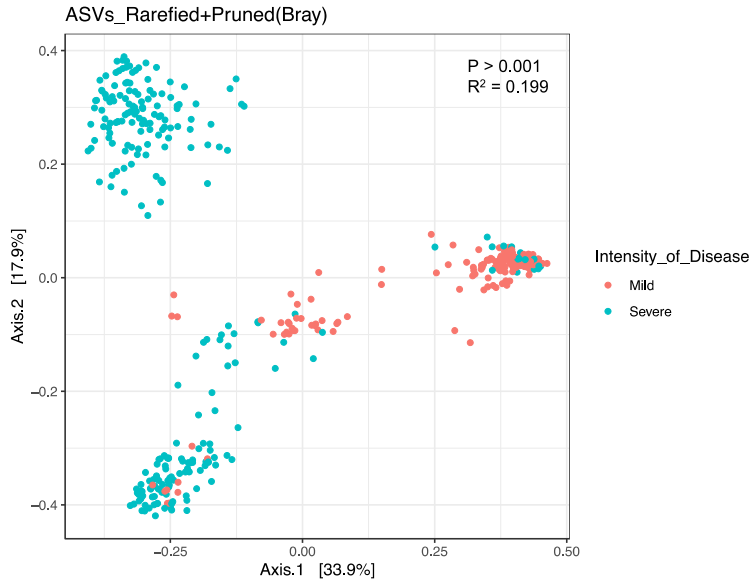
before and after cohousing. Mice were treated with DSS and exhibited different intensities of colitis symptoms. SPF-1, SPF-5 and SPF-6 mice were characterised by mild colitis while SPF-2, SPF-4, SPF-6 and DysN6 mice developed severe colitis.

Beta diversity analysis using Bray-Curtis dissimilarity metric (figure 3.18) of this large dataset showed that microbiotas causing similar severities of colitis significantly clustered together. DESeq2 analysis identified differentially abundant taxa among the groups of mice that experienced mild and severe colitis. *Adlercreutzia*, a genus known to be depleted in UC patients (Shaw *et al*., 2016; Bandenas *et al*., 2022), was greatly enriched (25 log2 fold change) in the mild colitis group. The high abundance of this taxa could be a reason to why SPF-1, SPF-5, and SPF-6 mice had mild colitis symptoms. *E. lenta* was increased in the mild colitis group. Different ASVs of the genus *Paramuribaculum* were abundant in both mild and severe colitis groups. Other members of Bacteroidales with positive (*Marseilla* aka *Prevotella marseillensis*), negative (*Alloprevotella*, *Duncaniella*, *Parabacteroides distasonis*) and mixed (Muribaculaceae) associations with UC were predominantly increased in mice experiencing severe colitis. The severe colitis group had relatively more increases in proinflammatory taxa, including *H. typhlonius* and *Mucispirillum*.

A


ASVs_Rarefied+Pruned(Bray)
Axis.2 [17.9%]
Axis.1 [33.9%]
Condition
Colitis
Colitogenic

B


ASVs_Rarefied+Pruned(Bray)
P > 0.001
R² = 0.199
Axis.2 [17.9%]
Axis.1 [33.9%]
Intensity_of_Disease
Mild
Severe

C

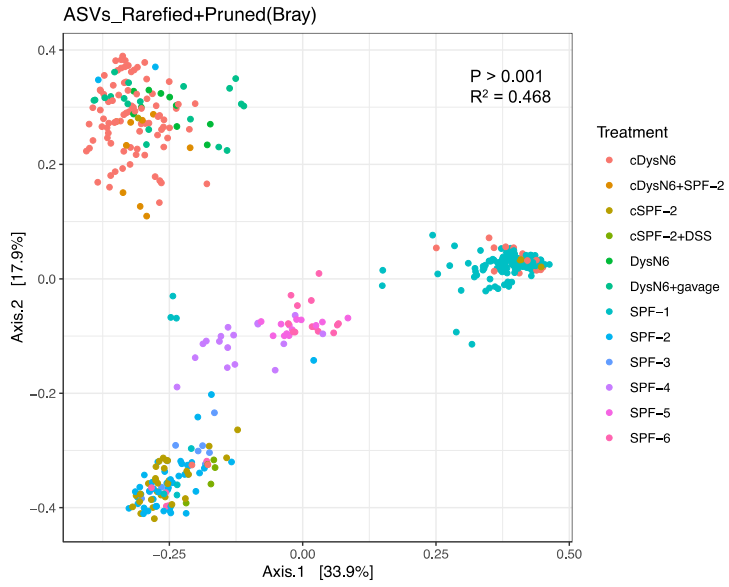

ASVs_Rarefied+Pruned(Bray)
P > 0.001
R² = 0.468
Axis.2 [17.9%]
Axis.1 [33.9%]
Treatment
cDysN6
cDysN6+SPF−2
cSPF−2
cSPF−2+DSS
DysN6
DysN6+gavage
SPF−1
SPF−2
SPF−3
SPF−4
SPF−5
SPF−6

Figure 3.18: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Roy *et al.* (2017) using PCoA plots coloured based on (A) condition (healthy = blue, colitogenic = green, colitis = red) and (B) intensity of disease and (C) treatment shows differences in microbial composition at the different disease intensities.

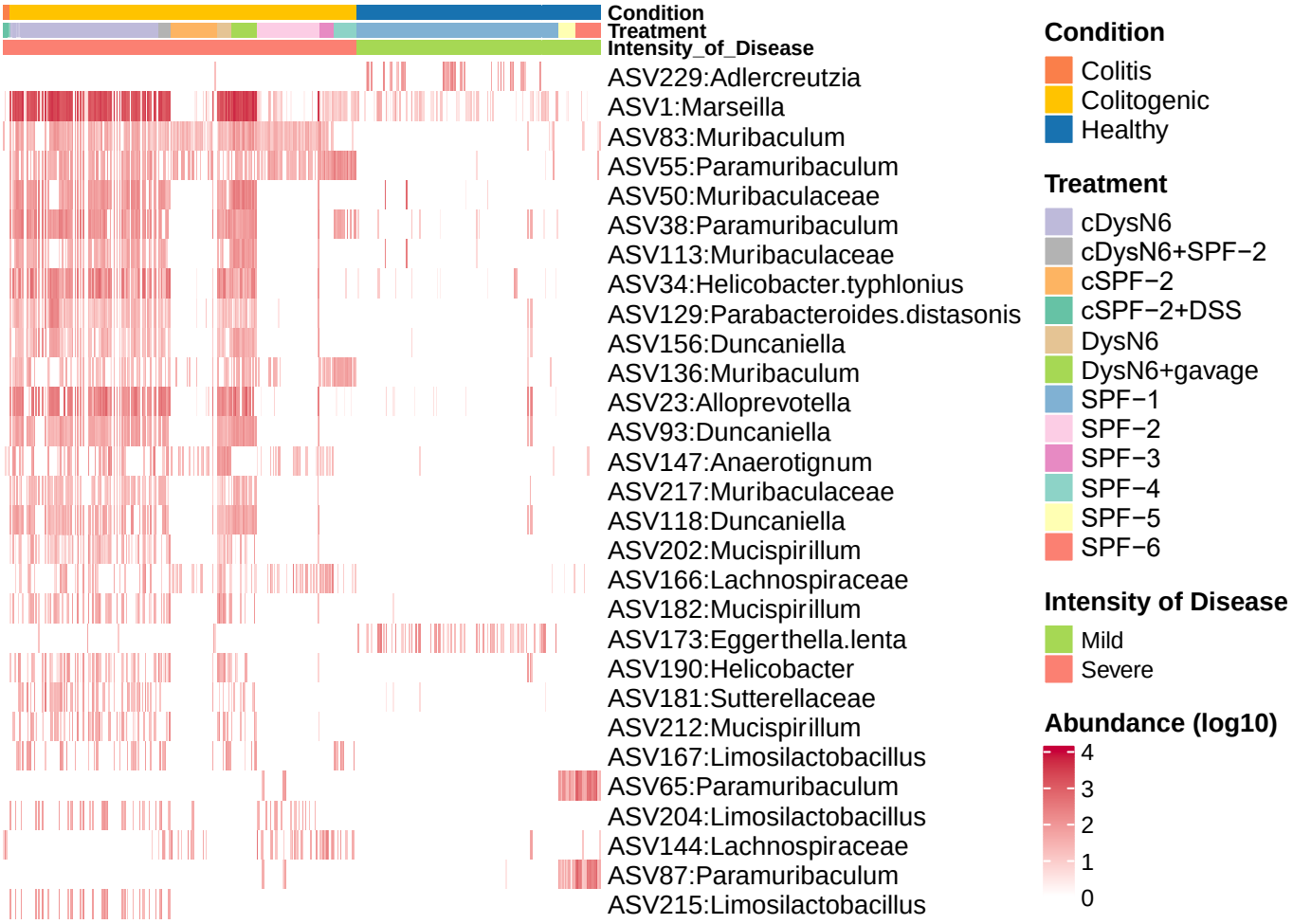


Figure 3.19: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa (at or above a ± 5 log2 fold change) at the lowest taxonomic classification from Roy *et al.* (2017) samples, grouped by condition (blue = healthy, yellow = colitogenic, orange = colitis), treatment and disease intensity (green = mild, red = severe). Plotted taxa were identified by a pairwise comparison of samples from both mild and severe colitis groupings using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.5 Song *et al.*

Authors of this study (Song et al., 2020) discovered that bile acid metabolites produced by members of the gut microbiota modulate regulatory T cell homeostasis. They also found that a diet rich in bile

acids ameliorates gut inflammation via commensal metabolic activities and analysed the microbiomes of mice put on either a nutrient-rich diet or a minimal diet, pre-DSS treatment.

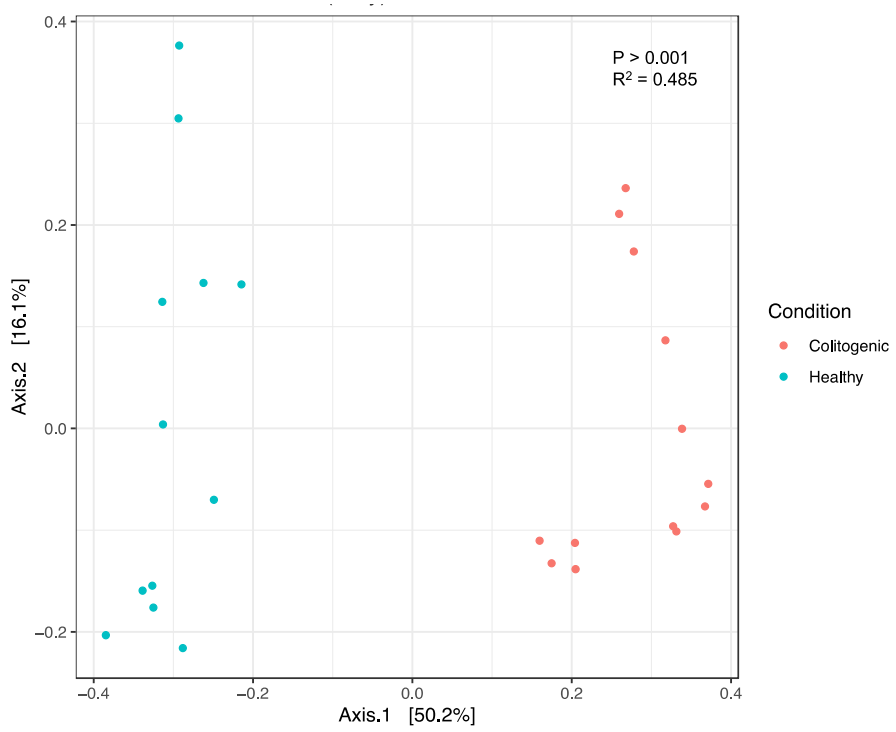


Figure 3.20: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Song *et al.* (2020) using PCoA plots coloured based on condition (healthy = blue, colitogenic = red) shows differences in microbial populations across condition groups.

Analysis of the Song *et al*. (2020) dataset showed that microbiotas of mice on the nutrient-rich diet were significantly different from those on the minimal diet (p = 2.47E-05). This can be visualised in the PCoA from figure 3.20. Further analysis corroborated the findings of the authors showing the abundance of taxa associated with gut inflammation, like *Parasutterella* and the genus *Romboutsia* of the Peptostreptococcaceae family (detected in the meta-analysis) in mice put on the minimal diet (figure 3.21). *Anaerotignum*, a bacteria associated with fibre degradation and bile acid metabolism was enriched in the colitogenic group (on the minimal diet) (Lesniak *et al*., 2022). Clostridiales species (*Kineothrix*, *Dysosmobacter*) generally associated with healthy microbiotas were of high abundance in microbiotas of mice fed the enriched diet.

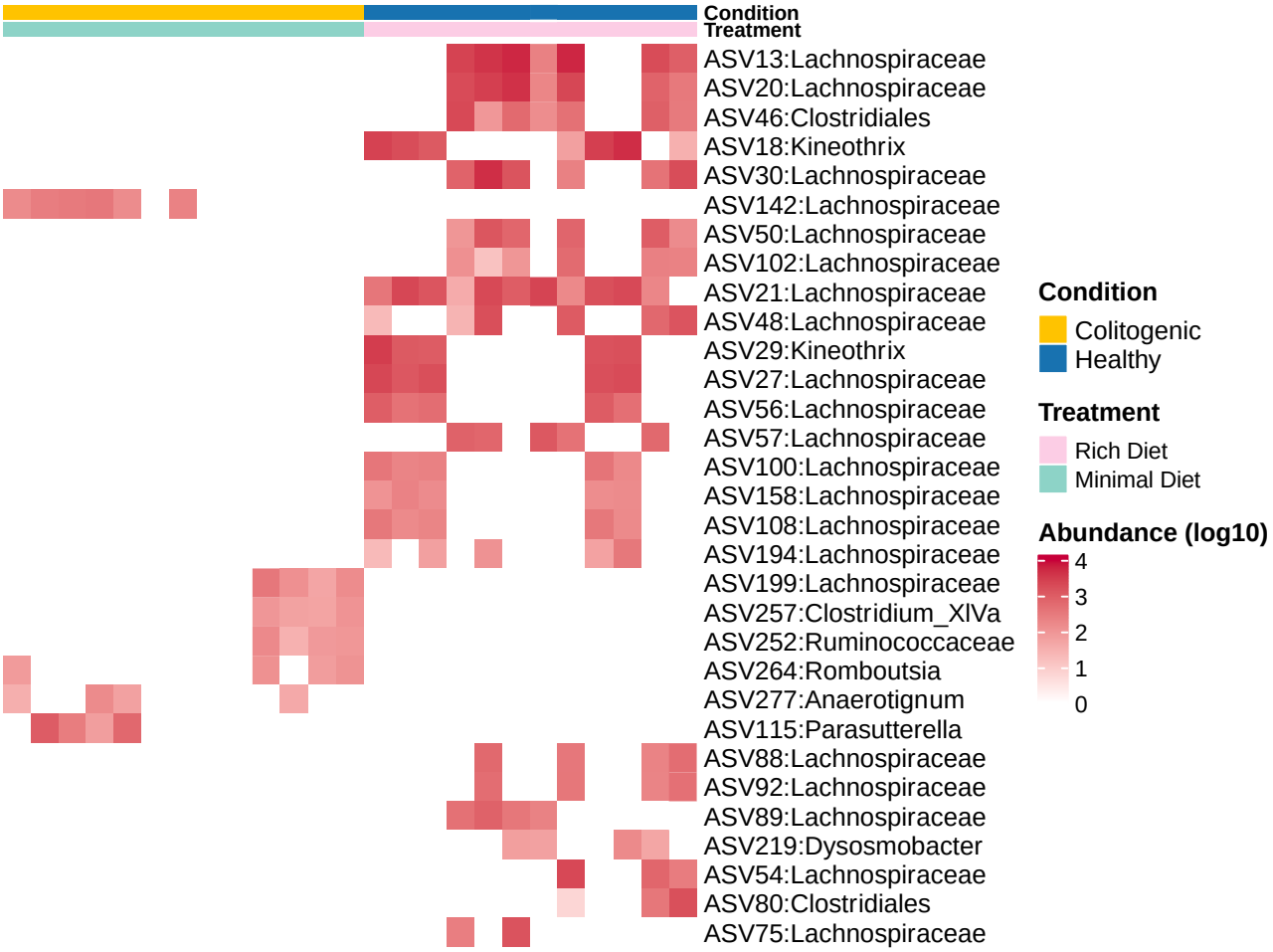


Figure 3.21: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa (at or above a ± 12 log2 fold change) at the lowest taxonomic classification from Song *et al.* (2020) samples, grouped by condition (blue = healthy, yellow = colitogenic) and treatment (pink = rich diet, green = minimal diet). Plotted taxa were identified by a pairwise comparison of samples from both healthy and colitogenic condition groups using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.6 Nakajima *et al.*

Nakajima *et al.* (2018) found that IgA promotes the symbiotic relationship among bacteria, contributing to gut homeostasis through modulation of Mucus-Associated Functional Factors (MAFFs) in gut commensal. They created different lines of SPF mice with the commensals *B. thetaiotaomicron* and *B. vulgatus* carrying different variants of MAFF with, then sequenced their gut microbiomes prior DSS treatment.

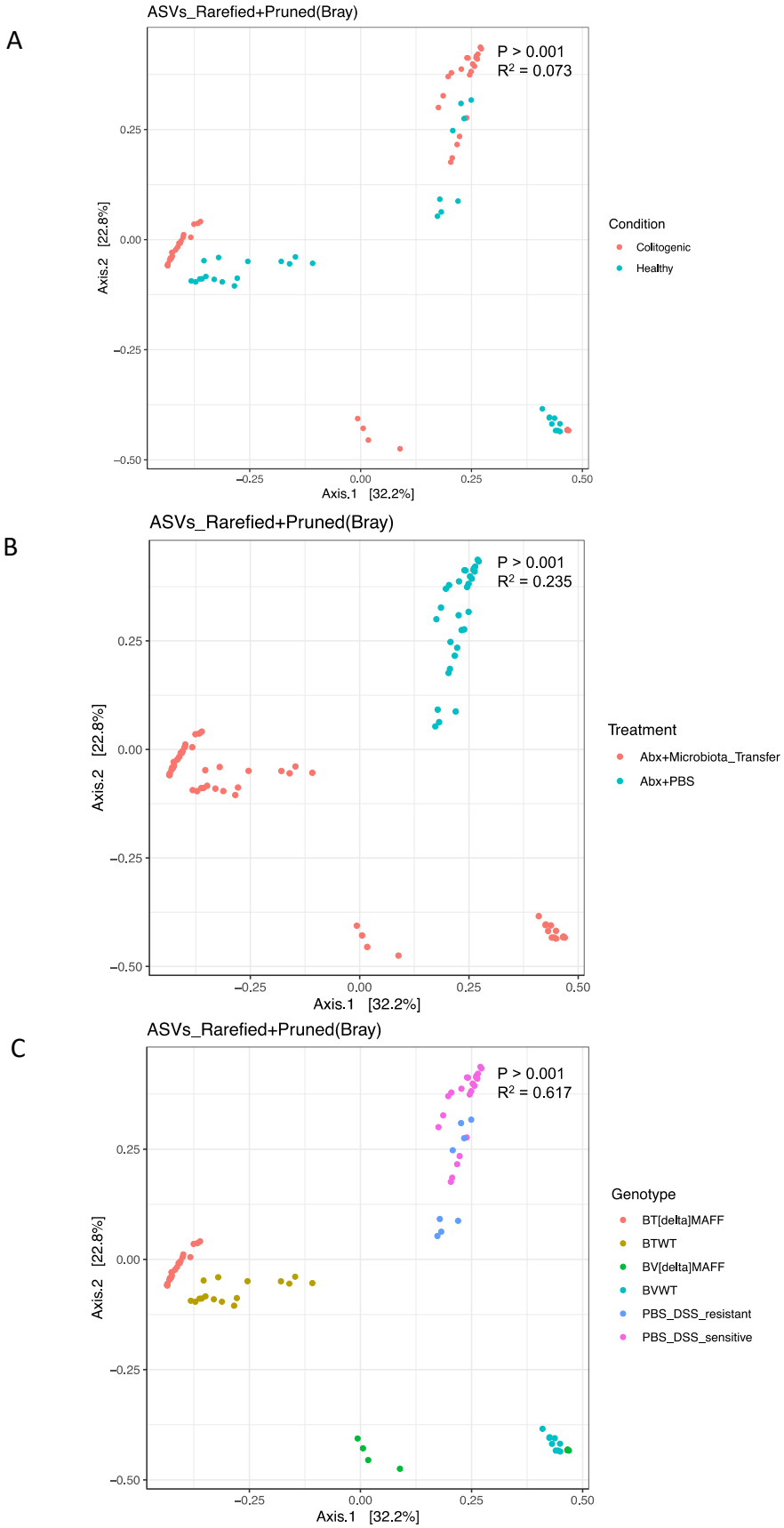


Figure 3.22: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Nakajima *et al.* (2018) using PCoA plots coloured based on (A) condition (healthy = blue, colitogenic =

red), (B) treatment and (C) genotype shows that microbial profiles varied across condition, treatment and genotypic groups.

Wilcoxon's test shows that there is a significant association between gut microbial composition and both the condition of the mice and the treatment they received (whether it was a microbial transfer or PBS). The latter is clearly demonstrated in PCoA plot (B) from figure 3.22. Proinflammatory genera like, *Enterocloster*, *Robinsoniella*, and *Escherishia*/*Shigella* spp. were highly abundant in microbiotas of colitogenic mice, two of which were also flagged in non-healthy mice from the meta-analysis (figure 3.23 & 3.6). *Eisenbergiella*, *Hungatella* and *Dysosmobacter* were enriched in healthy microbiotas in this study and group A meta-analysis. These genera have been previously negatively correlated with UC in mice and/or humans (Gryaznova *et al*., 2021; Schaubeck *et al*., 2016; Le Roy *et al*., 2022).

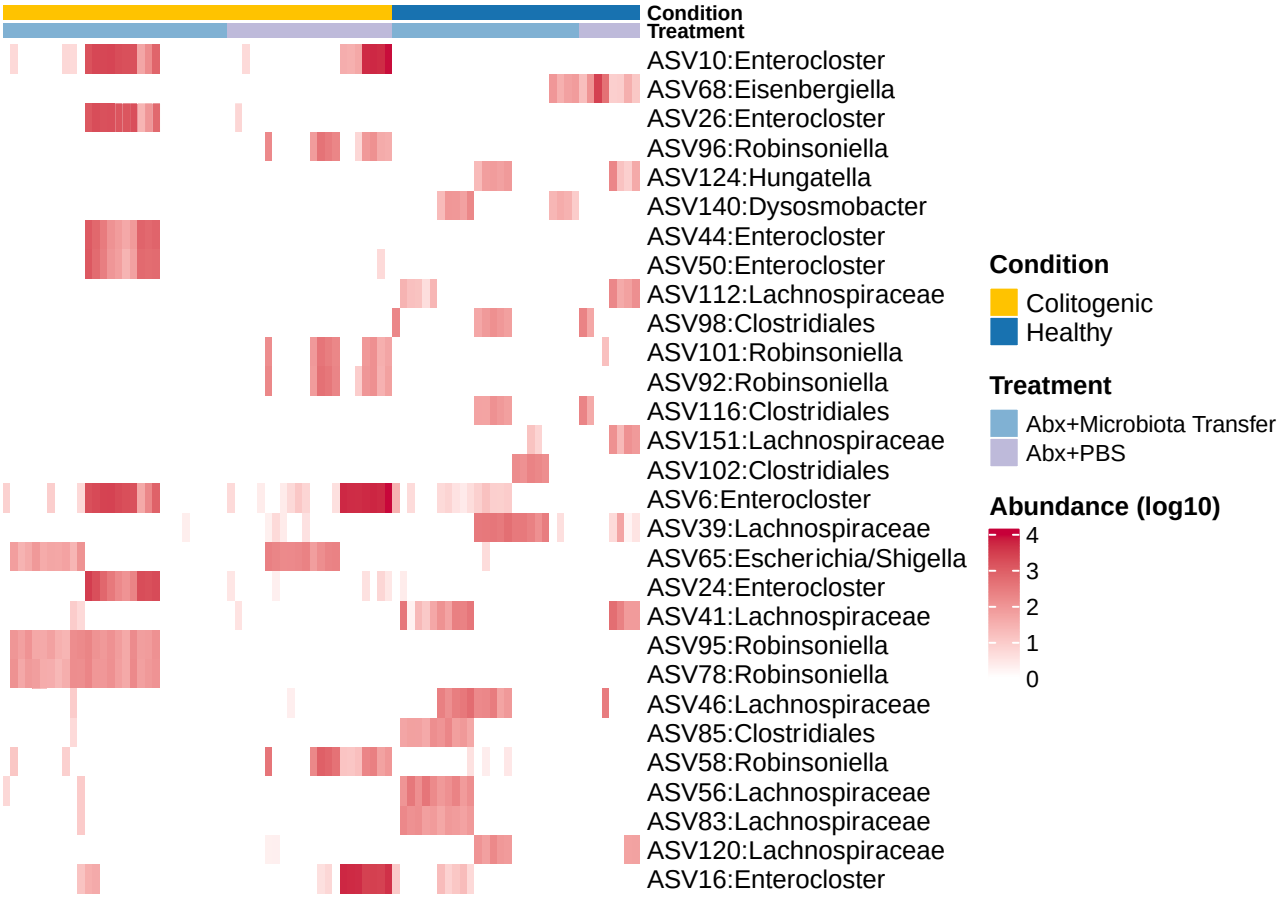


Figure 3.23: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa (at or above a ± 8 log2 fold change) at the lowest taxonomic classification from Nakajima *et al.* (2018)

samples, grouped by condition (blue = healthy, yellow = colitogenic) and treatment (sky blue = antibiotics and microbiota transfer, lilac = antibiotics and PBS). Plotted taxa were identified by a pairwise comparison of samples from both healthy and colitogenic condition groups using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.7 Kazakevych *et al.*

While investigating the role of chromatin remodelling factor *Smarcad1* in murine colitis, the authors (Kazakevych et al., 2020) discovered the critical role played by the microbiome in DSS-induced colitis. Both WT and *Smarcad1*-KO mice developed DSS-induced colitis after being cohoused with donor mice known to be sensitive to DSS. The intestinal microbiota was characterised before and after cohousing, prior to DSS administration.

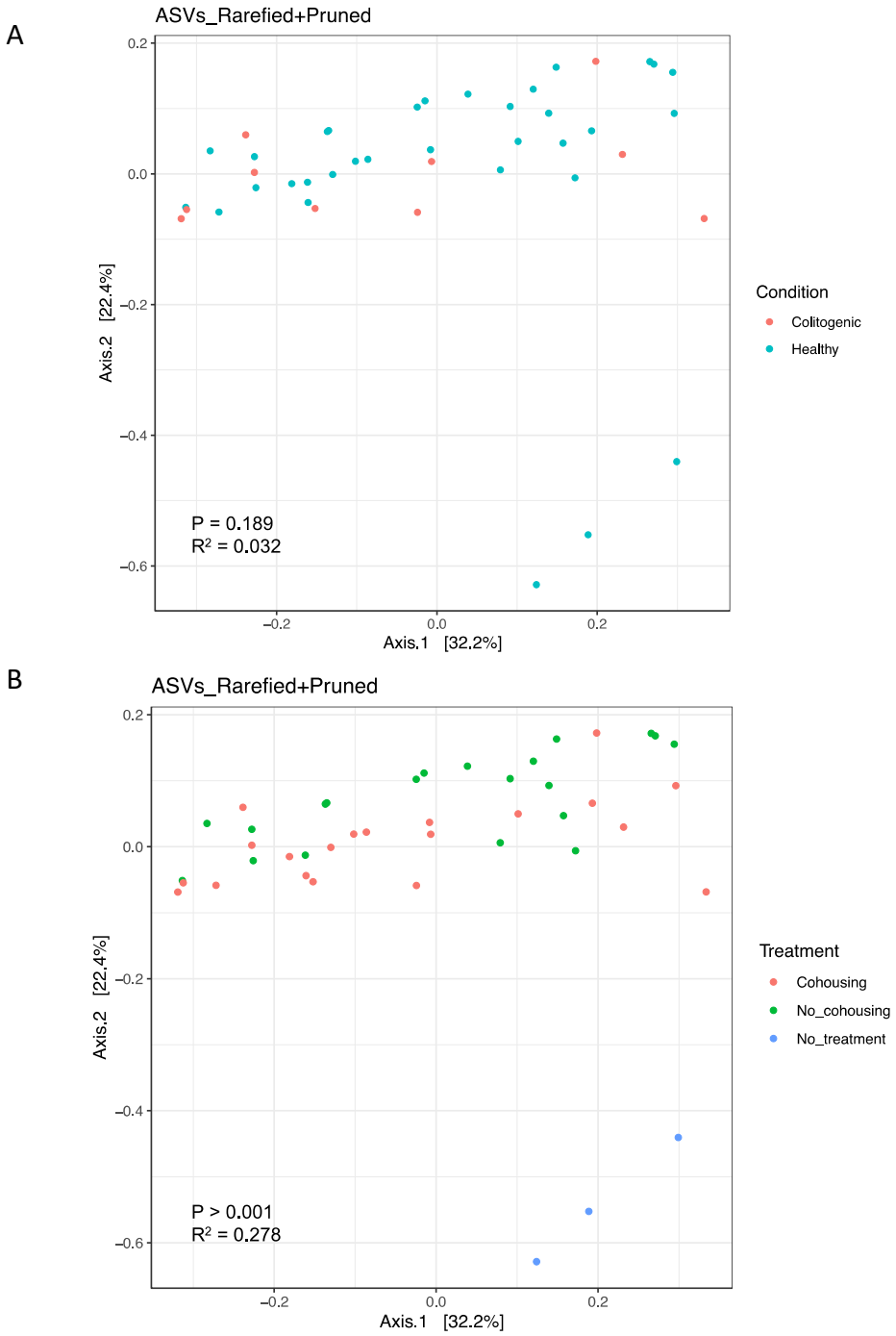


Figure 3.24: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Kazakevych *et al.* (2020) using PCoA plots coloured based on (A) condition (healthy = blue, colitogenic = red) and (B) treatment shows that microbial profiles did not vary between condition groups and only slightly among treatment groups.

Pairwise comparison testing on this dataset showed that the changes incurred on the host's microbiome by the treatment (cohousing) were more substantial ($p = 0.038$) than those by the condition of the host

(p = 0.61). However, microbiotas among treatments did not vary enough to form separate clusters as shown in figure 3.24B. Further testing highlighted genera like *Alistipes* and *Paramuribaculum* as well as different ASVs assigned to the Lachnospiraceae family that were found more abundant in healthy or DSS-resistant mice (figure 3.25). While the previously mentioned taxa generally confer health, an IBD positively correlated genus, *Faecalibaculum* (Erysipelotrichales spp.), was also found in high numbers in DSS-resistant mice. Muribaculaceae species, *Paramuribaculum* and *Muribaculum*, were enriched in DSS-sensitive mice. Microbiotas of colitogenic mice also had a differential abundance of genera that are commonly found in healthy microbiomes like, *Odoribacter*, *Agathobacter* and *Monoglobus*.

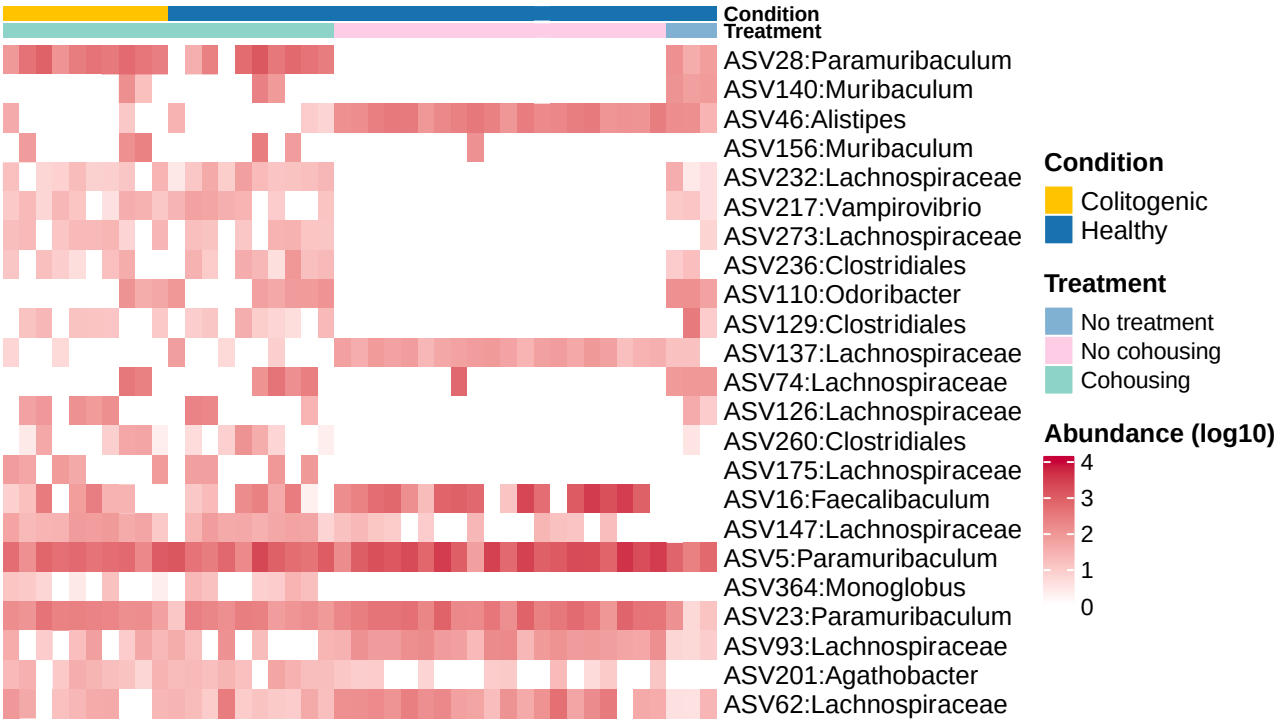


Figure 3.25: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa at the lowest taxonomic classification from Kazakevych *et al.* (2020) samples, grouped by condition (blue = healthy, yellow = colitogenic) and treatment (sky blue = no treatment, pink = no cohousing, green = cohousing). Plotted taxa were identified by a pairwise comparison of samples from both cohousing and non-cohousing treatment groups using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.8 Qi *et al.*

The authors (Qi *et al.*, 2020) examined the role of Enterobacteriaceae, like *E. coli*, in gut inflammation and found that deficiencies in Reg4 and complement factor D (CFD) led to intestinal inflammation due to overgrowth of *E. coli*. They characterised the gut microbial composition of WT mice and mice with the Reg4 gene deleted as well as CFD deficient mice before and after DSS exposure.

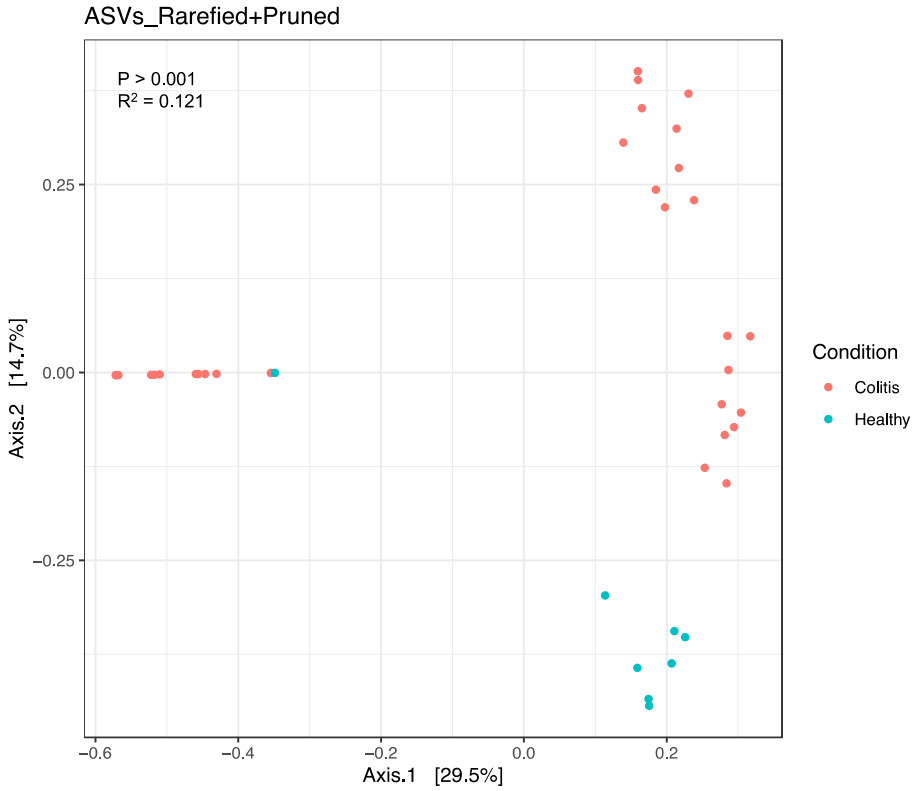


Figure 3.26: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Qi *et al.* (2020) using PCoA plots coloured based on condition (healthy = blue, colitis = red) shows differences in microbial populations across condition groups.

In this study, the condition of the mouse highly affected its gut microbial composition (p = 8.8E-0.5, Wilcoxon) with gut communities from each condition group forming clear separate clusters (figure 3.26). Bacterial taxa commonly associated with inflammation like Peptostreptococcaceae, Erysipelotrichales (*Faecalibaculum*, *Turicibacter*, *C. cocleatum*) and the opportunistic Bacteroidales species, *B. thetaiotaomicron* were enriched in the colitis group. Lachnospiraceae members with conflicting roles in UC were present in high numbers in colitic microbiotas. For instance, *Blautia* and

Mediterraneibacter are known SCFA producers, however, they also produce aldehydes usually harmful to hosts as they trigger oxidation stress (Abdeulgheni *et al*., 2022). Other taxa common in non-healthy groups in the meta-analysis were also differentially abundant in the colitis group of this study, some of which are *Mailhella*, *Escherichia*/*Shigella* spp., *Bifidobacterium* spp., *Helicobacter* spp., Desulfovibrionaceae spp., and *Parasutterella* spp. Species negatively correlated with IBD were high in numbers in healthy microbiotas.

Figure 3.27: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa (at or above a ± 12 log2 fold change) at the lowest taxonomic classification from Qi *et al.* (2020) samples, grouped by condition (blue = healthy, orange = colitis). Plotted taxa were identified by a pairwise comparison of samples from both healthy and colitis condition groups using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.9 Mamantopoulos *et al.*

Mamantopoulos *et al.* (2017) dismissed the suggested role for inflammasomes in intestinal inflammation through regulation of the gut microbiome by controlling for non-genetic confounders. They did so by characterising the microbial ecology of littermate-controlled Nlrp-6 deficient mice and previously germfree (GF) ASC-deficient mice with naturally formed gut microbiotas post birth, in two animal different animal facilities, before DSS treatment (i.e., all samples were considered as "colitogenic").

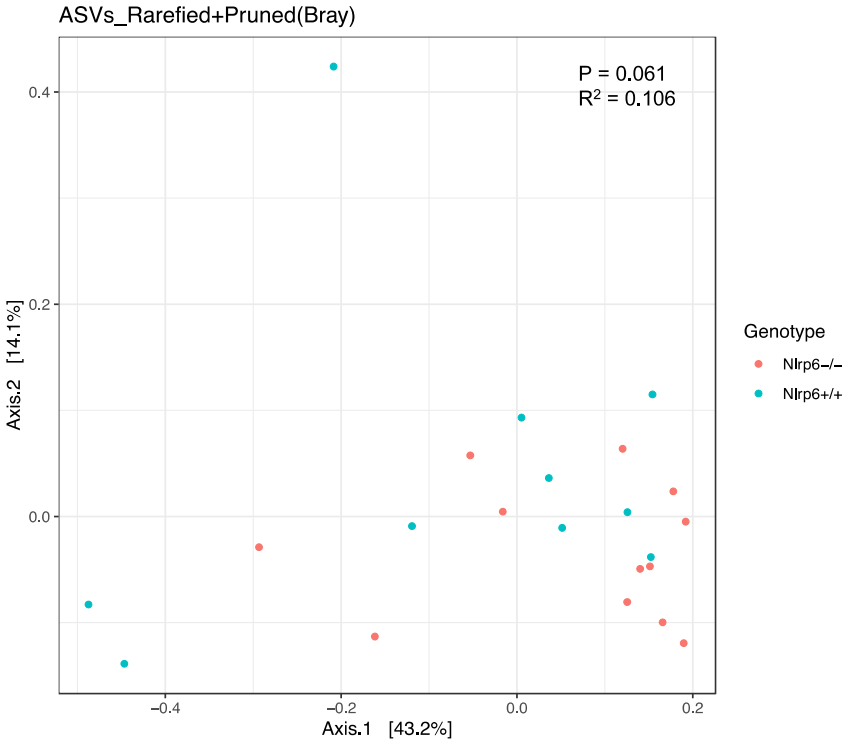


Figure 3.28: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Mamantopoulos *et al.* (2017) using PCoA plots coloured based on genotype (*Nlrp6* WT = blue, *Nlrp6* KO = red) showed the absence of distinct microbial profiles in both genotypes.

Analysis of this dataset showed that the gut microbial ecologies of Nlrp6 variants differed (0.018, Wilcoxon test) but were not distinctly different to form separate clusters as shown in figure 3.28. PERMANOVA testing showed that samples between both groups of genetically variant mice do not differ significantly. DESeq2 highlighted ASVs that were differentially abundant between both genotypes, most of which belonged to Bacteroidales (Muribaculaceae, *Muribaculum* and *Duncaniella*)

and Lachnospiraceae (figure 3.29). Along with Erysipelotrichaceae, these two bacterial taxonomic groups have been associated with mice of non-healthy conditions in the meta-analyses from section 3. As opposed to meta-analysis findings, Eggerthellaceae (genus *Slackia*) were highly abundant in colitogenic microbiomes of this study, while Deltaproteobacteria was enriched in both group B and this study.

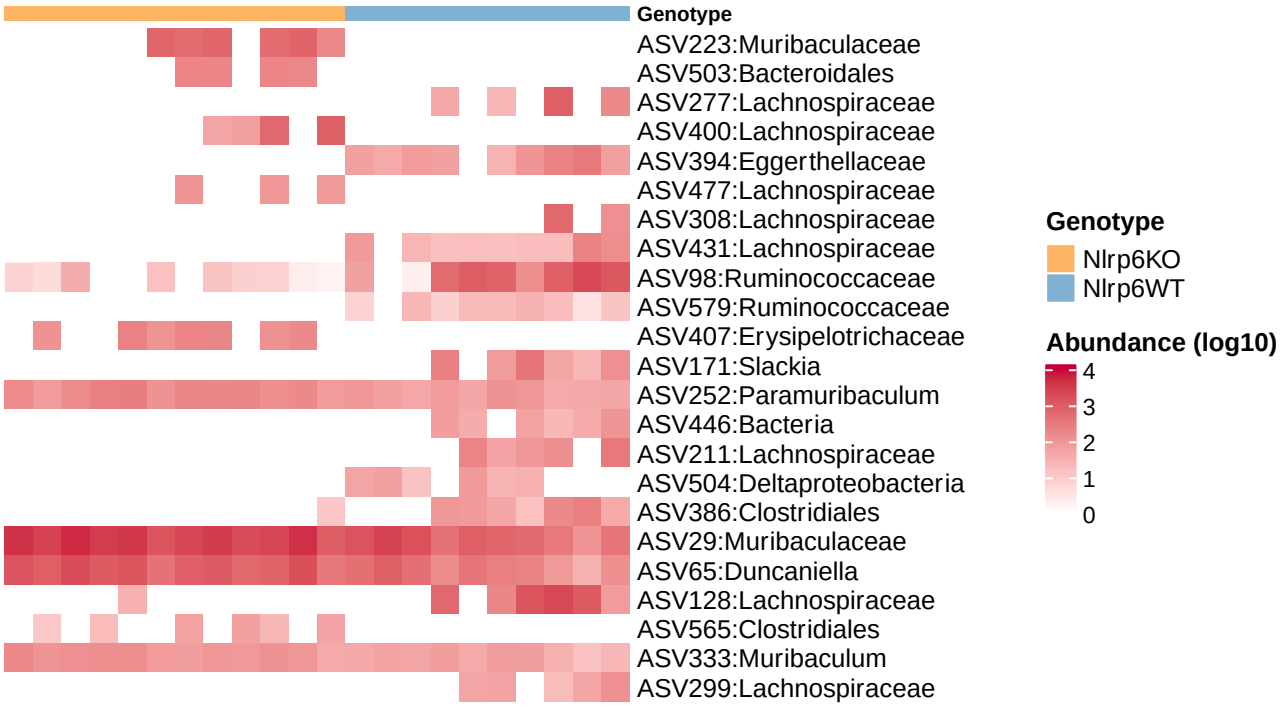


Figure 3.29: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa at the lowest taxonomic classification from Mamantopoulos *et al.* (2017) samples, grouped by condition (blue = *Nlrp6* WT, orange = *Nlrp6* KO). Plotted taxa were identified by a pairwise comparison of samples from both WT and KO genotype groups using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.10 Schmidt *et al.*

Authors from this study (Schmidt et al., 2019) inoculated GF mice with specific pathogen free (SPF) microbiota, before DSS exposure, to measure differences in barrier function, particularly in macrophage compartments. They found that the presence of a gut microbial community is a requirement for colonic macrophage infiltration post DSS treatment.

Both pairwise comparison testing, and beta diversity analysis (figure 3.30) of this dataset show no differences in microbiotas between different condition groups and treatment groups. DESeq2 identified Lachnospiraceae as the only differentially abundant taxa, which was massively enriched (25 log2 fold change) in the healthy group (figure 3.31).

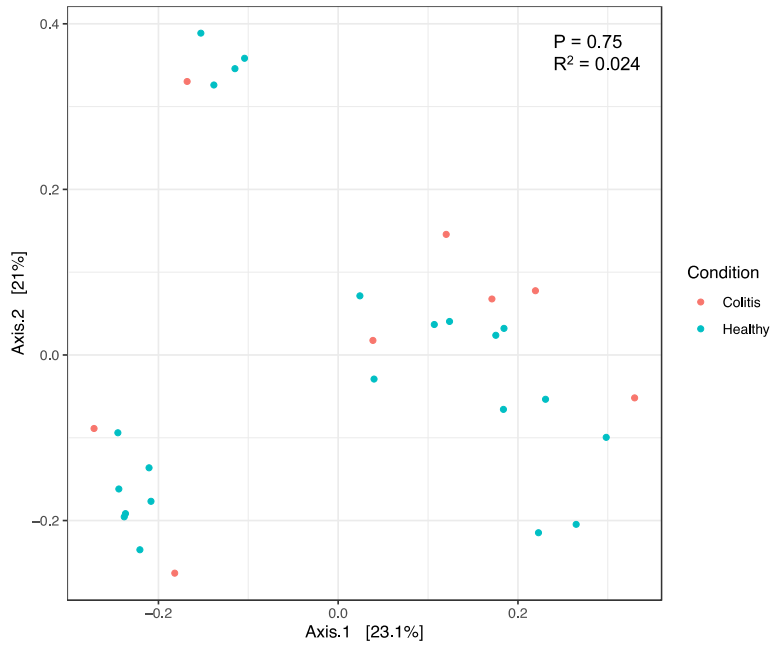


Figure 3.30: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Schmidt *et al.* (2019) using PCoA plots coloured based on condition (Healthy = blue, Colitis = red) showed the absence of distinct microbial profiles in both conditions.

Figure 3.31: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa at the lowest taxonomic classification from Schmidt *et al.* (2019) samples, grouped by condition (blue = Healthy, orange = colitis). Plotted taxa were identified by a pairwise comparison of samples from both healthy and colitis condition groups using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.11 Kimura *et al.*

Kimura *et al.* (2020) investigated the role of the protein osteoprotegrin (OPG) in gut immunity and IBD. They generated *Opg*$^{-/-}$ mice which, along with a group of WT mice, were exposed to DSS. The authors concluded that deletion of *Opg* alleviated DSS-induced colitis. Gut ecology of both variants was characterised prior to DSS treatment.

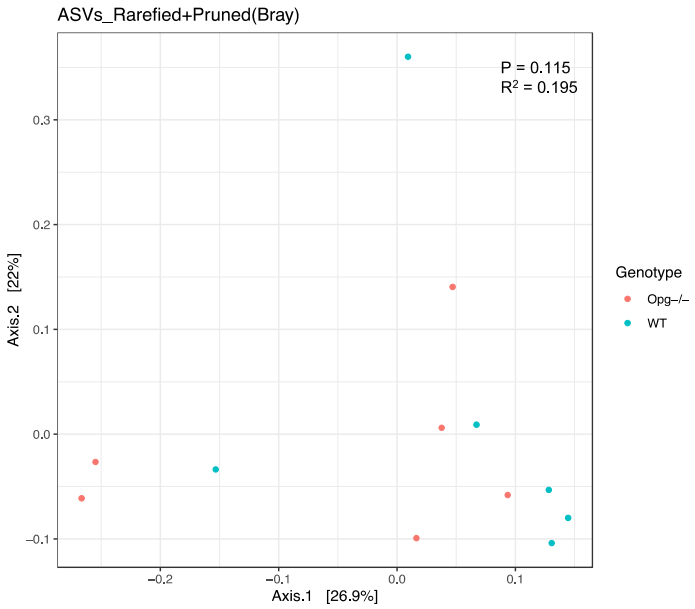


Figure 3.32: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Kimura *et al.* (2020) using PCoA plots coloured based on genotype (*Opg* WT = blue, *Opg* KO = red) showed no significant differences in microbial profiles across both genotypes.

Analysis of the data provided by Kimura *et al.* (2020) showed that gut microbiota, likely, does not play a role in the reduced symptoms of DSS-induced colitis in *Opg*$^{-/-}$ mice. Wilcoxon's rank sum test showed no statistical difference between microbiotas of WT and *Opg*$^{-/-}$ mice, collectively. Beta diversity analysis using Bray-Curits dissimilarity as a distance metric also showed no significant difference in the microbiomes across samples of both mouse groups. However, a few taxa stood out in both groups with the differential abundance analysis. The bile acid producer, *Phocea* (reclassified as *Merdimmobilis*) was abundant in *Opg*$^{-/-}$ mice known for its negative correlation with IBD in

humans could be behind the amelioration of colitis symptoms (Thomas *et al*., 2022). WT group microbiotas were enriched with the beneficial commensal *Hungatella* (Rawat *et al*., 2022), the ambivalent *Parabacteroides distasonis* and the phylum Saccharibacteria which was positively associated with colitis in group B and Kazakevych *et al*. (2020) study.

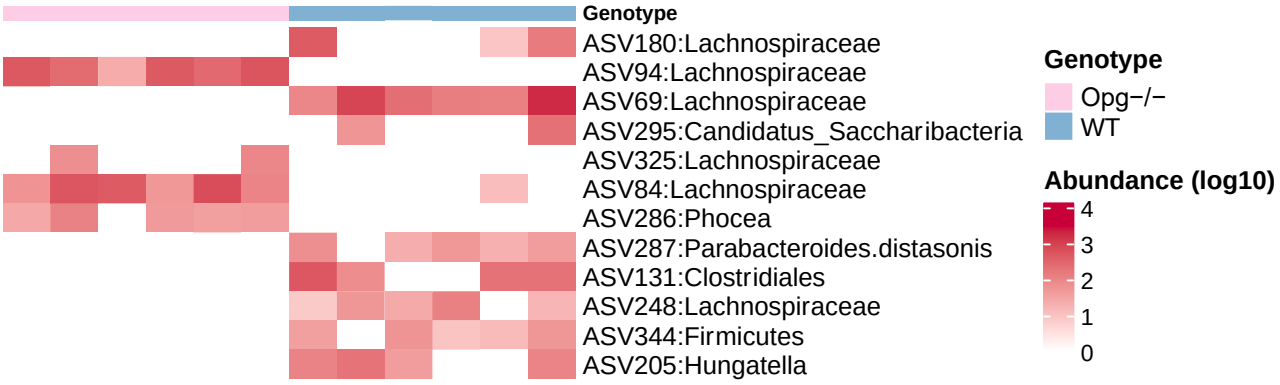


Figure 3.33: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa at the lowest taxonomic classification from Kimura *et al*. (2020) samples, grouped by genotype (blue = *Opg* WT, pink = *Opg* KO). Plotted taxa were identified by a pairwise comparison of samples from both *Opg* WT and *Opg* KO genotype groups using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.5.12 Khan *et al*.

The authors (Khan *et al*., 2020) studied the effect of simple sugars, including glucose, sucrose, and fructose, on the pathogenesis of colitis using the DSS model in WT and *Il10*$^{-/-}$ mice. Gut microbiota was characterised in WT mice before and after dietary changes. They noticed a worse colitis response to DSS in glucose-fed mice and an abundance of mucus-degrading bacteria, like *A. muciniphila* and *B. fragilis*.

Wilcoxon's rank sum test as well as beta diversity analysis done on this dataset (figure 3.34) showed no difference between microbiotas of "healthy" and colitogenic mice. However, the gut ecology of mice on the control diet varied considerably between that of mice on the fructose and sucrose diet ($p = 0.002$).

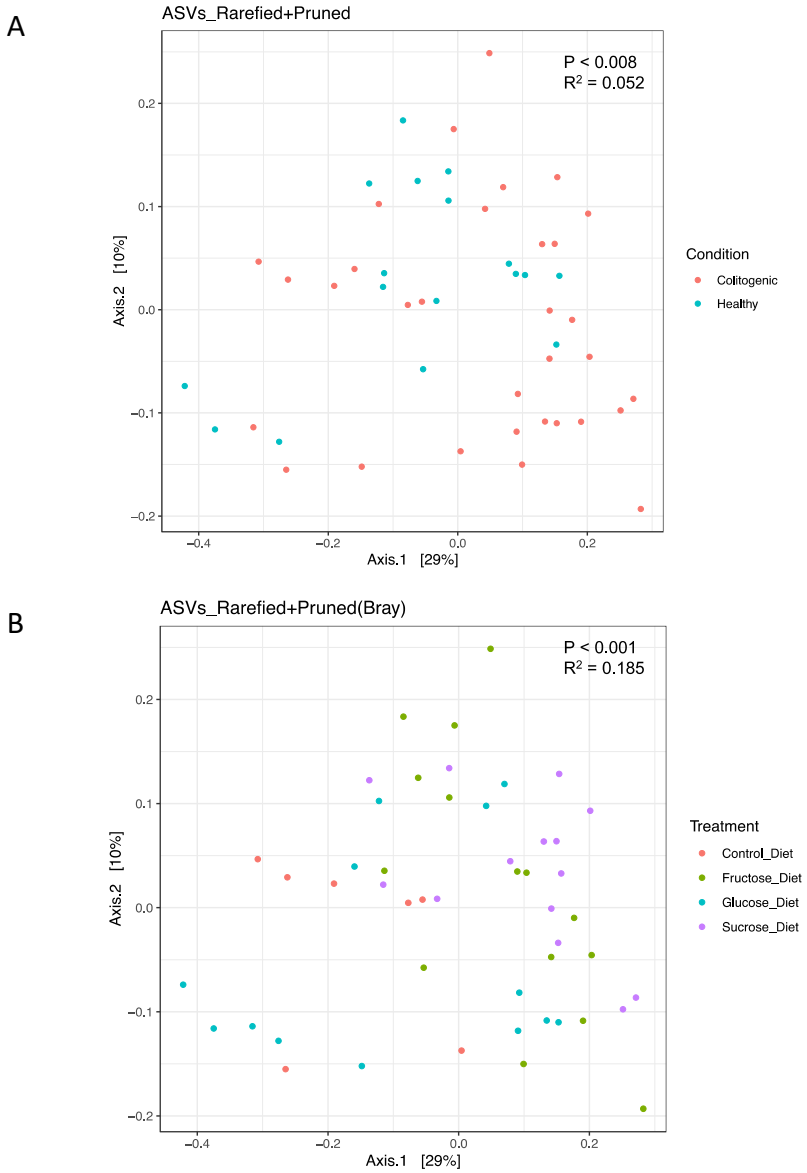


Figure 3.34: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Khan *et al.* (2020) using PCoA plots coloured based on (A) condition (healthy = blue, colitogenic = red) and (B) treatment shows that microbial profiles did not vary significantly between condition groups and only slightly among treatment groups (sucrose and control).

Due to the observed negative phenotypic effects of the glucose diet on colitis symptoms, a pairwise comparison was done between microbial profiles of mice on the control diet and the glucose diet. This analysis highlighted the increase in abundance of 16.2 log2 fold in colitis-associated *Helicobacter* species in mice on the glucose diet (Dingemanse *et al.*, 2015). Bacteria of the less understood Muribaculaceae, like *Paramuribaculum*, were present in high numbers with increased dietary glucose

intake. The same bacterial species have been observed in relatively high abundance in non-healthy groups of the meta-analysis.

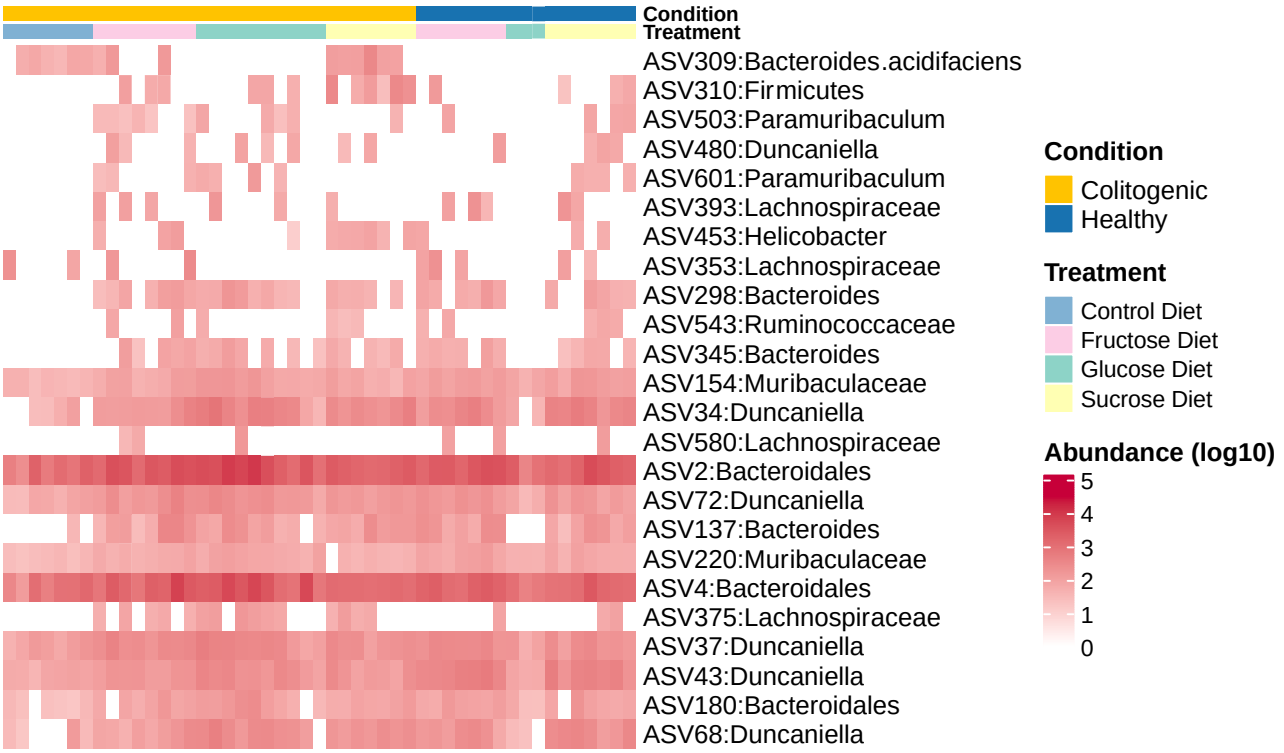


Figure 3.35: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa at the lowest taxonomic classification from Khan *et al.* (2020) samples, grouped by condition (blue = healthy, yellow = colitogenic) and treatment (sky blue = control diet, pink = fructose diet, green = glucose diet, yellow = sucrose diet). Plotted taxa were identified by a pairwise comparison of samples from both control and glucose diet groups using DESeq2 due to investigate the microbes behind the observed phenotype of worsened colitis (only FDR-adjusted results with q > 0.01 are shown).

3.5.13 Ma *et al.*

Ma *et al*. (2020) investigated the role of the protein Gasdermin D (GSDMD) in colitis and found it to be highly activated in DSS-treated mice, independently of the microbiota. They discovered this by cohousing WT mice with GSDMD and ASC deficient mice. ASC is a protein present upstream of GSDMD in the inflammatory signalling cascade. They characterised the gut microbiota of WT, $GSDMD^{-/-}$, and $ASC^{-/-}$ mice pre DSS treatment.

There was no statistically significant difference in microbiotas across healthy and colitogenic groups as well as between WT, $GSDMD^{-/-}$, and $ASC^{-/-}$ mice. There is also no distinct clustering between the

mentioned groups as demonstrated in figure 38. However, there are differentially abundant taxa among the genetically variant mouse groups underlined by DESeq2 (figure 3.37B). Bacteroidales and Lachnospiraceae (except for ASV37) were enriched in $ASC^{-/-}$ mice and depleted in the other two mouse groups. Differential abundance analysis on the condition groups highlighted an association between colitis and the two taxa, Lachnospiraceae and *Mucispirillum* (figure 3.37A). The enriched taxa found in both heatmaps of figure 3.37 match the findings of the meta-analysis, specifically groups A and B.

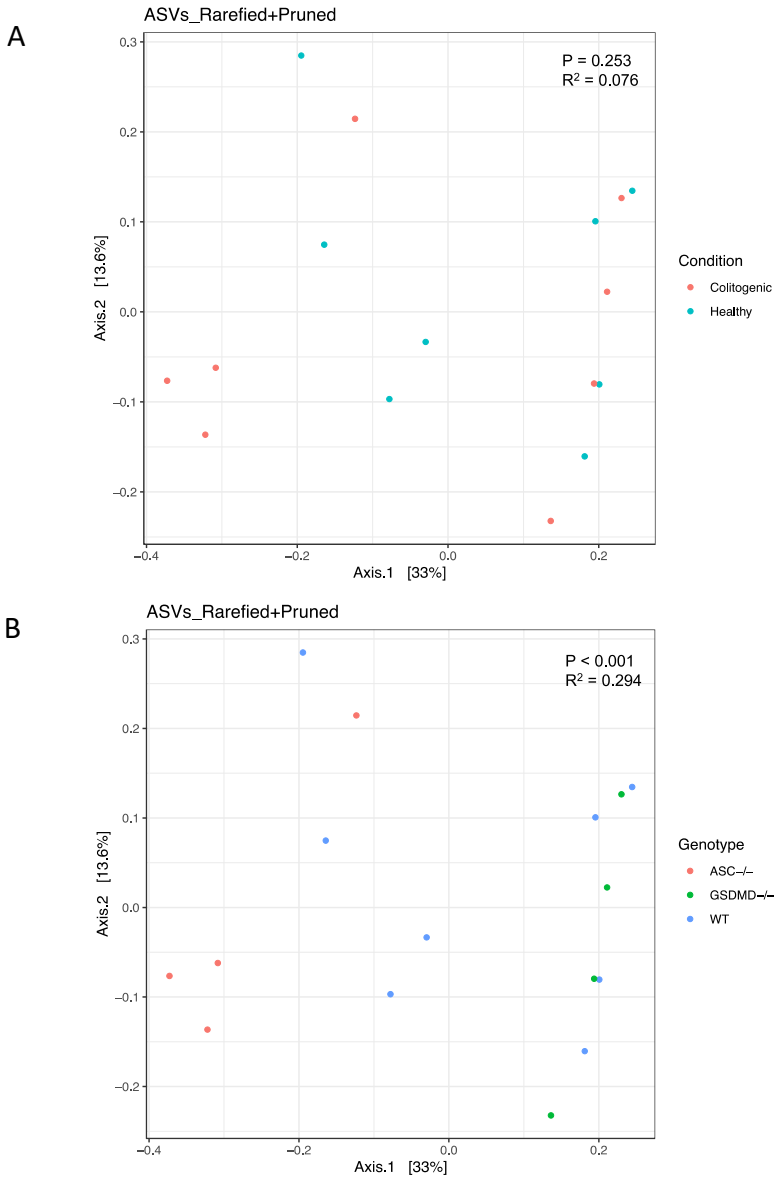


Figure 3.36: Visualisation of beta diversity (Bray-Curtis dissimilarity) across microbial profiles from Ma *et al.* (2020) using PCoA plots coloured based on (A) condition (healthy = blue, colitogenic = red) and (B) genotype shows that microbial profiles did not vary significantly between condition groups and genotype groups.

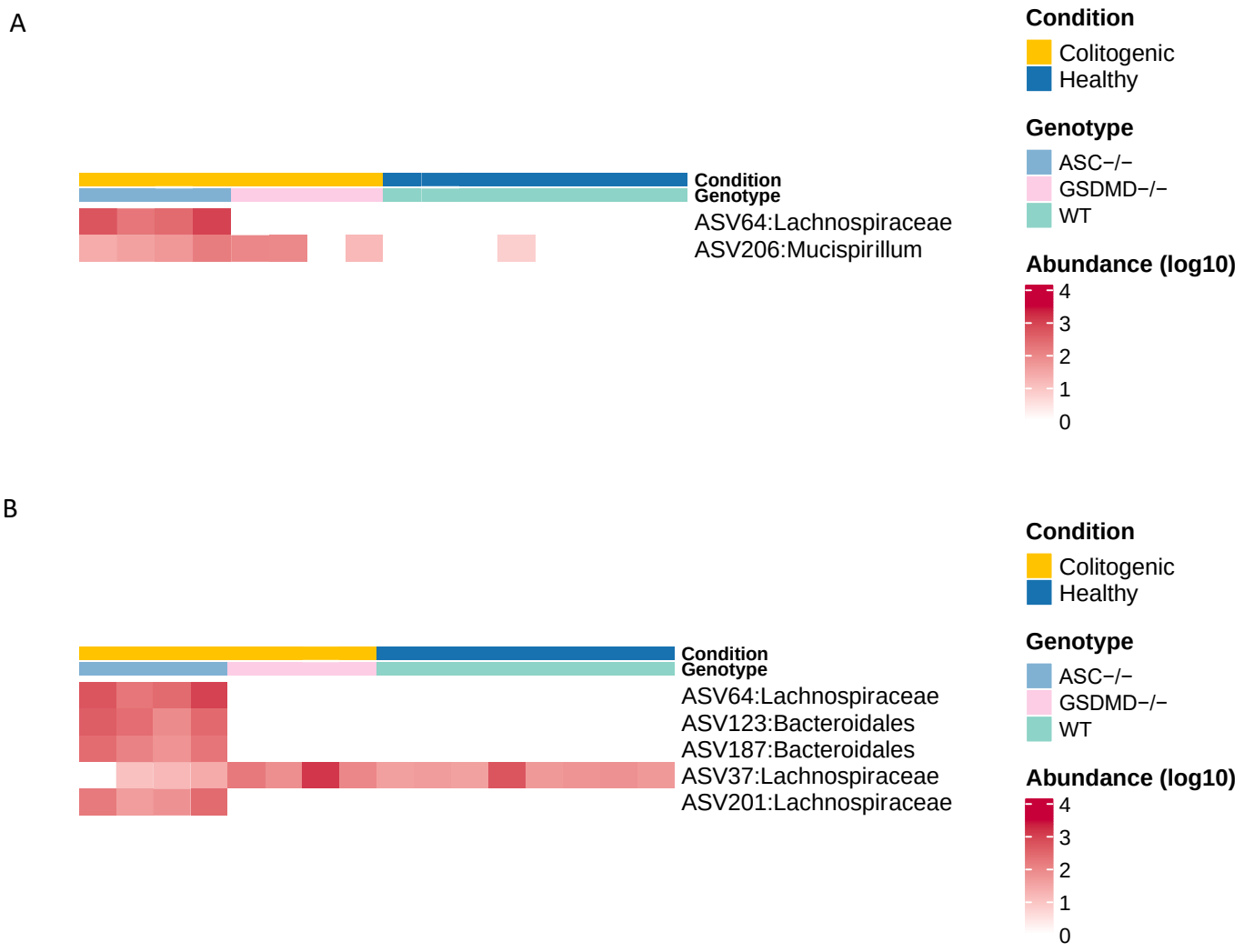


Figure 3.37: Heatmap presenting the abundance (log10 transformed) of the differentially abundant taxa at the lowest taxonomic classification from Ma *et al.* (2020) samples, grouped by condition (blue = healthy, yellow = colitogenic) and genotype (sky blue = *ASC* KO, pink = *GSDMD* KO, green = *GSDMD* WT) using DESeq2 (only FDR-adjusted results with q > 0.01 are shown).

3.6 Results Summary

Based on the combined results of the meta-analysis and single datasets it can be concluded that there are microbial patterns associated with an inflammatory (colitis) state as well as a pre inflammatory state (colitogenic), unified with a low Firmicutes/Bacteroidetes ratio. As highlighted in all 3 runs of the meta-analysis, the condition of the mouse plays a significant role in its gut microbial composition, except for the comparison between healthy and colitis mice investigated in group A. However, in some

single datasets, gut compositional changes were highly influenced by other extrinsic and intrinsic factors like, genotype, treatment, and the length of the treatment, rather than condition. This substantiates that gut flora structure is transient as it is perpetually subjected to environmental shifts. The majority of observed alterations in taxa abundances across conditions were in Bacteroidales and Clostridiales bacteria, as detected by the meta-analysis (tables 3.6 & 3.7) and across the individual studies. Clostridiales species were mainly associated with healthy or pre-inflammatory states (colitogenic) apart from a few taxa mostly belonging to the Peptostreptococcaceae family, like *Rombutsia* and *Terrisporobacter*, as well as *Robinsoniella* and *Clostridium* species. Taxa belonging to the order Bacteroidales were differentially abundant predominantly in colitogenic microbiotas, particularly *Paramuribaculum*, *Duncaniella*, *Bacteroides* and *Muribaculum*, except for *Alistipes*. These specific genera were only flagged in the single datasets and not the meta-analysis. The pattern of differential abundance described in Bacteroidales and Clostridiales species reflects the low Firmicutes/Bacteroidetes ratio observed in the investigated gut floras. Erysipelotrichales bacteria, including *Faecalibaculum*, *C. cocleatum* and *Turicibacter*, were increased in either colitogenic or colitic microbiotas. *E. lenta* was enriched in healthy microbiotas based on two of the meta-analysis runs. Other taxa including *Bifidobacteria*, *Esherichia*/*Shigella spp*., Sutterellaceae, and *Helicobacter* were linked with colitis in the meta-analysis and/or at least 3 studies. Details of the taxa highlighted above can be found in tables 3.6-3.8 below.

Table 3.6: Summary of differentially abundant taxa (detected by DESeq2) from the order Clostridiales (also known as Eubacteriales) down to the lowest taxonomy level detected (genus) across the meta-analysis and single studies. The '*' denotes the presence of the taxa and its colour represents the condition group the taxa was abundant in (blue = healthy, yellow = colitogenic, orange = colitis). Other species include ASVs/taxa assigned to each of the listed families as well as other genera/species that fall under the same family but only detected in one study. Other Clostridiales includes ASVs assigned to the order or other comprising families that are not listed due to their low prevalence across studies.

	Clostridiales																			
	Lachnospiraceae									**Clostridiaceae**			**Peptostreptococcaceae**			**Ruminococcaceae**		**Oscillospiraceae**	**Other Clostridiales**
	Robinsoniella	***Eisenbergiella***	***Blautia***	***Mediterraneibacter***	***Herbinix***	***Anaerotignum***	***Kineothrix***	***Schaedlerella***	**Other species**	***Hungatella***	***Clostridium XIVa***	***Clostridium spp.***	***Romboutsia***	***Terrisporobacter***	**Other species**	***Ruminococcus***	**Other species**	***Dysosmobacter***	
MMUPHin (all DSS)	*								***			**	*	*	*	*			***
MMUPHin (group A)		*	*	*					***	*		*			*	*		*	***
MMUPHin (group B)	*				*				***			**	*		*	*			***
Surana & Kasper (2017)			*						**										**
Tran *et al*. (2019)								*	**		*								*
Spalinger *et al*. (2019)						*	*	*	***		*	*							
Roy *et al*. (2017)						*			*										
Song *et al*. (2020)						*	*		**		*		*				*	*	*
Nakajima *et al*. (2018)	*	*							*	*								*	**
Kazakevych *et al*. (2020)									**										
Qi *et al*. (2020)			*	*					**			*		*	*				*
Mamantopoulos *et al*. (2017)									*								*		*
Schmidt *et al*. (2019)									*										
Kimura *et al*. (2020)									*	*									*
Khan *et al*. (2020)									**								*		
Ma *et al*. (2020)									**										

Table 3.7: Summary of differentially abundant taxa (detected by DESeq2) from the order Bacteroidales down to the lowest taxonomy level detected (genus) across the meta-analysis and single studies. The '*' denotes the presence of the taxa and its colour represents the condition group the taxa was abundant in (blue = healthy, yellow = colitogenic, orange = colitis). Other Bacteroidales includes ASVs assigned to the order or other comprising families that are not listed due to their low prevalence across studies.

	Bacteroidales												
	Muribaculaceae		**Bacteroidaceae**			**Odoribacteriaceae**	**Prevotellaceae**		**Rikenellaceae**		**Tennerellaceae**		**Other Bacteroidales**
	Paramuribaculum	***Duncaniella***	***Bacteroides***	***B. acidafaciens***	***B. thetaiotaomicron***	***Odoribacter***	***Marseilla***	***Alloprevotella***	***Muribaculum***	***Alistipes***	***Parabacteroides***	***P. distasonis***	
MMUPHin (all DSS)					*		*						***
MMUPHin (group A)											*		*
MMUPHin (group B)					*								**
Surana & Kasper (2017)	*	*											
Tran *et al*. (2019)	**	**	**	**					**	**			
Spalinger *et al*. (2019)	*	*	**	**		**	*	*		**			**
Roy *et al*. (2017)	*	*						*	*			*	
Song *et al*. (2020)													
Nakajima *et al*. (2018)													
Kazakevych *et al*. (2020)	**					*			*	*			
Qi *et al*. (2020)		*	*		*						*		*
Mamantopoulos *et al*. (2017)		*							*				*
Schmidt *et al*. (2019)													
Kimura *et al*. (2020)												*	
Khan *et al*. (2020)	*	*	*	*									**
Ma *et al*. (2020)													*

Table 3.8: Summary of differentially abundant taxa (detected by DESeq2) from the order Erysipelotrichales and other taxa down to the lowest taxonomy level detected (species) across the meta-analysis and single studies. The '*' denotes the presence of the taxa and its colour represents the condition group the taxa was abundant in (blue = healthy, yellow = colitogenic, orange = colitis). Other species include ASVs/taxa assigned to each of the listed orders as well as other genera/species that fall under the same family but only detected in one study.

	Erysipelotrichales						Others						
	Eggerthellaceae		Erysipelotrichaceae	Coprobacillaceae	Turicibacteraceae	Other Species							
	Eggerthella lenta	*Other Species*	*Faecalibaculum*	*Clostridium cocleatum*	*Turicibacter*		*Bifidobacteria*	*Akkermansia*	*Escherichia/Shigella spp.*	Sutterellaceae	*Helicobacter*	*Paludicola*	Candidatus Saccharibacteria
MMUPHin (all DSS)	*		*	*	*		*	*	*		*		
MMUPHin (group A)						*			*				*
MMUPHin (group B)	*		*	*	*		*	*	*	*	*	*	*
Surana & Kasper (2017)													
Tran *et al*. (2019)						*					*		
Spalinger *et al*. (2019)			*	*	*			*					
Roy *et al*. (2017)	*	*								*			
Song *et al*. (2020)										*			
Nakajima *et al*. (2018)									*				
Kazakevych *et al*. (2020)			*										
Qi *et al*. (2020)			*	*	*	*	*		*	*	*	*	
Mamantopoulos *et al*. (2017)		*											
Schmidt *et al.* (2019)													
Kimura *et al*. (2020)													*
Khan *et al.* (2020)													
Ma *et al.* (2020)													

Chapter 4: Discussion

4.1 Thesis Summary and Importance

Meta-analysis of studies investigating gut microbial associations in murine UC using the DSS model suggests that there are specific taxa associated with UC, both the state of the disease itself and the period leading up to inflammation. In fact, most of the bacterial taxa that were differentially abundant in the colitis group were also present in colitogenic mice, when compared to the healthy group. Overall, there were no taxa unique to the colitis group and the relative abundance of bacterial taxa in general was lower. The harsh environment in the gut brought about by colitis-related inflammation probably made it inhospitable for commensals, thus depleting bacterial communities. Meta-analysis of all the DSS datasets as well as group B datasets (with samples clustering based on variables like condition, genotype, treatment, length of treatment and intensity of disease) showed a clear link between Erysipelotrichales and colitis. Species from that bacteria order were increased in microbiotas of either colitogenic or colitic mice in the meta-analysis and in 7 out of the 13 datasets (this might not be reflected in heatmap due to the different log2 fold change threshold applied to each dataset when plotting). Members of Erysipelotrichia are known for their involvement in polyol metabolism, SCFA production, as well as host bile acid and lipid metabolism (Zhong et al., 2015; Tiffany et al., 2021; Lynch et al., 2021). Erysipelotrichaceae, Coprobacillaceae, and Turicibacteraceae are the three most prominent families, from this bacterial class, present in high abundance in colitogenic and colitic mice. While the other two families are less described, bacteria belonging to Erysipelotrichaceae are known to be highly immunogenic with their heavy coating of IgA and causing severe colitis symptoms in mice (Palm et al., 2014). Shifts in abundance of this family in IBD patients and animal models have been reported. Lower levels of Erysipelotrichaceae were observed in CD patients while higher levels of the same family were

detected in DSS-treated mice (Ren et al., 2020; Kaakoush, 2015). Despite the enrichment of Erysipelotrichaceae species reported in mice with UC, a decrease of *F. rodentium* (Erysipelotrichales spp.) was observed in an inflammation-induced model of colitis associated colorectal cancer (CRC) (Zagato et al., 2015). *F. rodentium* was shown to metabolically impose an anti-tumorigenic effect by inhibiting HDAC activity through butyrate production (Zagato et al., 2015). UC is a known risk factor for developing CRC in humans and *Holdemanella bifomi*s (human equivalent for *F. rodentium*) was also shown to have similar effects in humans (Zagato et al., 2015). Thus, the association of this species with colitis does not imply an inflammatory nature but possibly a beneficial one. The few available studies on links between IBD and *Turicibacter* (Erysipelotrichales spp.) presented a reduction of the genus in DSS treated and TNBS (other colitis model) treated mice as well as CD patients (Ren et al., 2021; Hall, Kozik & Nakatsu, 2015; Ma et al., 2022). The genus seems to have difficulty thriving during inflammation based on their enrichment in mice with depleted CD8+ T cells (Presley et al., 2010). Eventhough, Turicibacter was enriched in the colitis group in this meta-analysis it is impossible to determine the level of inflammation during the reported "colitis" state in each study. The condition colitis was defined by sequencing samples from mice post DSS treatment, regardless of sampling time, introducing large variability between "colitis" samples. A very recent study investigated the metabolic activity of *Turicibacter* and described how modifications to host bile acids and lipid metabolism by the genus is strain dependant (Lynch et al., 2023). The strain-specific mechanistic differences could be another potential explanation to why *Turicibacter* was found highly abundant in this meta-analysis, opposite to what is described in the literature, since functional properties of taxa rather than their sole presense is what affects the host. The inconsistencies in the nature of the links between Erysipelotrichales species and colitis can be attributed to the poor characterisation and lack of functional studies on this group (Kaakoush, 2015). Based on my results and the current literature, the increased

presence of this group in colitogenic and colitic microbiotas does not necessarily imply that Erysipelotrichales are pro inflammatory species.

The enrichment of Bacteroidales (belonging to Bacteroidetes phylum) species in colitogenic microbiotas found in this meta-analysis agrees with the widely accepted low Firmicutes/Bacteroidetes ratio found in IBD patients (Stojanov, Berlec, & Štrukelj, 2020). Many different Bacteroidales species were found differentially abundant across all 13 studies and *B. thetaiotamicron*, *Marseilla* (*aka P. marseillensis*) and *Parabacteroides* were the ones that stood out across the 3 meta-analyses. *B. thetaiotaomicron* has shown to be increased in UC providing benefit by ameliorating symptoms in UC patients and mice with CD (Delday et al., 2019; Nomura et al., 2021). Very little is known about *P. marseillensis* except for its discovery in a patient with recurrent *C. difficile* infection (Yimagou et al. 2019). Generally, *Prevotella* species are prevalent in the gut flora, however, contradictory conclusions have been drawn about their effect on the host. Some members of the genus have been implicated in gut mucosal inflammation (Prasoodanan et al., 2021; Ijazovic et al., 2021) while others have been described for providing beneficial functions to their host, like improved glucose metabolism (Kovatcheva-Datchary et al., 2015). Thus, it is hard to draw a conclusion on the enrichment of *P. marseillensis* in colitis. It is also difficult to draw a conclusion on the role of *P. distasonis* in colitis. This ambivalent species has been described for its dichotomous role in IBD (Ejezi et al., 2021) which is another testament to the importance of investigating the functional properties of the gut microbiota. *Acetobacteroides* was found enriched in healthy mice and depleted in colitogenic and colitic mice. This meta-analysis is the first to describe an association between *Acetobacteroides* and UC. This genus is not well studied, however, a study characterising its only described species, *A. hydrogenigenes*, reported its carbohydrate fermentation activities and its involvement of acetate production (Su et al., 2014).

Changes in members of Clostridiales (belonging to Firmicutes phylum), particularly Lachnospiraceae bacteria, were highlighted by the meta-analyses. Alterations in the abundance of Lachnospiraceae species were observed in all 3 condition groups. Lachnospiraceaea are a family of obligate anaerobes that are core members of the gut microbial communities. They are considered to be one of the main producers of SCFAs in the gut (Vacca er al., 2020). *Robinsoniella* species as well as *Herbinix* were enriched in colitogenic microbiotas. *R. peoriensis* is a rising pathogen previously mistaken for *C. difficile* infections (Gomez et al., 2011; Roberts et al., 2016; Krueger et al., 2022). Again, this genus/species is very poorly described, however, it has been reported for its SCFA production particularly acetate and succinate but not butyrate (Cotta et al., 2009). Two studies have reported a link between *Robinsoniella* and gut inflammation in an *Il10*$^{-/-}$ and DSS colitis mouse model (Wohlgemuth et al., 2011; Xie et al., 2022). *Herbinix* has been described for its cellulolytic activities as well as potential butyrate production, and this meta-analysis is the first evidence linking it to murine colitis (Koeck et al., 2015; Jung et al., 2021). Generally, members of the Lachnospiraceae family play conflicting roles in health and disease (Vacca et al., 2020). Numbers of *Romboutsia* and *Terrisporobacter* of the Peptostreptococcaceae family were also increased in microbiotas of mice with colitis. An increase in Peptostreptococcaceae has been linked with UC in humans (Alam et al., 2020). *Romboutsia* was first described in association with colitis only recently. Following that, another study reported the enrichment of the genus in UC patients (Rausch et al., 2023; Dahal et al., 2023). *Terrisporobacter* (previously classified as a *Clostridium* species) is yet another poorly represented species in IBD, though a recent study reported the decrease in abundance of the genera in treated CD paediatric patients (Sprockett et al., 2019). This meta-analysis is probably the first to associate the genus with experimental UC.

Other genera like *Bifidobacteria*, *Akkermansia*, *Helicobacter*, and unclassified *Escherichia/Shigella* species were also found in higher numbers in colitis and colitogenic groups. *Bifidobacterium* species (*B. bifidum* and *B. pseudolongum*) are members of the gut and oral flora, usually with a protective effect against colitis and are typically found in lower abundance in UC patients as well as colitic mice (De Caro et al., 2016; Duranti et al., 2016; Singh et al., 2020; Guo et al., 2022). However, pathogenic roles that may lead to septicaemia in immunocompromised hosts have been described (Pathak, Trilligan & Rapose, 2014; Esaiassen et al., 2017). These reports are a possible explanation to why *Bifidobacterium* species were found enriched in the disease groups from this meta-analysis. *A. muciniphila* is a mucin degrader, SCFA producing species and is known for the beneficial role it plays in the gut (Luo et al., 2022). It is usually highly abundant in healthy individuals and reduced during colitis. However, a study reported that the effects of *A. muciniphila* on DSS-induced colitic mice are strain specific, some strains were able to reverse DSS outcomes while another exacerbated it (Liu et al., 2021). Since there are no universal guidelines detailing phenotypic (besides weightloss and bloody stools) and histological colitis indicators in mice, samples from the colitis group might not necessarily reflect the inflammatory stage of UC accurately. Sampling time for colitis microbiomes could be equivalent to the remission state in humans, when inflammation has subsided and the mucosal lining is on the mend shifting the gut flora with an increase in mucosa-dwelling species. This could be a reason for the enrichment of the mucin-degrading *Akkermansia* in the colitis group . On the other hand, the *Helicobacter* genus is known for its infectious repertoire. *H. typhlonius* and *H. rodentium* infections are very common in animal facilities and have been shown to affect inflammation severity in Il10$^{-/-}$ mice, even trigger IBD development in some cases (Chichlowski et al., 2008). *Escherichia/Shigella* sp. are unclassified Enterobacteriaceae members. Bacteria belonging to the family

Enterobacteriaceae are known to be harmful or pathobionts and are abundant in IBD patients (Baldelli et al., 2021).

Despite the presence of thousands of studies linking gut microbes to health and IBD, there are many inconsistencies in the findings. These inconsistencies could be context dependent, however, lacking guidelines and experimental plans are two main reasons. There is no universal guide or set criteria for colitis severity in mice, leaving it open for personal interpretation which tends to be subjective. Mouse diets can be very different across animal facilities which introduces variability when comparing microbial composition. Discrepancies also arise when mice of both sexes are used, and samples are not annotated nor compared for sex differences. Furthermore, variations in computation methods when analysing microbiome data is widespread. The field is yet undecided on the appropriate methods to use, whether is it during the pre-processing step (OTU vs ASV) or downstream analysis (i.e., taxonomic profiling and microbial community analysis). To identify consistent microbial markers, scientists should work towards controlling the mentioned variables. Due to the nature of this thesis, I was only able to control for computational variation like 16S rRNA sequence processing, taxonomic assignment, differential abundance testing as well as multiple testing corrections. Firstly, there are different bioinformatics pipelines used to analyse microbial amplicon sequences which are divided into two groups depending on the feature output, OTU or ASV. Each pipeline differs in sensitivity and specificity of output feature detection giving rise to varaitions in analysed 16S rRNA sequences (Prodan et al., 2020). Taxonomic classification can be done using 4 different databases with any of their multiple available training sets, which can affect the specificity of the taxonomic assignment resulting in more/less unclassified taxa. Downstream analysis methods like differential abundance testing can be done using different tools that have been shown to produce different outputs of with some overlap as well as variable statistical

robustness and false discovery rates, depending on subsequent pre-processing steps (mentioned above) (Nearing et al., 2022). Changes in the discussed steps can introduce biological variations making it difficult to simply compare results from different studies. This meta-analysis eliminated the chances of any microbial changes that could have risen from different computational methods, which are known to affect biological interpretations (Nearing et al., 2022). While the taxa described in this work are just associations with UC in mice, they raise awareness about gaps in the field. Many of these taxa are poorly described in the literature and need to be investigated. Extensive work on defining the functional profiles of gut commensals as well as microbial co-dependencies is needed to understand the role they play in health and disease. This work sheds light on the complexity of assigning a cause-and-effect relationship between microbiome and IBD and the unattainability of defining a microbial signature for IBD due to confounding variables.

4.2 Gut Microbial Associations with UC in Humans and Mice

Results from this meta-analysis show a reduction in the butyrate-producing Firmicutes and an increase in acetate and propionate producing Bacteroidetes in colitogenic mice. This is also observed in human UC patients (Basha et al., 2022). Enterobacteriaceae were found in higher levels in UC patients which corresponds with the increase of an unclassified *Escherichia/Shigella* species in colitis and colitogenic mouse groups investigated (Ma et al., 2022). *E. lenta* is likely beneficial to the host with a report of its abundance being a predictor of FMT success in UC patients as well as its abundance in healthy mouse group and depletion in colitis and colitogenic groups. (Li et al., 2020). The presence of Peptostreptococaceae as well as pathogenic species like *Mycobacterium*, adherent-invasive *E. coli* (AIEC), *C. difficile*, *Helicobacter* and *Salmonella* is more prominent in UC patients compared to healthy controls (Shen et al., 2018; Alam et al., 2020). *Helicobacter* and *Clostridium* species as well as Peptostreptococaceae species were differentially abundant in colitis and colitogenic groups indicating that there is a positive correlation between those pathogenic bacteria and UC. While *Bacteroides* species in general are depleted in UC patients, *B. thetaiotaomicron* is abundant in both UC patients and mice with a seemingly positive role (Wexler et al., 2007). *B. thetaiotaomicron* is highly adaptable and can switch between metabolising dietary and host polysaccharides depending on nutrient availability, which explains why it can thrive under harsh conditions like during inflammation (Wexler, 2007). This resourceful species has the ability of modulating the gut flora via the immune system (Cash et al., 2006). The often-depleted Lactobacilli were shown to prolong the remission state of UC in patients which suggests that they play an anti-inflammtory role in the host (Zocco et al., 2006). This is also reflected in their reduced abundance in colitic mice

from this meta-analysis. Most of these patterns in bacterial abundances were by the IBD Transcriptome and Metatranscriptome Meta-Analysis (TaMMA) framework (Massimino et al., 2021). This platform has a set of 26 publicly available transcriptomic and metatranscriptomic datasets from different human tissues, all batch-corrected and analysed using the same pipeline. Using this resource, I was able to confirm some of the patterns I found in individual datasets described earlier, like the increase of Enterobacteriaceae in UC stool samples, particularly *Escherichia* and *Shigella* species. *E. lenta* was also found to be enriched by 1-fold in UC patients. Depending on the species, *Clostridium* bacteria were also differentially abundant in UC patients. However, IBD TaMMA reports an unsignificant changes in *Bacteroides* species between healthy subjects and those with colitis. Moreover, Peptostreptococcaceae seem to be only slightly more abundant in healthy individuals opposite to what was reported in this meta-analysis. This can be due to variations introduced by different pipelines or environmental differences. Overall, it is safe to say that microbial patterns detected by this meta-analysis in relation to UC can also, mostly, be found in humans.

4.3 Limitations and Challenges

There are several limitations of the work presented in this thesis imposed by study design and others by the framework of the available datasets. The first few potential biases introduced by study design were in the study selection step of this meta-analysis. To have a manageable list, I employed the addition an impact factor of 7 or greater as additional filter in the study selection. This choice does not confer journal bias but merely a mean to produce a compressed list of datasets for further screening as manually screening over 1800 publications would not have been feasible with the project's timeline. Most of the studies generated by the search query were published in journals with an impact factor between 5 and 7, which means that many relevant datasets could have been included. However, with the recent developments in natural language processing (NLP) models, investigating that large number of studies would be possible and could be explored in future research. Another potential bias could be the different variables tested in some individual dataset. The meta-analysis tested for changes in gut microbiome structure across condition groups only. When analysing individual datasets, the effect of the condition on gut microbial composition was not always statistically significant and, in some studies, like Kimura et al. (2020) and Mamantopoulos et al. (2017), samples from

only one condition group was available making impossible to test for condition introducing variation when testing for differentially abundant taxa.

As briefly mentioned in the previous section, a lack of standardised experimental system for murine colitis was a major issue faced when pre-processing the datasets for meta-analysis. As discussed in great length in chapter 1, diet plays a huge role in the structure of the microbiome. Unless diet was the main area of investigation it is often overlooked. The exact composition of standard chow diet is considered proprietary information and is not available to the public. It varies across manufacturers and can even change from one batch to another with no prior notice creating protential bias in the involved research (Pellizon, 2016). There should be an established control diet with a set composition for universal use in all non-diet related microbiome investigations. Colitis grading is another factor that needs to be standardised in UC mouse models. There are no formal criteria defining colitis in mice nor grading of symptoms outlining colitis severity (equivalent to flare-ups and remission in humans), leaving it for personal interpretation by both authors of the study when interpreting results and myself during meta-analysis design. In addition to that, the criteria used by researchers to define the condition of the mice (healthy vs colitis), or the severity of the disease (if even mentioned) is often not described in the study. Therefore, condition grouping and intensity of disease was left to my personal interpretation (as described in the methods chapter) of the authors' lacking description and likely introduced variability in results across mouse groups. Another issue encountered was poor curation of the metadata assigned to sequenced samples, where sample names were improperly annotated, the sequenced variable region of the 16S gene and the sex of mice were not always mentioned In fact, the majority of the studies included in this meta-analysis conducted their experiments on male mice only when both human and animal studies have shown significant sex differences in intestinal microbial communities (Kim et al., 2020).

This again introduces bias and presents incomplete characterisation of associative links between UC and the gut flora. Finally, the shortage of studies investigating functional roles of gut bacteria in health and disease made it difficult to postulate the type of associative link between taxa and colitis.

4.4 Future work

The work done throughout this thesis only draws associative links between the gut microbiome and UC in the DSS mouse model. As highlighted in 3.3.1, the colitis murine model has a great effect on the structure of the gut microbial community. Thus, to be able to generalise the microbial patterns associated with UC a meta-analysis should include datasets employing other mouse models, like the *Il10* KO model and the T cell transfer model. Other variables like, sex and extraction/sequencing kits, should also be considered when analysing the data to eliminate unwanted biases. Co-occurrence network analysis of bacterial associations with UC should be done to explore patterns and connections within and across microbial communities. The findings should also be validated in the wet lab to determine whether these associations are of causal nature or mere correlations. To better understand the role of the taxa in UC, colitis would be induced using DSS in gnotobiotic mice colonised with a synthetic microbiota including species like, *Turicibacter*, *Faecalibaculum*, *C. cocleatum*, *Bifidobacteria*, *Helicobacter*, *Terrisporobacter* and *Herbinix*, and microbial compositional changes would be characterised using 16S sequencing. Once a link has been established between microbes and the condition,

their functional profiles would be identified using metagenomic and transcriptomic profiling followed by metabolomic assays. Functional profiling can be confirmed in vitro once appropriate culturing protocols are established. The meta-analysis workflow should be applied on human datasets and the results should be cross-checked to demonstrate the translatability of the data.

4.5 Conclusions

This meta-analysis drew links between microbes and colitogenic/colitic states. The main pattern observed was changes in taxa abundance from the orders Bacteroidales and Clostridiales (particularly Lachnospiraceae), with an increased abundance of Bacteroidales species in the colitogenic and/or colitis groups. This observation agrees with the widely accepted low Firmicutes/Bacteroidetes ratio detected in UC cases, whether in humans or mice (Stojanov, Berlec, & Štrukelj, 2020). Erysipelotrichales are predominant in colitogenic and colitis groups, as well as *Bifidobacteria*, *Esherichia*/*Shigella spp*., and *Helicobacter*. An associative link between two taxa, *Terrisporobacter* and *Herbinix*, and experimental UC was first reported in this thesis. I hypothesize that the presence of pro-inflammatory taxa like *Escherichia/Shigella* spp., *Helicobacter*, *Robinsoniella* and Peptrostreptococcaceae species is a predisposition to murine UC. A causal relationship between these taxa and UC cannot be deduced due to the lack of functional experiments. However, this work raises questions and areas in the field that need to be investigated in the future.

Chapter 5: Faecal matter transfer in GF shows that phenotypic changes in intestinal epithelial lining modulated by the immune system are microbiota dependent

5.1 Chapter Overview

In this chapter, the findings of another project I worked on, in collaboration with the Laboratory of Immunoinflammation at University of Campinas, Brazil, during my PhD are presented. These findings have been published in the journal *Microbiome,* in April 2023, under the title "Inulin diet uncovers complex diet-microbiota-immune cell interactions remodelling the gut epithelium". My contribution to this study, detailed in the methods section of this chapter, includes 16S rRNA data processing and analysis for all microbiota characterisation experiments, except for the ones employing gnotobiotic mice. Briefly, the experimental design involved grouping mice based on their diet, inulin-rich and control diet, then identifying the composition of their microbiome 30 days post diet changes. This experiment was repeated and expanded on by inoculating germ-free mice with faeces from mice of either groups. Mice fed an inulin-rich diet and the germ-free mice inoculated from the latter mice exhibited phenotypic changes in the gut epithelium. I discuss the results of the analysis I did on the 16S rRNA sequencing data from those two experiments linking it to the project as a whole in the manuscript summary. Since this work does not investigate gut microbial associations with UC, I was not able to include it in my meta-analysis and is instead described here in a separate chapter.

5.2 Manuscript Summary

In summary, dose-dependent phenotypic changes in the gut epithelium, including a 25% deepening of colonic crypts and lengthening of the caecum and colon, were observed in mice

fed an inulin (fermentable fibre) rich diet (10%) for 30 days. EdU incorporation assay, flow cytometry, immunostaining and clonogenicity assays all confirmed that these modulations were due to increased cell proliferation rates in colonic crypts. Lineage tracing demonstrated the enhanced proliferative activity of colonic crypt dwelling $Lgr5^{+}$ stem cells in inulin fed mice. Bulk RNA sequencing of intestinal epithelial cells (IECs) showed an upregulation of genes involved in cell cycle pathways as well as DNA replication and repair pathways while single-cell RNA sequencing (scRNAseq) analysis indicated a two-fold increase in proliferative cells in inulin fed mice compared to the control group. Further analysis of scRNAseq data showed that inulin has an impact on proliferation and differentiation of colonic epithelial cells with an increase in cycling cells and goblet cells as well as their expression of mucus-associated genes. The described modulatory phenotype elicited by an inulin diet seemed to be microbiota dependent. The increased proliferative activity in colonic crypts was only observed in SPF and was absent in GF and antibiotic treated mice. 16S analysis of faecal samples from SPF mice showed an increase in Bacteroidetes species, particularly *Bacteroides* and *Duncaniella*. FMT from inulin fed SPF to GF mice induced the same phenotypic even when dietary inulin was not supplied. To reduce variation and complexity of the gut microbiome the diet experiment was repeated using gnotobiotic mice colonised with a synthetic microbiota of 13 fully sequenced human commensal strains (including *B. ovatus*, *B. uniformis*, *B. thetaiotamicron*, *B. caccae*, *Barnesiella intestinihominis*, *Roseburia intestinalis*, *Eubacterium rectale*, *Marvinbryantia formatexigens*, *Clostridium symbiosum*, *Collinsella aerofaciens*, *E. coli* HS, *Desulfovibrio piger* and *A. muciniphila*). The phenotype was again observed in the gnotobiotic mice with alterations in the microbiome, particularly an enrichment of two *Bacteroides* species and *C. aerofaciens*. Bacteroidales was the only overlapping taxa when comparing the microbiotas of the three mouse groups (SPF, GF + FMT, and gnotobiotic mice), with a significant increase in abundance. Further experimental work showed that dietary inulin increased expression of *Il22*

as well as the number of proinflammatory Th17 cells. The inulin-induced phenotype was not just microbiota-dependent but also reliant on interactions with γδ T cells and their secreted cytokine *Il22*. This study shows the complexity of defining the role or assigning a causative association between microbes and phenotypic changes let alone disease manifestations.

5.3 Methods

The pipelines used to process and analyse the data in this study are mostly similar to those covered in the chapter 2. The main differences were in the analysis methods, which will be highlighted in the 5.3.2.

5.3.1 Data Processing

The provided de-multiplexed 16S rRNA sequences were processed using the DADA2 pipeline (Callahan et al., 2016) after passing FastQC quality checks. The reverse reads were of poor quality and failed the checks, preventing paired read merging, thus only the forward reads were used in downstream analysis. The default DADA2 parameters used and only modified at the trimming step to account for primer sequence contamination. The FASTQ files were then filtered based on the error rates and quality scores generated by the DADA2 algorithm. Following this, the reads were denoised and summarised into amplicon sequence variants (ASVs) rather than operational taxonomic units (OTUs), then filtered for chimeric sequences. Taxonomy was assigned using the RDP Classifier training set 18 (Wang et al., 2017).

5.3.2 Data Analysis and Taxonomic Profiling

The *phyloseq* R package was used for downstream analysis, including the heatmap plots (McMurdie & Holmes, 2013). To explore the variability among sample groups, UniFrac distances between the samples were calculated and visualised using a PCoA plot. The statistical

significance between differences microbial in communities were tested using DESeq2 which is based on a negative binomial distribution (Wang et al., 2007). LEfSe analysis was used to determine the ASVs that were statistically different among the two experimental groups. LDA threshold of 2 was used and a significance alpha value of 0.5 was set for both the Kruskal-Wallis and Wilcoxon tests. The top 20 ASVs were sorted by LDA scores and plotted to show the differentially abundant taxa that are statistically significant.

5.4 Results

5.4.1 Data Processing

Processing the data from the first mouse experiment which included two experimental groups: inulin rich diet and control diet. In experiment 1, mice were placed on either of the two diets for 30 days, sampled for microbiota composition testing then euthanised. Experiment 2 was an expansion of experiment 1 where the sampled faeces were used for FMT in germ-free mice. The germ-free mice were then put on a chow diet for 21 days before their microbiomes were sampled and analysed. The sequences from both experiments were processed using the DADA2 pipeline as described in the methods section of this chapter. Read tracking, feature counts and taxonomy assignment were used to measure the success of the pipeline (table 5.1). In the first experiment, using pair-ended reads yielded either 0 reads or 15% of retained reads that were failing to align to taxonomy (table 5.1). The same was observed with experiment 2 where the use of merged sequences resulted in low retained reads and failed taxonomy alignment, in some DADA2 runs. 76% of the sequences were retained with successful taxonomy alignment when the DADA2 pipeline was run on the forward reads only, thus only single reads were used to generate the ASV table (described in detail in chapter 3).

Table 5.1: Summary of the trimming parameters tested in the DADA2 pipeline and their corresponding read tracking, feature counts and taxonomy assignment information.

Trimming Parameters	Total Reads	Retained Reads	Number of ASVs	Assigned Taxonomy
Experiment 1				
F=250; R=225	2258086	15.4%	3212	No
F=250; R=175		0%	0	No
F=250		76.7%	871	Yes
F=235; R=175 Pooled		59.1%	1580	Yes
Experiment 2				
F=225; R=225	1669462	0%	56	No
F=250; R=250		13.8%	56	Yes
F=285; F=285		13.7%	1938	Yes
F=285; F=285 (Pooled)		72%	2002	No
F=235		76%	2706	Yes

5.4.2 Data Analysis and Taxonomic Profiling

As detailed in chapter 3, the ASV tables, taxonomy table and FASTA file were combined into a phyloseq object then used to generate a phylogenetic tree using ape and phangorn v. 2.11.1 (Paradis & Schliep, 2019; Schliep *et al*., 2017). Distances between the sample points were calculated using the UniFrac distance metric and plotted into a PCoA plot (figure 5.1). PCoA plot shows that the microbiotas from mice on the inulin-rich diet were different from those put on the control diet. PERMANOVA testing showed this to be statistically significant with about 37% of the variation between samples is due to the chosen grouping i.e. the diet tested (figure 5.1). *Duncaniella*, Bacteroidales and *Akkermansia* were particularly enriched in mice fed an inulin-rich diet (figure 5.2A). These observations were not exactly replicated in the second experiment. FMT from mice of the different diet groups did not see to have a great effect on the overall gut microbial composition with no major shifts detected, despite the consistent

phenotypic changes observed (figure 5.1B). However, like the first experiment, the bacterial taxa *Duncaniella* and Bacteroidales were more abundant in mice receiving FMT from mice put on an inulin-rich diet (figure 5.2B). This implies that these bacterial taxa are associated with the fibre inulin.

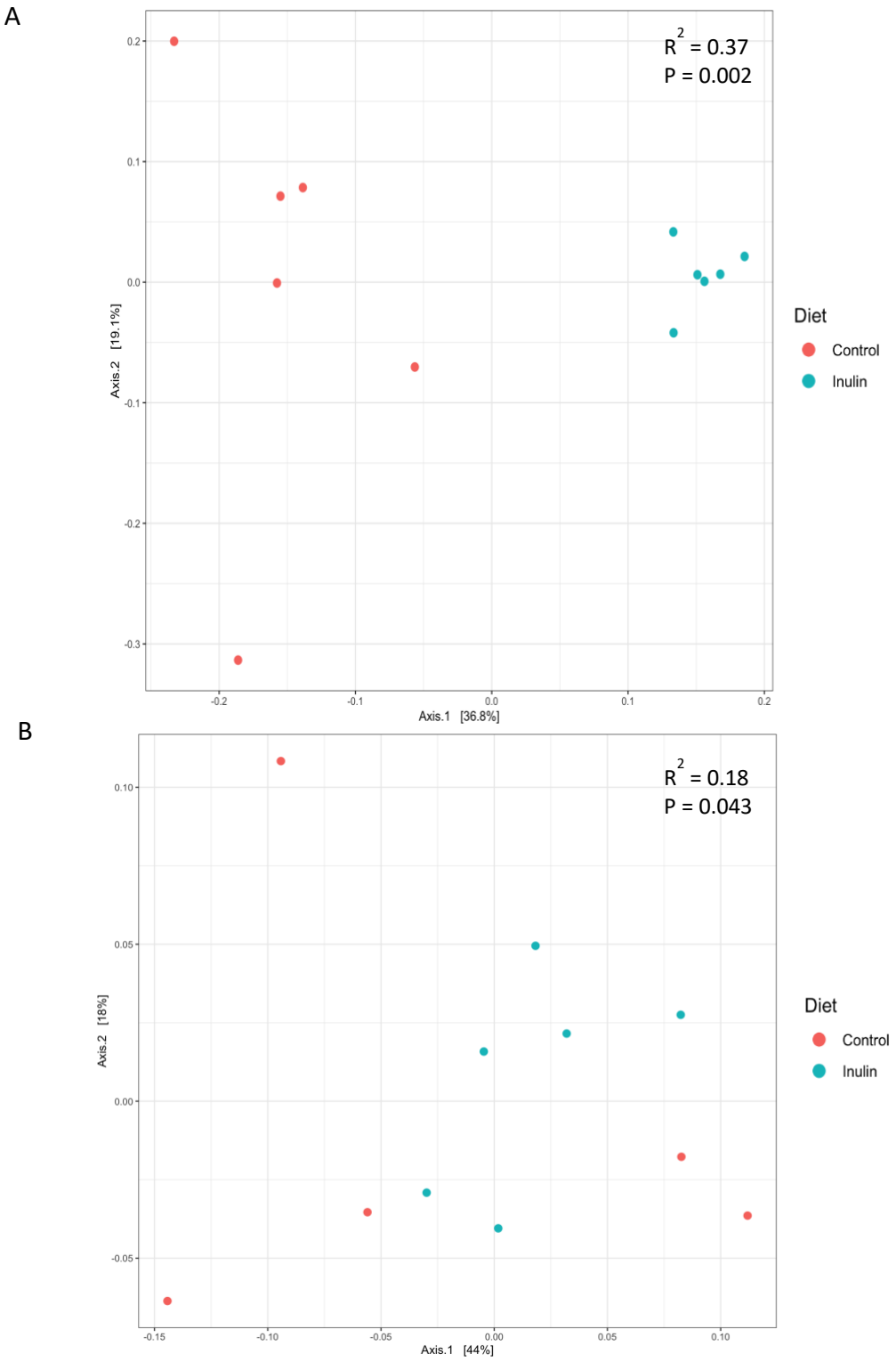


Figure 5.1: Visualisation of beta diversity (UniFrac) across microbial profiles from Corrêa et al. (2023) using PCoA plots coloured based on diet (Inulin = blue, Control = red) shows that microbial profiles varied significantly between diet groups in experiment 1 (A) but the same was not observed in the second experiment (B).

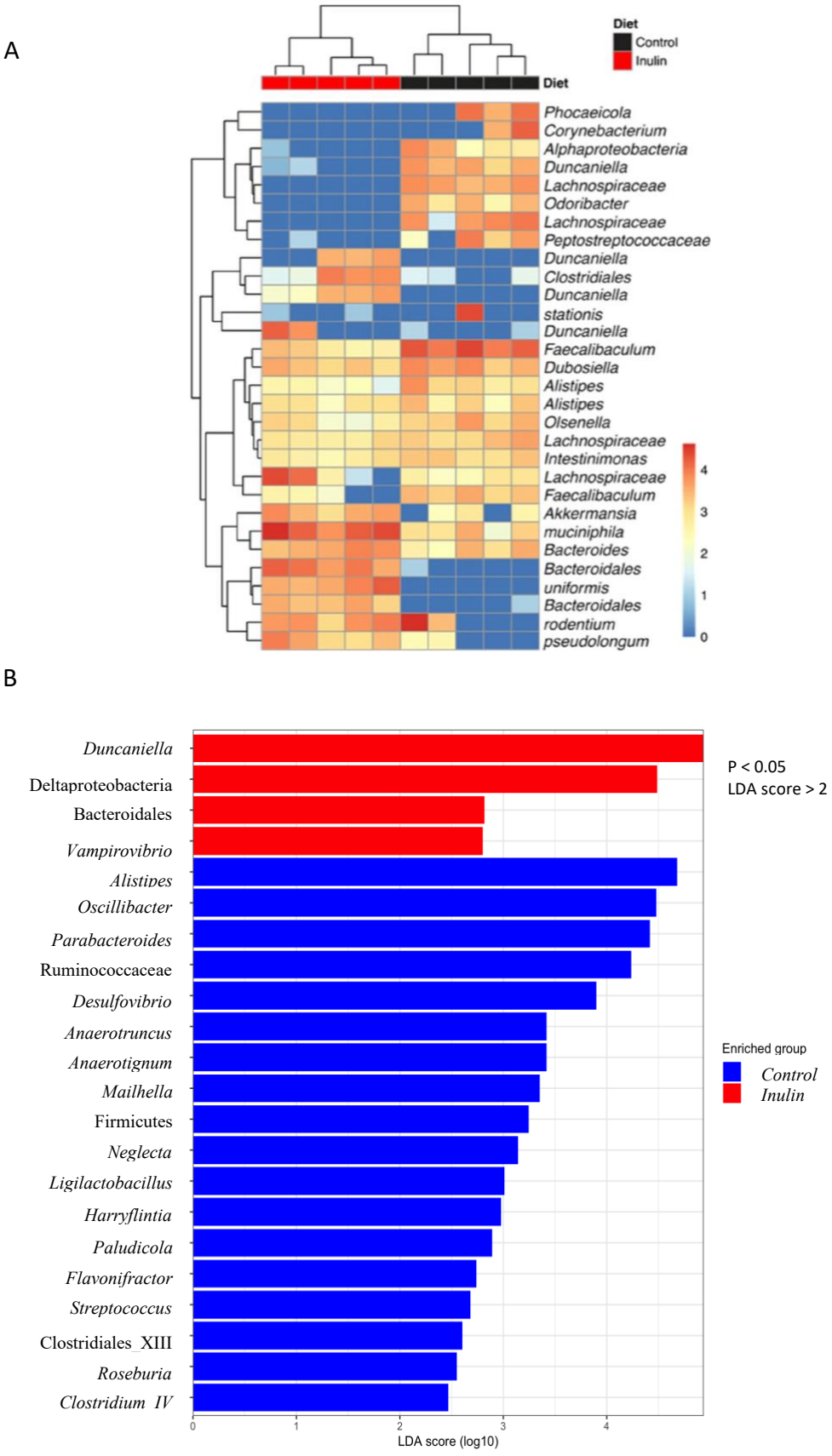


Figure 5.2: Inulin-rich diet implicated in gut microbial shifts. (A) Heatmap showing relatively abundant taxa across the two diets in experiment 1. (B) Barplot showing LDA scores from LEfSe analysis on differentially abundant taxa from mice in experiment 2 post chow diet.

5.5 Conclusion

Based on my analysis combined with the other data presented in associated manuscript, it can be concluded that a diet high in inulin triggers microbiota-dependent phenotypic changes modulated by the immune system. Antibiotic treated mice as well as germ-free mice did not experience the gut epithelial changes with inulin supplementation. *Duncaniella* and Bacteroidales species have been associated with these modifications in the gut epithelium. Results from the experiment with the SPF mince further confirmed the enrichment of Bacteroidales members, specifically *Bacteroidetes*, when inulin was involved. *Duncaniella* was not one of the 13 strains present in the SPF mice, hence it was not flagged in the results. Further investigations highlighted the role of γδ T cells and their secreted *Il22* in the observed phenotype. This further confirms that studying the microbiome in relation to a phenotype or disease in isolation is too simplistic and will not paint an accurate picture.

References

Abdugheni, R. *et al.* (2022) “Metabolite profiling of human-originated Lachnospiraceae at the strain level,” *iMeta*, 1(4). doi: 10.1002/imt2.58.

Agus, A. *et al.* (2016) ‘Western diet induces a shift in microbiota composition enhancing susceptibility to Adherent-Invasive E. coli infection and intestinal inflammation .’, *Sci. Rep.*. doi: 10.1038/srep19032.

Ahn, S. *et al.* (2008) ‘Hepatocyte nuclear factor 4α in the intestinal epithelial cells protects against inflammatory bowel disease’, *Inflamm. Bowel Dis.,* 14(7), pp. 908–920. doi: 10.1002/ibd.20413.

Alam, M. T. *et al.* (2020) “Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels,” *Gut pathogens*, 12, p. 1. doi: 10.1186/s13099-019-0341-6.

Alexander, M. *et al.* (2020) “Human Gut Bacterial Metabolism Drives Th17 Activation and Colitis,” *SSRN Electronic Journal*. doi: 10.2139/ssrn.3737642.

Alzarhani, A. K. *et al.* (2019) ‘Are drivers of root-associated fungal community structure context specific?’, *ISME J.*, 13(5), pp. 1330–1344. doi:10.1038/s41396-019-0350-y.

Ananthakrishnan, A. N. *et al.* (2012) ‘Higher predicted vitamin D status is associated with reduced risk of Crohn's disease’, Gastroenterology, 142(3), pp. 482–489. doi: 10.1053/j.gastro.2011.11.040.

Ananthakrishnan, A. N. *et al.* (2017) ‘Environmental triggers in IBD: A review of progress and evidence’, *Nat. Rev. Gastroenterol. Hepatol*., 15(1), pp. 39–49. doi: 10.1038/nrgastro.2017.136.

Artis, D. *et al.* (2004) 'RELMβ/FIZZ2 is a goblet cell-specific immune-effector molecule in the gastrointestinal tract', *Proc Natl Acad Sci USA*, 101(37), pp. 13596-13600. doi:10.1073/pnas.0404034101.

Arumugam, M. *et al.* (2011) 'Enterotypes of the human gut microbiome', *Nature*, 473(7346), pp. 174–180. doi: 10.1038/nature09944.

Astó, E. *et al.* (2019) 'The Efficacy of Probiotics, Prebiotic Inulin-Type Fructans, and Synbiotics in Human Ulcerative Colitis: A Systematic Review and Meta-Analysis', *Nutrients,* 11(2). doi: 10.3390/nu11020293.

Azad, M. B. *et al.* (2016) 'Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: a prospective cohort study', *BJOG,* 123(6), pp. 983–993. doi: 10.1111/1471-0528.13601.

Bäckhed, F. *et al.* (2012) 'Defining a healthy human gut microbiome: Current concepts, future directions, and clinical applications', *Cell Host Microbe*, 12(5), pp. 611–622. doi: 10.1016/j.chom.2012.10.012.

Baldelli, V. *et al.* (2021) "The Role of Enterobacteriaceae in Gut Microbiota Dysbiosis in Inflammatory Bowel Diseases," *Microorganisms*, 9(4), p. 697. doi: 10.3390/microorganisms9040697.

Bankevich, A. *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), pp. 455–77. doi: 10.1089/cmb.2012.0021.

Barthels, C. *et al.* (2017) "CD40-signalling abrogates induction of RORγt^{+} Treg cells by intestinal CD103^{+} DCs and causes fatal colitis," *Nature communications*, 8, p. 14715. doi: 10.1038/ncomms14715.

Barrett, J. C. *et al.* (2009) 'Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region', *Nat. Genet.*, 41(12), pp. 1330–1334. doi: 10.1038/ng.483.

Basha, O. M. *et al.* (2022) "Impact of gut Microbiome alteration in Ulcerative Colitis patients on disease severity and outcome," *Clinical and Experimental Medicine*. doi: 10.1007/s10238-022-00917-x.

Bassaganya-Riera, J. *et al.* (2011) 'Soluble fibers and resistant starch ameliorate disease activity in interleukin-10-deficient mice with inflammatory bowel disease', *J. Nutr.*, 141(7), pp. 1318–1325. doi: 10.3945/jn.111.139022.

Battistini, C. *et al.* (2021) 'Vitamin D Modulates Intestinal Microbiota in Inflammatory Bowel Diseases', *Int. J. of Mol.Sci.*, 22(1). doi: 10.3390/ijms22010362.

Baydi, Z. *et al.* (2021) 'An Update of Research Animal Models of Inflammatory Bowel Disease', *ScientificWorldJournal*. doi: 10.1155/2021/7479540.

Benchimol, E. I. *et al.* (2017) 'Rural and Urban Residence during Early Life is Associated with a Lower Risk of Inflammatory Bowel Disease: A Population-Based Inception and Birth Cohort Study', *Am. J. Gastroenterol.,* 112(9), pp. 1412–1422. doi: 10.1038/ajg.2017.208.

Benjamin, J. *et al*. (2011) ' Randomised, double-blind, placebo-controlled trial of fructo-oligosaccharides in active Crohn's disease', *Gut*, 60(7), pp. 923-929. doi:10.1136/gut.2010.232025.

Berry, D. *et al.* (2015) “Intestinal Microbiota Signatures Associated with Inflammation History in Mice Experiencing Recurring Colitis,” *Frontiers in Microbiology*, 6. doi: 10.3389/fmicb.2015.01408.

Bibbò, S. *et al.* (2020) ‘Fecal Microbiota Transplantation: Screening and Selection to Choose the Optimal Donor’, *J. Clin. Med.,* 9(6). doi: 10.3390/jcm9061757.

Bjarnason, I. *et al.* (2018) ‘Mechanisms of Damage to the Gastrointestinal Tract From Non-steroidal Anti-inflammatory Drugs’, *Gastroenterology*, 154(3), pp. 500-514. doi: 10.1053/j.gastro.2017.10.049.

Bojanova, D. P. and Bordenstein, S. R. (2016) ‘Fecal Transplants: What Is Being Transferred ?’, *PLoS Biol.*, 14(7),. doi: 10.1371/journal.pbio.1002503.

Borrelli, O. *et al.* (2006) ‘Polymeric Diet Alone Versus Corticosteroids in the Treatment of Active Pediatric Crohn’s Disease: A Randomized Controlled Open-Label Trial', *Clin. Gastroenterol. Hepatol.,* 4(6), pp. 744–753. doi:10.1016/j.cgh.2006.03.010.

Britton, G. J. *et al.* (2019) “Microbiotas from Humans with Inflammatory Bowel Disease Alter the Balance of Gut Th17 and RORγt+ Regulatory T Cells and Exacerbate Colitis in Mice,” *Immunity*, 50(1). doi: 10.1016/j.immuni.2018.12.015.

Browne, H. P., Shao, Y. and Lawley, T. D. (2022) ‘Mother – infant transmission of human microbiota’, *Curr. Opin. Microbiol.*, 69. doi:10.1016/j.mib.2022.102173.

Burrello, C. *et al.* (2018) “Therapeutic faecal microbiota transplantation controls intestinal inflammation through IL10 secretion by immune cells,” *Nature communications*, 9(1), p. 5184. doi: 10.1038/s41467-018-07359-8.

Busbee, P. B. *et al.* (2020) “Indole-3-carbinol prevents colitis and associated microbial dysbiosis in an IL-22-dependent manner,” *JCI insight*, 5(1), p. e127551. doi: 10.1172/jci.insight.127551.

Callahan, B. J. *et al.* (2016) “DADA2: High-resolution sample inference from Illumina amplicon data,” *Nature Methods*, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Canesso, M. C. C. *et al.* (2018) “The cytosolic sensor STING is required for intestinal homeostasis and control of inflammation - PubMed,” *Mucosal immunology*, 11(3). doi: 10.1038/mi.2017.88.

Caporaso, J. G. *et al.* (2011) ‘Moving pictures of the human microbiome’, *Genome Biol.,* 12(5). doi: 10.1186/gb-2011-12-5-r50.

Caro, C. D. *et al.* (2019) “Intestinal inflammation increases convulsant activity and reduces antiepileptic drug efficacy in a mouse model of epilepsy,” *Scientific Reports*, 9(1), pp. 1–10. doi: 10.1038/s41598-019-50542-0.

Carter, C. S. (2021) “A ‘Gut Feeling’ to Create a 10th Hallmark of Aging,” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 76(11). doi: 10.1093/gerona/glab191.

Cash, H. L. *et al.* (2006) “Symbiotic bacteria direct expression of an intestinal bactericidal lectin,” *Science (New York, N.Y.)*, 313(5790), pp. 1126–30. doi: 10.1126/science.1127119.

Chai, J. N. *et al.* (2017) “*Helicobacter* species are potent drivers of colonic T cell responses in homeostasis and inflammation,” *Science immunology*, 2(13), p. eaal5068. doi: 10.1126/sciimmunol.aal5068.

Chang, C.-S. *et al.* (2021) "Identification of a gut microbiota member that ameliorates DSS-induced colitis in intestinal barrier enhanced Dusp6-deficient mice," *Cell Reports*, 37(8), p. 110016. doi: 10.1016/j.celrep.2021.110016.

Chassaing, B. *et al.* (2015) 'Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome', *Nature*, 519(7541), pp. 92–96. doi: 10.1038/nature14232.

Chassaing, B. *et al.* (2017) 'Dietary emulsifiers directly alter human microbiota composition and gene expression ex vivo potentiating intestinal inflammation', *Gut*, 66(8), pp. 1414–1427. doi: 10.1136/gutjnl-2016-313099.

Chen, L. *et al.* (2017) "NLRP12 attenuates colon inflammation by maintaining colonic microbial diversity and promoting protective commensal bacterial growth," *Nature immunology*, 18(5), pp. 541–551. doi: 10.1038/ni.3690.

Cheng, F. *et al.* (2020) 'Probiotic mixture VSL#3: An overview of basic and clinical studies in chronic diseases', *World J. Clin. Cases*, 8(8), pp. 1361–1385. doi:10.12998/wjcc.v8.i8.1361.

Chermesh, I. *et al.* (2007) 'Failure of Synbiotic 2000 to prevent postoperative recurrence of Crohn's disease', *Dig. Dis. Sci.,* 52(2)*,* pp. 385–389. doi: 10.1007/s10620-006-9549-7.

Chichlowski, M. *et al.* (2008) "Helicobacter typhlonius and Helicobacter rodentium Differentially Affect the Severity of Colon Inflammation and Inflammation-Associated Neoplasia in IL10-Deficient Mice," *Comparative Medicine*, 58(6).

Chudy-Onwugaje, K. O. *et al.* (2019) 'A State-of-the-Art Review of New and Emerging Therapies for the Treatment of IBD', *Inflamm. Bowel Dis.*, 25(5), pp. 820–830. doi:10.1093/ibd/izy327.

Contijoch, E. J. *et al.* (2019) "Gut microbiota density influences host physiology and is shaped by host and microbial factors," *eLife*, 8, p. e40553. doi: 10.7554/eLife.40553.

Cornick, S. *et al.* (2019) "VAMP8-mediated MUC2 mucin exocytosis from colonic goblet cells maintains innate intestinal homeostasis," *Nature communications*, 10(1), p. 4306. doi: 10.1038/s41467-019-11811-8.

Corrêa, R.O., Castro, P.R., Fachi, J.L. *et al.* Inulin diet uncovers complex diet-microbiota-immune cell interactions remodeling the gut epithelium. *Microbiome* **11**, 90 (2023). https://doi.org/10.1186/s40168-023-01520-2

Costello, S. P. *et al.* (2019) 'Effect of Fecal Microbiota Transplantation on 8-Week Remission in Patients With Ulcerative Colitis A Randomized Clinical Trial', *JAMA,* 321(2), pp. 156–164. doi: 10.1001/jama.2018.20046.

Cox, S. R. *et al.* (2020) 'Effects of Low FODMAP Diet on Symptoms, Fecal Microbiome, and Markers of Inflammation in Patients With Quiescent Inflammatory Bowel Disease in a Randomized Trial', *Gastroenterology*., 158(1), pp. 176-188.e7. doi: 10.1053/j.gastro.2019.09.024.

Cresci, G. A. and Bawden, E. (2015) 'Gut microbiome: What we do and don't know', *Nutri. Clin. Pract.*, 30(6), pp. 734–746. doi: 10.1177/0884533615609899.

Cresci, G. A. M. and Izzo, K. (2019) 'Gut Microbiome', in Corrigan, L., Roberts, K., and Steiger, E. (eds.) *Adult Short Bowel Syndrome*, pp. 45–54. doi: 10.1016/b978-0-12-814330-8.00004-4.

Crothers, J. W. *et al.* (2021) 'Daily, oral FMT for long-term maintenance therapy in ulcerative colitis: results of a single-center, prospective, randomized pilot study', *BMC Gastroenterol.*, 21(1). doi: 10.1186/s12876-021-01856-9.

Cummings, J. H. and Macfarlane, G. T. (1991) 'The control and consequences of bacterial fermentation', *J. Appl. Bacteriol*., 70(6), pp. 443–459. doi: 10.1111/j.1365-2672.1991.tb02739.x.

Cummings, J. H. and Macfarlane, G. T. (2002) 'Gastrointestinal effects of prebiotics', *Br. J. Nutr.,* 87(2), pp. S145-S151. doi: 10.1079/BJN/2002530.

D'Haens, G. *et al.* (2022) 'Risankizumab as induction therapy for Crohn's disease: results from the phase 3 ADVANCE and MOTIVATE induction trials', *The Lancet*, 339(10340), pp. 2015–2030.

Cotta, M. A. *et al.* (2009) "Robinsoniella peoriensis gen. nov., sp. nov., isolated from a swine-manure storage pit and a human clinical source - PubMed," *International journal of systematic and evolutionary microbiology*, 59(Pt 1). doi: 10.1099/ijs.0.65676-0.

Dahal, R. H. *et al.* (2023) "Insight into gut dysbiosis of patients with inflammatory bowel disease and ischemic colitis," *Frontiers in Microbiology*, 14. doi: 10.3389/fmicb.2023.1174832.

Daniel, H. *et al.* (2014) 'High-fat diet alters gut microbiota physiology in mice', *ISME J.*, 8(2), pp. 295–308. doi: 10.1038/ismej.2013.155.

Dao, M. C. *et al.* (2016) 'Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: Relationship with gut microbiome richness and ecology', *Gut*, 65(3), pp. 426–436. doi: 10.1136/gutjnl-2014-308778.

Das, B. *et al.* (2010) ‘Tissue distibution of murine Muc19/smgc gene products’, *J. Histochem. Cytochem.*, 58(2), pp. 141–156. doi: 10.1369/jhc.2009.954891.

David, L. A. *et al.* (2014) ‘Diet rapidly and reproducibly alters the human gut microbiome’, *Nature*, 505(7484), pp. 559–563. doi: 10.1038/nature12820.

Dethlefsen, L. *et al.* (2008) ‘The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing', *PLoS Biol.,* 6(11). doi:10.1371/journal.pgen.1000255.

Delday, M. *et al.* (2019) “Bacteroides thetaiotaomicron Ameliorates Colon Inflammation in Preclinical Models of Crohn’s Disease,” *Inflammatory bowel diseases*, 25(1), pp. 85–96. doi: 10.1093/ibd/izy281.

Desai, M. S. *et al.* (2016) “A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility,” *Cell*, 167(5), pp. 1339-1353.e21. doi: 10.1016/j.cell.2016.10.043.

Dethlefsen, L. *et al.* (2008) ‘The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing', *PLoS Biol.,* 6(11). doi:10.1371/journal.pgen.1000255.

Dethlefsen, L. and Relman, D. A. (2011) ‘Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation’, *Proc. Natl. Acad. Sci USA,* 108(Suppl1), pp. S4554-S4561. doi: 10.1073/pnas.1000087107.

Devkota, S. *et al.* (2012) ‘Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice’, *Nature*, 487(7405), pp. 104–108. doi:10.1038/nature11225.

Dingemanse, C. *et al.* (2015) “Akkermansia muciniphila and Helicobacter typhlonius modulate intestinal tumor development in mice - PubMed,” *Carcinogenesis*, 36(11). doi: 10.1093/carcin/bgv120.

Drekonja, D. *et al.* (2015) ‘Fecal Microbiota Transplantation for Clostridium difficile Infection: A Systematic Review’, *Ann. Intern. Med.,* 162(9). pp. 630-638. doi:10.7326/M14-2693.

Duerkop, B. A. *et al.* (2018) “Murine colitis reveals a disease-associated bacteriophage community,” *Nature microbiology*, 3(9), pp. 1023–1031. doi: 10.1038/s41564-018-0210-y.

Dumbrell, A. J., Ferguson, R. M. W. and Clark, D. R. (2016) *Microbial Community Analysis by Single-Amplicon High-Throughput Next Generation Sequencing: Data Analysis – From Raw Output to Ecology*, *Springer Berlin Heidelberg*. Available at: https://link.springer.com/protocol/10.1007/8623_2016_228 (Accessed: July 7, 2023).

Duranti, S. *et al.* (2016) “Elucidating the gut microbiome of ulcerative colitis: bifidobacteria as novel microbial biomarkers - PubMed,” *FEMS microbiology ecology*, 92(12). doi: 10.1093/femsec/fiw191.

Edgar, R. C. *et al.* (2011) “UCHIME improves sensitivity and speed of chimera detection,” *Bioinformatics*, 27(16), pp. 2194–2200. doi: 10.1093/bioinformatics/btr381.

Eeckhaut, V. *et al.* (2013) ‘Butyricicoccus pullicaecorum in inflammatory bowel disease’, *Gut,* 62(12)*,* pp. 1745–1752. doi: 10.1136/gutjnl-2012-303611.

Eindor-Abarbanel, A., Healey, G. R. and Jacobson, K. (2021) ‘Therapeutic Advances in Gut Microbiome Modulation in Patients with Inflammatory Bowel Disease from Pediatrics to Adulthood’, *Int. J. Mol. Sci.*, 22(22). doi: 10.3390/ijms222212506.

Esaiassen, E. *et al.* (2017) “Bifidobacterium Bacteremia: Clinical Characteristics and a Genomic Approach To Assess Pathogenicity,” *Journal of Clinical Microbiology*, 55(7). doi: 10.1128/JCM.00150-17.

Ezeji, J. C. *et al.* (2021) “Parabacteroides distasonis: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health,” *Gut Microbes*, 13(1). doi: 10.1080/19490976.2021.1922241.

Fachi, J. L. *et al.* (2019) “Butyrate Protects Mice from Clostridium difficile-Induced Colitis through an HIF-1-Dependent Mechanism - PubMed,” *Cell reports*, 27(3). doi: 10.1016/j.celrep.2019.03.054.

Fan, Y. *et al.* (2019) “IDDF2019-ABS-0252 Effect of multidonor intensive fecal microbiota transplantation by capsules for active uncreative colitis: a prospective trial,” in *Clinical Gastroenterology*. BMJ Publishing Group Ltd and British Society of Gastroenterology. Available at: http://dx.doi.org/10.1136/gutjnl-2019-iddfabstracts.210 (Accessed: July 7, 2023).

Fan, Y. and Pedersen, O. (2020) “Gut microbiota in human metabolic health and disease,” *Nature Reviews Microbiology*, 19(1), pp. 55–71. doi: 10.1038/s41579-020-0433-9.

Fedorak, R. N. *et al.* (2015) ‘The probiotic VSL#3 Has anti-inflammatory effects and could reduce endoscopic recurrence after surgery for Crohn's disease’, *Clin. Gastroenterol. Hepatol.,* 13(5), pp. 928-935.e2. doi: 10.1016/j.cgh.2014.10.031.

Fernandez-Banares, F. *et al.* (1999) ‘Randomized clinical trial of Plantago ovata seeds (Dietary Fiber) as compared with mesalamine in maintaining remission in ulcerative colitis. Spanish Group for the Study of Crohn's Disease and Ulcerative Colitis (GETECCU)’, *Am. J. Gastroenterol.,* 94(2), pp. 427–433. doi: 10.1111/j.1572-0241.1999.872_a.x.

Filippo, C. De *et al.* (2010) 'Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa', *Proc. Natl. Acad. Sci USA,*107(33), pp. 14691–14696. doi: 10.1073/pnas.1005963107.

Flores, G. E. *et al.* (2014) 'Temporal variability is a personalized feature of the human microbiome', *Genome Biol.*, 15(12), p. 531. doi: 10.1186/s13059-014-0531-y.

Forsyth, C. B. *et al.* (2017) 'Alcohol Feeding in Mice Promotes Colonic Hyperpermeability and Changes in Colonic Organoid Stem Cell Fate',*Alcohol Clin. Exp. Res.*, 41(12), pp. 2100-2113. doi: 10.1111/acer.13519..

Fowler, E. V. *et al.* (2008) 'ATG16L1 T300A shows strong associations with disease subgroups in a large Australian IBD population: further support for significant disease heterogeneity', *Am. J. Gastroenterol.*, 103(10), pp. 2519–2526. doi:10.1111/j.1572-0241.2008.02023.x.

Frank, D. N. *et al.* (2007) "Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases," *Proceedings of the National Academy of Sciences of the United States of America*, 104(34), pp. 13780–5. doi: 10.1073/pnas.0706625104.

Franzosa, E. A. *et al.* (2019) 'Gut microbiome structure and metabolic activity in inflammatory bowel disease', *Nat. Microbiol.*, 4(2), pp. 293–305. doi:10.1038/s41564-018-0306-4.

Freeman, K. *et al.* (2021) 'The incidence and prevalence of inflammatory bowel disease in UK primary care: a retrospective cohort study of the IQVIA Medical Research Database', *BMC Gastroenterol.*, 21(1). doi:10.1186/s12876-021-01716-6.

Frost, G. *et al.* (2014) 'The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism', *Nat. Commun*. doi: 10.1038/ncomms4611.

Gacesa, R. *et al.* (2022) 'Environmental factors shaping the gut microbiome in a Dutch population', *Nature*, 604(7907), pp. 732–739. doi: 10.1038/s41586-022-04567-7.

Ganji-Arjenaki, M. and Rafieian-Kopaei, M. (2018) 'Probiotics are a good choice in remission of inflammatory bowel diseases: A meta analysis and systematic review', *J. Cell. Physiol.*, 233(3), pp. 2091–2103. doi: 10.1002/jcp.25911.

Gerassy-Vainberg, S. *et al.* (2018) "Radiation induces proinflammatory dysbiosis: transmission of inflammatory susceptibility by host cytokine induction - PubMed," *Gut*, 67(1). doi: 10.1136/gutjnl-2017-313789.

Ghosh, N. and Premchand, P. (2015) "A UK cost of care model for inflammatory bowel disease," *Frontline gastroenterology*, 6(3), pp. 169–174. doi: 10.1136/flgastro-2014-100514.

Gilbert, J. A. and Stephens, B. (2018) 'Microbiology of the built environment', *Nat. Rev. Microbiol.*, 16(11), pp. 661–670. doi: 10.1038/s41579-018-0065-5.

Gomez, E. *et al.* (2011) "Isolation of Robinsoniella peoriensis from Four Human Specimens," *Journal of Clinical Microbiology*, 49(1). doi: 10.1128/JCM.01305-10.

Graham, D. B. and Xavier, R. J. (2020) 'Pathway paradigms revealed from the genetics of inflammatory bowel disease', *Nature*, 578(7796), pp. 527–539. doi:10.1038/s41586-020-2025-2.

Grootjans, J. *et al.* (2019) 'Epithelial endoplasmic reticulum stress orchestrates a protective IgA response', *Science*, 363(6430), pp. 993–998. doi:10.1126/science.aat7186.

Gruber, L. *et al.* (2013) 'High Fat Diet Accelerates Pathogenesis of Murine Crohn's Disease-Like Ileitis Independently of Obesity', *PLoS One,* 8(8). doi:10.1371/journal.pone.0071661.

Gryaznova, M. V. *et al.* (2021) "Study of microbiome changes in patients with ulcerative colitis in the Central European part of Russia," *Heliyon*, 7(3), p. e06432. doi: 10.1016/j.heliyon.2021.e06432.

Gu, W. *et al.* (2022) "Dynamic Changes in Gut Microbiome of Ulcerative Colitis: Initial Study from Animal Model," *Journal of Inflammation Research*, Volume 15, pp. 2631–2647. doi: 10.2147/jir.s358807.

Gu, Z. (2022) "Complex heatmap visualization," *iMeta*, 1(3). doi: 10.1002/imt2.43.

Guan, Q. (2019) 'A Comprehensive Review and Update on the Pathogenesis of Inflammatory Bowel Disease', *J. Immunol. Res*. doi:10.1155/2019/7247238.

Guo, A. *et al.* (2022) "High oncostatin M predicts lack of clinical remission for patients with inflammatory bowel disease on tumor necrosis factor α antagonists," *Scientific Reports*, 12(1), pp. 1–8. doi: 10.1038/s41598-022-05208-9.

Hafer, A. *et al.* (2007) 'Effect of oral lactulose on clinical and immunohistochemical parameters in patients with inflammatory bowel disease: a pilot study', *BMC Gastroenterol.,* 7(1). doi: 10.1186/1471-230X-7-36.

Hallert, C. *et al.* (1991) 'Ispaghula Husk May Relieve Gastrointestinal Symptoms in Ulcerative Colitis in Remission Ispaghula Husk May Relieve Gastrointestinal Symptoms in Ulcerative Colitis in Remission', *Scand. J. Gastroenterol.,* 26(7), pp. 747-750. doi:10.3109/00365529108998594.

Hamilton, M. J. *et al.* (2013) 'High-throughput DNA sequence analysis reveals stable engraftment of gut microbiota following transplantation of previously frozen fecal bacteria' *Gut Microbes*, 4(2), pp. 125–135. doi:10.4161/gmic.23571.

Hernández-Chirlaque, C. *et al.* (2016) 'Germ-free and Antibiotic-treated Mice are Highly Susceptible to Epithelial Injury in DSS Colitis', *J. Crohns. Colitis.,* 10(11*),* pp. 1324–1335. doi:10.1093/ecco-jcc/jjw096.

Hernandez-sanabria, E. *et al.* (2020) 'Short-term supplementation of celecoxib-shifted butyrate production on a simulated model of the gut microbial ecosystem and ameliorated in vitro inflammation', *NPJ Biofilms and Microbiomes*., 6(1). doi:10.1038/s41522-020-0119-0.

Herp, S. *et al.* (2019) "Mucispirillum schaedleri Antagonizes Salmonella Virulence to Protect Mice against Colitis - PubMed," *Cell host & microbe*, 25(5). doi: 10.1016/j.chom.2019.03.004.

Hiippala, K. *et al.* (2020) "Novel Odoribacter splanchnicus Strain and Its Outer Membrane Vesicles Exert Immunoregulatory Effects in vitro," *Frontiers in Microbiology*, 11. doi: 10.3389/fmicb.2020.575455.

Hillman, E. T. *et al.* (2017) 'Microbial ecology along the gastrointestinal tract', *Microbes Environ.*, 32(4), pp. 300–313. doi: 10.1264/jsme2.ME17017.

Hills, R. D. *et al.* (2019) 'Gut microbiome: Profound Implications for Diet and Disease', *Nutrients*, 11(7). doi: 10.3390/nu11071613.

Hoshi, N. *et al.* (2012) 'MyD88 signalling in colonic mononuclear phagocytes drives colitis in IL-10 deficient mice', *Nat. Commun*., 3. doi:10.1038/ncomms2113.

Hu, B. *et al.* (2020) "Amyloid-Polyphenol Hybrid Nanofilaments Mitigate Colitis and Regulate Gut Microbial Dysbiosis - PubMed," *ACS nano*, 14(3). doi: 10.1021/acsnano.9b09125.

Huang, H. *et al.* (2017) 'Fine-mapping inflammatory bowel disease loci to single-variant resolution', *Nature*, 547(7662), pp. 173–178. doi:10.1038/nature22969.

Hugot, J. P. *et al.* (2001) 'Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease', *Nature*, 411(6837), pp. 599–603. doi:10.1038/35079107.

Hugot, J. P. *et al.* (2007) 'Prevalence of CARD15/NOD2 mutations in Caucasian healthy people', *Am. J. Gastroenterol.*, 102(6), pp. 1259–1267. doi:10.1111/j.1572-0241.2007.01149.x.

Hung, T. Van and Suzuki, T. (2016) 'Dietary Fermentable Fiber Reduces Intestinal Barrier Defects and Inflammation in Colitic Mice', *J. Nutr.,* 146(10), pp. 1970-1979. doi:10.3945/jn.116.232538.

Huttenhower, C. *et al.* (2012) 'Structure, function and diversity of the healthy human microbiome', *Nature*, 486(7402), pp. 207–214. doi:10.1038/nature11234.

Ihekweazu, F. D. *et al.* (2019) "*Bacteroides ovatus* ATCC 8483 monotherapy is superior to traditional fecal transplant and multi-strain bacteriotherapy in a murine colitis model," *Gut microbes*, 10(4), pp. 504–520. doi: 10.1080/19490976.2018.1560753.

Iljazovic, A. *et al.* (2020a) "Perturbation of the gut microbiome by Prevotella spp. enhances host susceptibility to mucosal inflammation," *Mucosal Immunology*, 14(1), pp. 113–124. doi: 10.1038/s41385-020-0296-4.

Iljazovic, A. *et al.* (2020b) “Perturbation of the gut microbiome by Prevotella spp. enhances host susceptibility to mucosal inflammation,” *Mucosal Immunology*, 14(1), pp. 113–124. doi: 10.1038/s41385-020-0296-4.

Integrative, T. (2014) ‘The integrative human microbiome project: Dynamic analysis of microbiome-host omics profiles during periods of human health and disease corresponding author’, *Cell Host Microbe*, 16(3), pp. 276–289. doi:10.1016/j.chom.2014.08.014.

Iyer, N. (2016) ‘Methods in microbiome research’, *Lab. Anim. (NY)*, 45(9), pp. 323–326. doi:10.1038/laban.1093.

Jackson, M. A. *et al.* (2021) ‘Accurate identification and quantification of commensal microbiota bound by host immunoglobulins’. *Microbiome*, 9(1). doi:10.1186/s40168-020-00992-w.

James, S. L. *et al.* (2015) ‘Abnormal fibre usage in UC in remission’, *Gut*, 64(4), pp. 562–570. doi:10.1136/gutjnl-2014-307198.

Jacobs, J. P. *et al.* (2017) “Microbial, metabolomic, and immunologic dynamics in a relapsing genetic mouse model of colitis induced by T-synthase deficiency,” *Gut microbes*, 8(1), pp. 1–16. doi: 10.1080/19490976.2016.1257469.

Jochum, S. B. *et al.* (2022) “Colonic Epithelial Circadian Disruption Worsens Dextran Sulfate Sodium–Induced Colitis,” *Inflammatory Bowel Diseases*, 29(3), pp. 444–457. doi: 10.1093/ibd/izac219.

Johansson, J. *et al.* (2010) ‘Bacteria Penetrate the Inner Mucus Layer before Inflammation in the Dextran Sulfate Colitis Model’, *PLoS One,* 5(8). doi: 10.1371/journal.pone.0012238.

Johnson, J. S. *et al.* (2019) 'Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis', *Nat. Commun*., 10(1). doi:10.1038/s41467-019-13036-1.

Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files
(Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

Jostins, L. *et al.* (2012) 'Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease', *Nature*, 491(7422), pp. 119–124. doi: 10.1038/nature11582.

Jung, D. H. *et al.* (2021) "An efficient system for intestinal on-site butyrate production using novel microbiome-derived esterases," *Journal of Biological Engineering*, 15(1), pp. 1–12. doi: 10.1186/s13036-021-00259-4.

K., V. P. P. *et al.* (2021) "Western and non-western gut microbiomes reveal new roles of Prevotella in carbohydrate metabolism and mouth–gut axis," *npj Biofilms and Microbiomes*, 7(1), pp. 1–17. doi: 10.1038/s41522-021-00248-x.

Kaakoush, N. O. (2015) "Insights into the Role of Erysipelotrichaceae in the Human Host," *Frontiers in Cellular and Infection Microbiology*, 5. doi: 10.3389/fcimb.2015.00084.

Kabat, A. M. *et al.* (2016) 'The autophagy gene Atg16l1 differentially regulates Treg and TH2 cells to control intestinal inflammation', *Elife*, 5. doi:10.7554/eLife.12444.

Kanauchi, O. *et al.* (2003) 'Treatment of ulcerative colitis patients by long-term administration of germinated barley foodstuff: multi-center open trial', *Int. J. Mol. Med.,* 12(5), pp. 701–704.

Kang, J.-H. *et al.* (2015) "Rapid Origin Determination of the Northern Mauxia Shrimp (Acetes chinensis) Based on Allele Specific Polymerase Chain Reaction of Partial

Mitochondrial 16S rRNA Gene," *Asian-Australasian Journal of Animal Sciences*, 28(4). doi: 10.5713/ajas.14.0613.

Kang, Y. *et al.* (2015) 'Single prokaryotic cell isolation and total transcript amplification protocol for transcriptomic analysis', *Nat. Protoc*., 10(7), pp. 974–984. doi:10.1038/nprot.2015-058.

Kang, Y. *et al.* (2022) 'The Role and Function of Mucins and Its Relationship to Inflammatory Bowel Disease', *Front. Med.* doi:10.3389/fmed.2022.848344.

Kaplan, G. G. *et al.* (2010) 'The inflammatory bowel diseases and ambient air pollution: a novel association'. *Am. J. Gastroenterol.*, 105(11), pp. 2412–2419. doi:10.1038/ajg.2010.252.

Kaser, A. *et al.* (2008) 'XBP1 Links ER Stress to Intestinal Inflammation and Confers Genetic Risk for Human Inflammatory Bowel Disease', *Cell,* 134(5), pp. 743–756. doi:10.1016/j.cell.2008.07.021.

Kazakevych, J. *et al.* (2020) "Smarcad1 mediates microbiota-induced inflammation in mouse and coordinates gene expression in the intestinal epithelium," *Genome biology*, 21(1), p. 64. doi: 10.1186/s13059-020-01976-7.

Khan, S. *et al.* (2020) "Dietary simple sugars alter microbial ecology in the gut and promote colitis in mice - PubMed," *Science translational medicine*, 12(567). doi: 10.1126/scitranslmed.aay6218.

Kim, Y. *et al.* (2020) "Dietary cellulose prevents gut inflammation by modulating lipid metabolism and gut microbiota," *Gut microbes*, 11(4), pp. 944–961. doi: 10.1080/19490976.2020.1730149.

Khoruts, A. *et al.* (2010) ‘Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea', *J. Clin. Gastroenterol.,* 44(5), pp. 354–360. doi: 10.1097/MCG.0b013e3181c87e02.

Kiesler, P., Fuss, I. J. and Strober, W. (2015) ‘Experimental Models of Inflammatory Bowel Diseases’, Cell. Mol. Gastroenterol. Hepatol., 1(2), pp. 154-170. doi:10.1016/j.jcmgh.2015.01.006.

Kim, Y. S. *et al.* (2020) “Sex Differences in Gut Microbiota,” *The World Journal of Men’s Health*, 38(1). doi: 10.5534/wjmh.190009.

Kimura, S. *et al.* (2020) “Osteoprotegerin-dependent M cell self-regulation balances gut infection and immunity,” *Nature communications*, 11(1), p. 234. doi: 10.1038/s41467-019-13883-y.

Kish, L. *et al.* (2013) ‘Environmental Particulate Matter Induces Murine Intestinal Inflammatory Responses and Alters the Gut Microbiome’, *PLoS One*, 8(4). doi:10.1371/journal.pone.0062220.

Knights, D. *et al.* (2014) ‘Rethinking "enterotypes"', *Cell Host Microbe*, 16(4), pp. 433–437. doi:10.1016/j.chom.2014.09.013.

Koeck, D. E. *et al.* (2015) “Herbinix hemicellulosilytica gen. nov., sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor,” *International Journal of Systematic and Evolutionary Microbiology*, 65(Pt_8), pp. 2365–2371. doi: 10.1099/ijs.0.000264.

Kotlarz, D. *et al.* (2012) 'Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy', *Gastroenterology*, 143(2), pp. 347–355. doi:10.1053/j.gastro.2012.04.045.

Kovatcheva-Datchary, P. *et al.* (2015) "Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella - PubMed," *Cell metabolism*, 22(6). doi: 10.1016/j.cmet.2015.10.001.

Kozik, A. J. *et al.* (2019) 'Comparison of the fecal, cecal, and mucus microbiome in male and female mice after TNBS-induced colitis', *PLoS One*, 14(11).

Krueger, C. *et al.* (2022) "Robinsoniella peoriensis: An emerging pathogen and rare cause of wound infection in children," *Journal of the Association of Medical Microbiology and Infectious Disease Canada*, 7(3). doi: 10.3138/jammi-2021-0038.

Kuehbacher, T. *et al.* (2008) 'Intestinal TM7 bacterial phylogenies in active inflammatory bowel disease', *J. Med. Microbiol.,* 57(12), pp. 1569–1576. doi:10.1099/jmm.0.47719-0.

Kühn, R. K. *et al.* (1993) 'Interleukin-10-deficient mice develop chronic enterocolitis', *Cell,* 75(2), pp. 263–274. doi:10.1016/0092-8674(93)80068-p.

Kyo, K. *et al.* (1999) 'Association of ulcerative colitis with rare VNTR alleles of the human intestinal mucin gene, MUC3', *Hum. Mol. Genet.,* 8(2), pp. 307–311. doi:10.1093/hmg/8.2.307.

Kyo, K. *et al.* (2001) 'Associations of distinct variants of the intestinal mucin gene MUC3A with ulcerative colitis and Crohn's disease', *J. Hum. Genet.*, 46(1), pp. 5–20. doi: 10.1007/s100380170118.

Lamb, C. A. *et al.* (2019) 'British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults', *Gut*, 68(Suppl 3), pp. S1–S106. doi: 10.1136/gutjnl-2019-318484.

Laudisi, F. *et al.* (2019) "The Food Additive Maltodextrin Promotes Endoplasmic Reticulum Stress-Driven Mucus Depletion and Exacerbates Intestinal Inflammation," *Cellular and molecular gastroenterology and hepatology*, 7(2), pp. 457–473. doi: 10.1016/j.jcmgh.2018.09.002.

Ledder, O. *et al.* (2014) 'Clinical Patterns and Outcome of Early-Onset Inflammatory Bowel Disease', *J. Pediatr. Gasteroenterol. Mutr.,* 59(5), pp. 562–564. doi:10.1097/MPG.0000000000000465.

Lee, M. and Chang, E. B. (2021) 'Inflammatory Bowel Diseases (IBD) and the Microbiome—Searching the Crime Scene for Clues', *Gastroenterology*, 160(2), pp. 524–537. doi:10.1053/j.gastro.2020.09.056.

Lemire, P. *et al.* (2017) "The NLR Protein NLRP6 Does Not Impact Gut Microbiota Composition - PubMed," *Cell reports*, 21(13). doi: 10.1016/j.celrep.2017.12.026.

Leonardi, I. *et al.* (2018) "CX3CR1^{+} mononuclear phagocytes control immunity to intestinal fungi," *Science (New York, N.Y.)*, 359(6372), pp. 232–236. doi: 10.1126/science.aao1503.

Lesniak, N. A. *et al.* (2022) "The Gut Bacterial Community Potentiates Clostridioides difficile Infection Severity," *mBio*, 13(4). doi: 10.1128/mbio.01183-22.

Levast, B., Li, Z. and Madrenas, J. (2015) 'The role of IL-10 in microbiome-associated immune modulation and disease tolerance', *Cytokine*, 75(2), pp. 291–301. doi: 10.1016/j.cyto.2014.11.027.

Levine, A., Sigall Boneh, R. and Wine, E. (2018) 'Evolving role of diet in the pathogenesis and treatment of inflammatory bowel diseases', *Gut*, 67(9), pp. 1726–1738. doi: 10.1136/gutjnl-2017-315866.

Levy, M. *et al.* (2015) "Microbiota-Modulated Metabolites Shape the Intestinal Microenvironment by Regulating NLRP6 Inflammasome Signaling," *Cell*, 163(6), pp. 1428–43. doi: 10.1016/j.cell.2015.10.048.

Liu, J. Z. *et al.* (2015) 'Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations', *Nat. Genet.*, 47(9), pp. 979–986. doi: 10.1038/ng.3359.

Li, Q. *et al.* (2020) "Fecal Microbiota Transplantation for Ulcerative Colitis: The Optimum Timing and Gut Microbiota as Predictors for Long-Term Clinical Outcomes," *Clinical and Translational Gastroenterology*, 11(8). doi: 10.14309/ctg.0000000000000224.

Li, Q. *et al.* (2022) "Enterobacter ludwigii protects DSS-induced colitis through choline-mediated immune tolerance," *Cell Reports*, 40(9), p. 111308. doi: 10.1016/j.celrep.2022.111308.

Liu, F. *et al.* (2020) "Mechanistic insights into the attenuation of intestinal inflammation and modulation of the gut microbiome by krill oil using in vitro and in vivo models," *Microbiome*, 8(1), p. 83. doi: 10.1186/s40168-020-00843-8.

Liu, M.-J. *et al.* (2022) "Recent findings in Akkermansia muciniphila-regulated metabolism and its role in intestinal diseases," *Clinical Nutrition*, 41(10), pp. 2333–2344. doi: 10.1016/j.clnu.2022.08.029.

Llewellyn, S. R. *et al.* (2018) "Interactions Between Diet and the Intestinal Microbiota Alter Intestinal Permeability and Colitis Severity in Mice," *Gastroenterology*, 154(4), pp. 1037-1046.e2. doi: 10.1053/j.gastro.2017.11.030.

Low, D., Nguyen, D. D. and Mizoguchi, E. (2013) 'Animal models of ulcerative colitis and their application in drug research', *Drug Des. Devel. Ther.*, 7, pp. 1341–1356. doi: 10.2147/DDDT.S40107.

Luo, Y. *et al.* (2022) "Rational consideration of Akkermansia muciniphila targeting intestinal health: advantages and challenges," *npj Biofilms and Microbiomes*, 8(1), pp. 1–11. doi: 10.1038/s41522-022-00338-4.

Lynch, J. B. *et al.* (2023) "Gut microbiota Turicibacter strains differentially modify bile acids and host lipids," *Nature Communications*, 14(1), pp. 1–15. doi: 10.1038/s41467-023-39403-7.

Ma, X., Hua, J. and Li, Z. (2008) 'Probiotics improve high fat diet-induced hepatic steatosis and insulin resistance by increasing hepatic NKT cells', *J. Hepatol.,* 49(5), pp. 821–830. doi: 10.1016/j.jhep.2008.05.025.

Ma, C. *et al.* (2020) "Gasdermin D in macrophages restrains colitis by controlling cGAS-mediated inflammation," *Science advances*, 6(21), p. eaaz6717. doi: 10.1126/sciadv.aaz6717.

Maccaferri, S. *et al.* (2010) 'Rifaximin modulates the colonic microbiota of patients with Crohn's disease: an in vitro approach using a continuous culture colonic model system', J. Antimicrob. Chemother., 65(12), pp. 2556–2565. doi:10.1093/jac/dkq345.

Maclellan, A. *et al.* (2017) 'The Impact of Exclusive Enteral Nutrition (EEN) on the Gut Microbiome in Crohn's Disease: A Review', *Nutrients,* 9(5),. doi:10.3390/nu9050447.

Maloy, K. J. and Powrie, F. (2011) 'Intestinal homeostasis and its breakdown in inflammatory bowel disease', *Nature*, 474(7351), pp. 298–306. doi: 10.1038/nature10208.

Mamantopoulos, M. *et al.* (2017) "Nlrp6- and ASC-Dependent Inflammasomes Do Not Shape the Commensal Gut Microbiota Composition - PubMed," *Immunity*, 47(2). doi: 10.1016/j.immuni.2017.07.011.

Manichanh, C. *et al.* (2006) 'Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach', *Gut*, 55(2), pp. 205–211. doi: 10.1136/gut.2005.073817.

Martinez-Medina, M. *et al.* (2014) 'Western diet induces dysbiosis with increased E. coli in CEABAC10 mice, alters host barrier function favouring AIEC colonisation', *Gut,* 63(1), pp. 116–124. doi: 10.1136/gutjnl-2012-304119.

Martyniak, A. *et al*. (2021) 'Prebiotics, Probiotics, Synbiotics, Paraprobiotics and Postbiotic Compounds in IBD', *Biomolecules*, 11(12). doi: 10.3390/biom11121903.

Maseda, D. and Ricciotti, E. (2020) 'NSAID–Gut Microbiota Interactions', *Front. Pharmacol.*, 11. doi: 10.3389/fphar.2020.01153.

Masella, A. P. *et al.* (2012) "PANDAseq: paired-end assembler for illumina sequences," *BMC Bioinformatics*, 13(1). doi: 10.1186/1471-2105-13-31.

Massimino, L., Lamparelli, L.A., Houshyar, Y. *et al.* The Inflammatory Bowel Disease Transcriptome and Metatranscriptome Meta-Analysis (IBD TaMMA) framework. *Nat Comput Sci* **1**, 511–515 (2021). https://doi.org/10.1038/s43588-021-00114-y

Matsuoka, K. *et al.* (2018) 'Evidence-based clinical practice guidelines for inflammatory bowel disease', *J. Gastroenterol*., 53(3), pp. 305–353. doi: 10.1007/s00535-018-1439-1.

McBurney, M. I. *et al.* (2019) 'Establishing What Constitutes a Healthy Human Gut Microbiome: State of the Science, Regulatory Considerations, and Future Directions', *J. Nutr*., 149(11), pp. 1882–1895. doi: 10.1093/jn/nxz154.

McCole, D. F. (2014) 'IBD candidate genes and intestinal barrier regulation', *Inflammatory Bowel Diseases*, 20(10), pp. 1829–1849. doi:10.1097/MIB.0000000000000090.

McGovern, D. P. B., Kugathasan, S. and Cho, J. H. (2015) 'Genetics of Inflammatory Bowel Diseases', *Gastroenterology*, 149(5), pp. 1163-1176.e2. doi:10.1053/j.gastro.2015.08.001.

McGovern, D. and Powrie, F. (2007) 'The IL23 axis plays a key role in the pathogenesis of IBD', *Gut*, 56(10), pp. 1333–1336. doi: 10.1136/gut.2006.115402.

McMurdie, P. J. and Holmes, S. (2013) "phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data," *PLoS ONE*, 8(4), p. e61217. doi: 10.1371/journal.pone.0061217.

Meisel, M. *et al.* (2017) "Interleukin-15 promotes intestinal dysbiosis with butyrate deficiency associated with increased susceptibility to colitis," *The ISME journal*, 11(1), pp. 15–30. doi: 10.1038/ismej.2016.114.

Methé, B. A. *et al.* (2012) 'A framework for human microbiome research', *Nature*, 486(7402), pp. 215–221. doi: 10.1038/nature11209.

Miranda, P. M. *et al.* (2018) "High salt diet exacerbates colitis in mice by decreasing Lactobacillus levels and butyrate production," *Microbiome*, 6(1), p. 57. doi: 10.1186/s40168-018-0433-4.

Miyazato, S. *et al.* (2016) 'Continuous intake of resistant maltodextrin enhanced intestinal immune response through changes in the intestinal environment in mice', *Biosci. Microbiota Food Health,* 35(1), pp. 1–7. doi: 10.12938/bmfh.2015-009.

Muise, A. M. *et al.* (2009) 'Polymorphisms in E-cadherin (CDH1) result in a mislocalised cytoplasmic protein that is associated with Crohn's disease', *Gut*, 58(8), pp. 1121–1127. doi: 10.1136/gut.2008.175117.

Miyoshi, J. *et al.* (2017) "Peripartum Antibiotics Promote Gut Dysbiosis, Loss of Immune Tolerance, and Inflammatory Bowel Disease in Genetically Prone Offspring," *Cell reports*, 20(2), pp. 491–504. doi: 10.1016/j.celrep.2017.06.060.

Moschen, A. R. *et al.* (2016) "Lipocalin 2 Protects from Inflammation and Tumorigenesis Associated with Gut Microbiota Alterations - PubMed," *Cell host & microbe*, 19(4). doi: 10.1016/j.chom.2016.03.007.

Mutlu, E. A. *et al.* (2011) 'Particulate matter air pollution causes oxidant-mediated increase in gut permeability in mice', *Part. Fibre Toxicol.*, 8. doi: 10.1186/1743-8977-8-19.

Nagao-Kitamoto, H. *et al.* (2020) "Interleukin-22-mediated host glycosylation prevents Clostridioides difficile infection by modulating the metabolic activity of the gut microbiota," *Nature medicine*, 26(4), pp. 608–617. doi: 10.1038/s41591-020-0764-0.

Nair, M. G. *et al.* (2009) 'Alternatively activated macrophage-derived RELM-α is a negative regulator of type 2 inflammation in the lung', *J. Exp. Med.*, 206(4), pp. 1200–1207. doi: 10.1084/jem.20082048.

Nakajima, A. *et al.* (2018) "IgA regulates the composition and metabolic function of gut microbiota by promoting symbiosis between bacteria," *The Journal of experimental medicine*, 215(8), pp. 2019–2034. doi: 10.1084/jem.20180427.

Natividad, J. M. *et al.* (2018) 'Bilophila wadsworthia aggravates high fat diet induced metabolic dysfunctions in mice', *Nat. Commun.*, 9(1). doi: 10.1038/s41467-018-05249-7.

Nearing, J. T. *et al.* (2022) "Microbiome differential abundance methods produce different results across 38 datasets," *Nature Communications*, 13(1), pp. 1–16. doi: 10.1038/s41467-022-28034-z.

Ng, S. C. *et al.* (2013) 'Incidence and phenotype of inflammatory bowel disease based on results from the Asia-pacific Crohn's and colitis epidemiology study', *Gastroenterology*, 145(1), pp. 158-165.e2. doi: 10.1053/j.gastro.2013.04.007.

Ng, S. C. and Hart, A. L. (2014) 'Probiotics in inflammatory bowel disease: Modulation of the pathological immune or inflammatory activation. Their potential benefits in the different phases of the disease.'.

Ni, J. *et al.* (2017) 'Gut microbiota and IBD: Causation or correlation?', *Nat. Rev. Gastroenterol. Hepatol.*, 14(10), pp. 573–584. doi: 10.1038/nrgastro.2017.88.

Nickerson, K. P. and Mcdonald, C. (2012) 'Crohn's disease-associated adherent-invasive Escherichia coli adhesion is enhanced by exposure to the ubiquitous dietary polysaccharide maltodextrin', *PLoS One,* 7(12),. doi: 10.1371/journal.pone.0052132.

Nikolaus, S. *et al.* (2017) 'Increased Tryptophan Metabolism Is Associated With Activity of Inflammatory Bowel Diseases', *Gastroenterology*, 153(6), pp. 1504-1516.e2. doi: 10.1053/j.gastro.2017.08.028.

Nikolenko, S. I., Korobeynikov, A. I. and Alekseyev, M. A. (2013) "BayesHammer: Bayesian clustering for error correction in single-cell sequencing," *BMC genomics*, 14 Suppl 1(Suppl 1), p. S7. doi: 10.1186/1471-2164-14-S1-S7.

Nishida, A. *et al.* (2017) 'Gut microbiota in the pathogenesis of inflammatory bowel disease', *Clin. J.Gastroenterol.*, 11(1). doi: 10.1007/s12328-017-0813-5.

Nomura, M. *et al.* (2020) "Association of Short-Chain Fatty Acids in the Gut Microbiome With Clinical Response to Treatment With Nivolumab or Pembrolizumab in Patients With Solid Cancer Tumors," *JAMA network open*, 3(4), p. e202895. doi: 10.1001/jamanetworkopen.2020.2895.

Ogura, Y. (2001) 'A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease', *Nature*, 411(6837), pp. 603-606. doi: 10.3964/j.issn.1000-0593(2009)02-0418-05.

Okai, S. *et al.* (2016) "High-affinity monoclonal IgA regulates gut microbiota and prevents colitis in mice - PubMed," *Nature microbiology*, 1(9). doi: 10.1038/nmicrobiol.2016.103.

Oksanen, J. *et al.* (no date) *Community Ecology Package [R package vegan version 2.6-4]*, *CRAN - R-Project*. Available at: https://CRAN.R-project.org/package=vegan.

Okumura, R. and Takeda, K. (2018) 'Maintenance of intestinal homeostasis by mucosal barriers', *Inflamm. Regen.,* 38(1). doi: 10.1186/s41232-018-0063-z.

Opstelten, J. L. *et al.* (2016) 'Exposure to Ambient Air Pollution and the Risk of Inflammatory Bowel Disease: A European Nested Case–Control Study', *Dig. Dis. Sci.*, 61(10), pp. 2963–2971. doi: 10.1007/s10620-016-4249-4.

Opstelten, J. L. *et al.* (2018) 'Prediagnostic Serum Vitamin D Levels and the Risk of Crohn's Disease and Ulcerative Colitis in European Populations: A Nested Case-Control Study', *Inflamm. Bowel Dis.*, 24(3), pp. 633–640. doi: 10.1093/ibd/izx050.

Palm, N. W. *et al.* (2014) "Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease," *Cell*, 158(5), pp. 1000–1010. doi: 10.1016/j.cell.2014.08.006.

Palumbo, V. D. *et al.* (2016) 'The long-term effects of probiotics in the therapy of ulcerative colitis: A clinical study', *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech Repub.,* 160(3), pp. 372–377.

Paradis, E. and Schliep, K. (2018) "ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R," *Bioinformatics*, 35(3), pp. 526–528. doi: 10.1093/bioinformatics/bty633.

Paramsothy, S. *et al.* (2017) 'Faecal Microbiota Transplantation for Inflammatory Bowel Disease: A Systematic Review and Meta-analysis', *J. Crohns Colitis*, 11(10), pp. 1180–1199. doi: 10.1093/ecco-jcc/jjx063.

Parker, B. J. *et al.* (2020) "The Genus Alistipes: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health," *Frontiers in Immunology*, 11. doi: 10.3389/fimmu.2020.00906.

Pathak, P., Trilligan, C. and Rapose, A. (2014) "Bifidobacterium--friend or foe? A case of urinary tract infection with Bifidobacterium species," *BMJ case reports*, 2014, p. bcr2014205122. doi: 10.1136/bcr-2014-205122.

Pellizzon, M. (2016) *Choice of laboratory animal diet influences intestinal health*, *Nature*. Available at: https://www.nature.com/articles/laban.1014.

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

Pigneur, B. *et al.* (2019) 'Mucosal Healing and Bacterial Composition in Response to Enteral Nutrition Mucosal Healing and Bacterial Composition in Response to Enteral Nutrition Vs Steroid-based Induction Therapy — A Randomised Prospective Clinical Trial in Children With Crohn's Disease', *J. Crohns Colitis,* 13(7), 846-855. doi: 10.1093/ecco-jcc/jjy207.

Piovani, D. *et al.* (2019) 'Environmental Risk Factors for Inflammatory Bowel Diseases: An Umbrella Review of Meta-analyses', *Gastroenterology*, 157(3), pp. 647-659.e4. doi: 10.1053/j.gastro.2019.04.016.

Powell, N. *et al.* (2015) "Interleukin 6 Increases Production of Cytokines by Colonic Innate Lymphoid Cells in Mice and Patients With Chronic Intestinal Inflammation," *Gastroenterology*, 149(2), pp. 456–67.e15. doi: 10.1053/j.gastro.2015.04.017.

Prescott, S. L. (2017) 'History of medicine: Origin of the term microbiome and why it matters', *Hum. Microbiome J.*, 4(1), pp. 24–25. doi: 10.1016/j.humic.2017.05.004.

Presley, L. L. *et al.* (2010) "Bacteria Associated with Immunoregulatory Cells in Mice," *Applied and Environmental Microbiology*, 76(3), pp. 936–941. doi: 10.1128/aem.01561-09.

Presti, A. L. *et al.* (2019) "Fecal and Mucosal Microbiota Profiling in Irritable Bowel Syndrome and Inflammatory Bowel Disease," *Frontiers in Microbiology*, 10. doi: 10.3389/fmicb.2019.01655.

Prideaux, L. *et al.* (2013) 'Impact of ethnicity, geography, and disease on the microbiota in health and inflammatory bowel disease, *Inflamm. Bowel Dis.,* 19(13), pp. 2906-2918. doi: 10.1097/01.MIB.0000435759.05577.12.

Proctor, L. M. *et al.* (2019) 'The Integrative Human Microbiome Project', *Nature*, 569(7758), pp. 641–648. doi: 10.1038/s41586-019-1238-8.

Prodan, A. *et al.* (2020) "Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing," *PLOS ONE*, 15(1). doi: 10.1371/journal.pone.0227434.

Qi, H. *et al.* (2020) "Reg4 and complement factor D prevent the overgrowth of E. coli in the mouse gut," *Communications biology*, 3(1), p. 483. doi: 10.1038/s42003-020-01219-2.

Queipo-Ortuño, M. I. *et al.* (2012) 'Influence of red wine polyphenols and ethanol on the gut microbiota ecology and biochemical biomarkers', *Am. J. Clin. Nutr.*, 95(6), pp. 1323–1334. doi: 10.3945/ajcn.111.027847.

Rakoff-Nahoum, S., Hao, L. and Medzhitov, R. (2006) 'Role of Toll-like Receptors in Spontaneous Commensal-Dependent Colitis', *Immunity*, 25(2), pp. 319–329. doi: 10.1016/j.immuni.2006.06.010.

Ramakrishnan, S. K. *et al.* (2019) "Intestinal non-canonical NFκB signaling shapes the local and systemic immune response," *Nature communications*, 10(1), p. 660. doi: 10.1038/s41467-019-08581-8.

Rausch, P. *et al.* (2023) "Microbial Dynamics in Newly Diagnosed and Treatment Naïve IBD Patients in the Mediterranean," *Inflammatory Bowel Diseases*, 29(7), pp. 1118–1132. doi: 10.1093/ibd/izad004.

Rawat, P. S. *et al.* (2022) "Hungatella hathewayi, an Efficient Glycosaminoglycan-Degrading *Firmicutes* from Human Gut and Its Chondroitin ABC Exolyase with High Activity and Broad Substrate Specificity," *Applied and environmental microbiology*, 88(22), p. e0154622. doi: 10.1128/aem.01546-22.

Reber, S. O. *et al.* (2016) "Immunization with a heat-killed preparation of the environmental bacterium Mycobacterium vaccae promotes stress resilience in mice," *Proceedings of the National Academy of Sciences of the United States of America*, 113(22), pp. E3130-9. doi: 10.1073/pnas.1600324113.

Reinoso Webb, C. *et al.* (2018) "Differential Susceptibility to T Cell-Induced Colitis in Mice: Role of the Intestinal Microbiota," *Inflammatory Bowel Diseases*, 24(2), pp. 361–379. doi: 10.1093/ibd/izx014.

Ren, J. *et al.* (2020) "Acacetin Ameliorates Experimental Colitis in Mice via Inhibiting Macrophage Inflammatory Response and Regulating the Composition of Gut Microbiota," *Frontiers in Physiology*, 11. doi: 10.3389/fphys.2020.577237.

Rivas, M. A. *et al.* (2011) 'Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease', *Nat. Genet*., 43(11), pp. 1066–1073. doi: 10.1038/ng.952.

Roager, H. M. *et al.* (2016) 'Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut', *Nat. Microbiol*., 1(9). doi: 10.1038/nmicrobiol.2016.93.

Roberts, C. H. *et al.* (2016) "Draft Genome Sequence of Robinsoniella peoriensis 6600698, a Confounder of Clostridium difficile Diagnosis," *Genome announcements*, 4(6), pp. e01275-16. doi: 10.1128/genomeA.01275-16.

Rogers, M. A. M. and Aronoff, D. M. (2016) 'The Influence of Nonsteroidal Anti-Inflammatory Drugs on the Gut Microbiome', *Clin.Microbiol. Infect*., 22(2). doi: 10.1016/j.cmi.2015.10.003.

Rognes, T. *et al.* (2016) “VSEARCH: a versatile open source tool for metagenomics,” *PeerJ*, 4, p. e2584. doi: 10.7717/peerj.2584.

Rothschild, D. *et al.* (2018) ‘Environment dominates over host genetics in shaping human gut microbiota’, *Nature*, 555(7695), pp. 210–215. doi: 10.1038/nature25973.

Roy, T. L. *et al.* (2022) “Dysosmobacter welbionis is a newly isolated human commensal bacterium preventing diet-induced obesity and metabolic disorders in mice,” *Gut*, 71(3), pp. 534–543. doi: 10.1136/gutjnl-2020-323778.

Roy, U. *et al.* (2017) “Distinct Microbial Communities Trigger Colitis Development upon Intestinal Barrier Damage via Innate or Adaptive Immune Cells,” *Cell reports*, 21(4), pp. 994–1008. doi: 10.1016/j.celrep.2017.09.097.

Salim, S. Y., Kaplan, G. G. and Madsen, K. L. (2014) ‘Air pollution effects on the gut microbiota: A link between exposure and inflammatory disease’, *Gut Microbes*, 5(2), pp. 215–219. doi: 10.4161/gmic.27251.

Santiago Badenas, A. *et al.* (2020) “A210 ADLERCREUTZIA IS DEPLETED IN UC PATIENTS, EVEN BEFORE CLINICAL DIAGNOSIS, AND INVERSELY CORRELATES WITH FECAL ELASTOLYTIC ACTIVITY,” *Journal of the Canadian Association of Gastroenterology*, 3(Supplement_1), pp. 82–83. doi: 10.1093/jcag/gwz047.209.

Sarashina-Kida, H. *et al.* (2017) “Gallbladder-derived surfactant protein D regulates gut commensal bacteria for maintaining intestinal homeostasis,” *Proceedings of the National Academy of Sciences of the United States of America*, 114(38), pp. 10178–10183. doi: 10.1073/pnas.1712837114.

Savin, Z. *et al.* (2018) 'Smoking and the intestinal microbiome', *Arch. Microbiol*., 200(5), pp. 677–684. doi: 10.1007/s00203-018-1506-2.

Schaubeck, M. *et al.* (2016) "Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence," *Gut*, 65(2), pp. 225–37. doi: 10.1136/gutjnl-2015-309333.

Schliep, K. *et al.* (2017) "Intertwining phylogenetic trees and networks," *Methods in Ecology and Evolution*, 8(10), pp. 1212–1220. doi: 10.1111/2041-210x.12760.

Schmidt, T. S. B., Raes, J. and Bork, P. (2018) 'The Human Gut Microbiome: From Association to Modulation', *Cell*, 172(6), pp. 1198–1215. doi: 10.1016/j.cell.2018.02.044.

Schmidt, F. *et al.* (2019) "Microbial Colonization in Adulthood Shapes the Intestinal Macrophage Compartment - PubMed," *Journal of Crohn's & colitis*, 13(9). doi: 10.1093/ecco-jcc/jjz036.

Schroeder, B. O. *et al.* (2018) "Bifidobacteria or Fiber Protects against Diet-Induced Microbiota-Mediated Colonic Mucus Deterioration," *Cell host & microbe*, 23(1), pp. 27-40.e7. doi: 10.1016/j.chom.2017.11.004.

Schulfer, A. F. *et al.* (2018) "Intergenerational transfer of antibiotic-perturbed microbiota enhances colitis in susceptible mice," *Nature microbiology*, 3(2), pp. 234–242. doi: 10.1038/s41564-017-0075-5.

Seishima, J. *et al.* (2019) "Gut-derived Enterococcus faecium from ulcerative colitis patients promotes colitis in a genetically susceptible mouse host," *Genome biology*, 20(1), p. 252. doi: 10.1186/s13059-019-1879-9.

Sellon, R. K. *et al.* (1998) 'Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice', *Infect. Immun.*, 66(11), pp. 5224–5231. doi: 10.1128/iai.66.11.5224-5231.1998.

Sender, R., Fuchs, S. and Milo, R. (2016) 'Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans', *Cell*, 164(3), pp. 337–340. doi: 10.1016/j.cell.2016.01.013.

Seregin, S. S. *et al.* (2017) "NLRP6 Protects Il10$^{-/-}$ Mice from Colitis by Limiting Colonization of Akkermansia muciniphila," *Cell reports*, 19(4), pp. 733–745. doi: 10.1016/j.celrep.2017.03.080.

Shadnoush, M. *et al.* (2015) 'Effects of Probiotics on Gut Microbiota in Patients with Inflammatory Bowel Disease: A Double-blind, Placebo-controlled Clinical Trial', *Korean J. Gastroenterol.,* 65(4), pp. 215–221. doi: 10.4166/kjg.2015.65.4.215.

Shahi, S. K., Freedman, S. N. and Mangalam, A. K. (2017) "Gut microbiome in multiple sclerosis: The players involved and the roles they play," *Gut Microbes*, 8(6), pp. 607–615. doi: 10.1080/19490976.2017.1349041.

Shang, L. *et al.* (2021) "Core Altered Microorganisms in Colitis Mouse Model: A Comprehensive Time-Point and Fecal Microbiota Transplantation Analysis," *Antibiotics*, 10(6), p. 643. doi: 10.3390/antibiotics10060643.

Shaw, S. Y., Blanchard, J. F. and Bernstein, C. N. (2010) 'Association Between the Use of Antibiotics in the First Year of Life and Pediatric Inflammatory Bowel Disease', *Am. J. Gastroenterol*., 105(12), pp. 2687-2692. doi: 10.1038/ajg.2010.398.

Shaw, K. A. *et al.* (2016) “Dysbiosis, inflammation, and response to treatment: a longitudinal study of pediatric subjects with newly diagnosed inflammatory bowel disease,” *Genome Medicine*, 8(1), pp. 1–13. doi: 10.1186/s13073-016-0331-y.

Shen, Z.-H. *et al.* (2018) “Relationship between intestinal microbiota and ulcerative colitis: Mechanisms and clinical application of probiotics and fecal microbiota transplantation,” *World Journal of Gastroenterology*, 24(1). doi: 10.3748/wjg.v24.i1.5.

Shorten, A. and Shorten, B. (2013) “What is meta-analysis? - PubMed,” *Evidence-based nursing*, 16(1). doi: 10.1136/eb-2012-101118.

Silveira, A. L. M. *et al.* (2017) ‘Preventive rather than therapeutic treatment with high fiber diet attenuates clinical and inflammatory markers of acute and chronic DSS- induced colitis in mice’, *Eur. J. Nutr*., 56(1), pp. 179-191. doi: 10.1007/s00394-015-1068-x.

Singh, S. *et al.* (2020) “Anti-inflammatory Bifidobacterium strains prevent dextran sodium sulfate induced colitis and associated gut microbial dysbiosis in mice,” *Scientific Reports*, 10(1), pp. 1–14. doi: 10.1038/s41598-020-75702-5.

Sinha, S. R. *et al.* (2020) “Dysbiosis-Induced Secondary Bile Acid Deficiency Promotes Intestinal Inflammation,” *Cell host & microbe*, 27(4), pp. 659-670.e5. doi: 10.1016/j.chom.2020.01.021.

Sokol, H. *et al.* (2020) ‘Fecal microbiota transplantation to maintain remission in Crohn’s disease: a pilot randomized controlled study’. *Microbiome*, 8(1). doi: 10.1186/s40168-020-0792-5.

Song, X. *et al.* (2020) “Microbial bile acid metabolites modulate gut RORγ^{+} regulatory T cell homeostasis,” *Nature*, 577(7790), pp. 410–415. doi: 10.1038/s41586-019-1865-0.

Sood, A. *et al.* (2019) 'Role of Faecal Microbiota Transplantation for Maintenance of Remission in Patients With Ulcerative Colitis: A Pilot Study', *J. Crohns Colitis,* 13(10), pp. 1311–1317. doi: 10.1093/ecco-jcc/jjz060.

Spalinger, M. R. *et al.* (2019) "Protein tyrosine phosphatase non-receptor type 22 modulates colitis in a microbiota-dependent manner," *The Journal of clinical investigation*, 129(6), pp. 2527–2541. doi: 10.1172/JCI123263.

Sprockett, D. *et al.* (2019) "Treatment-Specific Composition of the Gut Microbiota Is Associated With Disease Remission in a Pediatric Crohn's Disease Cohort," *Inflammatory Bowel Diseases*, 25(12). doi: 10.1093/ibd/izz130.

Steed, H. *et al.* (2010) 'Clinical trial: the microbiological and immunological effects of synbiotic consumption - a randomized double-blind placebo-controlled study in active Crohn's disease', *Ailment Pharmacol. Ther.,* 32(7), pp. 872-883. doi: 10.1111/j.1365-2036.2010.04417.x.

Stojanov, S., Berlec, A. and Štrukelj, B. (2020) "The Influence of Probiotics on the Firmicutes/Bacteroidetes Ratio in the Treatment of Obesity and Inflammatory Bowel disease," *Microorganisms*, 8(11), p. 1715. doi: 10.3390/microorganisms8111715.

Su, X.-L. *et al.* (2014) "Acetobacteroides hydrogenigenes gen. nov., sp. nov., an anaerobic hydrogen-producing bacterium in the family Rikenellaceae isolated from a reed swamp - PubMed," *International journal of systematic and evolutionary microbiology*, 64(Pt 9). doi: 10.1099/ijs.0.063917-0.

Sun, J. *et al.* (2016) 'Therapeutic potential to modify the mucus barrier in inflammatory bowel disease', *Nutrients*, 8(1). doi: 10.3390/nu8010044.

Sun, D. *et al.* (2021) “Angiogenin maintains gut microbe homeostasis by balancing α-Proteobacteria and Lachnospiraceae,” *Gut*, 70(4), pp. 666–676. doi: 10.1136/gutjnl-2019-320135.

Sun, L. *et al.* (2023) “Cold stress induces colitis-like phenotypes in mice by altering gut microbiota and metabolites,” *Frontiers in Microbiology*, 14. doi: 10.3389/fmicb.2023.1134246.

Sun, Q. *et al.* (2022) “Enhancing nutritional niche and host defenses by modifying the gut microbiome,” *Molecular Systems Biology*, 18(11). doi: 10.15252/msb.20209933.

Sun, X. *et al.* (2018) “Microbiota-Derived Metabolic Factors Reduce Campylobacteriosis in Mice,” *Gastroenterology*, 154(6), pp. 1751-1763.e2. doi: 10.1053/j.gastro.2018.01.042.

Surana, N. K. and Kasper, D. L. (2017) “Moving beyond microbiome-wide associations to causal microbe identification,” *Nature*, 552(7684). doi: 10.1038/nature25019.

Suzuki, T. and Hara, H. (2010) ‘Dietary fat and bile juice, but not obesity, are responsible for the increase in small intestinal permeability induced through the suppression of tight junction protein expression in LETO and OLETF rats’, *Nutr. Metab.,* 7(1). *doi: 10.1186/1743-7075-7-19.*

Swidsinski, A. *et al.* (2009) ‘Bacterial Overgrowth and Inflammation of Small Intestine After Carboxymethylcellulose Ingestion in Genetically’, *Inflamm. Bowel Dis.,* 15(3), pp. 359–364. doi: 10.1002/ibd.20763.

Tamaki, H. *et al.* (2016) ‘Efficacy of probiotic treatment with Bifidobacterium longum 536 for induction of remission in active ulcerative colitis: A randomized, double-blinded,

placebo-controlled multicenter trial', *Dig. Endosc.,* 28(1), pp. 67–74. doi: 10.1111/den.12553.

Tang, Q. *et al.* (2020) 'Current Sampling Methods for Gut Microbiota: A Call for More Precise Devices', *Front. Cell. Infect. Microbiol.*, 10(1). doi:10.3389/fcimb.2020.00151.

Tannock, G. W. *et al.* (2000) 'Analysis of the Fecal Microflora of Human Subjects Consuming a Probiotic Product Containing Lactobacillus rhamnosus DR20'. *Appl. Environ. Microbiol.,* 66(6), pp. 2578-2588. doi: 10.1128/AEM.66.6.2578-2588.2000.

Tao, Z. *et al.* (2021) "Xi Lei San Attenuates Dextran Sulfate Sodium-Induced Colitis in Rats and TNF-α-Stimulated Colitis in CACO2 Cells: Involvement of the NLRP3 Inflammasome and Autophagy," *Mediators of inflammation*, 2021, p. 1610251. doi: 10.1155/2021/1610251.

Thomas, J. P. *et al.* (2022) "The Emerging Role of Bile Acids in the Pathogenesis of Inflammatory Bowel Disease," *Frontiers in Immunology*, 13. doi: 10.3389/fimmu.2022.829525.

Tiffany, C. R. *et al.* (2021) "The metabolic footprint of Clostridia and Erysipelotrichia reveals their role in depleting sugar alcohols in the cecum," *Microbiome*, 9(1), pp. 1–13. doi: 10.1186/s40168-021-01123-9.

Toubai, T. *et al.* (2019) "Host NLRP6 exacerbates graft-versus-host disease independent of gut microbial composition," *Nature microbiology*, 4(5), pp. 800–812. doi: 10.1038/s41564-019-0373-1.

Tran, H. Q. *et al.* (2019) "Flagellin-elicited adaptive immunity suppresses flagellated microbiota and vaccinates against chronic inflammatory diseases," *Nature communications*, 10(1), p. 5650. doi: 10.1038/s41467-019-13538-y.

Turnbaugh, P. J. *et al.* (2007) 'The Human Microbiome Project', *Nature*, 449(7164), pp. 804–810. doi: 10.1038/nature06244.

Turner, D. *et al.* (2019) 'Antibiotic Cocktail for Pediatric Acute Severe Colitis and the Microbiome : The PRASCO Randomized Controlled Trial', *Inflamm. Bowel Dis.*, 26(11), pp. 1733-1742. doi: 10.1093/ibd/izz298.

Turner, J. R. (2009) 'Intestinal mucosal barrier function in health and disease', *Nat. Rev. Immunol.,* 9(11), pp. 799–809. doi: 10.1038/nri2653.

Ungaro, R. *et al.* (2014) 'Antibiotics associated with increased risk of new-onset Crohn's disease but not ulcerative colitis: a meta-analysis', *Am. J. Gastroenterol*., 109(11), pp. 1728-1738. doi: 10.1038/ajg.2014.246.

Vacca, M. *et al.* (2020) "The Controversial Role of Human Gut Lachnospiraceae," *Microorganisms*, 8(4). doi: 10.3390/microorganisms8040573.

Vaga, S. *et al.* (2020) 'Compositional and functional differences of the mucosal microbiota along the intestine of healthy individuals', *Sci. Rep.*, 10(1). doi: 10.1038/s41598-020-71939-2.

Vaishnava, S. *et al.* (2011) 'The antibacterial lectin RegIIIγ promotes the spatial segregation of microbiota and host in the intestine', *Science*, 334(6053), pp. 255–258. doi: 10.1126/science.1209791.

Valdes, A. M. *et al.* (2018) 'Role of the gut microbiota in nutrition and health', *BMJ*, 361. doi: 10.1136/bmj.k2179.

Van den Abbeele, P. *et al.* (2012) "Incorporating a mucosal environment in a dynamic gut model results in a more representative colonization by lactobacilli," *Microbial biotechnology*, 5(1), pp. 106–15. doi: 10.1111/j.1751-7915.2011.00308.x.

Vandeputte, D. *et al.* (2016) 'Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates', *Gut*, 65(1), pp. 57–62. doi: 10.1136/gutjnl-2015-309618.

Van Sommeren, S. *et al.* (2011) 'HNF4α and CDH1 are associated with ulcerative colitis in a Dutch cohort', *Inflamm. Bowel Dis.*, 17(8), pp. 1714–1718. doi:10.1002/ibd.21541.

Venegas, D. P. *et al.* (2019) 'Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases', *Front. Immunol.* doi: 10.3389/fimmu.2019.00277.

Viennois, E. *et al.* (2019) "Host-derived fecal microRNAs can indicate gut microbiota healthiness and ability to induce inflammation," *Theranostics*, 9(15), pp. 4542–4557. doi: 10.7150/thno.35282.

Viladomiu, M. *et al.* (2017) "IgA-coated *E. coli* enriched in Crohn's disease spondyloarthritis promote T_H17-dependent inflammation," *Science translational medicine*, 9(376), p. eaaf9655. doi: 10.1126/scitranslmed.aaf9655.

Visschedijk, M. C. *et al.* (2016) 'Pooled resequencing of 122 ulcerative colitis genes in a large Dutch cohort suggests population-Specific associations of rare variants in MUC2', *PLoS One*, 11(8). doi: 10.1371/journal.pone.0159609.

Wan, Z. *et al.* (2022) "Intermediate role of gut microbiota in vitamin B nutrition and its influences on human health," *Frontiers in Nutrition*, 9. doi: 10.3389/fnut.2022.1031502.

Wang, H. *et al.* (2016) 'Dietary Non-digestible Polysaccharides Ameliorate Intestinal Epithelial Barrier Dysfunction in IL-10 Knockout Mice', *J. Crohns Colitis,* 10(9), pp. 1076–1086. doi: 10.1093/ecco-jcc/jjw065.

Wang, J. *et al.* (2018) 'Meta-analysis of human genome-microbiome association studies: The MiBioGen consortium initiative', *Microbiome*, 6(1). doi: 10.1186/s40168-018-0479-3.

Wang, L. *et al.* (2020) "A purified membrane protein from *Akkermansia muciniphila* or the pasteurised bacterium blunts colitis associated tumourigenesis by modulation of $CD8^+$ T cells in mice," *Gut*, 69(11), pp. 1988–1997. doi: 10.1136/gutjnl-2019-320105.

Wang, Y. and Kasper, L. H. (2014) "The role of microbiome in central nervous system disorders," *Brain, behavior, and immunity*, 38, pp. 1–12. doi: 10.1016/j.bbi.2013.12.015.

Webb, C. R. *et al.* (2018) 'Differential Susceptibility to T Cell-Induced Colitis in Mice : Role of the Intestinal Microbiota', *Inflamm. Bowel Dis.,* 24(2), pp. 361–379. doi: 10.1093/ibd/izx014.

Weingarden, A. R. and Vaughn, B. P. (2017) 'Intestinal microbiota, fecal microbiota transplantation, and inflammatory bowel disease', *Gut Microbes*, 8(3), pp. 238–252. doi: 10.1080/19490976.2017.1290757.

Weinstock, G. M. (2012) "Genomic approaches to studying the human microbiota," *Nature*, 489(7415), pp. 250–6. doi: 10.1038/nature11553.

Wexler, H. M. (2007) "Bacteroides: the Good, the Bad, and the Nitty-Gritty," *Clinical Microbiology Reviews*, 20(4). doi: 10.1128/CMR.00008-07.

Wheeler, M. L. *et al.* (2016) "Immunological Consequences of Intestinal Fungal Dysbiosis," *Cell host & microbe*, 19(6), pp. 865–73. doi: 10.1016/j.chom.2016.05.003.

Wilmanski, T. *et al.* (2021) 'Gut microbiome pattern reflects healthy ageing and predicts survival in humans', *Nat. Metab.*, 3(1), pp. 274-286. doi: 10.1038/s42255-021-00348-0.

Wirtz, S. and Neurath, M. F. (2007) 'Mouse models of inflammatory bowel disease ☆', *Adv. Drug Deliv. Rev.,* 59(11), pp. 1073–1083. doi: 10.1016/j.addr.2007.07.003.

Wohlgemuth, S. *et al.* (2011) "Intestinal steroid profiles and microbiota composition in colitic mice - PubMed," *Gut microbes*, 2(3). doi: 10.4161/gmic.2.3.16104.

Wright, E. S. (2020) "RNAconTest: comparing tools for noncoding RNA multiple sequence alignment based on structural consistency," *RNA*, 26(5), pp. 531–540. doi: 10.1261/rna.073015.119.

Wu, G. D. *et al.* (2011) 'Linking long-term dietary patterns with gut microbial enterotypes', *Science,* 334(6052), pp. 105–108. doi: 10.1126/science.1208344.

Xavier, R. J. and Podolsky, D. K. (2007) 'Unravelling the pathogenesis of inflammatory bowel disease', *Nature*, 448(7152), pp. 427–434. doi: 10.1038/nature06005.

Xiao, L. *et al.* (2021) 'Deterministic transition of enterotypes shapes the infant gut microbiome at an early age', *Genome Biol*., 22(1). doi: 10.1186/s13059-021-02463-3.

Xie, M. *et al.* (2022) "Effects of arabinoxylan and chlorogenic acid on the intestinal microbiota in dextran sulfate sodium–treated mice," *Frontiers in Nutrition*, 9. doi: 10.3389/fnut.2022.950446.

Yang, J.-Y. *et al.* (2016) "Enteric Viruses Ameliorate Gut Inflammation via Toll-like Receptor 3 and Toll-like Receptor 7-Mediated Interferon-β Production - PubMed," *Immunity*, 44(4). doi: 10.1016/j.immuni.2016.03.009.

Yang, Y. *et al.* (2020) “Cross-talk between the gut microbiota and monocyte-like macrophages mediates an inflammatory response to promote colitis-associated tumourigenesis,” *Gut*, 70(8), p. gutjnl-2020-320777. doi: 10.1136/gutjnl-2020-320777.

Yang, Z. *et al.* (2020) ‘Fecal Microbiota Transplant via Endoscopic Delivering Through Small Intestine and Colon: No Difference for Crohn's Disease’, *Dig. Dis. Sci*., 65(1), pp. 150-157. doi: 10.1007/s10620-019-05751-y.

Yao, M. *et al.* (2021) “Improved functionality of Ligilactobacillus salivarius Li01 in alleviating colonic inflammation by layer-by-layer microencapsulation,” *npj Biofilms and Microbiomes*, 7(1). doi: 10.1038/s41522-021-00228-1.

Yatsunenko, T. *et al.* (2012) ‘Human gut microbiome viewed across age and geography’, *Nature*, 486(7402), pp. 222–227. doi: 10.1038/nature11053.

Yimagou, E. K. *et al.* (2019) “Prevotella marseillensis sp. nov., a new bacterium isolated from a patient with recurrent Clostridium difficile infection,” *New Microbes and New Infections*, 32. doi: 10.1016/j.nmni.2019.100606.

Yoshimatsu, Y. *et al.* (2015) ‘Effectiveness of probiotic therapy for the prevention of relapse in patients with inactive ulcerative colitis’, *World J. Gastroenterol.,* 21(19), pp. 5985–5994. doi: 10.3748/wjg.v21.i19.5985.

Young, R. B. *et al.* (2021) ‘Key Technologies for Progressing Discovery of Microbiome-Based Medicines’, *Front. Microbiol.,* 12(1). doi: 10.3389/fmicb.2021.685935.

Younis, N., Zarif, R. and Mahfouz, R. (2020) ‘Inflammatory bowel disease: between genetics and microbiota’, *Mol. Biol. Rep.,* 47(4), pp. 3053–3063. doi: 10.1007/s11033-020-05318-5.

Zagato, E. *et al.* (2020) “Endogenous murine microbiota member Faecalibaculum rodentium and its human homologue protect from intestinal tumour growth,” *Nature Microbiology*, 5(3), pp. 511–524. doi: 10.1038/s41564-019-0649-5.

Zhang, D. *et al.* (2022) “Interactions between polysaccharides and gut microbiota: A metabolomic and microbial review,” *Food Research International*, 160, p. 111653. doi: 10.1016/j.foodres.2022.111653.

Zhang, Y. S. *et al.* (2019) “STAT4 activation by leukemia inhibitory factor confers a therapeutic effect on intestinal inflammation,” *The EMBO journal*, 38(6), p. e99595. doi: 10.15252/embj.201899595.

Zhang, Z. H. *et al.* (2014) “A comparative study of techniques for differential expression analysis on RNA-Seq data,” *PloS one*, 9(8), p. e103207. doi: 10.1371/journal.pone.0103207.

Zhang, X.-F. *et al.* (2021) ‘Clinical effects and gut microbiota changes of using probiotics, prebiotics or synbiotics in inflammatory bowel disease: a systematic review and meta-analysis’, *Eur. J. Nutr.,* 60(5), pp. 2855-2875. doi: 10.1007/s00394-021-02503-5.

Zhao, M. *et al.* (2018) “Deficiency in intestinal epithelial O-GlcNAcylation predisposes to gut inflammation,” *EMBO molecular medicine*, 10(8), p. e8736. doi: 10.15252/emmm.201708736.

Zhu, W. *et al.* (2018) “Precision editing of the gut microbiota ameliorates colitis,” *Nature*, 553(7687), pp. 208–211. doi: 10.1038/nature25172.

Zhu, W. *et al.* (2019) “Editing of the gut microbiota reduces carcinogenesis in mouse models of colitis-associated colorectal cancer,” *The Journal of experimental medicine*, 216(10), pp. 2378–2393. doi: 10.1084/jem.20181939.

Zimmermann, M. *et al.* (2019) ‘Mapping human microbiome drug metabolism by gut bacteria and their genes’, *Nature*, 570(7762), pp. 462–467. doi: 10.1038/s41586-019-1291-3.

Zocco, M. A. *et al.* (2006) “Efficacy of Lactobacillus GG in maintaining remission of ulcerative colitis - PubMed,” *Alimentary pharmacology & therapeutics*, 23(11). doi: 10.1111/j.1365-2036.2006.02927.x.

Zuo, T. *et al.* (2018) ‘va’, *Nat. Rev. Gastroenterol. Hepatol.*, 15(7), pp. 440–452. doi: 10.1038/s41575-018-0003-z.