Gut microbiome insights into the pathophysiology of inflammatory bowel disease

Gina Lucille Guzzo

Department of Ecology and Evolutionary Biology
School of Biological Sciences
Faculty of Sciences
University of Adelaide

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

August 2022

I acknowledge the Kaurna people, the original custodians of the Adelaide Plains and the land on which I worked and lived during my PhD. I extend my respects to Elders past, present, and emerging. I acknowledge that sovereignty was never ceded.

Table of Contents

Thesis Abstract

The global prevalence of inflammatory bowel disease (IBD) has been steadily rising since the turn of the century. IBD is a chronic gastrointestinal disease with no cure, and it therefore has a significant lifelong burden for people with the disease. The causal mechanisms of IBD are not fully understood, and genetic factors do not entirely explain its development. One factor, the gut microbiome (i.e., the microorganisms that inhabit the gut together with their genes and metabolic products), has been implicated in IBD, and studying it offers an opportunity to further our understanding of how the disease develops and persists. Individuals with IBD have gut microbiomes that look different from healthy individuals, a phenomenon often described as 'dysbiosis'. It has been difficult to define specifically what dysbiosis means for people with the disease as there is wide variation between study designs, while also having methodological limitations and oversights. Despite this, therapies for IBD are being developed around altering the gut microbiome. An encouraging treatment option is faecal microbiota transplantation (FMT), wherein a gut microbiome sample (stool) from a healthy donor is transferred to a recipient with IBD. The efficacy of FMT is currently comparable to other therapeutics for IBD, and continued research of the gut microbiome may lead to improvements in FMT.

This thesis takes an interdisciplinary approach to studying the gut microbiome in IBD to offer insights into the pathophysiology of the disease. Most gut microbiome studies of IBD have centred around bacteria, and this thesis provides an alternative perspective by highlighting the necessity of including nonbacterial gut microbes—namely fungi, protozoa, and viruses—in IBD research. This point is then demonstrated with the use of computational and statistical methods to show that intestinal fungi and protozoa have an altered distribution in IBD. Through these methods, it is also shown how nonbacterial microbes can be used to improve what is known of FMT. In addition to the inclusion of nonbacterial microbes, longitudinal studies are necessary to improve the efficacy of FMT by revealing the microbial changes that lead to remission. These studies are more accessible when flexible sampling types are used to enable participants to collect their own samples. The comparability of an alternative sample type (dry stool swabs) collected by participants is evaluated in this thesis against its reference sample type (whole stool) in a longitudinal study of FMT for IBD. Lastly, this thesis provides

a guide to designing longitudinal microbiome studies in clinical and public health research that is built on critical statistical considerations.

Overall, this thesis synthesises insights from microbiome research with clinical and biostatistical expertise to bridge the gap between basic and translational science. The knowledge presented in this thesis paves the way for advancements and alternative approaches to gut microbiome research that will ultimately improve outcomes for individuals with the disease.

Thesis Declaration

I, Gina L. Guzzo, certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Gina Guzzo

21 August 2022

Acknowledgements

I thank the participants, researchers, clinicians, and all others involved in the research presented in this thesis. My sincerest hope is that this work will improve the lives of those with IBD in a meaningful way.

To my principal supervisor—Laura—you've pushed me to a level of independence and research ability that I never imagined I could have. I'm incredibly grateful for your ability to keep me on track through all the highs and lows, and for somehow leaving me with an even bigger passion for microbiome research than when I began my PhD. To my co-supervisors—Murthy, Bastien, and Jane—thank you for allowing me the opportunity to share in each of your expertise and passions. It has been a privilege to learn from the best.

A huge 'thank you' goes to my vibrant network of peers and friends of the Metagenomics Team and ACAD more broadly who have supported, motivated, and inspired me along the way. The ACAD Nature Appreciation Society—Matilda, Caitlin M, Evelyn, Raph, Robbi, Xavi, and Gludi—gets a special mention for all the hikes, bicycle rides, boardgames, and slacklining adventures that helped keep me laughing even at my most stressful times.

Thank you to the Plum Chums—Matilda, Evelyn, Caitlin M, Caitlin S, and Felicia—and to the Microbiome Support Group—Matilda and Caitlin S—for being there to share all our PhD and life experiences. I'm incredibly inspired by each of you intelligent and strong women.

A special thank you to Evelyn, my PhD partner-in-crime. I'm so lucky to have found you when I did. You kept me sane throughout the process, with all the laughs and goofy times, but also for your shoulder to cry on. I can't wait to see what's next for us both!

To my family—mom, dad, and Chris—thank you for sending so much love and support from afar to keep me going. Mom, I know it's been difficult having your only child all the way on the other side of the world for so many years. I appreciate all you've sacrificed to get me here. A 'thank you' to my Australian parents as well—Anne and Robert—for always being there for me when my Canadian family couldn't be.

Vin, a 'thank you' is just not enough for the unconditional love and support you have shown me over these years. I really feel like I would not have made it through without you. For all the second-hand emotions you endured, this really is a 'we' project rather than my own. The good news is—we did it!!!

Thesis Introduction

The gut microbiome in human health

The human body is occupied by trillions of microorganisms that form micro-ecosystems in several of its anatomical regions: the gastrointestinal tract, mouth, vagina, and skin constitute major sites of microbial colonisation (Turnbaugh et al. 2007). These microbes carry out various metabolic functions, including some that are necessary for our survival such as vitamin K synthesis in the gut (Ellis et al. 2021). They are also intimately connected with the immune system; human-associated microbes can defend against invasive pathogens and help 'educate' the immune system to discriminate between these pathogens and beneficial, commensal microorganisms (Wang et al. 2017). The microorganisms that colonise the human body include bacteria, archaea, fungi, protozoa, and viruses (Turnbaugh et al. 2007; Nash et al. 2017; Tisza & Buck 2021). Coupled with their metabolic products, genes, and surrounding micro-environment, these are known as the 'human microbiome' (Marchesi & Ravel 2015; Berg et al. 2020). The microbiome has been coined the 'forgotten organ' because of its late discovery, and studying it provides key opportunities to advance our understanding of human health (O'Hara & Shanahan 2006).

We can now more fully appreciate the human microbiome thanks to developments in high-throughput methods, including amplicon sequencing and shotgun metagenomic sequencing. Amplicon sequencing is a targeted approach that amplifies common marker genes found in microbes, such as the 16S ribosomal RNA (rRNA) gene found in bacteria and archaea and 18S rRNA genes and internal transcribed spacer (ITS-1 and ITS-2) genes found in fungi and protozoa (Amaral-Zettler et al. 2009; Stoeck et al. 2010; Caporaso et al. 2012). This approach gives a picture of the taxonomic makeup of a microbial community. Shotgun metagenomic sequencing is indiscriminate; it does not rely on marker genes when sequencing the DNA of samples (Sharpton 2014). Therefore, shotgun sequencing can capture genes that tell us about the functional information of a community. Both methods allow us to characterise a plethora of microbial species, instead of being limited to culturing a few at a time (Young et al. 2021).

Our ability to describe the vastness of the human microbiome over the past few decades has fundamentally changed how the human body is viewed in medical research. Though estimates are still being revised, it is understood that the human body hosts roughly as many microbial cells as human cells (Sender, Fuchs & Milo 2016). More than half of these microbial cells and their genetic content are harboured in the gastrointestinal tract (i.e., the 'gut microbiome') where they perform important functions, such as digest food, metabolise drugs, synthesise vitamins, and interact with the immune system (Ramotar et al. 1984;

McMahon 2002; Hooper, Littman & Macpherson 2012). Given its intimate relationship with the body, it should come as no surprise that deviations to the gut microbiome have been implicated in various pathologies (Wang et al. 2017). In fact, it is now difficult to find a condition that has not been associated with the gut microbiome (Kho & Lal 2018). Multifactorial diseases have especially benefited from this research, and the gut microbiome is considered another 'missing link' to understand the nongenetic contributors to disease (Kho & Lal 2018; Vijay & Valdes 2022). Changes to the gut microbiome have been associated with many of these conditions ranging from metabolic disorders, (e.g., obesity and heart disease), mental disorders (e.g., depression and schizophrenia), autoimmune conditions (e.g., multiple sclerosis and systemic lupus erythematosus), and immune-mediated (e.g., inflammatory bowel disease and allergic conditions) (Hevia et al. 2014; Boulangé et al. 2016; Rothhammer & Quintana 2016; Pascal et al. 2018; Glassner, Abraham & Quigley 2020; Limbana, Khan & Eskander 2020; Szeligowski et al. 2020; Witkowski, Weeks & Hazen 2020).

Introduction to inflammatory bowel disease

The incentive to map out environmental factors that contribute to these diseases is greater than ever; the global incidence of multifactorial, noncommunicable diseases is on the rise as the incidence of infectious diseases declines (Bach 2018). One such disease, inflammatory bowel disease (IBD), has been steadily increasing in cases across the world, with developed nations having the highest burden of IBD (Ng et al. 2018; Alatab et al. 2020). In Australia, a 2013 report projected that between 80,000 to 100,000 people would be diagnosed with the disease by the current year, 2022 (PwC 2013). IBD is a blanket term for several subtypes; the two main diseases are Crohn's disease (CD) and ulcerative colitis (UC) (Podolsky 2002). CD constitutes inflammatory symptoms anywhere along the gastrointestinal tract, from the mouth to anus, whereas UC is restricted to the colon. The inflammation in UC is also restricted to the mucosa, but in CD, inflammation can exist across the entire organ wall (i.e., transmurally) (Xavier & Podolsky 2007). Both subtypes are characterised by a chronic, relapsing-remitting pattern of disease. That is, patients may go through periods of increased disease activity (flare) followed by a reduction or resolution of symptoms (Liverani et al. 2016). Long-term remission can be achieved in some cases, but there is no definitive cure for the disease.

While stakeholder reports highlight the financial burden of IBD (at least $3 billion AUD per annum in Australia) (PwC 2013), this is far from the only burden to consider. As there is no cure, IBD confers an impaired quality of life and high morbidity for patients (Andrews et al. 2010). The main symptom of UC is frequent bloody diarrhoea which is often coupled with abdominal pain and urgency to defecate (Carter, Lobo & Travis 2004; GESA 2018). The symptoms of CD vary due its heterogeneity, but can include diarrhoea, pain, general malaise, and serious complications such as intestinal obstruction (Carter, Lobo & Travis 2004; GESA 2018). In Australia, IBD is commonly diagnosed in adults by the age of 30, but 10-20% of cases are diagnosed during childhood (Grover, De Nardi & Lewindon 2017). A large proportion of newly diagnosed individuals are thus young and spend many of their productive years managing their chronic disease (GESA 2018). Current treatment options sit at a moderate level of efficacy to induce remission and, with no guarantee of lifelong remission, can leave patients with a worry of relapse (Cai, Wang & Li 2021). Therefore, the development of therapies with improved efficacies is of great importance.

The link between inflammatory bowel disease and the gut microbiome

The difficulty of finding effective treatments for IBD stems from its multifactorial nature. At the molecular level, the characteristic inflammation of IBD results from a dysfunctional immune response that aberrantly targets microorganisms in the intestinal tract (Khor, Gardet & Xavier 2011; Xu et al. 2014). However, it is not fully understood why this immune activity arises. There is a genetic component to IBD and over 200 genetic risk loci have been identified, but none of these are fully penetrant (Gordon et al. 2015; Cleynen et al. 2016; Uniken Venema et al. 2017). Additionally, studies of twins have shown that identical twins are more likely to develop the same subtype of the disease than fraternal twins (Halfvarson et al. 2006). However, not all twin pairs develop the disease, illustrating the contribution of nongenetic factors to its aetiology (Halfvarson et al. 2006; Halme et al. 2006). The global incidence of IBD is also rising at a level that cannot solely be explained by genetics, particularly in developing nations and regions which have seen a rapid increase in IBD (Zuo et al. 2018).

A prominent theory on the cause and persistence of IBD is the gut microbiome (Glassner, Abraham & Quigley 2020). Many of the genetic risk loci are found in or adjacent to immune genes associated with microbial interactions (Cleynen et al. 2016). Other supporting evidence

includes how colitis does not spontaneously develop in germ-free mice when they are kept in germ-free conditions, and that faecal stream diversion can induce remission in CD (Harper et al. 1985; Rutgeerts et al. 1991; Sellon et al. 1998; Veltkamp et al. 2001; Gkouskou et al. 2014). A key indication lies in the fact that industrialisation (the process of social and economic change that occurs when a society shifts from agrarian to industrial life) is associated with a greater risk of IBD; developing regions of the world are seeing much higher incidences of IBD than neighbouring rural areas (Kaplan & Ng 2016; Ng et al. 2018). In addition to IBD, industrialisation has been linked to an increase in several immune-mediated diseases including type 1 diabetes, multiple sclerosis, and allergic diseases (Patterson et al. 2014; Fogarty 2015; Noorimotlagh et al. 2021).

A leading hypothesis to explain this phenomenon is that an industrialised lifestyle causes changes to the microbiome which then leads to immune dysfunction (Bach 2018; Zuo et al. 2018). Industrialisation is associated with reduced diversity of the gut microbiome (De Filippo et al. 2010; Schnorr et al. 2014), and several lifestyle factors linked to industrialisation have both been associated with a reduction in microbial diversity and independently confer a risk of IBD, including antibiotic use during childhood, a diet high in saturated fat and refined sugar and low in dietary fibre, and exposure to pollution (De Filippo et al. 2010; Shaw, Blanchard & Bernstein 2010; Kronman et al. 2012; Ng et al. 2015; Jin et al. 2017; Lewis & Abreu 2017). Further, Al Nabhani et al. (2019) showed that, in mice, the period between birth and weaning, when solid foods are gradually introduced, is a critical window of immune development through early exposure to microbes. This exposure triggers an immune reaction, deemed the ‘weaning reaction’, which was protective against development of murine allergy, colitis, and colorectal cancer. This hypothesis suggests that, without a ‘weaning reaction’ the immune becomes dysregulated and results in disease due to environmental triggers later in life.

Alternatively, the gut microbiome observed in IBD may be a consequence of disease. It has been hypothesised that chronic intestinal inflammation causes increased levels of oxygen in the gut (Rigottier-Gois 2013). The availability of oxygen in the gut affects the distribution of microbial species, and increased oxygen would support a shift towards facultative anaerobes from the dominant obligate anaerobes that normally live in a healthy gut. This scenario could support interventions that limit oxygen in the affected intestinal region to shift the microbiome toward obligate anaerobes. Overall, understanding the gut microbiome changes

that occur in IBD and the potential environmental triggers of the disease is crucial to understand how it can be prevented and treated.

i. Findings of the gut microbiome in IBD

The overarching finding of the gut microbiome in IBD is that it is reduced in diversity from unaffected individuals, often referred to as 'dysbiosis' of the gut microbiome (Glassner, Abraham & Quigley 2020; Aldars-García, Chaparro & Gisbert 2021). This reduced diversity has been measured in different ways across high-throughput sequencing studies, including within-sample diversity (alpha diversity) (e.g., Imhann et al. 2018), between-sample diversity (beta diversity) (e.g., Schirmer et al. 2018), and specific microbial taxa such as a decrease in *Roseburia* (Chen et al. 2014; Laserna-Mendieta et al. 2017) and *Faecalibacterium prausnitzii* (Laserna-Mendieta et al. 2017; Zakrzewski et al. 2018), and an increase in *Bifidobacterium* species (Mar et al. 2016; Takahashi et al. 2016; Nishino et al. 2018). However, these findings are not always consistent. For example, *Roseburia* has been observed as being both increased and decreased in UC (Chen et al. 2014; Nishino et al. 2018). It has also been difficult to ascertain whether observed differences are a cause of the disease, an effect, or both (Glassner, Abraham & Quigley 2020).

ii. Current limitations of microbiome studies

It has been difficult to pin down a precise description of 'dysbiosis' in IBD because studies vary wildly in their designs. A scan through systematic reviews will reveal the plethora of studies that have already been conducted on this topic, and the variation in their methods and findings (Pittayanon et al. 2020; Aldars-García, Chaparro & Gisbert 2021). For example, sampling the microbiome can be performed with different methods, the two most common being stool sampling and mucosal biopsy, and these methods capture a different microbial makeup from each other (Bassis et al. 2017; Vandeputte et al. 2017a). There is also variability in the cohorts that have been sampled; geographic location, treatment status (e.g., naïve versus not naïve to treatment), sample size, and various other participant demographics and lifestyle factors differ between studies and can all affect the gut microbiome (Pittayanon et al. 2020; Vujkovic-Cvijin et al. 2020; Aldars-García, Chaparro & Gisbert 2021; Wilkinson et al. 2021). A lack of quantitative approaches in microbiome research hinders our ability to accurately compare microbiomes between individuals, as sequencing results give relative

abundance measures and the true microbial cell load (i.e., 'microbial load') can vary greatly between individuals (Vandeputte et al. 2017b). Integrating microbial load via cell counting with sequencing results can overcome this issue (Vandeputte et al. 2017b), but this method has not been widely adopted, likely due to resource limitations. Thus, the comparability between existing studies is limited and has likely prevented us from uncovering broad trends.

The nascency of gut microbiome research means that previous studies have been limited by methodological constraints. Most high-throughput sequencing studies of the microbiome are conducted with 16s rRNA sequencing because of its affordability. As sequencing costs have decreased, more shotgun metagenomic studies are now appearing in the literature (Pittayanon et al. 2020; Aldars-García, Chaparro & Gisbert 2021). This method allows more precise species- and strain-level identification across more species which may increase the comparability of studies and give better functional insight into the microbiome in IBD (Fang et al. 2018). Additionally, shotgun sequencing permits the survey of different types of microorganisms. Most studies in IBD to date have focussed on bacteria as they are the most abundant microorganism of the gastrointestinal tract, while leaving out other less common, but potentially important, microbial types such as fungi, protozoa, and viruses (Sartor & Wu 2017; Richard & Sokol 2019). Shotgun sequencing facilitates the identification of bacteria while also surveying other microbes in one sampling and sequencing effort. The reduced cost of sequencing has also made it easier to conduct studies with greater sampling sizes, thereby increasing the ability of researchers to conduct different study designs (Minich et al. 2018; Wilkinson et al. 2021). Cross-sectional studies have dominated the literature of IBD and the gut microbiome, as they are easier and cheaper to conduct than longitudinal studies (Aldars-García, Chaparro & Gisbert 2021). However, longitudinal studies are needed to grasp the temporal dynamics of the microbiome to elucidate important trends in IBD, such as how the microbiome changes with the development of disease or increased disease activity (Halfvarson et al. 2017). These studies will help us determine the causal role of the gut microbiome.

iii. The gut microbiome as a therapy for IBD

Even though the role of the gut microbiome in IBD is still not fully understood, researchers and clinicians are trialling therapies dependent on altering the patient's microbiome to treat the disease (Sharma et al. 2020). Current first-line treatments for IBD include

immunomodulators, immunosuppressants, and biologics, but as previously stated, their efficacy is far from ideal (Girardin et al. 2012; GESA 2018). Additionally, these therapies come with undesirable side effects (Quezada, McLean & Cross 2018). As most therapies for IBD rely on targeting the immune response, there has been interest in expanding therapeutics into new territories. One area of active research is faecal microbiota transplantation (FMT), which is the transplantation of a microbiome sample, in the form of stool, from a healthy donor to a recipient with IBD (Sadowsky & Khoruts 2016). The purpose of FMT is to induce remission by restoring the recipient's microbiome to a healthy state and it relies on the assumption that healthy individuals have microbiomes with anti-inflammatory properties. FMT has mostly been trialled in UC with moderate success rates (Costello et al. 2017; Caldeira et al. 2020), though one trial for CD has also been published (Sokol et al. 2020). A Cochrane library systematic review of trials of FMT in UC found that the therapy may induce remission in some recipients, but more high-quality studies are needed before its efficacy can be confirmed (Imdad et al. 2018). Studies of FMT often include analyses to characterise the microbiomes of donor samples and recipients, and they are plagued by similar study design and methodological limitations outlined above.

The efficacy of FMT can likely be improved as the field of microbiome research progresses towards new, improved, and standardised methods and studies. Our ability to refine FMT relies on us understanding several characteristics of the microbiome, including what constitutes a healthy donor microbiome that is beneficial for transplantation, and how the patient microbiome interacts with the transplanted microbiome. As the cost of high-throughput sequencing continues to decrease, this interaction can be measured with higher resolution, repeated sampling in the same individuals over the course of clinical trials.

Thesis overview

In this thesis, I bring together multidisciplinary methods of studying the microbiome to advance our understanding of the pathophysiology of inflammatory bowel disease (IBD). As intestinal bacteria are commonly studied in the context of IBD, I highlight the need to incorporate other, rarer intestinal microbes, such as eukaryotes and viruses, into this research. I use bioinformatic and statistical approaches to identify eukaryotes in the gut microbiomes of individuals with IBD and associate the identified microbes with patient factors. I also explore longitudinal gut microbiome samples from a clinical trial of faecal microbiota transplantation (FMT) for individuals with IBD. Lastly, I build on the limitations of the previous chapter and provide recommendations and considerations for designing longitudinal gut microbiome studies in clinical and public health research.

Chapter 1

The neglected gut microbiome: fungi, protozoa, and bacteriophages in inflammatory bowel disease

Bacteria are most commonly studied in gut microbiome research of IBD due to their high abundance over other microbes. Researchers often claim that 'dysbiosis' is present in the gut microbiome of individuals with IBD when only the intestinal bacteria have been studied. This published review summarises what is known of nonbacterial microbes that are often neglected in IBD research and how they can be studied with high-throughput DNA sequencing method. The aim of this review is to encourage further exploration into this understudied area and reminds researchers to consider the comprehensiveness of gut microbiome studies before making sweeping conclusions on the behaviour of the microbiome in IBD.

Chapter 2

Individuals with inflammatory bowel disease have an altered gut microbiome composition of fungi and protozoa

This chapter advances the existing knowledge of the nonbacterial microbiome in IBD summarised in Chapter 1. I use two bioinformatic tools outlined in the previous chapter to

detect eukaryotes in published gut microbiome samples of people with IBD and compare these findings to people without IBD. I also detect eukaryotes in longitudinal gut microbiome samples from individuals with IBD who received FMT and their FMT donors. I show that eukaryotes of the gut microbiome differ in IBD from healthy individuals and vary over time.

Chapter 3

The utility of dry swabs over stool samples: a case study from a faecal microbiota transplant trial

FMT is a promising therapeutic option for IBD, but success rates are currently moderate due to a lack of understanding of the underlying mechanisms that instigate remission. One option to increase our understanding is through longitudinal sampling of the FMT recipient's gut microbiome. This sampling could allow us to witness the daily or weekly changes that lead to remission. However, the typical method for gut microbiome sampling, whole stool collection, is not always feasible at such a high frequency of collection. Here, I assess an alternative to whole stool—dry swabs of stool—to determine whether they are an appropriate proxy to whole stool collection in an FMT trial for IBD.

Chapter 4

Recommendations and considerations for longitudinal studies of the human microbiome

Longitudinal studies are crucial in clinical research to account for the time-dependent variation of the microbiome in patients. The longitudinal study design of Chapter 3 suffered from several limitations, which I build upon in this chapter to provide a guide for researchers and clinicians intending to embark on a longitudinal microbiome study. I approach this from two angles: considerations for study design and available statistical methods for clinical research. I emphasise how researchers should recognise the most resource-efficient and appropriate way to conduct a longitudinal study to answer their questions of interest. This chapter is presented with multidisciplinary insights to be accessible to individual with backgrounds from basic and clinical science, and biostatistics.

References

Al Nabhani, Z, Dulauroy, S, Marques, R, Cousu, C, Al Bounny, S, Déjardin, F, Sparwasser, T, Bérard, M, Cerf-Bensussan, N & Eberl, G 2019, 'A Weaning Reaction to Microbiota Is Required for Resistance to Immunopathologies in the Adult', *Immunity*, vol. 50, no. 5, May 21, pp. 1276-1288.e1275.

Alatab, S, Sepanlou, SG, Ikuta, K, Vahedi, H, Bisignano, C, Safiri, S, Sadeghi, A, Nixon, MR, Abdoli, A, Abolhassani, H, Alipour, V, Almadi, MAH, Almasi-Hashiani, A, Anushiravani, A, Arabloo, J, Atique, S, Awasthi, A, Badawi, A, Baig, AAA, Bhala, N, Bijani, A, Biondi, A, Borzì, AM, Burke, KE, Carvalho, F, Daryani, A, Dubey, M, Eftekhari, A, Fernandes, E, Fernandes, JC, Fischer, F, Haj-Mirzaian, A, Haj-Mirzaian, A, Hasanzadeh, A, Hashemian, M, Hay, SI, Hoang, CL, Househ, M, Ilesanmi, OS, Jafari Balalami, N, James, SL, Kengne, AP, Malekzadeh, MM, Merat, S, Meretoja, TJ, Mestrovic, T, Mirrakhimov, EM, Mirzaei, H, Mohammad, KA, Mokdad, AH, Monasta, L, Negoi, I, Nguyen, TH, Nguyen, CT, Pourshams, A, Poustchi, H, Rabiee, M, Rabiee, N, Ramezanzadeh, K, Rawaf, DL, Rawaf, S, Rezaei, N, Robinson, SR, Ronfani, L, Saxena, S, Sepehrimanesh, M, Shaikh, MA, Sharafi, Z, Sharif, M, Siabani, S, Sima, AR, Singh, JA, Soheili, A, Sotoudehmanesh, R, Suleria, HAR, Tesfay, BE, Tran, B, Tsoi, D, Vacante, M, Wondmieneh, AB, Zarghi, A, Zhang, Z-J, Dirac, M, Malekzadeh, R & Naghavi, M 2020, 'The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017', *The Lancet Gastroenterology & Hepatology*, vol. 5, no. 1, pp. 17-30.

Aldars-García, L, Chaparro, M & Gisbert, JP 2021, 'Systematic Review: The Gut Microbiome and Its Potential Clinical Application in Inflammatory Bowel Disease', *Microorganisms*, vol. 9, no. 5, Apr 30.

Amaral-Zettler, LA, McCliment, EA, Ducklow, HW & Huse, SM 2009, 'A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes', *PLoS One*, vol. 4, no. 7, p. e6372.

Andrews, JM, Mountifield, RE, Van Langenberg, DR, Bampton, PA & Holtmann, GJ 2010, 'Un-promoted issues in inflammatory bowel disease: opportunities to optimize care', *Intern Med J*, vol. 40, no. 3, Mar, pp. 173-182.

Bach, JF 2018, 'The hygiene hypothesis in autoimmunity: the role of pathogens and commensals', *Nat Rev Immunol*, vol. 18, no. 2, Feb, pp. 105-120.

Bassis, CM, Moore, NM, Lolans, K, Seekatz, AM, Weinstein, RA, Young, VB, Hayden, MK & for the, CDCPEP 2017, 'Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles', *BMC Microbiology*, vol. 17, no. 1, 2017/03/31, p. 78.

Berg, G, Rybakova, D, Fischer, D, Cernava, T, Vergès, M-CC, Charles, T, Chen, X, Cocolin, L, Eversole, K, Corral, GH, Kazou, M, Kinkel, L, Lange, L, Lima, N, Loy, A, Macklin, JA, Maguin, E, Mauchline, T, McClure, R, Mitter, B, Ryan, M, Sarand, I, Smidt, H, Schelkle, B, Roume, H, Kiran, GS, Selvin, J, Souza, RSCd, van Overbeek, L, Singh, BK, Wagner, M, Walsh, A, Sessitsch, A & Schloter, M 2020, 'Microbiome definition re-visited: old concepts and new challenges', *Microbiome*, vol. 8, no. 1, 2020/06/30, p. 103.

Boulangé, CL, Neves, AL, Chilloux, J, Nicholson, JK & Dumas, M-E 2016, 'Impact of the gut microbiota on inflammation, obesity, and metabolic disease', *Genome Medicine*, vol. 8, no. 1, 2016/04/20, p. 42.

Cai, Z, Wang, S & Li, J 2021, 'Treatment of Inflammatory Bowel Disease: A Comprehensive Review', *Frontiers in Medicine*, vol. 8, 2021-December-20.

Caldeira, LdF, Borba, HH, Tonin, FS, Wiens, A, Fernandez-Llimos, F & Pontarolo, R 2020, 'Fecal microbiota transplantation in inflammatory bowel disease patients: A systematic review and meta-analysis', *PLoS One*, vol. 15, no. 9, pp. e0238910-e0238910.

Caporaso, JG, Lauber, CL, Walters, WA, Berg-Lyons, D, Huntley, J, Fierer, N, Owens, SM, Betley, J, Fraser, L, Bauer, M, Gormley, N, Gilbert, JA, Smith, G & Knight, R 2012, 'Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms', *The ISME Journal*, vol. 6, no. 8, 2012/08/01, pp. 1621-1624.

Carter, MJ, Lobo, AJ & Travis, SPL 2004, 'Guidelines for the management of inflammatory bowel disease in adults', *Gut*, vol. 53, no. suppl 5, pp. v1-v16.

Chen, L, Wang, W, Zhou, R, Ng, SC, Li, J, Huang, M, Zhou, F, Wang, X, Shen, B, M, AK, Wu, K & Xia, B 2014, 'Characteristics of fecal and mucosa-associated microbiota in Chinese patients with inflammatory bowel disease', *Medicine (Baltimore)*, vol. 93, no. 8, Aug, p. e51.

Cleynen, I, Boucher, G, Jostins, L, Schumm, LP, Zeissig, S, Ahmad, T, Andersen, V, Andrews, JM, Annese, V, Brand, S, Brant, SR, Cho, JH, Daly, MJ, Dubinsky, M, Duerr, RH, Ferguson, LR, Franke, A, Gearry, RB, Goyette, P, Hakonarson, H, Halfvarson, J, Hov, JR, Huang, H, Kennedy, NA, Kupcinskas, L, Lawrance, IC, Lee, JC, Satsangi, J, Schreiber, S, Theatre, E, van der Meulen-de Jong, AE, Weersma, RK, Wilson, DC, Parkes, M, Vermeire, S, Rioux, JD, Mansfield, J, Silverberg, MS, Radford-Smith, G, McGovern, DP, Barrett, JC & Lees, CW 2016, 'Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study', *Lancet*, vol. 387, no. 10014, Jan 9, pp. 156-167.

Costello, SP, Soo, W, Bryant, RV, Jairath, V, Hart, AL & Andrews, JM 2017, 'Systematic review with meta-analysis: faecal microbiota transplantation for the induction of remission for active ulcerative colitis', *Aliment Pharmacol Ther*, vol. 46, no. 3, Aug, pp. 213-224.

De Filippo, C, Cavalieri, D, Di Paola, M, Ramazzotti, M, Poullet, JB, Massart, S, Collini, S, Pieraccini, G & Lionetti, P 2010, 'Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa', *Proceedings of the National Academy of Sciences*, vol. 107, no. 33, pp. 14691-14696.

Ellis, JL, Karl, JP, Oliverio, AM, Fu, X, Soares, JW, Wolfe, BE, Hernandez, CJ, Mason, JB & Booth, SL 2021, 'Dietary vitamin K is remodeled by gut microbiota and influences community composition', *Gut Microbes*, vol. 13, no. 1, Jan-Dec, pp. 1-16.

Fang, X, Monk, JM, Nurk, S, Akseshina, M, Zhu, Q, Gemmell, C, Gianetto-Hill, C, Leung, N, Szubin, R, Sanders, J, Beck, PL, Li, W, Sandborn, WJ, Gray-Owen, SD, Knight, R, Allen-Vercoe, E, Palsson, BO & Smarr, L 2018, 'Metagenomics-Based, Strain-Level Analysis of Escherichia coli From a Time-Series of Microbiome Samples From a Crohn's Disease Patient', *Front Microbiol*, vol. 9, 2018-October-30.

Fogarty, AW 2015, 'What have studies of non-industrialized countries told us about the cause of allergic disease?', *Clin Exp Allergy*, vol. 45, no. 1, Jan, pp. 87-93.

GESA 2018, *Australian Guidelines for General Practitioners and Physicians: Inflammatory Bowel Disease*, The Gastroenterological Society of Australia (GESA), Melbourne, Victoria, Australia.

Girardin, M, Manz, M, Manser, C, Biedermann, L, Wanner, R, Frei, P, Safroneeva, E, Mottet, C, Rogler, G & Schoepfer, AM 2012, 'First-line therapies in inflammatory bowel disease', *Digestion*, vol. 86 Suppl 1, pp. 6-10.

Gkouskou, K, Deligianni, C, Tsatsanis, C & Eliopoulos, A 2014, 'The gut microbiota in mouse models of inflammatory bowel disease', *Frontiers in cellular and infection microbiology*, vol. 4, 2014-February-28.

Glassner, KL, Abraham, BP & Quigley, EMM 2020, 'The microbiome and inflammatory bowel disease', *Journal of Allergy and Clinical Immunology*, vol. 145, no. 1, pp. 16-27.

Gordon, H, Trier Moller, F, Andersen, V & Harbord, M 2015, 'Heritability in inflammatory bowel disease: from the first twin study to genome-wide association studies', *Inflamm Bowel Dis*, vol. 21, no. 6, Jun, pp. 1428-1434.

Grover, Z, De Nardi, A & Lewindon, PJ 2017, 'Inflammatory bowel disease in adolescents', *Aust Fam Physician*, vol. 46, no. 8, pp. 565-571.

Halfvarson, J, Brislawn, CJ, Lamendella, R, Vázquez-Baeza, Y, Walters, WA, Bramer, LM, D'Amato, M, Bonfiglio, F, McDonald, D, Gonzalez, A, McClure, EE, Dunklebarger, MF, Knight, R & Jansson, JK 2017, 'Dynamics of the human gut microbiome in inflammatory bowel disease', *Nat Microbiol*, vol. 2, Feb 13, p. 17004.

Halfvarson, J, Jess, T, Magnuson, A, Montgomery, SM, Orholm, M, Tysk, C, Binder, V & Järnerot, G 2006, 'Environmental factors in inflammatory bowel disease: a co-twin control study of a Swedish-Danish twin population', *Inflamm Bowel Dis*, vol. 12, no. 10, Oct, pp. 925-933.

Halme, L, Paavola-Sakki, P, Turunen, U, Lappalainen, M, Farkkila, M & Kontula, K 2006, 'Family and twin studies in inflammatory bowel disease', *World J Gastroenterol*, vol. 12, no. 23, Jun 21, pp. 3668-3672.

Harper, PH, Lee, EC, Kettlewell, MG, Bennett, MK & Jewell, DP 1985, 'Role of the faecal stream in the maintenance of Crohn's colitis', *Gut*, vol. 26, no. 3, Mar, pp. 279-284.

Hevia, A, Milani, C, López, P, Cuervo, A, Arboleya, S, Duranti, S, Turroni, F, González, S, Suárez, A, Gueimonde, M, Ventura, M, Sánchez, B & Margolles, A 2014, 'Intestinal dysbiosis associated with systemic lupus erythematosus', *MBio*, vol. 5, no. 5, Sep 30, pp. e01548-01514.

Hooper, LV, Littman, DR & Macpherson, AJ 2012, 'Interactions between the microbiota and the immune system', *Science*, vol. 336, no. 6086, Jun 8, pp. 1268-1273.

Imdad, A, Nicholson, MR, Tanner-Smith, EE, Zackular, JP, Gomez-Duarte, OG, Beaulieu, DB & Acra, S 2018, 'Fecal transplantation for treatment of inflammatory bowel disease', *Cochrane Database Syst Rev*, vol. 11, no. 11, Nov 13, p. Cd012774.

Imhann, F, Vich Vila, A, Bonder, MJ, Fu, J, Gevers, D, Visschedijk, MC, Spekhorst, LM, Alberts, R, Franke, L, van Dullemen, HM, Ter Steege, RWF, Huttenhower, C, Dijkstra, G, Xavier, RJ, Festen, EAM, Wijmenga, C, Zhernakova, A & Weersma, RK 2018, 'Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease', *Gut*, vol. 67, no. 1, Jan, pp. 108-119.

Jin, Y, Wu, S, Zeng, Z & Fu, Z 2017, 'Effects of environmental pollutants on gut microbiota', *Environmental Pollution*, vol. 222, 2017/03/01/, pp. 1-9.

Kaplan, GG & Ng, SC 2016, 'Globalisation of inflammatory bowel disease: perspectives from the evolution of inflammatory bowel disease in the UK and China', *Lancet Gastroenterol Hepatol*, vol. 1, no. 4, Dec, pp. 307-316.

Kho, ZY & Lal, SK 2018, 'The Human Gut Microbiome – A Potential Controller of Wellness and Disease', *Front Microbiol*, vol. 9, 2018-August-14.

Khor, B, Gardet, A & Xavier, RJ 2011, 'Genetics and pathogenesis of inflammatory bowel disease', *Nature*, vol. 474, no. 7351, Jun 15, pp. 307-317.

Kronman, MP, Zaoutis, TE, Haynes, K, Feng, R & Coffin, SE 2012, 'Antibiotic exposure and IBD development among children: a population-based cohort study', *Pediatrics*, vol. 130, no. 4, Oct, pp. e794-803.

Laserna-Mendieta, EJ, Clooney, AG, Carretero-Gomez, JF, Moran, C, Sheehan, D, Nolan, JA, Hill, C, Gahan, CGM, Joyce, SA, Shanahan, F & Claesson, MJ 2017, 'Determinants of Reduced Genetic Capacity for Butyrate Synthesis by the Gut Microbiome in Crohn's Disease and Ulcerative Colitis', *Journal of Crohn's and Colitis*, vol. 12, no. 2, pp. 204-216.

Lewis, JD & Abreu, MT 2017, 'Diet as a Trigger or Therapy for Inflammatory Bowel Diseases', *Gastroenterology*, vol. 152, no. 2, Feb, pp. 398-414.e396.

Limbana, T, Khan, F & Eskander, N 2020, 'Gut Microbiome and Depression: How Microbes Affect the Way We Think', *Cureus*, vol. 12, no. 8, pp. e9966-e9966.

Liverani, E, Scaioli, E, Digby, RJ, Bellanova, M & Belluzzi, A 2016, 'How to predict clinical relapse in inflammatory bowel disease patients', *World J Gastroenterol*, vol. 22, no. 3, pp. 1017-1033.

Mar, JS, LaMere, BJ, Lin, DL, Levan, S, Nazareth, M, Mahadevan, U, Lynch, SV & Blaser, MJ 2016, 'Disease Severity and Immune Activity Relate to Distinct Interkingdom Gut Microbiome States in Ethnically Distinct Ulcerative Colitis Patients', *MBio*, vol. 7, no. 4, pp. e01072-01016.

Marchesi, JR & Ravel, J 2015, 'The vocabulary of microbiome research: a proposal', *Microbiome*, vol. 3, pp. 31-31.

McMahon, RJ 2002, 'Biotin in metabolism and molecular biology', *Annu Rev Nutr*, vol. 22, pp. 221-239.

Minich, JJ, Humphrey, G, Benitez, RAS, Sanders, J, Swafford, A, Allen, EE & Knight, R 2018, 'High-Throughput Miniaturized 16S rRNA Amplicon Library Preparation Reduces Costs while Preserving Microbiome Integrity', *mSystems*, vol. 3, no. 6, Nov-Dec, pp. e00166-00118.

Nash, AK, Auchtung, TA, Wong, MC, Smith, DP, Gesell, JR, Ross, MC, Stewart, CJ, Metcalf, GA, Muzny, DM, Gibbs, RA, Ajami, NJ & Petrosino, JF 2017, 'The gut mycobiome of the Human Microbiome Project healthy cohort', *Microbiome*, vol. 5, no. 1, Nov 25, p. 153.

Ng, SC, Shi, HY, Hamidi, N, Underwood, FE, Tang, W, Benchimol, EI, Panaccione, R, Ghosh, S, Wu, JCY, Chan, FKL, Sung, JJY & Kaplan, GG 2018, 'Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies', *Lancet*, vol. 390, no. 10114, Dec 23, pp. 2769-2778.

Ng, SC, Tang, W, Leong, RW, Chen, M, Ko, Y, Studd, C, Niewiadomski, O, Bell, S, Kamm, MA, de Silva, HJ, Kasturiratne, A, Senanayake, YU, Ooi, CJ, Ling, KL, Ong, D, Goh, KL, Hilmi, I, Ouyang, Q, Wang, YF, Hu, P, Zhu, Z, Zeng, Z, Wu, K, Wang, X, Xia, B, Li, J, Pisespongsa, P, Manatsathit, S, Aniwan, S, Simadibrata, M, Abdullah, M, Tsang, SW, Wong, TC, Hui, AJ, Chow, CM, Yu, HH, Li, MF, Ng, KK, Ching, J, Wu, JC, Chan, FK & Sung, JJ 2015, 'Environmental risk factors in inflammatory bowel disease: a population-based case-control study in Asia-Pacific', *Gut*, vol. 64, no. 7, Jul, pp. 1063-1071.

Nishino, K, Nishida, A, Inoue, R, Kawada, Y, Ohno, M, Sakai, S, Inatomi, O, Bamba, S, Sugimoto, M, Kawahara, M, Naito, Y & Andoh, A 2018, 'Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease', *Journal of Gastroenterology*, vol. 53, no. 1, 2018/01/01, pp. 95-106.

Noorimotlagh, Z, Azizi, M, Pan, H-F, Mami, S & Mirzaee, SA 2021, 'Association between air pollution and Multiple Sclerosis: A systematic review', *Environmental Research*, vol. 196, 2021/05/01/, p. 110386.

O'Hara, AM & Shanahan, F 2006, 'The gut flora as a forgotten organ', *EMBO reports*, vol. 7, no. 7, pp. 688-693.

Pascal, M, Perez-Gordo, M, Caballero, T, Escribese, MM, Lopez Longo, MN, Luengo, O, Manso, L, Matheu, V, Seoane, E, Zamorano, M, Labrador, M & Mayorga, C 2018, 'Microbiome and Allergic Diseases', *Frontiers in immunology*, vol. 9, 2018-July-17.

Patterson, C, Guariguata, L, Dahlquist, G, Soltész, G, Ogle, G & Silink, M 2014, 'Diabetes in the young - a global view and worldwide estimates of numbers of children with type 1 diabetes', *Diabetes Res Clin Pract*, vol. 103, no. 2, Feb, pp. 161-175.

Pittayanon, R, Lau, JT, Leontiadis, GI, Tse, F, Yuan, Y, Surette, M & Moayyedi, P 2020, 'Differences in Gut Microbiota in Patients With vs Without Inflammatory Bowel Diseases: A Systematic Review', *Gastroenterology*, vol. 158, no. 4, Mar, pp. 930-946.e931.

Podolsky, DK 2002, 'Inflammatory bowel disease', *N Engl J Med*, vol. 347, no. 6, Aug 8, pp. 417-429.

PwC 2013, Improving inflammatory bowel disease care across Australia, PA (PwC).

Quezada, SM, McLean, LP & Cross, RK 2018, 'Adverse events in IBD therapy: the 2018 update', *Expert Rev Gastroenterol Hepatol*, vol. 12, no. 12, Dec, pp. 1183-1191.

Ramotar, K, Conly, JM, Chubb, H & Louie, TJ 1984, 'Production of menaquinones by intestinal anaerobes', *J Infect Dis*, vol. 150, no. 2, Aug, pp. 213-218.

Richard, ML & Sokol, H 2019, 'The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases', *Nature Reviews Gastroenterology & Hepatology*, vol. 16, no. 6, Jun, pp. 331-345.

Rigottier-Gois, L 2013, 'Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis', *Isme j*, vol. 7, no. 7, Jul, pp. 1256-1261.

Rothhammer, V & Quintana, FJ 2016, 'Environmental control of autoimmune inflammation in the central nervous system', *Curr Opin Immunol*, vol. 43, Dec, pp. 46-53.

Rutgeerts, P, Goboes, K, Peeters, M, Hiele, M, Penninckx, F, Aerts, R, Kerremans, R & Vantrappen, G 1991, 'Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum', *Lancet*, vol. 338, no. 8770, Sep 28, pp. 771-774.

Sadowsky, MJ & Khoruts, A 2016, 'Faecal microbiota transplantation is promising but not a panacea', *Nat Microbiol*, vol. 1, no. 3, 2016/02/24, p. 16015.

Sartor, RB & Wu, GD 2017, 'Roles for Intestinal Bacteria, Viruses, and Fungi in Pathogenesis of Inflammatory Bowel Diseases and Therapeutic Approaches', *Gastroenterology*, vol. 152, no. 2, Feb, pp. 327-339.e324.

Schirmer, M, Franzosa, EA, Lloyd-Price, J, McIver, LJ, Schwager, R, Poon, TW, Ananthakrishnan, AN, Andrews, E, Barron, G, Lake, K, Prasad, M, Sauk, J, Stevens, B, Wilson, RG, Braun, J, Denson, LA, Kugathasan, S, McGovern, DPB, Vlamakis, H, Xavier, RJ & Huttenhower, C 2018, 'Dynamics of metatranscription in the inflammatory bowel disease gut microbiome', *Nat Microbiol*, vol. 3, no. 3, 2018/03/01, pp. 337-346.

Schnorr, SL, Candela, M, Rampelli, S, Centanni, M, Consolandi, C, Basaglia, G, Turroni, S, Biagi, E, Peano, C, Severgnini, M, Fiori, J, Gotti, R, De Bellis, G, Luiselli, D, Brigidi, P, Mabulla, A, Marlowe, F, Henry, AG & Crittenden, AN 2014, 'Gut microbiome of the Hadza hunter-gatherers', *Nature Communications*, vol. 5, no. 1, 2014/04/15, p. 3654.

Sellon, RK, Tonkonogy, S, Schultz, M, Dieleman, LA, Grenther, W, Balish, E, Rennick, DM & Sartor, RB 1998, 'Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice', *Infect Immun*, vol. 66, no. 11, Nov, pp. 5224-5231.

Sender, R, Fuchs, S & Milo, R 2016, 'Revised Estimates for the Number of Human and Bacteria Cells in the Body', *PLoS Biol*, vol. 14, no. 8, Aug, p. e1002533.

Sharma, A, Das, P, Buschmann, M & Gilbert, JA 2020, 'The Future of Microbiome-Based Therapeutics in Clinical Applications', *Clinical Pharmacology & Therapeutics*, vol. 107, no. 1, pp. 123-128.

Sharpton, TJ 2014, 'An introduction to the analysis of shotgun metagenomic data', *Front Plant Sci*, vol. 5, p. 209.

Shaw, SY, Blanchard, JF & Bernstein, CN 2010, 'Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease', *Am J Gastroenterol*, vol. 105, no. 12, Dec, pp. 2687-2692.

Sokol, H, Landman, C, Seksik, P, Berard, L, Montil, M, Nion-Larmurier, I, Bourrier, A, Le Gall, G, Lalande, V, De Rougemont, A, Kirchgesner, J, Daguenel, A, Cachanado, M, Rousseau, A, Drouet, É, Rosenzwajg, M, Hagege, H, Dray, X, Klatzman, D, Marteau, P, Beaugerie, L & Simon, T 2020, 'Fecal microbiota transplantation to maintain remission in Crohn's disease: a pilot randomized controlled study', *Microbiome*, vol. 8, no. 1, Feb 3, p. 12.

Stoeck, T, Bass, D, Nebel, M, Christen, R, Jones, MD, Breiner, HW & Richards, TA 2010, 'Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water', *Mol Ecol*, vol. 19 Suppl 1, Mar, pp. 21-31.

Szeligowski, T, Yun, AL, Lennox, BR & Burnet, PWJ 2020, 'The Gut Microbiome and Schizophrenia: The Current State of the Field and Clinical Applications', *Frontiers in Psychiatry*, vol. 11, 2020-March-12.

Takahashi, K, Nishida, A, Fujimoto, T, Fujii, M, Shioya, M, Imaeda, H, Inatomi, O, Bamba, S, Sugimoto, M & Andoh, A 2016, 'Reduced Abundance of Butyrate-Producing Bacteria Species in the Fecal Microbial Community in Crohn's Disease', *Digestion*, vol. 93, no. 1, pp. 59-65.

Tisza, MJ & Buck, CB 2021, 'A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases', *Proceedings of the National Academy of Sciences*, vol. 118, no. 23, p. e2023202118.

Turnbaugh, PJ, Ley, RE, Hamady, M, Fraser-Liggett, CM, Knight, R & Gordon, JI 2007, 'The Human Microbiome Project', *Nature*, vol. 449, no. 7164, 2007/10/01, pp. 804-810.

Uniken Venema, WT, Voskuil, MD, Dijkstra, G, Weersma, RK & Festen, EA 2017, 'The genetic background of inflammatory bowel disease: from correlation to causality', *J Pathol*, vol. 241, no. 2, Jan, pp. 146-158.

Vandeputte, D, Tito, RY, Vanleeuwen, R, Falony, G & Raes, J 2017a, 'Practical considerations for large-scale gut microbiome studies', *FEMS Microbiol Rev*, vol. 41, no. Supp_1, Aug 1, pp. S154-s167.

Vandeputte, D, Kathagen, G, D'hoe, K, Vieira-Silva, S, Valles-Colomer, M, Sabino, J, Wang, J, Tito, RY, De Commer, L, Darzi, Y, Vermeire, S, Falony, G & Raes, J 2017b, 'Quantitative microbiome profiling links gut community variation to microbial load', *Nature*, vol. 551, no. 7681, 2017/11/01, pp. 507-511.

Veltkamp, C, Tonkonogy, SL, de Jong, YP, Albright, C, Grenther, WB, Balish, E, Terhorst, C & Sartor, RB 2001, 'Continuous stimulation by normal luminal bacteria is essential for the development and perpetuation of colitis in Tgε26 mice', *Gastroenterology*, vol. 120, no. 4, 2001/03/01/, pp. 900-913.

Vijay, A & Valdes, AM 2022, 'Role of the gut microbiome in chronic diseases: a narrative review', *European Journal of Clinical Nutrition*, vol. 76, no. 4, 2022/04/01, pp. 489-501.

Vujkovic-Cvijin, I, Sklar, J, Jiang, L, Natarajan, L, Knight, R & Belkaid, Y 2020, 'Host variables confound gut microbiota studies of human disease', *Nature*, vol. 587, no. 7834, 2020/11/01, pp. 448-454.

Wang, B, Yao, M, Lv, L, Ling, Z & Li, L 2017, 'The Human Microbiota in Health and Disease', *Engineering*, vol. 3, no. 1, 2017/02/01/, pp. 71-82.

Wilkinson, JE, Franzosa, EA, Everett, C, Li, C, Hu, FB, Wirth, DF, Song, M, Chan, AT, Rimm, E, Garrett, WS & Huttenhower, C 2021, 'A framework for microbiome science in public health', *Nat Med*, vol. 27, no. 5, May, pp. 766-774.

Witkowski, M, Weeks, TL & Hazen, SL 2020, 'Gut Microbiota and Cardiovascular Disease', *Circulation Research*, vol. 127, no. 4, pp. 553-570.

Xavier, RJ & Podolsky, DK 2007, 'Unravelling the pathogenesis of inflammatory bowel disease', *Nature*, vol. 448, no. 7152, Jul 26, pp. 427-434.

Xu, X-R, Liu, C-Q, Feng, B-S & Liu, Z-J 2014, 'Dysregulation of mucosal immune response in pathogenesis of inflammatory bowel disease', *World J Gastroenterol*, vol. 20, no. 12, pp. 3255-3264.

Young, RB, Marcelino, VR, Chonwerawong, M, Gulliver, EL & Forster, SC 2021, 'Key Technologies for Progressing Discovery of Microbiome-Based Medicines', *Front Microbiol*, vol. 12, 2021-June-22.

Zakrzewski, M, Simms, LA, Brown, A, Appleyard, M, Irwin, J, Waddell, N & Radford-Smith, GL 2018, 'IL23R-Protective Coding Variant Promotes Beneficial Bacteria and Diversity in the Ileal Microbiome in Healthy Individuals Without Inflammatory Bowel Disease', *Journal of Crohn's and Colitis*, vol. 13, no. 4, pp. 451-461.

Zuo, T, Kamm, MA, Colombel, JF & Ng, SC 2018, 'Urbanization and the gut microbiota in health and inflammatory bowel disease', *Nature Reviews Gastroenterology & Hepatology*, vol. 15, no. 7, Jul, pp. 440-452.

Chapter 1

Statement of Authorship

Title of Paper	The neglected gut microbiome: fungi, protozoa, and bacteriophages in inflammatory bowel disease
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Guzzo, GL, Andrews, JM & Weyrich, LS 2022, 'The Neglected Gut Microbiome: Fungi, Protozoa, and Bacteriophages in Inflammatory Bowel Disease', Inflamm Bowel Dis, p. izab343. doi: https://doi.org/10.1093/ibd/izab343

Principal Author

Name of Principal Author (Candidate)	Gina L. Guzzo		
Contribution to the Paper	Conceptualised the review, conducted the research, wrote the manuscript, and generated the figures.		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	01/06/2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Conceptualised the review, provided substantial discussion and feedback to the research, and edited all versions of the manuscript.		
Signature		Date	24/06/2022

Name of Co-Author	Jane M. Andrews		
Contribution to the Paper	Provided discussion and feedback to the research and edited all versions of the manuscript.		
Signature		Date	17/08/2022

Inflammatory Bowel Diseases, 2022, **28**, 1112–1122
https://doi.org/10.1093/ibd/izab343
Advance access publication 29 January 2022
Review Article - Basic Science

The Neglected Gut Microbiome: Fungi, Protozoa, and Bacteriophages in Inflammatory Bowel Disease

Gina L. Guzzo, BSc,[*] Jane M. Andrews, MD, PhD,[†] and Laura S. Weyrich, PhD[*,‡]

From the [*]School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia
[†]Inflammatory Bowel Disease Service, Department of Gastroenterology and Hepatology, Royal Adelaide Hospital and School of Medicine, Faculty of Health Sciences, University of Adelaide, Adelaide, South Australia, Australia
[‡]Department of Anthropology and Huck Institutes of the Life Sciences, Pennsylvania State University, State College, PA, USA
Address correspondence to: Gina L. Guzzo, The University of Adelaide, Adelaide, South Australia, Australia (gina.guzzo@adelaide.edu.au).

The gut microbiome has been implicated in the pathogenesis of inflammatory bowel disease (IBD). Studies suggest that the IBD gut microbiome is less diverse than that of the unaffected population, a phenomenon often referred to as dysbiosis. However, these studies have heavily focused on bacteria, while other intestinal microorganisms—fungi, protozoa, and bacteriophages—have been neglected. Of the nonbacterial microbes that have been studied in relation to IBD, most are thought to be pathogens, although there is evidence that some of these species may instead be harmless commensals. In this review, we discuss the nonbacterial gut microbiome of IBD, highlighting the current biases, limitations, and outstanding questions that can be addressed with high-throughput DNA sequencing methods. Further, we highlight the importance of studying nonbacterial microorganisms alongside bacteria for a comprehensive view of the whole IBD biome and to provide a more precise definition of dysbiosis in patients. With the rise in popularity of microbiome-altering therapies for the treatment of IBD, such as fecal microbiota transplantation, it is important that we address these knowledge gaps to ensure safe and effective treatment of patients.

Lay Summary

Fungi, protozoa, and bacteriophages are often neglected in gut microbiome research of inflammatory bowel disease. Here, we review what is currently known of these microbes in inflammatory bowel disease and how they can be studied using high-throughput DNA sequencing methods.

Key Words: inflammatory bowel disease (IBD), microbiome, fungi, protozoa, bacteriophage

Introduction

Inflammatory bowel disease (IBD) is an umbrella diagnosis for a group of chronic inflammatory disorders of the gastrointestinal tract; the 2 most commonly diagnosed forms are ulcerative colitis (UC) and Crohn's disease (CD).[1] IBDs are multifactorial diseases that arise from complex interactions between genetic, environmental, and microbial factors.[2-4] Among the microbial factors, the gut microbiome (Box 1) has been implicated in the disease.[5,6] Patients with IBD generally have gut microbiomes that are less diverse in species and function compared with unaffected individuals,[6,7] a microbial signature often referred to as dysbiosis.[8] Research has not arrived at a consensus on the role of dysbiosis in IBD—whether it is a causal factor in IBD development, a perpetuating factor, or simply a result of the disease or lifestyle changes in response to the disease.[9,10]

Bacteria are easily identified in gut microbiome studies due to their abundance and have become the focal point of most IBD research (Figure 1), to the exclusion of other intestinal microorganisms such as fungi, protozoa, and bacteriophages.[11,12] This myopic investigation of the microbiome has caused the usage of terminology to shift to a point of imprecision and opaqueness; the terms *microbiota* and *microbiome* are often used to denote only the bacterial portion of a microbial community.[13-15] This usage becomes increasingly ambiguous when researchers describe IBD patients with dysbiosis, yet only the bacterial microbiome has been investigated.[9,16,17] Currently, nonbacterial microorganisms are mainly associated with pathogenicity.[12] Despite this, many are found in the gastrointestinal tracts of healthy individuals, and there is still conflicting evidence on whether these species have direct and consistent proinflammatory effects.[6,12,18-21] There is even evidence to suggest that the absence of some nonbacterial species is associated with disease.[22] In this review, we seek to address this knowledge gap by discussing what is known of nonbacterial gut microorganisms, namely fungi, protozoa, and viruses, in IBD. We draw attention to new high-throughput sequencing methods used to study these microbes to develop a more comprehensive understanding of the gut microbiome in IBD.

Gut Fungi in IBD

Are gut fungi pathogenic or protective?

Fungi are found on every skin and mucosal surface of the human body,[26,27] with the skin, vagina, oral cavity, small intestine, and large intestine harboring the highest abundance and diversity of fungal species. Most of these species are yeasts such as *Candida*, *Malassezia*, and *Saccharomyces*,[27,28]

Received for publication: August 4, 2021. Editorial Decision: December 6, 2021

Box 1. Definitions of terms

Microbiota
A collection of microbes, including prokaryotes (bacteria and archaea), eukaryotes (microbial parasites and fungi), and viruses, found in a specified environment.[23] The term is often used as a shortened replacement for *bacterial microbiota*, which may cause confusion if not explicitly stated.

Microbiome
The combined microbiota, their genes and gene products, and their surrounding microenvironment.[23,24] Like *microbiota*, this term is often confusingly used to describe only bacterial populations. However, by default, it denotes the wider population of microbial types (prokaryotes, eukaryotes, and viruses), their genes, and environment.

Amplicon Sequencing
Sometimes referred to as metabarcoding, this high-throughput sequencing method can be used to survey the prokaryotic microbiome community by targeting the bacterial or archaeal 16S ribosomal RNA gene or to survey the eukaryotic community by targeting 18S and ITS (ITS1 and ITS2) ribosomal RNA genes.

Shotgun Sequencing
A high-throughput sequencing method in which the DNA of a sample is fragmented and sequenced at random. When used to explore the microbiome, it is often referred to as metagenomic sequencing.[25] Instead of targeting a small proportion of the total genes in a sample like amplicon sequencing, shotgun sequencing captures random fragments of any DNA in the sample, including both host and microbial DNA.

Mycobiome
Like *microbiome*, this term includes the fungal community in a specified environment, and its genetic and environmental information. It is synonymous with *fungal microbiome*.

Protozoa
Unicellular eukaryotic microbes often referred to as parasites in the context of human health.

Virus
A nonliving biological entity that infects cells in order to persist and replicate. Viruses are divided into 2 coarse groupings: those that infect bacteria are termed *bacteriophages* (shortened to *phages*), and those that infect eukaryotic cells, including host cells and microbial eukaryotes, known simply as viruses.

Virome
A term for all viral DNA in a specified environment.

and these species have been associated with IBD.[12,29] Elevated levels of anti-*S. cerevisiae* antibodies have been found in IBD patients,[30,31] and these antibodies have been associated with an early risk for surgery in CD patients,[32] and used to help differentiate[33] and predict the development of CD and UC.[34] Higher abundances of *Candida* species, particularly *C. albicans*, have been found in CD patients and their first-degree relatives, and in a mouse model of colitis.[35,36] This has sparked the hypothesis that intestinal inflammation leads to a compromised mucosal barrier that allows *C. albicans* and potentially other opportunistic bacteria and fungi to proliferate and/or penetrate the mucosal barrier, thus driving interaction with the host immune system.[37] Additionally, this fungal-induced inflammation may be exacerbated in IBD patients with inherited genetic mutations in antifungal immune genes (eg, *DECTIN*-1, *Card9*).[38]

However, there is also evidence that *Candida* and *Saccharomyces* are not exclusively pathogenic. For example, the presence of *Candida* species had an alleviating effect in a mouse model of acute colitis[39] and *S. cerevisiae* has been shown to have an attenuating effect on *Escherichia coli*–induced mouse colitis,[40] suggesting that fungi may have a protective effect against IBD. Another species, *S. boulardii*, also had an attenuating effect in mice with carcinogenic colitis, a condition that can develop from IBD.[41,42] The protective effect of *S. boulardii* has been trialed in CD patients with mixed success rates.[43,44] Thus, there is evidence that some fungi may be beneficial in IBD, but inconsistent findings exemplify the complexity of fungi-host interactions. Fungal species may have a spectrum of effects that depend on multiple factors related to an individual's physiology and microbiome.[11] These complexities cannot be gleaned from single-species interrogations, and instead, methods examining the whole microbial community are now being favored.

Insights into gut fungi in IBD from high-throughput sequencing

Initial studies on fungi were limited to species that could be isolated and cultured, wherein researchers characterized a cultured species by sequencing its genome, or generating antibodies to its cellular components.[27] This limitation historically biased the reporting of microbes to only culturable species. This may be one of the reasons why *Candida* and *Saccharomyces* are most often reported in IBD research, as many species from these genera are readily isolated, cultured, and identified.[45,46] Bacterial research previously suffered from similar limitations, and high-throughput DNA sequencing technologies, such as amplicon sequencing and shotgun sequencing, were developed to help overcome these challenges.[25] A major advantage of these methods is that they can indiscriminately capture DNA from several fungal taxonomic groups directly from a sample, without the tedious requirement of culturing each fungal species.[47] They can also capture DNA from unculturable fungal species in samples dominated by bacterial and human DNA, and thus may give a more representative depiction of the fungal community of a sample.[11]

With the rise of high-throughput sequencing, the past 10 years have seen a steady increase in studies of the human fungal microbiome, known as the mycobiome.[11] The 2 most common sampling types for surveying the mycobiome are fecal samples, either in the form of whole stool or swab, and mucosal biopsies. Fecal samples are used as a proxy for the intestinal microbiome due to the invasiveness of acquiring biopsies,[47] although it is expected that the microbiome composition of these 2 sample types will somewhat differ.[48] It is now apparent that the gut mycobiome can include species from several dozen genera of fungi (eg, *Alternaria*, *Aspergillus*, *Candida*, *Cladosporium*, *Cryptococcus*, *Debaryomyces*, *Fusarium*, *Galactomyces*, *Malassezia*, *Penicillium*, *Pichia*, *Rhodotorula*, *Saccharomyces*, *Trichosporon*), dominated by yeast species from the family *Saccharomycetaceae*.[49-52] Gut mycobiomes differ between individuals and seem to be more temporally variable than gut bacterial microbiomes.[49,51]

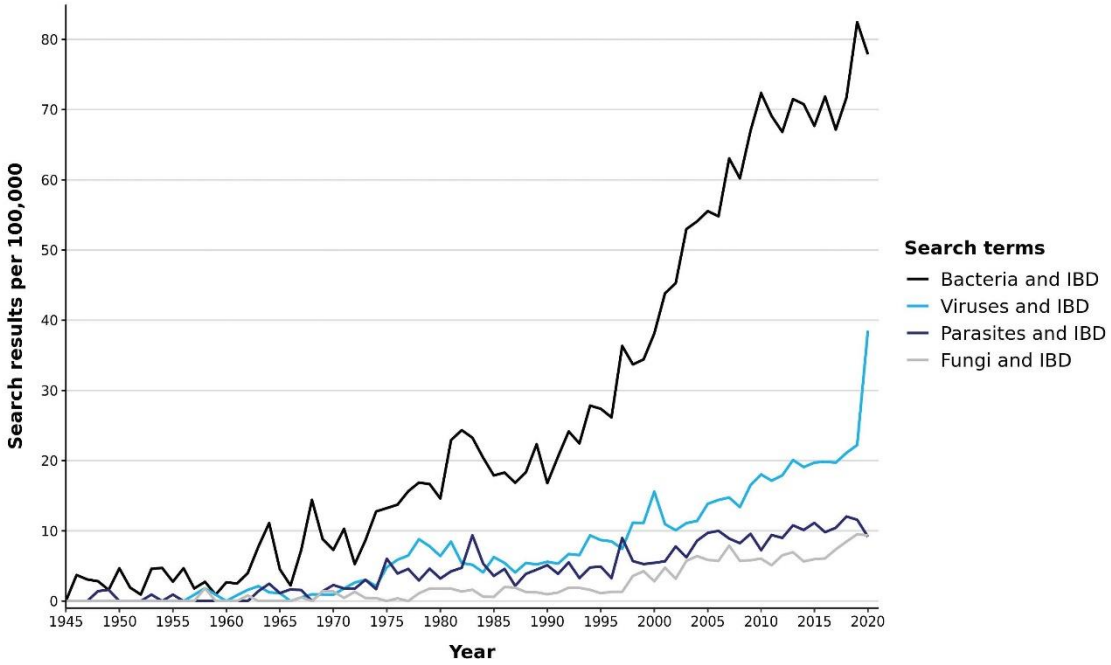

FIGURE 1. Search results of citations featuring microbial types and inflammatory bowel disease (IBD) on PubMed from 1945 to 2020. The total number of search results for bacteria and IBD (15 354) outnumbers results on IBD and viruses (4570), parasites (2314), and fungi (1516). Search conducted on November 1, 2021. Figure generated with PubMed by year (https://esperr.github.io/pubmed-by-year/).

Several mycobiome studies indicate that our intestinal mycobiome, like the bacterial microbiome, differs due to environmental factors such as mode of delivery during birth, age, diet, and geographical location.[53-56]

High-throughput sequencing approaches have revealed that the gut mycobiome differs in IBD patients, a microbial signature sometimes referred to as *fungal dysbiosis* (Table 1). Amplicon sequencing studies of colonic biopsies have shown that adult CD patients have a higher number of fungal species compared with UC patients and unaffected control subjects,[57] whereas adults with active UC have fewer species and less abundant mycobiomes in both colonic biopsies[58] and stool.[52] CD patients in flare also have a higher fungal load in both inflamed and uninflamed mucosa than CD patients in remission and healthy individuals.[59] Both CD and UC patients also have an altered abundance of different yeast species, namely *C. tropicalis* in stool of CD patients,[60] *D. hansenii* in inflamed mucosa of CD patients,[61] *Aspergillus* in colonic biopsies from UC patients,[58] and an increase in *C. albicans* and decrease in both *M. sympodialis* and *S. cerevisiae* in stool from a cohort of CD and UC patients in flare.[52] The fact that some yeasts are more abundant in IBD patients has been incorporated into the hypothesis that IBD may be caused or perpetuated by an overgrowth of opportunistic intestinal fungi.[12,29,37] These fungi may achieve this through interkingdom interactions, such as the biofilm formed between *C. tropicalis* and the bacterial species *E. coli* and *S. marcescens*, which induces the expression of pathogenic fungal hyphae.[60] Other fungal species may be able to exert their pathogenicity without the help of bacteria. For example, *D. hansenii* was shown to preferentially localize to inflamed mucosa in colonic tissue isolated from biopsy-injured mice and ileal tissue isolated from CD patients.[61] *D. hansenii* prevented repair of colonic mucosa in the absence of bacteria, which was established using gnotobiotic mice.

Because the mycobiome shows a marked alteration in IBD, it has the potential to be used as a diagnostic tool. For example, the fecal mycobiome was used to discriminate between CD and UC by combining fungal load with bacterial load, clinical biomarkers (fecal calprotectin and C-reactive protein), and demographic data (age, gender, BMI, and smoking habit) in a random forest predictive model.[65] The fecal mycobiome also differs in patients experiencing a flare[52,65] compared with patients in remission who may have gut mycobiomes that more closely resemble a healthy mycobiome,[64,65] and was better able to predict relapse in CD and UC patients when fungal load was incorporated into the predictive model described previously. Therefore, the fecal mycobiome may be a minimally invasive diagnostic tool for predicting IBD subtype and relapse.

Mycobiome shifts also exist in pediatric IBD patients (Table 1). Pediatric patients showed a reduction in overall gut fungal diversity and an increase in *Cyberlindnera jadinii* and *C. parapsilosis* in stool samples compared with healthy adult and pediatric control subjects.[62] In one of the few shotgun sequencing studies of the IBD mycobiome, pediatric patients with active CD undergoing a formula diet (exclusive enteral nutrition) or anti-tumor necrosis factor therapy also had elevated *C. jadinii*, as well as elevated *S. cerevisiae*, *Clavispora lusitaniae*, *C. albicans*, and *Kluyveromyces marxianus* at baseline compared with healthy pediatric control subjects.[63] This elevation in yeast species subsequently decreased following 8 weeks of nutrition therapy, suggesting that diet is an effective modifier of the mycobiome in patients.

It is clear that differences exist between the IBD mycobiome and unaffected individuals. However, a lack of research and studies incorporating different populations, sample types, and methods existing studies means that it is still too early to ascertain clear trends (Table 1). Research thus far suggests that the IBD mycobiome varies between CD and UC,[52,57,65]

TABLE 1. Current insights from high-throughput sequencing of the nonbacterial microbiome in IBD

Study	Study Population(s)	Microbiome Sample(s) and Methods	Key Findings
Fungi			
[57]	Active CD (n = 31), active UC (n = 26), non-IBD intestinal inflammation (n = 15), healthy individuals (n = 32)	18S rRNA–based amplification, denaturing gradient gel electrophoresis, and clone library analysis of stool and biopsies from inflamed colon	↑ fungal species in active CD biopsies compared with active UC and control subjects
[62]	Pediatric IBD (26 CD, 4 UC, and 2 indeterminate colitis), healthy adult and pediatric control subjects (n = 90)	ITS1 sequencing of stool	↓ fungal diversity (Shannon index) in IBD ↑ abundance of *Cyberlindnera jadinii* and *C. parapsilosis* and ↓ abundance of *Cladosporium cladosporioides* in IBD
[63]	Pediatric active CD (n = 90), healthy pediatric control subjects (n = 26)	Shotgun whole metagenome sequencing of stool	↑ abundance of *C. jadinii*, *S. cerevisiae*, *Clavispora lusitaniae*, *C. albicans*, and *Kluyveromyces marxianus* in active CD, which decreased following 8 weeks of exclusive enteral nutrition
[60]	Active and inactive CD (n = 20), CD relatives (n = 28), unrelated healthy individuals (n = 21)	ITS1, ITS2, and 16S rRNA sequencing of stool	↑ abundance of *C. tropicalis* in CD patients, positively correlated with *Serratia marcescens* and *Escherichia coli*
[59]	Active CD (n = 16), inactive CD (n = 7), healthy individuals (n = 10)	ITS2 rRNA sequencing of ileocolonic biopsies with quantitative PCR	↑ fungal load of both inflamed and inflamed mucosa in active CD compared with inactive CD and healthy control subjects
[58]	Active UC (n = 14), healthy individuals (n = 15)	18S and ITS2 rRNA sequencing of biopsies from inflamed colon; quantitative PCR of 18S rRNA for fungal load	↓ fungal species count and abundance in active UC ↑ abundance of *Aspergillus* in active UC
[52]	Active CD and UC patients (n = 106), inactive CD and UC patients (n = 129), healthy individuals (n = 38)	ITS2 rRNA sequencing of stool	↑ abundance of *C. albicans* and ↓ abundance of *S. cerevisiae* and *Malassezia sympodialis* in active IBD compared with remission ↓ fungal species count in UC
[64]	PSC patients with IBD in remission (n = 27), PSC patients without IBD (n = 22), IBD patients in remission without PSC (n = 33), and healthy individuals (n = 30)	ITS2 and 16S sequencing of stool	No difference in fungal diversity (Shannon and Chao1 indices) between IBD remission and healthy individuals ↑ fungal diversity (Shannon) index in PSC compared with IBD remission
[61]	Mice injured by colonic biopsies and treated with antibiotics to impair healing, control mice injured but not treated with antibiotics	Quantitative PCR of ITS of murine mucosal wounds and patient ileal biopsies	↑ *Debaryomyces hansenii* abundance in mucosal wounds of antibiotic-treated mice compared with control subjects
	Patients with active CD (n = 7) and healthy individuals (n = 10)		↑ *D. hansenii* abundance in inflamed mucosa of CD patients compared with uninflamed mucosa in same patients
[65]	Patients with 3-month remission of UC (n = 31), and ileal or ileocolonic CD (n = 34), patients with active CD (n = 55), UC relatives (n = 29), CD relatives (n = 29), healthy unrelated individuals (n = 28)	Quantitative PCR of ITS2 and 16S rRNA of stool, random forest predictive modelling	↑ fungal load in relapsed CD compared with patients who remained in remission ↑ fungal load in relapsed UC compared with UC remission, CD remission, and CD relapsed Fungal and bacterial load combined with clinical markers (C-reactive protein and fecal calprotectin) and demographic data distinguished UC from CD and could predict relapse
Protozoa			
[66]	Active and inactive CD and UC patients (n = 100), healthy individuals (n = 96)	Culture and PCR of stool	↓ *Blastocystis* and *Dientamoeba fragilis* prevalence in active CD and UC than inactive ↓ *Blastocystis* prevalence in both active and inactive IBD than control subjects

Table 1. Continued

Study	Study Population(s)	Microbiome Sample(s) and Methods	Key Findings
[67]	Patients with active CD (n = 76) and UC (n = 31), healthy individuals (n = 616)	18S rRNA sequencing of stool	↓ *Blastocystis* prevalence in IBD
Viruses			
[68]	Patients with active (n = 10) and inactive (n = 1) ileocolonic CD, healthy individuals (n = 8)	454 pyrosequencing of stool	↓ virome diversity (Shannon index) in CD
[69]	Pediatric patients with CD (n = 6), and healthy individuals (n = 6)	454 pyrosequencing of ileal and colonic biopsies, and gut washes	↑ abundances of viral species in CD, *Caudovirales* most abundant
[63]	Pediatric active CD (n = 90), healthy pediatric control subjects (n = 26)	Shotgun whole metagenome sequencing of stool	No difference in bacteriophage species between groups
[70]	Patients with CD (n = 18), UC (n = 42), healthy individuals (n = 12)	454 pyrosequencing of stool	↑ abundances of *Caudovirales* bacteriophage species in CD and UC
[71]	Patients with new-onset active CD (n = 12), healthy individuals (n = 12)	454 pyrosequencing of colonic biopsies	↑ viral species in active CD
[72]	C57BL6/J *Rag1*$^{-/-}$ mice with colitis induced by injection of CD4+ CD45RB^{High} T cells (n = 3) and control mice injected with saline (n = 3)	Shotgun whole metagenome sequencing of stool	↑ abundances of *Caudovirales* bacteriophages species in murine colitis

Abbreviations: CD, Crohn's disease; IBD, inflammatory bowel disease; PCR, polymerase chain reaction; PSC, primary sclerosing cholangitis; rRNA, ribosomal RNA; UC, ulcerative colitis.

between patients in remission and in flare,[52,64,65] and between sites of inflamed mucosa and adjacent uninflamed tissue.[61] Therefore, it is important that we continue to have different disease subtypes and disease states represented in future IBD mycobiome datasets. Longitudinal sampling will also help uncover whether compositional changes are a cause or effect of flare.[73]

Limitations and future directions of gut mycobiome research

Preliminary work shows that the IBD mycobiome differs from unaffected individuals, though inconsistencies in findings and, more importantly, an overall lack of research means that much work is still needed in this area. Fungi remain underexplored in sequence-based approaches, likely due to the low abundance of fungal DNA relative to bacterial DNA in gut microbiome samples.[51,65] It has been hypothesized that the ratio of fungal to bacterial cells changes throughout the gastrointestinal tract and that parts of the upper gastrointestinal tract (stomach and duodenum) have a higher ratio of fungi to bacteria than lower parts (jejunum, ileum, and colon).[11,74] The proportional influence of fungi may thus differ considerably throughout the gastrointestinal tract, and further research is needed to understand such differences in interactions. The anatomical variation in fungal interactions might indeed be one reason that we see an anatomical restriction in where CD and UC occur in individuals, and in the diseases themselves.

Owing to the low abundance of fungi in intestinal microbiome samples, deep sequencing is required to capture the fungal genomic component of these samples, which can be costly and time-consuming.[11] There are, however, methods to mitigate this difficulty. Samples can be enriched for fungal DNA prior to sequencing via several options of protocols and kits, to reduce the sequencing effort required to capture fungal DNA.[47,75] Additionally, computational tools have been developed that specifically recognize fungal DNA sequences. There are now several bioinformatic pipelines available to recover and taxonomically assign fungal DNA from amplicon data (eg, RiboTagger)[76] and shotgun data (eg, FindFungi, EukRep, HumanMycobiomeScan, EukDetect).[77-80]

High-throughput sequencing methods for studying the mycobiome can circumvent some of the limitations of previous technologies, but they also suffer from their own limitations. Though amplicon sequencing is a useful method for determining fungal abundances and coarse phylogenetic groupings, the approach does not always yield good resolution to the species level, or even to the genus level, and is generally less sensitive than 16S sequencing for bacteria.[27,81] Shotgun sequencing is more sensitive than amplicon sequencing; however, it is more expensive and computationally intensive. Because shotgun sequencing indiscriminately captures all the DNA in a sample, human DNA contamination is common and must be dealt with in the laboratory and computationally.[82] Another disadvantage of amplicon sequencing is that it does not allow for direct functional inference, as only ribosomal genes are sequenced with this method, and function is inferred with predictive tools.[83] Shotgun sequencing can recover partial or whole microbial genomes, so it enables direct functional inference. This method is additionally advantageous because one can profile both the bacterial and fungal portions of the microbiome in a single effort.

There are also limitations that affect both amplicon and shotgun sequencing. Gene and genome references available for fungi in databases are still biased toward already cultured organisms.[27] This is important to consider for both amplicon and shotgun studies wherein yeasts are still frequently reported over other fungi. Because fewer fungal species from the gastrointestinal tract have been cultured than bacteria, even less is known of their metabolic functions, and so predictive tools can be unreliable. Reference databases used to assign

taxonomy to fungi also contain thousands of unannotated and incorrectly annotated sequences.[81] Fortunately, tools that facilitate de novo assembly of fungal genomes are available to characterize the fraction of the mycobiome that is both unculturable and lacking reference genomes, such as a recent pipeline used to identify novel fungal genomes from premature infants,[84] and EukCC, a tool to estimate the quality of eukaryotic genome assembly.[85] With these tools, reference databases for fungi are expanding faster than ever before. Researchers can now retrospectively mine metagenomic data for fungal DNA, and this can similarly be done for IBD cohorts. New metagenomic approaches such as these are critically important to answer remaining questions regarding the role of the mycobiome in IBD: Is there a characteristic mycobiome signature of IBD or its subtypes? Is there a tipping point at which yeasts such as *Candida* become more abundant and, potentially, proinflammatory? Which fungal and bacterial species or strains are involved in this process? Is an increase in these species a cause or a result of IBD?

Gut Protozoa in IBD

Gut protozoa: Falsely villainized?

Intestinal parasites are typically known for causing dysenteric infections.[86] These parasites have gradually been depleted with industrial-associated lifestyle factors such as improved sanitation, hygiene, and health care.[87,88] However, an industrialized lifestyle has also been associated with the rise in incidences of IBD,[89,90] and some hypothesize that exposure to certain intestinal parasites may be beneficial for maintaining a healthy and diverse microbiome.[90-92] Macroparasites, namely helminths, and their purified antigens have been used to treat IBD in mice[93-95] and in controversial human trials with some success.[96-99] Protozoa have received far less attention in relation to IBD, although there are several protozoan species that are able to commensally colonize and reside in the human intestine.[18]

Blastocystis species and *Dientamoeba fragilis* are the most common protozoa found in human stool and are primarily transmitted through the fecal-oral route.[100,101] The prevalence of *Blastocystis* species in human stool ranges from 1% to 50% in developed nations and is generally >30% in developing nations.[67,101,102] Similarly, the prevalence of *D. fragilis* varies greatly between regions of the world, with a higher prevalence in developing regions.[103,104] *Blastocystis* and *D. fragilis* are often blamed for causing gastroenteritis-like symptoms, although they have been found in both symptomatic and asymptomatic individuals and their pathogenicity is thus still debated.[18] More recently, largescale controlled cohorts have not found an association between *Blastocystis*, *D. fragilis*, and gastroenteritis.[105-107] Rather, these protozoa were found to be more abundant in healthy individuals[108] and were are also associated with increased gut bacterial diversity.[51] These findings suggest that *Blastocystis* and *D. fragilis* may not be parasitic, but rather enteric commensals. In fact, the name "parasite" may be a misnomer for these species. This hypothesis is also supported in IBD patients, wherein both *Blastocystis* and *D. fragilis* have been found more frequently in unaffected individuals and UC patients with inactive disease than in UC and CD patients with active disease.[22,66,67] Whether the lower prevalence of *Blastocystis* in patients was a cause or effect of the disease was not addressed in these studies, but we should consider if these protozoa are a hallmark of a healthy gut, and whether administering antibiotics when they are found may be causing harm.

Limitations and future directions of gut protozoa research

The study of intestinal protozoa has experienced similar biases to fungal research—some species have been heavily studied whereas others are scarcely discussed (Table 1). There is an evident ascertainment bias toward the reporting of culturable parasites, and very little is known about unculturable protozoan members of the human gut microbiome.[109] For example, *Blastocystis* species, though anerobic, can be readily cultured and they are commonly detected with microscopy following in vitro culture from stool.[101,110] However, microscopic detection of *Blastocystis* subtypes in stool is less sensitive than sequencing methods, particularly when they are present in low abundances.[111] Capturing protozoan DNA can be achieved with 18S amplicon sequencing, and there are parasite-specific 18S primers that can capture DNA from several taxonomic groups.[112] Though as stated previously, this method is rarely sensitive enough for robust species-level resolution.[109] Shotgun sequencing can bypass some of the limitations of amplicon sequencing, but it is similarly limited by the low proportion of protozoa in the intestinal microbiome.[51] Thus, deep sequencing, even deeper than required to detect fungi, is necessary to capture enough protozoan DNA for species identifications. It is therefore critical that samples are enriched for eukaryotic cells prior to sequencing. Fortunately, decreases in the cost of sequencing, enrichment for eukaryotic DNA, and improvements to computational methods and reference databases may soon help to provide insights into protozoa in IBD.

Gut Bacteriophages in IBD

Bacteriophages: contributors to the IBD gut microbiome

An assortment of viral particles exists in the gastrointestinal environment of many animals, including humans.[21,113] Viruses of the gut microbiome include 2 major types: those that infect eukaryotic cells (eg, human cells) and phages that infect bacteria. While both types have been detected in the human gut,[114] phages comprise most of the viral species present in the gut. Phages can transfer genetic content, such as antibiotic resistance genes, between bacterial cells,[114,115] and cause rapid destruction of bacterial cells upon infection during the lytic cycle. Therefore, these viruses can regulate population levels of resident bacteria[114] and should be recognized as able contributors to microbiome composition shifts, such as those seen in IBD.[116]

The gut virome in IBD studies

The gut virome is an emerging area of study in IBD research, and to date, the field contains only a handful of studies (Table 1). One small study of stool samples from CD patients (n = 11) and unaffected control subjects (n = 8) found that virome and bacterial diversity in stool samples was lower in the patients.[68] Conversely, in another study, colonic biopsies of 12 CD patients had more viral species compared with the 12 control subjects.[71] This same study also found that the sample type and patient from which the sample originated had a greater impact on virome composition than the disease state, suggesting high inter-individual variation in virome composition. Bacterial composition was contrastingly less

variable within groups and was instead more affected by the disease state. Other studies suggest that *Caudovirales* phages, a grouping of over 350 double-stranded DNA viral species,[117] may be more abundant in murine colitis, in pediatric CD, and in adults patients with CD and UC.[69,70,72] However, not all studies have recapitulated this finding.[63] Given that virome research is newly emerging, discrepant findings between these studies may be largely influenced by methodological biases (discussed subsequently), in addition to confounding influences between cohorts.

Current limitations and future directions of virome research

Identifying and classifying viral DNA in microbial samples remains challenging.[118] As they have incredibly high diversity, tiny gene content, and acquire new mutations rapidly, viral species are not easily assigned to closely related species. There is also no gene common to all viruses that can be used as a viral identity marker, and thus, sequencing viral DNA cannot be achieved with a targeted amplicon-like sequencing method.[113] Additionally, viral DNA makes up a small proportion of the total DNA in a microbiome sample.[119] Culturing viruses is equally challenging. Viruses cannot make their own energy because they are parasitic and rely on host cells for resources, so these hosts must be identified and cultured as well. As many microbes of the gastrointestinal tract cannot be cultured, it is difficult to culture their associated viruses.[68]

Embarking on a metagenomic study of the virome may seem like a daunting task, but there are some methodological strategies that can assist in managing the challenge. Prior to sequencing, viral particles can be isolated and purified from a microbiome sample by size selection via centrifugation, filtering (0.2- to 0.45-μm filters), and particle precipitation with polyethylene glycol.[120] Newer computational tools can also reduce the difficulty of studying the human virome, such as METAVIR, an online resource for annotating virus genes from metagenomic data,[121] and VIP and VirFinder, which provide pipelines to map, filter, and identify viruses from metagenomic sequences.[122,123] There are also several databases to identify viral genes (eg, National Center for Biotechnology Information viral genomes resource, IMG/VR, and ACLAME).[124-126] Future gut virome studies can incorporate tools like these, following viral protein enrichment[127] or host DNA depletion, and high-throughput sequencing of patient microbiome samples.[118]

Fecal Microbiota Transplantation and the Neglected Microbiome

Given the observed link between the gut microbiome and IBD, researchers and clinicians have turned to microbiome-based therapies such as fecal microbiota transplantation (FMT) to treat the disease.[128-130] FMT initially received attention for its high efficacy in treating *Clostridioides difficile* infections,[131] and is a procedure that involves the transfer of stool or its microbial derivatives from a healthy donor to a patient, by means of enema, oral capsule, or nasogastric tube.[132] This therapy is presumed to work by restoring a patient's microbiome to a healthy state.[133] FMT is an attractive alternative to other standard therapies, such as immunosuppressants, biologics, or surgery, as successful engraftment of FMT offers the prospect of long-term symptom amelioration without the side effects of other treatment options. So far, FMT for IBD has seen moderate successes and low adverse events in UC patients.[134-136] The average rate of clinical remission achieved sits below 50% which is on par with many other IBD therapies, though this number varies depending on factors such as FMT type, mode of administration, donor type (related vs unrelated), IBD subtype, and geographic location.[134,135]

FMT success rates may also differ because the microbial composition of donor stools used in FMT is still poorly described.[9] Further, we have only successfully characterized a fraction of the total gut microbiome, leaving many species yet to be described.[137] It is therefore critical to include all microbial types in microbiome analyses of FMT studies to provide a more precise depiction of the biological material each patient receives. This might allow for better discrimination between effective and ineffective donor and recipient microbiome samples. Currently, prospective donor samples are screened for agents previously regarded as putative pathogens, such as *Blastocystis* and *D. fragilis*,[138] and donors who are positive for these protozoa may be ruled out.[139] As we have previously discussed, the growing body of literature would suggest that these protozoa are innocuous colonizers of the human gut and may in fact indicate a healthy microbiome. For example, one study did not find different outcomes between *C. difficile* patients receiving FMT that was positive and negative for *Blastocystis*.[140] Donor-derived viruses may also be an important factor for FMT efficacy. One study found that *C. difficile* patients who received donor stool with a higher content of *Caudovirales* phages were more likely to respond positively to their transplants.[141] Thus, further work on gut fungi, protozoa, and bacteriophages is required to reduce the likelihood of discounting commensal species in microbiome-based therapies.

Conclusions and Future Directions

Exploring the ill-defined, nonbacterial microbiome through high-throughput methods is the next logical step toward understanding the link between the gut microbiome and IBD. To this end, it is important that sequencing and computational methods for analyzing eukaryotes and viruses are accessible to clinicians, and that we continue to foster multidisciplinary collaborations to translate bioinformatic results to clinical diagnostics.[47] Future research should incorporate data from nonbacterial organisms with extensive patient information, such as disease state and lifestyle factors, to disentangle the interplay between microbial and host factors.[11,116,142] Last, IBD microbiome research, as in many other fields, will benefit from statistical modelling to disentangle relationships between eukaryotes, prokaryotes, viruses, and host genomic data.[143] As interest in the gut microbiome and microbiome-based therapies continues to rise, studying these relationships will ensure greater precision of diagnostics and treatments for IBD patients.

Acknowledgments

We thank Bastien Llamas and the Thesis Writing Group at the Australian Centre for Ancient DNA for providing edits and feedback on the manuscript.

The references for this Review were identified using PubMed and Web of Science with Boolean search terms for bacteria ("bacter*"), viruses ("virus* OR viral OR virome* OR bacteriophage* OR phage*"), parasites ("parasit* OR protoz* OR helminth* OR protist*"), and fungi ("fung* OR mycobio* OR yeast*") combined with the search terms for IBD ("IBD OR Crohn's OR ulcerative colitis OR inflammatory bowel"). The results were narrowed to articles and reviews published in English. The final reference list was chosen for its novelty and relevance to the scope of this Review topic.

Author Contributions

G.L.G. wrote and conducted the research for the manuscript. L.S.W. and J.M.A. provided substantial discussion and editing of the manuscript.

Support By

None to disclose.

Conflicts of Interest

The authors declare no competing interests.

REFERENCES

1. Podolsky DK. Inflammatory bowel disease (1). *N Engl J Med.* 1991;325:928-937.
2. Ho SM, Lewis JD, Mayer EA, et al. Challenges in IBD research: environmental triggers. *Inflamm Bowel Dis.* 2019;25:S13-S23.
3. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature.* 2011;474:307-317.
4. Loddo I, Romano C. Inflammatory bowel disease: genetics, epigenetics, and pathogenesis. *Front Immunol.* 2015;6:551.
5. Schirmer M, Garner A, Vlamakis H, Xavier RJ. Microbial genes and pathways in inflammatory bowel disease. *Nat Rev Microbiol.* 2019;17:497-511.
6. Zuo T, Ng SC. The gut microbiota in the pathogenesis and therapeutics of inflammatory bowel disease. *Front Microbiol.* 2018;9:2247.
7. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell.* 2012;148:1258-1270.
8. Hooks KB, O'Malley MA. Dysbiosis and its discontents. *MBio* 2017;8:e01492-17.
9. Khan I, Ullah N, Zha L, et al. Alteration of gut microbiota in inflammatory bowel disease (IBD): cause or consequence? IBD treatment targeting the gut microbiome. *Pathogens* 2019;8:126.
10. Ni J, Wu GD, Albenberg L, Tomov VT. Gut microbiota and IBD: causation or correlation? *Nat Rev Gastroenterol Hepatol.* 2017;14:573-584.
11. Richard ML, Sokol H. The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases. *Nat Rev Gastroenterol Hepatol.* 2019;16:331-345.
12. Sartor RB, Wu GD. Roles for intestinal bacteria, viruses, and fungi in pathogenesis of inflammatory bowel diseases and therapeutic approaches. *Gastroenterology* 2017;152:327-339.e4.
13. Clooney AG, Eckenberger J, Laserna-Mendieta E, et al. Ranking microbiome variance in inflammatory bowel disease: a large longitudinal intercontinental study. *Gut.* 2021;70:499-510.
14. Ryan FJ, Ahern AM, Fitzgerald RS, et al. Colonic microbiota is associated with inflammation and host epigenomic alterations in inflammatory bowel disease. *Nat Commun.* 2020;11:1512.
15. Vich Vila A, Imhann F, Collij V, et al. Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. *Sci Transl Med.* 2018;10:eaap8914.
16. DeGruttola AK, Low D, Mizoguchi A, Mizoguchi E. Current understanding of dysbiosis in disease in human and animal models. *Inflamm Bowel Dis.* 2016;22:1137-1150.
17. Fritsch J, Garces L, Quintero MA, et al. Low-fat, high-fiber diet reduces markers of inflammation and dysbiosis and improves quality of life in patients with ulcerative colitis. *Clin Gastroenterol Hepatol.* 2021;19:1189-1199.e30.
18. Chabé M, Lokmer A, Ségurel L. Gut protozoa: friends or foes of the human gut microbiota? *Trends Parasitol.* 2017;33:925-934.
19. Fritsch J, Abreu MT. Candida in IBD: friend or foe? *Cell Host Microbe.* 2020;27:689-691.
20. Lukeš J, Stensvold CR, Jirků-Pomajbíková K, Wegener Parfrey L. Are human intestinal eukaryotes beneficial or commensals? *PLoS Pathog.* 2015;11:e1005039.
21. Virgin HW. The virome in mammalian physiology and disease. *Cell.* 2014;157:142-150.
22. Rossen NG, Bart A, Verhaar N, et al. Low prevalence of Blastocystis sp. in active ulcerative colitis patients. *European Soc Clin Microbiol.* 2015;34:1039-1044.
23. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome.* 2015;3:31.
24. Turnbaugh PJ, Ley RE, Hamady M, et al. The human microbiome project. *Nature.* 2007;449:804-810.
25. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet.* 2004;38:525-552.
26. Parfrey LW, Walters WA, Lauber CL, et al. Communities of microbial eukaryotes in the mammalian gut within the context of environmental eukaryotic diversity. *Front Microbiol.* 2014;5:298.
27. Underhill DM, Iliev ID. The mycobiota: interactions between commensal fungi and the host immune system. *Nat Rev Immunol.* 2014;14:405-416.
28. Findley K, Oh J, Yang J, et al.; NIH Intramural Sequencing Center Comparative Sequencing Program. Topographic diversity of fungal and bacterial communities in human skin. *Nature.* 2013;498:367-370.
29. Miyoshi J, Sofia MA, Pierre JF. The evidence for fungus in Crohn's disease pathogenesis. *Clin J Gastroenterol.* 2018;11:449-456.
30. Main J, McKenzie H, Yeaman GR, et al. Antibody to Saccharomyces cerevisiae (bakers' yeast) in Crohn's disease. *BMJ.* 1988;297:1105-1106.
31. Quinton JF, Sendid B, Reumaux D, et al. Anti-Saccharomyces cerevisiae mannan antibodies combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: prevalence and diagnostic role. *Gut.* 1998;42:788-791.
32. Forcione DG, Rosen MJ, Kisiel JB, Sands BE. Anti-Saccharomyces cerevisiae antibody (ASCA) positivity is associated with increased risk for early surgery in Crohn's disease. *Gut.* 2004;53:1117-1122.
33. Reese GE, Constantinides VA, Simillis C, et al. Diagnostic precision of anti-Saccharomyces cerevisiae antibodies and perinuclear antineutrophil cytoplasmic antibodies in inflammatory bowel disease. *Am J Gastroenterol.* 2006;101:2410-2422.
34. van Schaik FD, Oldenburg B, Hart AR, et al. Serological markers predict inflammatory bowel disease years before the diagnosis. *Gut.* 2013;62:683-688.
35. Standaert-Vitse A, Sendid B, Joossens M, et al. Candida albicans colonization and ASCA in familial Crohn's disease. *Am J Gastroenterol.* 2009;104:1745-1753.
36. Zwolinska-Wcislo M, Brzozowski T, Budak A, et al. Effect of Candida colonization on human ulcerative colitis and the healing of inflammatory changes of the colon in the experimental model of colitis ulcerosa. *J Physiol Pharmacol.* 2009;60:107-118.
37. Yan L, Yang C, Tang J. Disruption of the intestinal mucosal barrier in Candida albicans infections. *Microbiol Res.* 2013;168:389-395.
38. Richard ML, Lamas B, Liguori G, et al. Gut fungal microbiota: the Yin and Yang of inflammatory bowel disease. *Inflamm Bowel Dis.* 2015;21:656-665.
39. Qiu X, Zhang F, Yang X, et al. Changes in the composition of intestinal fungi and their role in mice with dextran sulfate sodium-induced colitis. *Sci Rep.* 2015;5:10416.

40. Sivignon A, de Vallée A, Barnich N, et al. Saccharomyces cerevisiae CNCM I-3856 prevents colitis induced by AIEC bacteria in the transgenic mouse model mimicking Crohn's disease. *Inflamm Bowel Dis.* 2015;21:276-286.
41. Herszenyi L, Miheller P, Tulassay Z. Carcinogenesis in inflammatory bowel disease. *Digest Dis.* 2007;25:267-269.
42. Wang C, Li W, Wang H, et al. Saccharomyces boulardii alleviates ulcerative colitis carcinogenesis in mice by reducing TNF-α and IL-6 levels and functions and by rebalancing intestinal microbiota. *BMC Microbiol.* 2019;19:246.
43. Bourreille A, Cadiot G, Le Dreau G, et al.; FLORABEST Study Group. Saccharomyces boulardii does not prevent relapse of Crohn's disease. *Clin Gastroenterol Hepatol.* 2013;11:982-987.
44. Garcia Vilela E, De Lourdes De Abreu Ferrari M, Oswaldo Da Gama Torres H, et al. Influence of Saccharomyces boulardii on the intestinal permeability of patients with Crohn's disease in remission. *Scand J Gastroenterol.* 2008;43:842-848.
45. Duina AA, Miller ME, Keeney JB. Budding yeast for budding geneticists: a primer on the Saccharomyces cerevisiae model system. *Genetics.* 2014;197:33-48.
46. Graf B, Adam T, Zill E, Göbel UB. Evaluation of the VITEK 2 system for rapid identification of yeasts and yeast-like organisms. *J Clin Microbiol.* 2000;38:1782-1785.
47. Allaband C, McDonald D, Vázquez-Baeza Y, et al. Microbiome 101: studying, analyzing, and interpreting gut microbiome data for clinicians. *Clin Gastroenterol Hepatol.* 2019;17:218-230.
48. Tang Q, Jin G, Wang G, et al. Current sampling methods for gut microbiota: a call for more precise devices. *Front Cell Infect Microbiol.* 2020;10:151.
49. Hallen-Adams HE, Suhr MJ. Fungi in the healthy human gastrointestinal tract. *Virulence.* 2017;8:352-358.
50. Limon JJ, Skalski JH, Underhill DM. Commensal fungi in health and disease. *Cell Host Microbe.* 2017;22:156-165.
51. Nash AK, Auchtung TA, Wong MC, et al. The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome.* 2017;5:153.
52. Sokol H, Leducq V, Aschard H, et al. Fungal microbiota dysbiosis in IBD. *Gut.* 2017;66:1039-1048.
53. Bliss JM, Basavegowda KP, Watson WJ, et al. Vertical and horizontal transmission of Candida albicans in very low birth weight infants using DNA fingerprinting techniques. *Pediatr Infect Dis J.* 2008;27:231-235.
54. David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature.* 2014;505:559-563.
55. Kabwe MH, Vikram S, Mulaudzi K, et al. The gut mycobiota of rural and urban individuals is shaped by geography. *BMC Microbiol.* 2020;20:257.
56. Strati F, Di Paola M, Stefanini I, et al. Age and gender affect the composition of fungal population of the human gastrointestinal tract. *Front Microbiol.* 2016;7:1227.
57. Ott SJ, Kühbacher T, Musfeldt M, et al. Fungi and inflammatory bowel diseases: Alterations of composition and diversity. *Scand J Gastroenterol.* 2008;43:831-841.
58. Qiu X, Ma J, Jiao C, et al. Alterations in the mucosa-associated fungal microbiota in patients with ulcerative colitis. *Oncotarget.* 2017;8:107577-107588.
59. Liguori G, Lamas B, Richard ML, et al. Fungal dysbiosis in mucosa-associated microbiota of Crohn's disease patients. *J Crohns Colitis.* 2016;10:296-305.
60. Hoarau G, Mukherjee PK, Gower-Rousseau C, et al. Bacteriome and mycobiome interactions underscore microbial dysbiosis in familial Crohn's disease. *MBio* 2016;7:e01250-16.
61. Jain U, Ver Heul AM, Xiong S, et al. Debaryomyces is enriched in Crohn's disease intestinal tissue and impairs healing in mice. *Science.* 2021;371:1154-1159.
62. Chehoud C, Albenberg LG, Judge C, et al. Fungal signature in the gut microbiota of pediatric patients with inflammatory bowel disease. *Inflamm Bowel Dis.* 2015;21:1948-1956.
63. Lewis JD, Chen EZ, Baldassano RN, et al. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host Microbe.* 2015;18:489-500.
64. Lemoinne S, Kemgang A, Ben Belkacem K, et al.; Saint-Antoine IBD Network. Fungi participate in the dysbiosis of gut microbiota in patients with primary sclerosing cholangitis. *Gut.* 2020;69:92-102.
65. Sarrabayrouse G, Elias A, Yáñez F, et al. Fungal and bacterial loads: noninvasive inflammatory bowel disease biomarkers for the clinical setting. *mSystems* 2021;6:e01277-20.
66. Petersen AM, Stensvold CR, Mirsepasi H, et al. Active ulcerative colitis associated with low prevalence of Blastocystis and Dientamoeba fragilis infection. *Scand J Gastroenterol.* 2013;48:638-639.
67. Tito RY, Chaffron S, Caenepeel C, et al. Population-level analysis of Blastocystis subtype prevalence and variation in the human gut microbiota. *Gut.* 2019;68:1180-1189.
68. Pérez-Brocal V, García-López R, Vázquez-Castellanos JF, et al. Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach. *Clin Transl Gastroenterol.* 2013;4:e36.
69. Wagner J, Maksimovic J, Farries G, et al. Bacteriophages in gut samples from pediatric Crohn's disease patients: metagenomic analysis using 454 pyrosequencing. *Inflamm Bowel Dis.* 2013;19:1598-1608.
70. Norman JM, Handley SA, Baldridge MT, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell.* 2015;160:447-460.
71. Pérez-Brocal V, García-López R, Nos P, et al. Metagenomic analysis of crohn's disease patients identifies changes in the virome and microbiome related to disease status and therapy, and detects potential interactions and biomarkers. *Inflamm Bowel Dis.* 2015;21:2515-2532.
72. Duerkop BA, Kleiner M, Paez-Espino D, et al. Murine colitis reveals a disease-associated bacteriophage community. *Nat Microbiol.* 2018;3:1023-1031.
73. Lane ER, Zisman TL, Suskind DL. The microbiota in inflammatory bowel disease: current and therapeutic insights. *J Inflamm Res.* 2017;10:63-73.
74. Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* 2016;14:e1002533.
75. Halwachs B, Madhusudhan N, Krause R, et al. Critical issues in mycobiota analysis. *Front Microbiol.* 2017;8:180.
76. Xie C, Goi CL, Huson DH, et al. RiboTagger: fast and unbiased 16S/18S profiling using whole community shotgun metagenomic or metatranscriptome surveys. *BMC Bioinformatics.* 2016;17:508.
77. Donovan PD, Gonzalez G, Higgins DG, et al. Identification of fungi in shotgun metagenomics datasets. *PLoS One.* 2018;13:e0192898.
78. West PT, Probst AJ, Grigoriev IV, et al. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res.* 2018;28:569-580.
79. Soverini M, Turroni S, Biagi E, et al. HumanMycobiomeScan: a new bioinformatics tool for the characterization of the fungal fraction in metagenomic samples. *BMC Genom.* 2019;20:7.
80. Lind AL, Pollard KS. Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. *Microbiome.* 2021;9:58.
81. Köljalg U, Nilsson RH, Abarenkov K, et al. Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol.* 2013;22:5271-5277.
82. Quince C, Walker AW, Simpson JT, et al. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol.* 2017;35:833-844.
83. Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci.* 2014;5:209.
84. Olm MR, West PT, Brooks B, et al. Genome-resolved metagenomics of eukaryotic populations during early colonization of premature infants and in hospital rooms. *Microbiome.* 2019;7:26.
85. Saary P, Mitchell AL, Finn RD. Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. *Genome Biol.* 2020;21:244.

86. Kucik CJ, Martin GL, Sortor BV. Common intestinal parasites. *Am Fam Physician.* 2004;69:1161-1168.
87. Gizaw Z, Adane T, Azanaw J, et al. Childhood intestinal parasitic infection and sanitation predictors in rural Dembiya, northwest Ethiopia. *Environ Health Prev Med.* 2018;23:26.
88. Schmidlin T, Hürlimann E, Silué KD, et al. Effects of hygiene and defecation behavior on helminths and intestinal protozoa infections in Taabo, Côte d'Ivoire. *PLoS One.* 2013;8:e65722.
89. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2018;390:2769-2778.
90. Zuo T, Kamm MA, Colombel JF, Ng SC. Urbanization and the gut microbiota in health and inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol.* 2018a;15:440-452.
91. Bach JF. The hygiene hypothesis in autoimmunity: the role of pathogens and commensals. *Nat Rev Immunol.* 2018;18:105-120.
92. Loke P, Lim YA. Helminths and the microbiota: parts of the hygiene hypothesis. *Parasite Immunol.* 2015;37:314-323.
93. Cho MK, Lee CH, Yu HS. Amelioration of intestinal colitis by macrophage migration inhibitory factor isolated from intestinal parasites through toll-like receptor 2. *Parasite Immunol.* 2011;33:265-275.
94. Reyes JL, Lopes F, Leung G, et al. Treatment with cestode parasite antigens results in recruitment of CCR2+ myeloid cells, the adoptive transfer of which ameliorates colitis. *Infect Immun.* 2016;84:3471-3483.
95. Smith P, Mangan NE, Walsh CM, et al. Infection with a helminth parasite prevents experimental colitis via a macrophage-mediated mechanism. *J Immunol.* 2007;178:4557-4566.
96. Croese J, O'neil J, Masson J, et al. A proof of concept study establishing Necator americanus in Crohn's patients and reservoir donors. *Gut.* 2006;55:136-137.
97. Fortun P, Shepherd V, Moroz V, et al. OC-004 effect of hookworm treatment on active Crohn's disease. *Gut* 2010;59:A2.
98. Sandborn WJ, Elliott DE, Weinstock J, et al. Randomised clinical trial: the safety and tolerability of Trichuris suis ova in patients with Crohn's disease. *Aliment Pharmacol Ther.* 2013;38:255-263.
99. Summers RW, Elliott DE, Qadir K, et al. Trichuris suis seems to be safe and possibly effective in the treatment of inflammatory bowel disease. *Am J Gastroenterol.* 2003;98:2034-2041.
100. Garcia LS. Dientamoeba fragilis, one of the neglected intestinal protozoa. *J Clin Microbiol.* 2016;54:2243-2250.
101. Wawrzyniak I, Poirier P, Viscogliosi E, et al. Blastocystis, an unrecognized parasite: an overview of pathogenesis and diagnosis. *Ther Adv Infect Dis.* 2013;1:167-178.
102. El Safadi D, Gaayeb L, Meloni D, et al. Children of Senegal River Basin show the highest prevalence of Blastocystis sp. ever observed worldwide. *BMC Infect Dis.* 2014;14:164.
103. Cacciò SM. Molecular epidemiology of Dientamoeba fragilis. *Acta Trop.* 2018;184:73-77.
104. Stark D, Roberts T, Marriott D, et al. Detection and transmission of Dientamoeba fragilis from environmental and household samples. *Am J Trop Med Hyg.* 2012;86:233-236.
105. Brands MR, Van de Vijver E, Haisma SM, et al. No association between abdominal pain and Dientamoeba in Dutch and Belgian children. *Arch Dis Child.* 2019;104:686-689.
106. Holtman GA, Kranenberg JJ, Blanker MH, et al. Dientamoeba fragilis colonization is not associated with gastrointestinal symptoms in children at primary care level. *Fam Pract.* 2016;34:25-29.
107. Jokelainen P, Hebbelstrup Jensen B, Andreassen BU, et al. Dientamoeba fragilis, a commensal in children in Danish day care centers. *J Clin Microbiol.* 2017;55:1707-1713.
108. Dullaert-de Boer M, Schuurs TA, Vermeer M, et al. Distribution and relevance of Dientamoeba fragilis and Blastocystis species in gastroenteritis: results from a case-control study. *Eur J Clin Microbiol Infect Dis.* 2019;39:197-203.
109. Marzano V, Mancinelli L, Bracaglia G, et al. "Omic" investigations of protozoa and worms for a deeper understanding of the human gut "parasitome". *PLoS Negl Trop Dis.* 2017;11:e0005916.
110. Clark CG, Stensvold CR. Blastocystis: isolation, xenic cultivation, and cryopreservation. *Curr Protoc Microbiol.* 2016;43:20a.21.21-20a.21.28.
111. Roberts T, Barratt J, Harkness J, et al. Comparison of microscopy, culture, and conventional polymerase chain reaction for detection of Blastocystis sp. in clinical stool samples. *Am J Trop Med Hyg.* 2011;84:308-312.
112. Cannon MV, Bogale H, Rutt L, et al. A high-throughput sequencing assay to comprehensively detect and characterize unicellular eukaryotes and helminths from biological and environmental samples. *Microbiome* 2018;6:11.
113. Zhang YZ, Shi M, Holmes EC. Using metagenomics to characterize an expanding virosphere. *Cell.* 2018;172:1168-1172.
114. Carding SR, Davis N, Hoyles L. Review article: the human intestinal virome in health and disease. *Aliment Pharmacol Ther.* 2017;46:800-815.
115. Gregory AC, Zablocki O, Zayed AA, et al. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe.* 2020;28:724-740.e8.
116. Mukhopadhya I, Segal JP, Carding SR, et al. The gut virome: the 'missing link' between gut bacteria and host immunity? *Therap Adv Gastroenterol.* 2019;12:1756284819836620.
117. Lavigne R, Molineux IJ, Kropinski AM. Order—caudovirales. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, eds. *Virus Taxonomy.* San Diego: Elsevier; 2012:39-45.
118. Dutilh BE, Reyes A, Hall RJ, Whiteson KL. Editorial: virus discovery by metagenomics: the (im)possibilities. *Front Microbiol.* 2017;8:1710.
119. Qin J, Li R, Raes J, et al.; MetaHIT Consortium. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* 2010;464:59-65.
120. Kleiner M, Hooper LV, Duerkop BA. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics.* 2015;16:7.
121. Roux S, Tournayre J, Mahul A, et al. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics.* 2014;15:76.
122. Li Y, Wang H, Nie K, et al. VIP: an integrated pipeline for metagenomics of virus identification and discovery. *Sci Rep.* 2016;6:23774.
123. Ren J, Ahlgren NA, Lu YY, et al. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome.* 2017;5:69.
124. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. *Nucleic Acids Res.* 2015;43:D571-D577.
125. Paez-Espino D, Chen IA, Palaniappan K, et al. IMG/VR: a database of cultured and uncultured DNA Viruses and retroviruses. *Nucleic Acids Res.* 2017;45:D457-D465.
126. Leplae R, Hebrant A, Wodak SJ, Toussaint A. ACLAME: a CLAssification of Mobile genetic Elements. *Nucleic Acids Res.* 2004;32:D45-D49.
127. Metsky HC, Siddle KJ, Gladden-Young A, et al.; Viral Hemorrhagic Fever Consortium. Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. *Nat Biotechnol.* 2019;37:160-168.
128. Levine A, Wine E, Assa A, et al. Crohn's disease exclusion diet plus partial enteral nutrition induces sustained remission in a randomized controlled trial. *Gastroenterology* 2019;157:440-450.e8.
129. Sales-Campos H, Soares SC, Oliveira CJF. An introduction of the role of probiotics in human infections and autoimmune diseases. *Crit Rev Microbiol.* 2019;45:413-432.
130. Singh RK, Chang HW, Yan D, et al. Influence of diet on the gut microbiome and implications for human health. *J Transl Med.* 2017;15:73.

131. Kassam Z, Lee CH, Yuan Y, Hunt RH. Fecal microbiota transplantation for Clostridium difficile infection: systematic review and meta-analysis. *Am J Gastroenterol.* 2013;108:500-508.
132. Ramai D, Zakhia K, Fields PJ, et al. Fecal microbiota transplantation (FMT) with colonoscopy is superior to enema and nasogastric tube while comparable to capsule for the treatment of recurrent clostridioides difficile infection: a systematic review and meta-analysis. *Digest Dis Sci.* 2021;66:369-380.
133. Sadowsky MJ, Khoruts A. Faecal microbiota transplantation is promising but not a panacea. *Nat Microbiol.* 2016;1:16015.
134. Caldeira LdF, Borba HH, Tonin FS, et al. Fecal microbiota transplantation in inflammatory bowel disease patients: a systematic review and meta-analysis. *PLoS One.* 2020;15:e0238910.
135. Costello SP, Soo W, Bryant RV, et al. Systematic review with meta-analysis: faecal microbiota transplantation for the induction of remission for active ulcerative colitis. *Aliment Pharmacol Ther.* 2017;46:213-224.
136. Lopez J, Grinspan A. Fecal microbiota transplantation for inflammatory bowel disease. *Gastroenterol Hepatol (N Y).* 2016;12:374-379.
137. Thomas AM, Segata N. Multiple levels of the unknown in microbiome research. *BMC Biol.* 2019;17:48.
138. Terveer EM, van Beurden YH, Goorhuis A, et al. How to: establish and run a stool bank. *Clin Microbiol Infect.* 2017;23:924-930.
139. Costello SP, Hughes PA, Waters O, et al. Effect of fecal microbiota transplantation on 8-week remission in patients with ulcerative colitis: a randomized clinical trial. *JAMA.* 2019;321:156-164.
140. Terveer EM, van Gool T, Ooijevaar RE, et al.; Netherlands Donor Feces Bank (NDFB) Study Group. Human transmission of blastocystis by fecal microbiota transplantation without development of gastrointestinal symptoms in recipients. *Clin Infect Dis.* 2020;2630-2636.
141. Zuo T, Wong SH, Lam K, et al. Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome. *Gut.* 2018;67:634-643.
142. Laforest-Lapointe I, Arrieta MC. Microbial eukaryotes: a missing link in gut microbiome studies. *mSystems* 2018;3:e00201-17.
143. Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T, et al. Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. *Front Microbiol.* 2021;12:634511.

Chapter 2

Statement of Authorship

Title of Paper	Individuals with inflammatory bowel disease have an altered gut microbiome composition of fungi and protozoa
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Under review in MDPI Microorganisms.

Principal Author

Name of Principal Author (Candidate)	Gina L. Guzzo		
Contribution to the Paper	Study conceptualisation, bioinformatic and statistical analyses, visualisation and interpretation of the data, wrote the manuscript.		
Overall percentage (%)	75		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	01/06/2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Study conceptualisation, provided substantial discussion and feedback to the research and data interpretation, edited all versions of the manuscript.		
Signature		Date	24/06/2022

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Provided discussion and feedback to the research and data interpretation, reviewed the manuscript.		
Signature		Date	08/06/2022

Name of Co-Author	Jane M. Andrews		
Contribution to the Paper	Provided discussion and feedback to the research, reviewed and provided feedback to the manuscript.		
Signature		Date	19/08/2022

Name of Co-Author	Murthy N. Mittinty		
Contribution to the Paper	Data interpretation and statistical analysis, reviewed and provided feedback to the manuscript.		
Signature		Date	28/06/2022

microorganisms

MDPI

Article

Individuals with Inflammatory Bowel Disease Have an Altered Gut Microbiome Composition of Fungi and Protozoa

Gina L. Guzzo [1,*], Murthy N. Mittinty [2], Bastien Llamas [1,3,4], Jane M. Andrews [5] and Laura S. Weyrich [1,6,*]

1 School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia
2 School of Public Health, University of Adelaide, Adelaide, SA 5005, Australia
3 Indigenous Genomics Research Group, Telethon Kids Institute, Adelaide, SA 5001, Australia
4 National Centre for Indigenous Genomics, Australian National University, Canberra, ACT 2601, Australia
5 Inflammatory Bowel Disease Service, Department of Gastroenterology and Hepatology, Faculty of Health Sciences, Royal Adelaide Hospital & School of Medicine, University of Adelaide, Adelaide, SA 5000, Australia
6 Department of Anthropology and Huck Institutes of the Life Sciences, Pennsylvania State University, State College, PA 16802, USA
* Correspondence: gina.guzzo@adelaide.edu.au (G.L.G.); lsw132@psu.edu (L.S.W.)

Abstract: It is known that the bacterial gut microbiome is altered in inflammatory bowel disease (IBD), but far less is known about the role of eukaryotic microorganisms in IBD. While eukaryotes are rarer than bacteria within the gastrointestinal environment, the current literature suggests that they may also be implicated in IBD. In our study, we characterized these often-neglected eukaryotic microbial communities by identifying fungi and protozoa in published shotgun stool metagenomes from 355 people with IBD (206 with Crohn's disease, 126 with ulcerative colitis, and 23 with IBD-unclassified) and 471 unaffected healthy individuals. The individuals with IBD had a higher prevalence of fungi, particularly *Saccharomyces cerevisiae*, and a lower prevalence of protozoa, particularly *Blastocystis* species (subtypes 1, 2, 3, and 4). Regression analysis showed that disease state, age, and BMI were associated with the prevalence and abundance of these two genera. We also characterized the eukaryotic gut microbiome in a shotgun stool metagenomic dataset from people with IBD who received fecal transplants, with samples pre- and post-transplantation, and from their donors. We found that in some FMT recipients, a single eukaryotic species remained stable over time, while in other recipients, the eukaryotic composition varied. We conclude that the eukaryotic gut microbiome is altered and varies over time in IBD, and future studies should aim to include these microbes when characterizing the gut microbiome in IBD.

Keywords: inflammatory bowel disease (IBD); Crohn's disease (CD); ulcerative colitis (UC); gut microbiome; fungi; protozoa; eukaryote

Citation: Guzzo, G.L.; Mittinty, M.N.; Llamas, B.; Andrews, J.M.; Weyrich, L.S. Individuals with Inflammatory Bowel Disease Have an Altered Gut Microbiome Composition of Fungi and Protozoa. *Microorganisms* **2022**, *10*, 1910. https://doi.org/10.3390/microorganisms10101910

Academic Editor: Marcus J. Claesson

Received: 16 August 2022
Accepted: 22 September 2022
Published: 26 September 2022

1. Introduction

Inflammatory bowel disease (IBD) is a chronic relapsing-remitting condition of the gastrointestinal tract and includes Crohn's disease (CD) and ulcerative colitis (UC) [1]. IBD results from complex interactions between the host immune system, environment, and gut microbiome [2–4]. The gut microbiome refers to an assemblage of gastrointestinal microorganisms, their genes and gene products, and microenvironment [5], and it has become a popular topic in IBD research due to its observed links to the etiopathology of IBD [6]. However, a consensus as to what specifically constitutes a gut microbiome signature in IBD remains lacking, due to the high variability of microbiomes between individuals and within individuals over time, and different methodologies between studies [7,8]. Nevertheless, therapies to treat IBD by targeting the gut microbiome are being developed, including faecal microbiota transplantation (FMT), even though we do not yet understand the roles that all types of microbes play in this disease [9–12]. It is likely that a better understanding

of the role of the gut microbiome in IBD will enhance the development of novel diagnostics and therapies [8].

While the microbiome continues to offer an avenue for understanding and treating IBD, the majority of IBD research has focussed on intestinal bacteria [13]. Other intestinal microbes, such as eukaryotes (fungi and protozoa), are poorly studied in the context of IBD. Microbial eukaryotes make up a small proportion of gut microbiota relative to bacteria, and they are thus often overlooked or are not captured in microbiome studies [14]. For example, fungal DNA is estimated to comprise less than 1% of microbial DNA in the gut microbiome [15–17], and the ratio of fungi to bacteria may differ across the intestinal tract, resulting in highly variable functional contributions [16,18]. Fungi are generally thought to be pathogenic in IBD due to several reports of fungal overgrowth in the gut during disease, including increased abundances of yeasts, such as *Candida*, *Malassezia*, and *Saccharomyces* species, localisation of fungi to sites of intestinal inflammation in people with CD, and increased anti-fungal antibodies identified in IBD [17,19–22]. An overgrowth of *Candida* is hypothesised to promote intestinal inflammation and the proliferation of other opportunistic bacteria and fungi [23–25]. However, findings across studies of the fungal microbiome in IBD are inconsistent largely due to different methods, sample types, and an overall paucity of research [13,18,19].

Protozoa are the rarest intestinal eukaryotes and, consequently, the least studied eukaryotic types in IBD [13]. The two most common protozoa in the human gut are *Blastocystis* species and *Dientamoeba fragilis* [26,27]. While once thought to be pathogenic, more recent research suggests that these protozoa are asymptomatic colonisers of the gut, and may even be markers of health [15,28,29]. However, the commensal status of these species still remains controversial [30]. In IBD, a handful of studies have shown that affected individuals have a reduction in these protozoa compared to healthy individuals [31–33]. Therefore, a reduction in these species offers the prospect of being used as an indicator of IBD, and further investigation is needed to understand their prevalence in the disease and its subtypes.

In our study, we characterised the fungal and protozoan fraction of the gut microbiome in three IBD subtypes—CD, UC and IBD unclassified (IBDU)—and healthy individuals without IBD. We also identified microbial eukaryotes in a longitudinal FMT study of individuals with CD, which included samples from pre- and post-FMT time points and donor samples. We hypothesised that the eukaryotic microbiome, like the bacterial microbiome, would differ in IBD. To the best of our knowledge, this is the first large-cohort study of the eukaryotic microbiome in IBD to identify eukaryotic species using metagenomic sequencing data.

2. Materials and Methods

2.1. Study Populations and Design

We performed an observational case–control study to identify microbial eukaryotes in gut microbiome samples in publicly available datasets. This study included 826 individuals: 355 adults with IBD and 471 adult healthy control subjects (Table 1). The IBD samples were from the 1000IBD study [34] and consisted of 355 shotgun stool metagenome samples from 355 subjects with three subtypes of IBD: 206 with CD, 126 with UC and 23 with IBDU. The control samples were from the 500FG study [35] and consisted of 471 shotgun stool metagenome samples from 471 healthy subjects who did not have IBD. These two cohorts were selected for comparison for several reasons: both cohorts were from the Netherlands and both studies used the same microbial DNA extraction (Qiagen AllPrep DNA/RNA Mini Kit with mechanical lysis) and sequencing methods (whole metagenome shotgun sequencing with Illumina HiSeq).

Table 1. Summary characteristics of the study groups included in regression modelling.

Cohort Characteristics		IBD Group (n = 355)	Control Group (n = 471)
Sex	Female	214 (60.28%)	265 (56.26%)
	Male	141 (39.72%)	200 (42.46%)
	Unspecified	0 (0.00%)	6 (1.27%)
Age (years)	Median (IQR)	45.00 (34.25, 59.00)	23.00 (21.00, 27.00)
	18–40	145 (41.43%)	408 (87.74%)
	41–60	132 (37.71%)	31 (6.67%)
	61–80	69 (19.71%)	26 (5.59%)
	81+	4 (1.14%)	0 (0.00%)
	Missing	5	6
BMI	Median (IQR)	24.80 (21.70, 28.10)	22.30 (20.72, 24.39)
	Missing	0	14
Smoking Status	Current	78 (22.35%)	60 (14.12%)
	Past	146 (41.83%)	65 (15.29%)
	Never	125 (35.82%)	300 (70.59%)
	Missing	6	46
Diagnosis	CD	206 (58.03%)	NA
	UC	126 (35.49%)	
	IBDU	23 (6.48%)	

In addition to our observational case–control study, we investigated the eukaryotic gut microbiome composition of individuals with IBD and donors in an FMT trial. The dataset included 115 longitudinal shotgun stool metagenome samples from 17 adults with CD pre- and post-FMT, and single timepoint samples from their healthy FMT donors (n = 5), originally published in Kong et al. [12]. All participants in this study were recruited in France. The FMT recipients achieved clinical remission via oral corticosteroids prior to receiving FMT by colonoscopy. The recipients were randomised to receive either donor FMT (n = 8) or a sham FMT (n = 9), which consisted of the transplant serum (saline solution) alone. The five donors were divided amongst the eight recipients accordingly: three donors were allocated to a single recipient each, one donor was allocated to two recipients, and the last donor was allocated to three recipients. The donor who was allocated to two recipients only had one sample present in the dataset, thus there were a total of seven donor samples instead of an expected eight. A successful outcome of FMT was defined by Sokol et al. [36] as steroid-free clinical remission at 10 weeks post-FMT.

2.2. *Processing Metagenomic Sequences*

Samples from the 1000IBD and 500FG datasets were quality-controlled with fastp version 0.20.0 [37] with parameters to trim polyG (-g) and polyX tails (-x), filter low complexity reads (-y), reduce overrepresentation of reads (-p), and correct bases in overlapping regions (-c). KneadData version 0.7.2 (https://github.com/biobakery/kneaddata, accessed on 14 August 2022), a pipeline for processing metagenomic sequencing data, was then used to remove human DNA contamination from the samples by discarding all sequencing reads that aligned to the human reference genome GRCh37/hg19 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/, accessed on 14 August 2022) with Bowtie2 [38]. The pipeline was run with settings to bypass the read trimming step (–bypass-trim) and to remove intermediate output files (–remove-intermediate-output). Before quality control, the total read count per sample in the IBD dataset ranged from 4.1 to 26.2 million sequences, with a mean read count of 10.9 million (SD: 3.8 million). After quality control, the read counts ranged from 2.7 to 25.7 million sequences per sample, reducing the mean read count to 10.6 million (SD: 3.8 million). For the control dataset, the total read count per sample before quality control ranged from 2.4 to 34.2 million sequences and had an average of 15.1 million sequences (SD: 4.1 million). After processing the sequences, the

read count per sample ranged from 2.3 to 33.2 million sequences, reducing the average read count to 13.7 million (SD: 4.0 million).

The FMT study samples had previously been bioinformatically processed for sequencing quality control steps and to remove human DNA contamination, as described in [12], and we therefore did not perform these steps. The total read count per sample in the dataset ranged from 1.7 to 32.1 million sequences, with an average of 15.9 million (SD: 6.3 million) sequences per sample.

2.3. Identifying Eukaryotes in Gut Metagenomes

To identify eukaryotes, we used two metagenomic profiling tools: RiboTagger version 0.8.0 (https://github.com/xiechaos/ribotagger, accessed on 14 August 2022), which identifies eukaryotic DNA based on the 18S rRNA gene (v4, v5, v6, and v7 regions) [39], and EukDetect version 1.2 (https://github.com/allind/EukDetect, accessed on 14 August 2022), which identifies eukaryotic DNA by aligning sequences to a database of 521,824 microbial eukaryote marker genes [40]. Following the aforementioned sequence processing steps, RiboTagger was run on the 826 IBD and control samples with settings to identify sequences from any of the four 18S v regions (-r v4 v5 v6 v7), flags to exclude bacterial (-no-bacteria) and archaeal DNA (-no-archaea), and assigning taxonomy of putative eukaryotic DNA with the SILVA v119 database (https://www.arb-silva.de/documentation/release-119/, accessed on 14 August 2022). EukDetect was also run on the same 826 samples through the entire pipeline (–mode runall). The full EukDetect pipeline included a filtration step to remove secondary marker gene hits to reduce false positive detections [40]. After exploratory analysis of the results comparing RiboTagger and EukDetect, we decided to proceed with further analysis using only EukDetect due to its increased sensitivity for identifying eukaryotes. The 826 samples were then sorted by their FASTQ header read names (sort -k1,1 -t) to ensure matching order of forward and reverse reads and rarefied using seqtk version 1.3-r106 (https://github.com/lh3/seqtk, accessed on 14 August 2022) set to a random seed of 100 (-s100). All samples were normalised by rarefying them to 2.28 million sequences per sample, as this was the size of the smallest sample between both cohorts, ensuring that no samples were lost in the process. This rarefaction step was performed to enable abundance analysis by mitigating the effect of different sequencing library sizes on the abundances of detected eukaryotes.

We also used EukDetect to identify eukaryotes in the FMT study samples. Eukdetect was run on the samples, before and after rarefying the samples to 1.7 million sequences per sample, the size of the smallest sample (as described above). Before rarefying, 43 samples had eukaryotes detected in them, but rarefying reduced this sample size down to 13 samples. Detailed differential abundance analysis would not be possible with this small detection size, so we proceeded with unrarefied data for our analysis to maximize the recovery of any eukaryotes.

2.4. Statistical Analyses

2.4.1. Descriptive Taxonomic Analysis

Eukaryotic distributions were examined using prevalence and abundances. Prevalence was calculated from the proportion of individuals who had a eukaryote in their sample compared to the total size of their respective cohort groupings. Abundances were equivalent to read counts of each eukaryote, post-rarefaction. The analysis of eukaryote detections was conducted in R version 4.0.2 [41] with phyloseq version 1.38.0 [42], and plots were generated with *ggplot2* version 3.3.5 [43].

2.4.2. Cohort Analysis

The characteristics of the 355 IBD and 471 healthy group that were included in our regression analysis are described in Table 1. We used counts with percentages to describe categorical variables and medians with interquartile ranges to explain the distributions

of continuous variables. The summary table was generated with the R package *qwraps2* version 0.5.2 [44].

Both cohorts had a higher proportion of females than males (IBD cohort female proportion: 60.28%; healthy cohort female proportion: 56.26%) and had a median BMI in the normal range. The median age of the IBD cohort was higher than the control subject group. The majority of individuals in both cohorts were non-smokers at the time of study, with 77.65% non-smokers in the IBD group and 85.88% non-smokers in the healthy group. The healthy cohort included an additional variable for smoking status—household smoker—which was not recorded for the IBD cohort. Individuals who were recorded as having a household smoker were excluded from the analysis as there was no comparator information available. The highest number of missing cases was 14 BMI values in the healthy cohort, and since this was less than 5% of the total cohort (2.97%), we did not impute values and instead performed a complete-case analysis [45].

2.4.3. Regression Analysis

We performed a regression analysis to model the relationship between the abundances of eukaryotic genera found in the 355 IBD and 471 healthy groups with the demographic data of these cohorts (Table 2). Most species had too low of a prevalence to be included in the regression model; therefore, we only performed this analysis on the two most prevalent genera, *Blastocystis* and *Saccharomyces*. We used a generalised linear model (GLM), and our outcome variable was abundance, which was measured as a count variable. Usually when the outcome variable is a count variable, the family chosen for GLM is Poisson, with a *log* link function to estimate the incidence risk ratio [46]. However, many individuals in our cohort had no eukaryotes detected in their samples, and the abundance count was therefore zero for these individuals. We therefore used a zero-inflated negative binomial model (ZINB) over a Poisson to account for the surplus of zero abundance counts [47].

Table 2. Regression analysis results.

Model	Coefficient	*Blastocystis* spp. [1]			*Saccharomyces cerevisiae*		
		Estimate	Std. Error	*p*-Value	Estimate	Std. Error	*p*-Value
Count	IBD	−0.34	0.39	0.38	1.05	0.34	**<0.01**
	Sex—male	0.23	0.22	0.29	−1.15	0.41	**0.01**
	Age	0.01	0.01	0.35	−0.03	0.01	**<0.01**
	BMI—underweight	−0.85	0.40	**0.04**	−2.19	0.61	**<0.01**
	BMI—overweight	−0.81	0.28	**<0.01**	−0.69	0.45	0.13
	BMI—obese	−1.03	0.43	**0.02**	0.27	0.48	0.58
	Smoking—past	−0.15	0.33	0.64	0.52	0.50	0.30
	Smoking—current	−0.44	0.38	0.25	−0.68	0.38	0.08
Zero	IBD	1.84	0.37	**<0.01**	−1.40	0.44	**<0.01**
	Sex—male	0.32	0.24	0.18	−0.47	0.35	0.18
	Age	−0.02	0.01	**0.02**	0.01	0.01	0.37
	BMI—underweight	−0.03	0.68	0.97	−0.06	1.27	0.96
	BMI—overweight	0.46	0.34	0.17	0.01	0.43	0.98
	BMI—obese	0.88	0.78	0.26	−0.53	0.50	0.29
	Smoking—past	0.19	0.31	0.55	0.48	0.45	0.29
	Smoking—current	0.66	0.39	0.10	−0.42	0.39	0.27

[1] *Blastocystis* spp. includes *B. hominis* and *Blastocystis* sp. subtypes 1, 2, 3, and 4.

By using a ZINB, we assumed that an outcome of zero abundance was due to two reasons: (1) zero counts were due to technical effects, such as batch effects and inadequate sequencing depth, and (2) the eukaryote was not present in the individual (i.e., biological zero). ZINB was performed with the R package *pscl* version 1.5.5 [48] and consisted of two parts: a binary (logit) model to predict whether each zero outcome was due to technical or biological effects and a count model (negative binomial) to model the abundance counts of *Blastocystis* and *Saccharomyces*. We used a negative binomial count model instead of Poisson because we were expecting excess variation even when counts were greater than zero. To estimate robust standard errors, we used the R *sandwich* package version 3.0-1 [49]. Statistical analysis was performed in R version 4.0.2 [41].

3. Results

3.1. Detection of Eukaryotes in Gut Metagenomes

We profiled the gut microbiomes of 355 people with IBD and 471 healthy individuals with two metagenomic marker gene tools—RiboTagger and EukDetect—and found that eukaryotes were uncommon and only present in some of the individuals' samples. Most individuals did not have any eukaryotes identified in their sample (Figure 1A). RiboTagger identified eukaryotes in only 24 of the 355 people with IBD (6.8%) and in 74 of the 471 healthy individuals (15.7%). Though EukDetect identified more eukaryotes in the individuals than RiboTagger, the proportion of individuals with a eukaryote present was still in the minority with detection of eukaryotes in 100 of the people with IBD (28.2%) and 177 healthy individuals (37.6%). The finest resolution that could be obtained with RiboTagger was at the genus level because it identifies eukaryotes solely based on the 18S rRNA region. In contrast, EukDetect's database is better adapted to shotgun sequencing data and covers a broader range of eukaryotic diversity due to the inclusion of a large number of conserved marker genes. Therefore, EukDetect consistently identified more eukaryotic species than RiboTagger in both IBD and control subject groups, and within the IBD subtypes (Figure 1A,B; Tables S1 and S2). However, unlike RiboTagger, EukDetect did not identify *Dientamoeba* and *Galactomyces* because neither genus was present in the Eukdetect database.

Most individuals had only one eukaryotic genus detected in their sample. Ten individuals with IBD and 29 healthy individuals had more than one eukaryote detected in their sample with EukDetect, and no individuals with IBD and 14 healthy individuals had more than one eukaryote in their sample detected with RiboTaggger (Tables S1 and S2). A total of 13 eukaryotic genera were detected in the IBD group between the two marker gene tools (Figure 1C). These genera were: *Blastocystis, Candida, Clavispora*, *Cyberlindnera*, *Debaryomyces*, *Dientamoeba*, *Galactomyces*, *Malassezia*, *Meyerozyma*, *Nakaseomyces*, *Penicillium*, *Saccharomyces*, *Wickerhamomyces*. The IBDU group had the highest proportion of individuals with eukaryotes detected by EukDetect (47.8%) out of the three IBD subtypes (Figure 1B), and the proportions of individuals with eukaryotes in the CD and UC groups were similar to each other (CD: 26.2% and UC: 27.8%). The control subject group had 11 genera identified between both tools: *Blastocystis*, *Candida*, *Cyberlindnera*, *Debaryomyces*, *Dientamoeba*, *Giardia*, *Hanseniaspora*, *Malassezia*, *Penicillium*, *Pichia*, *Saccharomyces*. In total, the IBD cohort contained more fungal genera (11) than the control subject group (8). Conversely, three protozoan genera were present in the control group, whereas only one was found in the IBD group.

3.2. Distribution and Prevalence of Eukaryotes

3.2.1. Proportional Abundances

After confirming that eukaryotes were present and detectable in the samples, we proceeded to assess shifts in the distribution of eukaryotes using the EukDetect results on rarefied samples. All samples were rarefied to 2.28 million sequences per sample in order to compare abundances and ensure that sequencing library size did not affect the abundance results. Rarefaction resulted in the loss of three genera from the detection

threshold—*Giardia*, *Hanseniaspora*, and *Meyerozyma* (Table S3)—which was a side effect expected with this technique [50,51]. The total proportional abundances for each genera showed a difference between IBD and healthy subject group compositions; most notably a high proportional abundance of *Saccharomyces* in IBD (63.1%) and high proportional abundance of *Blastocystis* in the control subjects (93.0%) (Figure 2A). The composition for CD most closely resembled the total IBD composition, with a notable similarity in *Saccharomyces* abundances (63.1% in the total IBD group and 64.9% in CD) in contrast to UC (29.6%) and IBDU (88.8%) (Figure 2B). UC had the highest abundance of *Blastocystis* (45.2%), compared to CD (16.8%) and IBDU (11.2%). Across all groups, *Blastocystis* and *Saccharomyces* were the most abundant genera.

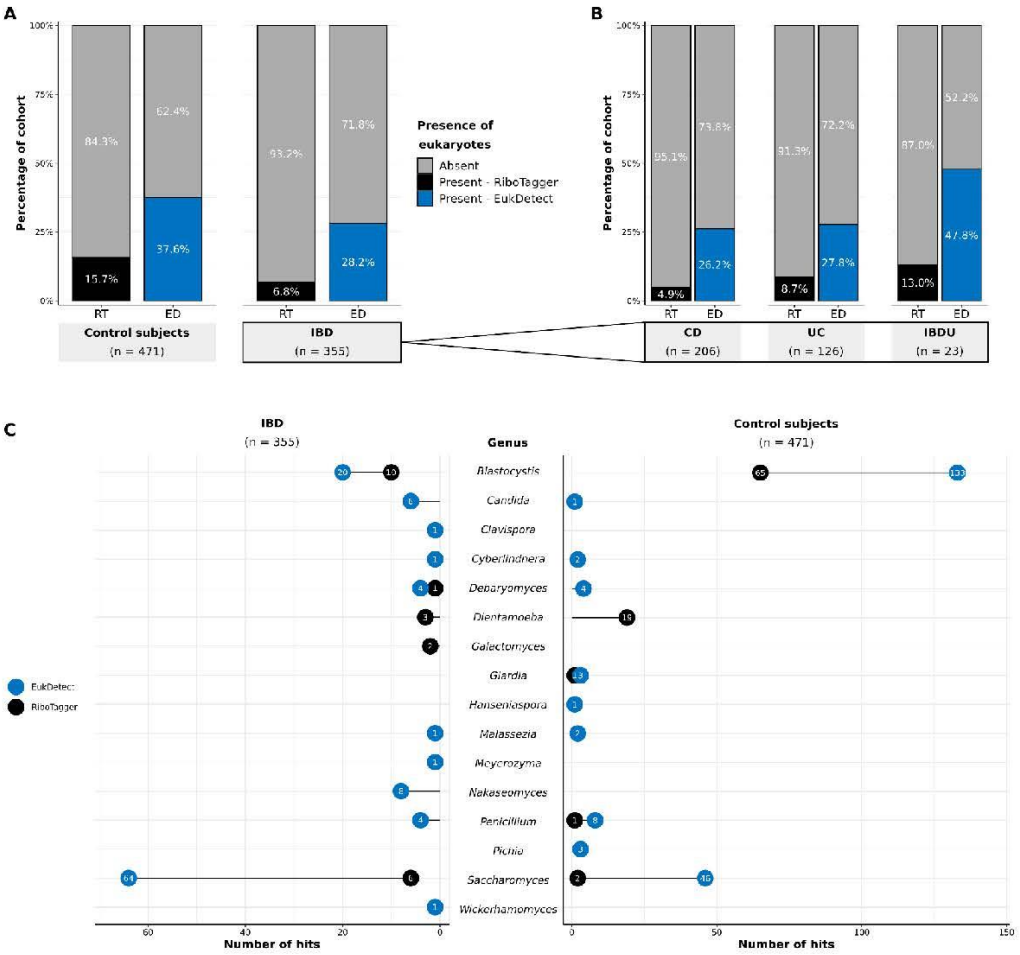

Figure 1. Prevalence of microbial eukaryotes in shotgun stool metagenomes of people with inflammatory bowel disease (IBD) (n = 355) and unaffected individuals (control subjects) (n = 471). (**A**) The proportion of individuals with microbial eukaryotes in their sample with at least family-level resolution is shown according to results detected by RiboTagger (RT) and EukDetect (ED) in control subject and IBD cohorts; (**B**) The same results are displayed as a proportion of samples with microbial eukaryotes detected in individuals with their IBD subtype—Crohn's disease (CD) (n = 206), ulcerative colitis (UC) (n = 126), and IBD unclassified (IBDU) (n = 23)—of the total IBD cohort; (**C**) The detection of microbial eukaryotes with at least genus-level resolution by RiboTagger and EukDetect in people with IBD and control subjects. A hit represents a single taxon identified in one sample. Individuals in all groups had only one sample per individual. Results are based on unrarefied sequencing data.

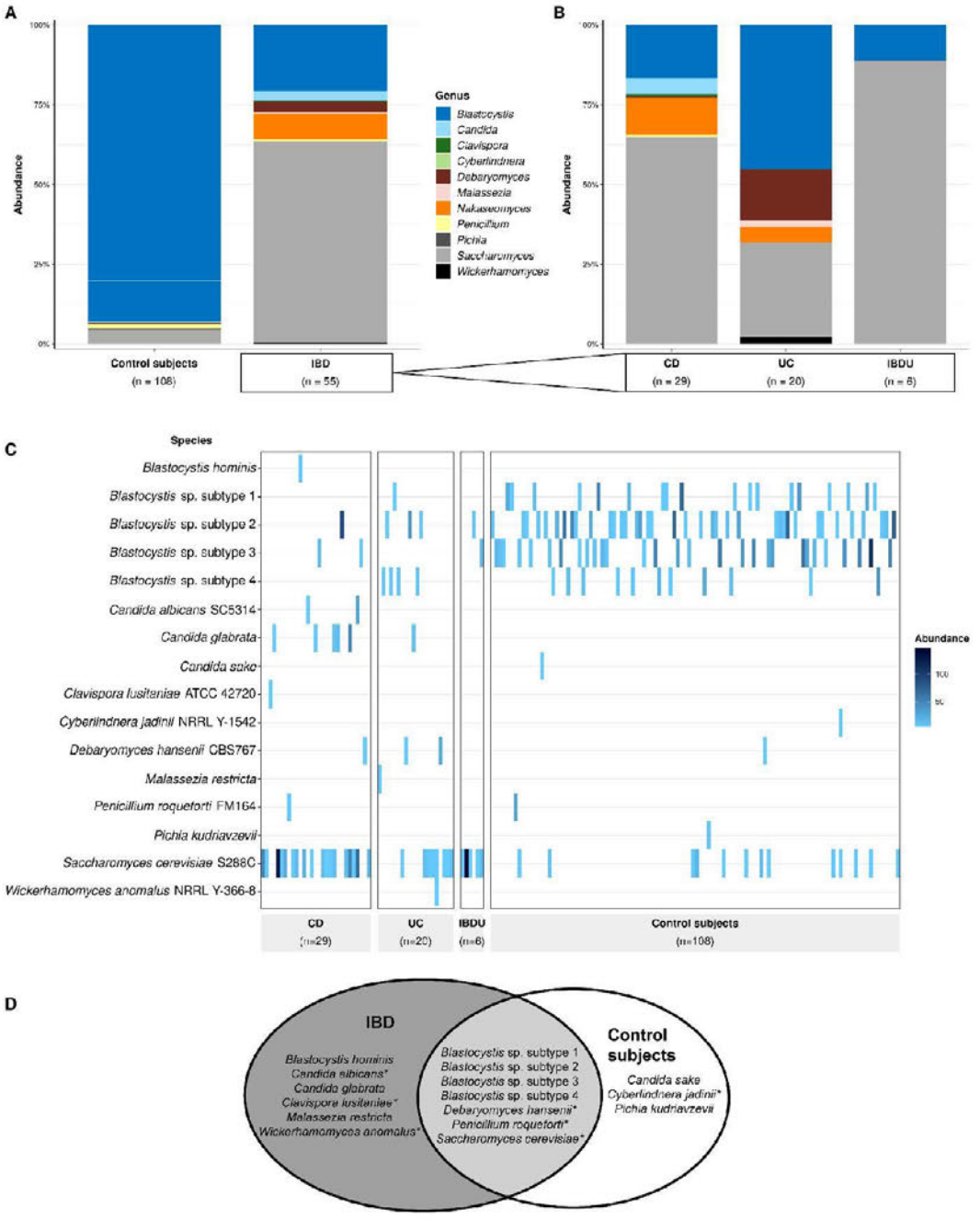

Figure 2. The prevalence and distribution of microbial eukaryotes in shotgun stool metagenomes is shown in people with inflammatory bowel disease (IBD) (n = 355) and unaffected healthy people (control subjects) (n = 471). (**A**) Total proportional abundances of microbial eukaryotes detected at the species-level in control subjects (n = 108) and individuals with IBD (n = 55); (**B**) Total proportional abundances are also broken down for all the people with IBD into each of their IBD subtypes—Crohn's disease (CD) (n = 29), ulcerative colitis (UC) (n = 20), and IBD unclassified (IBDU) (n = 6); (**C**) Distribution and abundances of microbial eukaryotes detected at the species-level in people with IBD, specified by their IBD subtype, compared to the control subject group (n = 108); (**D**) Eukaryotic species unique to

IBD and control subject groups and shared between groups. A species was considered shared if it was found in at least one individual in the IBD group and one individual in the control subject group. * Complete species names are: *Candida albicans* SC5314, *Clavispora lusitaniae* ATCC 42720, *Cyberlindnera jadinii* NRRL Y-1542, *Debaryomyces hansenii* CBS767, *Penicillium roqueforti FM164*, *Saccharomyces cerevisiae* S288C, and *Wickerhamomyces anomalus* NRRL Y-366-8. Results are based on EukDetect identifications in sequencing data rarefied to 2.28 million sequences per sample.

3.2.2. Species-Level Eukaryotic Prevalence

Since most individuals only had a single eukaryotic taxon in their sample, we were unable to perform detailed microbiome diversity analyses. Nevertheless, we were able to examine the distribution of eukaryotes across both cohorts at the species level due to using whole-genome marker genes with EukDetect. The IBD cohort had a greater number of species (13) than the control subject group (10) despite having a smaller overall cohort size (Figure 2C). *Candida* spp. were uncommon between both cohorts. Eight individuals with IBD (2.3%) had *Candida* spp. in their samples; two individuals (0.6%) with *C. albicans* SC5314 and six individuals (1.7%) with *C. glabrata*. Individuals with CD had most of the *Candida* spp. in the IBD cohort; seven individuals with CD (3.4%), compared to one individual with UC (0.8%) and no individuals with IBDU. In contrast, only one control subject (0.2%) had a single *Candida* species: *C. sake*, *Saccharomyces cerevisiae* S288C were notably more prevalent in the IBD group than the healthy group, being present in 31 people with IBD (8.7%) but only 13 healthy people (2.8%). *S. cerevisiae* S288C was present in 18 individuals with CD (8.7%), eight individuals with UC (6.3%), and five individuals with IBDU (21.7%). *Blastocystis* spp. were comparatively more prevalent in the healthy individuals than the individuals with IBD; 95 healthy individuals (20.2%) had at least one *Blastocystis* species in their sample whereas only 14 individuals with IBD (3.9%) had one *Blastocystis* species in their sample. *Blastocystis* spp. were present in four individuals with CD (1.9%), eight individuals with UC (6.4%) and most prevalent in two individuals with IBDU (8.7%). *Blastocystis* sp. subtype 2 was the most prevalent subtype between both cohorts—present in five individuals with IBD (1.4%) and 38 healthy individuals (8.1%).

3.2.3. Shared and Unique Species

Given that eukaryotes were rare among the cohorts, we used a generous threshold and considered a species shared if it was present in at least one individual in each cohort (Figure 2D). The shared eukaryotic microbiome between all individuals with IBD and healthy individuals included all *Blastocystis* sp. subtypes (1, 2, 3, and 4), and the yeasts *Debaryomyces hansenii* CBS767 and *Saccharomyces cerevisiae* S288C. Five of the six yeast species were uniquely identified in the IBD group. Additionally, *Blastocystis hominis* was found only in the IBD group, as it was identified in a single individual with CD (Figure 2C). The three species that were unique to the control subjects, *C. sake*, *Pichia kudriavzevii*, and *Cyberlindnera jadinii* NRRL Y-1542, were rare and each only identified in three discrete individuals (Figure 2C,D).

3.3. Effect of Faecal Microbiota Transplantation on Eukaryotes

We detected eukaryotes in longitudinal samples from individuals with CD who received FMT and their FMT donor samples, to investigate whether FMT has an effect on the eukaryotic microbiome. Twelve of the 17 people with CD had eukaryotes detected in their sample for at least one time point; six individuals who received donor FMT and six individuals who were in the control (sham) FMT group (Table S4). Only one of the seven donor samples had a eukaryotic species present (Figure 3). There did not appear to be a relationship between the type of species, FMT group, and FMT outcome (measured as steroid-free clinical remission at 10 weeks post-FMT) of the recipients. Several FMT recipients—recipients one, four, nine, and 15—maintained the same eukaryotic species both before and after receiving donor FMT. These individuals all experienced remission for 10 weeks post-FMT, suggesting that some eukaryotes may not interfere with success

following FMT. One eukaryotic species, *Blastocystis* sp. subtype 1, was also identified in a donor sample (donor 47) but this species was not present in their recipient's samples. Instead, their recipient (subject 15) maintained *Penicillium roqueforti* FM164 before and after FMT. Several fungal species including *P. roqueforti* FM164 were present in individuals with both success and failure outcomes—*D. hansenii* CBS767, *P. roqueforti* FM164, *P. kudriavzevii*, and *S. cerevisiae* S288C—which indicates a more complex relationship between intestinal fungi and FMT than can be resolved from this single, small study.

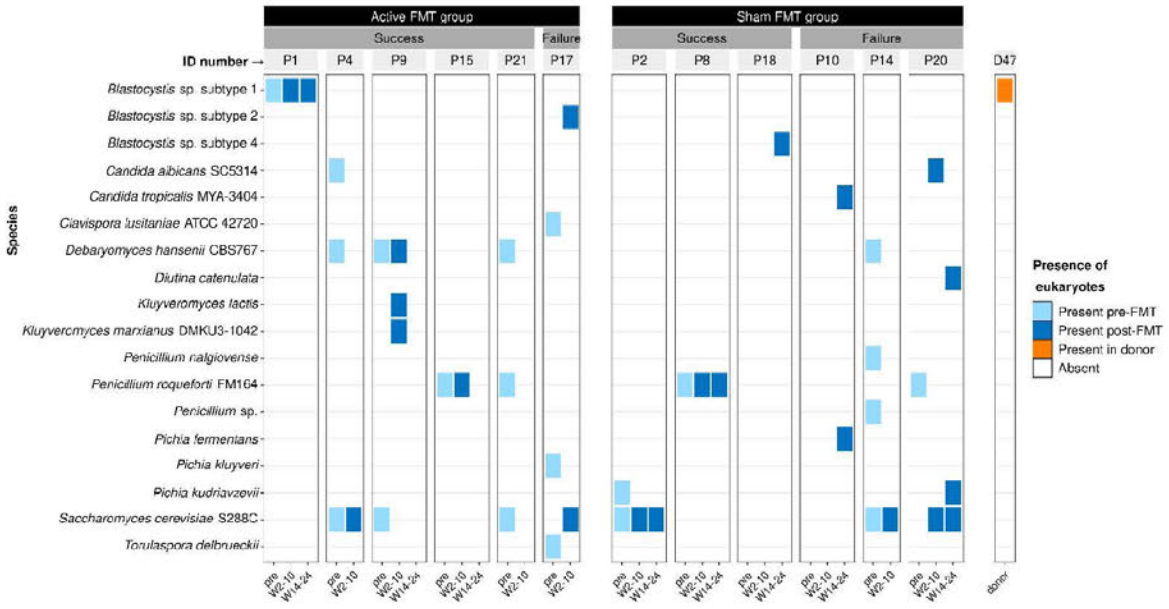

Figure 3. Longitudinal prevalence and distribution of microbial eukaryotes detected in shotgun stool metagenome samples from people with Crohn's disease (CD) who received faecal microbiota transplants (FMT). Eukaryotes were also identified in a single donor sample (donor 47) whose recipient was patient 15. FMT recipients in the Active FMT group received a donor sample, and recipients in the Sham FMT group received a sample of transplant serum (saline solution) only. Success refers to steroid-free clinical remission at ten weeks post-FMT. Individuals are identified by their ID numbers, with P preceding patient ID's and D preceding the donor ID. 'Pre' is pre-FMT and includes two weeks before and day of FMT, 'W2-10' includes results from weeks 2, 6, and 10 post-FMT, and 'W14-24' includes results from weeks 14, 18, and 24 post-FMT. Recipients 4, 14, 17, and 21 did not have any samples for weeks 14–24. Results are based on EukDetect identifications in unrarefied sequencing data.

3.4. Influence of Cohort Demographics on Gut Eukaryote Composition

We performed regression analysis to model the relationship between demographics of the IBD and healthy cohorts and the prevalence and abundance of microbial eukaryotes in their gut microbiomes. The cohort characteristics summarised in Table 1 were incorporated into a zero-inflated negative binomial regression model to determine the associations of these characteristics on the abundances of *Blastocystis* and *Saccharomyces*.

Due to the low number of individuals with each species of eukaryote, we grouped the *Blastocystis* spp. identified between both cohorts (*B. hominis*, and *Blastocystis* sp. subtypes 1, 2, 3, and 4) (Figure 2C) at the genus level and modelled the association between the presence and abundance of this genera and the cohort demographics summarised in Table 1. Disease state was statistically associated with the prevalence of *Blastocystis* spp. but not the abundance of *Blastocystis* spp. (Table 2). This result indicates that only the presence of *Blastocystis* spp. was linked to a healthy state, and within those who had *Blastocystis* spp., the abundance of this genus did not statistically differ between individuals with or without

IBD. BMI was associated with *Blastocystis* spp. counts; being underweight, overweight, and obese predicted a lower abundance of *Blastocystis*. Age was also negatively associated with the presence of *Blastocystis* spp., and therefore, older individuals were less likely to have *Blastocystis* spp. in their samples.

We also modelled the association between *S. cerevisiae* and the cohort demographics. In contrast to *Blastocystis* spp., we found that IBD was a predictor of *S. cerevisiae* being present and more abundant. BMI also had an association with *S. cerevisiae*, wherein being underweight predicted a lower abundance of this species. Lastly, age and sex were positively associated with *S. cerevisiae*, and an increase in age or being male predicted an increase in *S. cerevisiae* abundance.

4. Discussion

In this study, we explored the eukaryotic fraction of the gut microbiome in individuals with IBD and in healthy individuals without IBD. Previous studies of the eukaryotic microbiome in IBD have primarily used amplicon sequencing methods, and only one other study has examined fungi and protozoa together [13,40]. Our study is the first comprehensive investigation of intestinal eukaryotes using whole metagenome shotgun sequencing data that includes three IBD subtypes: CD, UC, and IBDU. Overall, we found that intestinal fungi and protozoa have a distinct distribution in IBD and its subtypes.

The IBD group in our study had a higher prevalence and abundance of fungi—notably *Saccharomyces*, *Nakaseomyces*, *Debaryomyces*, and *Candida* species—compared to the healthy cohort (Figure 2). Previous findings of fungal diversity in IBD have varied [19,52], but two studies have reported a higher abundance of *Candida* species in adults with IBD [25,53]. While *Candida* spp. were more prevalent in the IBD group in our study, they were only present in eight of the 355 total individuals with IBD. Further, a *Candida* species (*C. sake*) was only present in one of the healthy individuals, making it difficult to ascertain the overall relationship between *Candida* spp. and disease state. While this low prevalence of *Candida* was unexpected, the source studies of our datasets did not specifically target eukaryotes in their methodologies, which likely resulted in the loss of eukaryotic DNA from some species (discussed subsequently). Therefore, the absence of eukaryotic species in this study cannot be directly compared to previous studies.

We also found that *Saccharomyces cerevisiae* was more prevalent in all three IBD subtypes than in the healthy individuals and was statistically associated with IBD. This finding has not been reported in other microbiome studies, although one previous study reported the opposite finding: a decreased abundance of *S. cerevisiae* in active IBD [53]. However, we did not compare active and inactive IBD in our study, and differences between disease states would almost certainly affect the fungal microbiome [53–55]. Likewise, we did not have data on the current therapies, particularly immunosuppressants and steroids, and dietary composition of the IBD and control groups, factors that have been shown to affect intestinal fungal composition [56–59]. A larger sample size would also be required to stratify for these variables within the groups. While our study shows associative trends between fungi and IBD, we could not ascertain whether there may be any causal relationship between these microbes and the disease. This would entail a larger, longitudinal study with sampling of the microbiome in high-risk individuals prior to their development of IBD, such as the GEM project (www.gemproject.ca, accessed on 14 August 2022). It is a complex endeavour however, as the fungal microbiome is known to also be related to diet, BMI, and other lifestyle choices (e.g., smoking, alcohol, common medications such as proton pump inhibitors and antibiotics) and controlling for such a large variety of factors over a long period of time necessitates either a large sample size or a large effect size [56,57,60–64].

The fact that *S. cerevisiae* was more prevalent in individuals with IBD may have implications for microbial-based therapies, such as probiotics. A popular probiotic *S. cerevisiae* strain, *S. cerevisiae* var. *boulardii*, has been trialled for various gastrointestinal disorders [65] and is generally regarded as safe, though there have been several reports of adverse events [66–68]. This strain has been trialled for IBD as well, but its safety and

efficacy for the disease remain inconclusive [69]. While we only detected the strain *S. cerevisiae* S288C in our study, that may be because it was the only strain of *S. cerevisiae* included in the EukDetect database. The two strains share over 99% genome sequence similarity [70] and, therefore, individuals with IBD may be more susceptible to colonisation or adverse effects of *S. cerevisiae* strains. More research is needed to understand the eukaryotic microbiome in IBD to ensure the safety of probiotic strains such as *S. boulardii*.

In comparing IBD to a cohort of 471 unaffected individuals, we were able to characterise the healthy eukaryotic microbiome as well. Protozoa, namely *Blastocystis* and *Dientamoeba*, were notably more common in the healthy cohort than in IBD and associated with younger individuals with a healthy BMI (Figures 1 and 2, Table 2). Our findings confirm previous findings that *Blastocystis* spp. and *Dientamoeba* are more prevalent in the healthy gut microbiome than in IBD [31,32,71]. These protozoa have been historically attributed to gastroenteritis, but research in recent years, including our present study, is causing this paradigm to shift towards considering *Blastocystis* and *Dientamoeba* as common commensals of the healthy gut microbiome [28,29,33,72–74]. *Blastocystis* are specifically associated with higher bacterial diversity in healthy individuals [28,33]. This observation has been explained by *Blastocystis* functioning in a predator-prey relationship wherein they compete with opportunistic bacterial pathosymbionts and prevent them from overgrowing [28,75]. Therefore, the loss of this microbe may contribute to the reduced bacterial diversity commonly observed in IBD. Patients with gastroenteritis symptoms are occasionally prescribed antibiotics when *Blastocystis* or *Dientamoeba* are detected, which may further drive disturbance of the microbiome [76,77]. Given the new evidence, we encourage clinicians to now reconsider their practice in this regard and refrain from prescribing antibiotics due to the potential for long-term microbiome perturbations. Further, the recommendations for FMT in Australia do not support exclusion of donors based on the presence of *Blastocystis* and *Dientamoeba* [78]. Our findings add further support to this recommendation.

Comparing cohorts from different studies can inflate differences observed between groups [79], and we therefore tried to mitigate the biases that this approach introduces by only comparing studies that used similar methods and by keeping the studies that used different methods separate in our analyses. For example, it is known that geographical location has an influence on the composition of an individual's microbiome due to differences in environmental and lifestyle factors [80]. We controlled for this confounding effect by comparing two cohorts from the same country (Netherlands) in part of our study, and this comparison was kept separate from our analysis of the individuals in the FMT study who were recruited in France. The eukaryotic microbiome composition may differ in people with IBD in other regions of the world, akin to what has been observed in healthy individuals [81], and further research is needed that includes many individuals from a diversity of geographical locations.

Unlike geography, we could not properly age-match the participants between our studies, and there was a notable age difference between the two cohorts (Table 1). It is known that age can affect the diversity and composition of the gut microbiome [82], including intestinal fungi which can differ in diversity and composition from infancy to adulthood [83]. However, little is known about how fungal populations change throughout adulthood as previous work has not compared older adults to younger ones [84,85]. In our study, we found that increased age in adulthood was associated with an increased abundance of *S. cerevisiae* (Table 2), and this trend may also occur in other fungi we observed. We also found that older adults were less likely to have the *Blastocystis* spp. present in their gut microbiome, indicating that protozoa should be a point of inclusion in future age-stratified research.

We also chose two studies that used the same DNA extraction and sequencing methods since these can affect the resulting composition of metagenomic sequences [79]. However, we could not control for other biases introduced from different laboratory settings, such as different sample collection and reagents [79]. Another issue we could not account for was contamination, as neither study published negative controls (e.g., environmental or DNA

extraction blanks) to measure background levels of contaminant DNA. Contamination is normally considered a greater issue for low microbial biomass studies than for high biomass studies such as those involving stool samples [86]. However, the rarity of eukaryotes means that their detection may be more affected by contamination. Future research on the eukaryotic microbiome should endeavour to include negative controls where possible.

We also examined the longitudinal eukaryotic microbiome in individuals with CD who had received FMT, as well as single-timepoint samples from their FMT donors. Eukaryotes were variable and inconsistent across time between and within the FMT recipients (Figure 3). A high variability of eukaryotes over time has been observed in other studies, and fungi specifically show a higher intra- and inter-individual variability than bacteria in healthy individuals [15,87]. Only a subset of recipients maintained the same eukaryotic species over time, and this was irrespective of treatment group or outcome. Several recipients in the active FMT group maintained the same eukaryotic species before and after FMT, suggesting that the eukaryotic microbiome may remain stable even with FMT intervention in certain individuals. We expected to observe engraftment of donor eukaryotes in their recipients' microbiomes, but we did not observe this. This could be due to the small cohort size and would indicate that eukaryotic engraftment occurs at a low rate. However, the initial FMT study was not designed specifically for eukaryotic metagenomics and this likely limited our ability to detect eukaryotes. Although we did detect a single eukaryotic species, *Blastocystis* sp. subtype 1, in one donor, this species was not present in their recipient's samples. Interestingly, the recipient achieved a successful FMT outcome, which further supports investigating whether *Blastocystis* spp. are indeed safe to donate in FMT [78].

Across all the study groups, eukaryotes were rare or absent in the majority of samples (Figures 1 and 3). This finding was not surprising given that eukaryotes comprise a small proportion (less than 1%) of the total DNA of the gut microbiome [15–17]. Thus, deeper sequencing or alternative methodologies may be required to effectively detect eukaryotic DNA [18]. Additionally, all three of the source studies of our datasets were originally designed for investigating the bacterial microbiome and did not use methods to specifically preserve eukaryotic DNA for sequencing. Eukaryotic DNA can be enriched during microbial DNA extraction to maximise the likelihood of capturing eukaryotes in microbiome samples [14]. This may have biased our results towards more robust eukaryotic species whose DNA was not destroyed during the bacterial DNA extraction process, as mechanical lysis has been shown to significantly reduce fungal DNA yield compared to no lysis [14]. Despite these limitations, our results demonstrate that the eukaryotic microbiome can still be gleaned from metagenomic samples in some cases even when they are not enriched for eukaryotic DNA.

Since eukaryotes have begun to attract attention for their importance in gut microbiome research, we had several bioinformatic tools at our disposal to search for microbial eukaryotes [13]. We chose to use two different pipelines, RiboTagger and EukDetect, for their ease of use, lower demand for computational resources relative to other tools available, and the diversity provided using two separate databases [39,40]. RiboTagger identifies eukaryotes based on the 18S rRNA marker gene, whereas EukDetect's database includes 521,824 universal eukaryotic marker genes not limited to the 18S gene. Thus, we expected to identify more eukaryotes with greater resolution with EukDetect because eukaryotes could be identified by more than just their 18S genes. Indeed, EukDetect captured an overall higher prevalence and diversity of eukaryotes than RiboTagger (Figure 1). However, unlike RiboTagger, EukDetect failed to capture *Dientamoeba* and *Galactomyces*. *Dientamoeba* was unable to be detected as its genome has never been sequenced and is absent from EukDetect's database [40]. *Galactomyces* was also absent from EukDetect's database. The differing results between these two tools highlights how databases may limit findings based on how they were curated. Thus, we recommend using more than one database when using marker gene detection. Additionally, most available databases primarily consist of species that have been cultured and sequenced [88], and future work would benefit from identifying uncultured microbial eukaryotes via de novo genome assembly [89]. We were

also unable to explore strain variation within the species we identified in this study because we only used marker gene profilers due to computational limitations. Strain-level variation could further explain differences between IBD subtypes and healthy individuals, and it would be beneficial for future research to include this analysis.

5. Conclusions

The aim of our study was to explore the fungal and protozoan fraction of the gut microbiome in IBD. We were able to elucidate the IBD eukaryotic microbiome with greater precision than previous studies by using whole metagenome sequencing data to identify eukaryotic at the species level. Although these approaches need significant advancements in the future, we found that IBD and its subtypes have a eukaryotic microbiome composition distinct from individuals without IBD. Our findings highlight the need for more research that explores the nonbacterial microbiome in IBD, particularly studies that aim to preserve or enrich eukaryotic DNA and account for various factors that can affect the host eukaryotic microbiome (e.g., diet, medications, disease activity) [18,62,64]. Additionally, the results of our longitudinal analysis indicate that the eukaryotic microbiome varies over time, and future longitudinal sampling will be important to reveal the dynamics of intestinal eukaryotes. Our results also support the growing body of literature suggesting that *Blastocystis* are common in healthy individuals and associate with indicators of health. By including eukaryotes in our study, we provide a more comprehensive understanding of what constitutes a healthy microbiome from a diseased state such as IBD, and our results supplement findings on the bacterial microbiome in IBD. We advocate that a more inclusive approach to microbiome research not limited to bacteria is increasingly important as diagnostics and therapeutics for IBD are continuing to target the microbiome.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms10101910/s1, Table S1: Results of RiboTagger for the 1000IBD and 500FG studies; Table S2: Results of EukDetect for the 1000IBD and 500FG studies before rarefying; Table S3; Difference in EukDetect results for the 1000IBD and 500FG studies between pre- and post-rarefaction; Table S4: Results of EukDetect for the faecal microbiota transplant study data.

Author Contributions: Conceptualization, G.L.G. and L.S.W.; formal analysis, G.L.G.; visualization, G.L.G.; writing—original draft preparation, G.L.G.; writing—review and editing, G.L.G., L.S.W., M.N.M., B.L. and J.M.A.; supervision, L.S.W., M.N.M., B.L. and J.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Metagenomes and sample metadata for the 1000IBD study (https://1000ibd.org/, accessed on 14 August 2022) may be accessed subject to approval of the corresponding Data Access Committee at https://ega-archive.org/datasets/EGAD00001004194, accessed on 14 August 2022. The 500FG study (https://hfgp.bbmri.nl/menu/main/home, accessed on 14 August 2022) metagenomes and sample metadata are publicly available at https://www.ebi.ac.uk/ena/data/view/PRJNA319574, accessed on 14 August 2022. Metagenomes and sample metadata from the FMT study are publicly available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA625520/, accessed on 14 August 2022. The workflow for this study can be found at https://github.com/ginaguzzo/2021_gut_eukaryotes_in_IBD, accessed on 14 August 2022. As this study contains previously published datasets, we have adhered to the STORMS checklist where possible. The completed checklist for this research can be found at https://github.com/ginaguzzo/2021_gut_eukaryotes_in_IBD/blob/main/STORMS_Checklist.xlsx, accessed on 14 August 2022.

Acknowledgments: We thank the participants and researchers of the original studies for their contributions. We thank the Thesis Writing Group at the Australian Centre for Ancient DNA for providing edits and feedback on the manuscript, and the high-performance computing cluster (HPC1/Phoenix) team at the University of Adelaide for supporting the computational analysis of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Podolsky, D.K. Inflammatory bowel disease. *N. Engl. J. Med.* **1991**, *325*, 928–937. [CrossRef] [PubMed]
2. Khor, B.; Gardet, A.; Xavier, R.J. Genetics and pathogenesis of inflammatory bowel disease. *Nature* **2011**, *474*, 307–317. [CrossRef] [PubMed]
3. Ho, S.-M.; Lewis, J.D.; Mayer, A.E.; Bernstein, C.N.; Plevy, E.S.; Chuang, E.; Rappaport, S.M.; Croitoru, K.; Korzenik, J.R.; Krischer, J.; et al. Challenges in IBD Research: Environmental Triggers. *Inflamm. Bowel Dis.* **2019**, *25*, S13–S23. [CrossRef] [PubMed]
4. Loddo, I.; Romano, C. Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. *Front. Immunol.* **2015**, *6*, 551. [CrossRef]
5. Marchesi, J.R.; Ravel, J. The vocabulary of microbiome research: A proposal. *Microbiome* **2015**, *3*, 31. [CrossRef]
6. Kostic, A.D.; Xavier, R.J.; Gevers, D. The Microbiome in Inflammatory Bowel Disease: Current Status and the Future Ahead. *Gastroenterology* **2014**, *146*, 1489–1499. [CrossRef]
7. Ni, J.; Wu, G.D.; Albenberg, L.; Tomov, V.T. Gut microbiota and IBD: Causation or correlation? *Nat. Rev. Gastroenterol. Hepatol.* **2017**, *14*, 573–584. [CrossRef]
8. Khan, I.; Ullah, N.; Zha, L.; Bai, Y.; Khan, A.; Zhao, T.; Che, T.; Zhang, C. Alteration of Gut Microbiota in Inflammatory Bowel Disease (IBD): Cause or Consequence? IBD Treatment Targeting the Gut Microbiome. *Pathogens* **2019**, *8*, 126. [CrossRef]
9. Lopez, J.; Grinspan, A. Fecal Microbiota Transplantation for Inflammatory Bowel Disease. *Gastroenterol. Hepatol.* **2016**, *12*, 374–379.
10. Costello, S.P.; Soo, W.; Bryant, R.V.; Jairath, V.; Hart, A.L.; Andrews, J.M. Systematic review with meta-analysis: Faecal microbiota transplantation for the induction of remission for active ulcerative colitis. *Aliment. Pharmacol. Ther.* **2017**, *46*, 213–224. [CrossRef]
11. Caldeira, L.D.F.; Borba, H.H.; Tonin, F.S.; Wiens, A.; Fernandez-Llimos, F.; Pontarolo, R. Fecal microbiota transplantation in inflammatory bowel disease patients: A systematic review and meta-analysis. *PLoS ONE* **2020**, *15*, e0238910. [CrossRef] [PubMed]
12. Kong, L.; Lloyd-Price, J.; Vatanen, T.; Seksik, P.; Beaugerie, L.; Simon, T.; Vlamakis, H.; Sokol, H.; Xavier, R.J. Linking Strain Engraftment in Fecal Microbiota Transplantation With Maintenance of Remission in Crohn's Disease. *Gastroenterology* **2020**, *159*, 2193–2202.e5. [CrossRef] [PubMed]
13. Guzzo, G.L.; Andrews, J.M.; Weyrich, L.S. The Neglected Gut Microbiome: Fungi, Protozoa, and Bacteriophages in Inflammatory Bowel Disease. *Inflamm. Bowel Dis.* **2022**, *28*, 1112–1122. [CrossRef] [PubMed]
14. Halwachs, B.; Madhusudhan, N.; Krause, R.; Nilsson, H.; Moissl-Eichinger, C.; Högenauer, C.; Thallinger, G.G.; Gorkiewicz, G. Critical Issues in Mycobiota Analysis. *Front. Microbiol.* **2017**, *8*, 180. [CrossRef]
15. Nash, A.K.; Auchtung, T.A.; Wong, M.C.; Smith, D.P.; Gesell, J.R.; Ross, M.C.; Stewart, C.J.; Metcalf, G.A.; Muzny, D.M.; Gibbs, R.A.; et al. The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome* **2017**, *5*, 153. [CrossRef] [PubMed]
16. Sender, R.; Fuchs, S.; Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **2016**, *14*, e1002533. [CrossRef] [PubMed]
17. Sarrabayrouse, G.; Elias, A.; Yáñez, F.; Mayorga, L.; Varela, E.; Bartoli, C.; Casellas, F.; Borruel, N.; de Guise, C.H.; Machiels, K.; et al. Fungal and Bacterial Loads: Noninvasive Inflammatory Bowel Disease Biomarkers for the Clinical Setting. *mSystems* **2021**, *6*, e01277-20. [CrossRef] [PubMed]
18. Richard, M.L.; Sokol, H. The gut mycobiota: Insights into analysis, environmental interactions and role in gastrointestinal diseases. *Nat. Rev. Gastroenterol. Hepatol.* **2019**, *16*, 1–345. [CrossRef]
19. Miyoshi, J.; Sofia, M.A.; Pierre, J.F. The evidence for fungus in Crohn's disease pathogenesis. *Clin. J. Gastroenterol.* **2018**, *11*, 449–456. [CrossRef]
20. Van Schaik, F.D.M.; Oldenburg, B.; Hart, A.R.; Siersema, P.D.; Lindgren, S.; Grip, O.; Teucher, B.; Kaaks, R.; Bergmann, M.M.; Boeing, H.; et al. Serological markers predict inflammatory bowel disease years before the diagnosis. *Gut* **2012**, *62*, 683–688. [CrossRef]
21. Findley, K.; Oh, J.; Yang, J.; Conlan, S.; Deming, C.; Meyer, J.A.; Schoenfeld, D.; Nomicos, E.; Park, M.; Kong, H.H.; et al. Topographic diversity of fungal and bacterial communities in human skin. *Nature* **2013**, *498*, 367–370. [CrossRef] [PubMed]
22. Standaert-Vitse, A.; Sendid, B.; Joossens, M.; François, N.; Khoury, P.V.-E.; Branche, J.; Van Kruiningen, H.; Jouault, T.; Rutgeerts, P.; Gower-Rousseau, C.; et al. Candida albicans Colonization and ASCA in Familial Crohn's Disease. *Am. J. Gastroenterol.* **2009**, *104*, 1745–1753. [CrossRef]
23. Yan, L.; Yang, C.; Tang, J. Disruption of the intestinal mucosal barrier in Candida albicans infections. *Microbiol. Res.* **2013**, *168*, 389–395. [CrossRef]
24. Sartor, R.B.; Wu, G.D. Roles for Intestinal Bacteria, Viruses, and Fungi in Pathogenesis of Inflammatory Bowel Diseases and Therapeutic Approaches. *Gastroenterology* **2017**, *152*, 327–339.e4. [CrossRef] [PubMed]
25. Hoarau, G.; Mukherjee, P.K.; Gower, C.; Hager, C.; Chandra, J.; Retuerto, M.A.; Neut, C.; Vermeire, S.; Clemente, J.; Colombel, J.F.; et al. Bacteriome and Mycobiome Interactions Underscore Microbial Dysbiosis in Familial Crohn's Disease. *mBio* **2016**, *7*, e01250-16. [CrossRef] [PubMed]
26. Garcia, L.S. Dientamoeba fragilis, One of the Neglected Intestinal Protozoa. *J. Clin. Microbiol.* **2016**, *54*, 2243–2250. [CrossRef] [PubMed]
27. Wawrzyniak, I.; Poirier, P.; Viscogliosi, E.; Dionigia, M.; Texier, C.; Delbac, F.; El Alaoui, H. *Blastocystis*, an unrecognized parasite: An overview of pathogenesis and diagnosis. *Ther. Adv. Infect. Dis.* **2013**, *1*, 167–178. [CrossRef]
28. Nieves-Ramírez, M.E.; Partida-Rodríguez, O.; Laforest-Lapointe, I.; Reynolds, L.A.; Brown, E.M.; Valdez-Salazar, A.; Morán-Silva, P.; Rojas-Velázquez, L.; Morien, E.; Parfrey, L.W.; et al. Asymptomatic Intestinal Colonization with Protist *Blastocystis* Is Strongly Associated with Distinct Microbiome Ecological Patterns. *mSystems* **2018**, *3*, e00007-18. [CrossRef]

29. De Boer, M.D.; Schuurs, T.A.; Vermeer, M.; Ruijs, G.J.; van der Zanden, A.G.M.; Weel, J.F.; van Coppenraet, L.E.B. Distribution and relevance of Dientamoeba fragilis and *Blastocystis* species in gastroenteritis: Results from a case-control study. *Eur. J. Clin. Microbiol.* **2019**, *39*, 197–203. [CrossRef]
30. Chabé, M.; Lokmer, A.; Ségurel, L. Gut Protozoa: Friends or Foes of the Human Gut Microbiota? *Trends Parasitol.* **2017**, *33*, 925–934. [CrossRef]
31. Rossen, N.G.; Bart, A.; Verhaar, N.; Van Nood, E.; Kootte, R.; De Groot, P.F.; D'Haens, G.R.; Ponsioen, C.Y.; Van Gool, T. Low prevalence of *Blastocystis* sp. in active ulcerative colitis patients. *Eur. J. Clin. Microbiol.* **2015**, *34*, 1039–1044. [CrossRef] [PubMed]
32. Petersen, A.M.; Stensvold, C.R.; Mirsepasi, H.; Engberg, J.; Friis-Møller, A.; Porsbo, L.J.; Hammerum, A.M.; Nordgaard-Lassen, I.; Nielsen, H.V.; Krogfelt, K.A. Active ulcerative colitis associated with low prevalence of *Blastocystis* and *Dientamoeba fragilis* infection. *Scand. J. Gastroenterol.* **2013**, *48*, 638–639. [CrossRef] [PubMed]
33. Tito, R.Y.; Chaffron, S.; Caenepeel, C.; Lima-Mendez, G.; Wang, J.; Vieira-Silva, S.; Falony, G.; Hildebrand, F.; Darzi, Y.; Rymenans, L.; et al. Population-level analysis of Blastocystis subtype prevalence and variation in the human gut microbiota. *Gut* **2019**, *68*, 1180–1189. [CrossRef] [PubMed]
34. Imhann, F.; Van Der Velde, K.J.; Barbieri, R.; Alberts, R.; Voskuil, M.D.; Vila, A.V.; Collij, V.; Spekhorst, L.M.; Van Der Sloot, K.W.J.; Peters, V.; et al. The 1000IBD project: Multi-omics data of 1000 inflammatory bowel disease patients; data release 1. *BMC Gastroenterol.* **2019**, *19*, 1–10. [CrossRef]
35. Schirmer, M.; Smeekens, S.P.; Vlamakis, H.; Jaeger, M.; Oosting, M.; Franzosa, E.A.; Ter Horst, R.; Jansen, T.; Jacobs, L.; Bonder, M.J.; et al. Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* **2016**, *167*, 1125–1136.e8. [CrossRef]
36. Sokol, H.; Landman, C.; Seksik, P.; Berard, L.; Montil, M.; Nion-Larmurier, I.; Bourrier, A.; Le Gall, G.; Lalande, V.; De Rougemont, A.; et al. Fecal microbiota transplantation to maintain remission in Crohn's disease: A pilot randomized controlled study. *Microbiome* **2020**, *8*, 12. [CrossRef] [PubMed]
37. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
38. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
39. Xie, C.; Goi, C.L.W.; Huson, D.H.; Little, P.F.R.; Williams, R.B.H. RiboTagger: Fast and unbiased 16S/18S profiling using whole community shotgun metagenomic or metatranscriptome surveys. *BMC Bioinform.* **2016**, *17*, 277–282. [CrossRef]
40. Lind, A.L.; Pollard, K.S. Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. *Microbiome* **2021**, *9*, 1–18. [CrossRef]
41. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
42. McMurdie, P.J.; Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **2013**, *8*, e61217. [CrossRef] [PubMed]
43. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
44. DeWitt, P.; Bennett, T. qwraps2: Quick Wraps 2,0.5.2; CRAN. 2021. Available online: https://cran.r-project.org/web/packages/qwraps2/index.html (accessed on 14 August 2022).
45. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*; John Wiley & Sons, Inc.: New York, NY, USA, 2002.
46. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; CRC Press LLC: Boca Raton, FL, USA, 1989.
47. Zeileis, A.; Kleiber, C.; Jackman, S. Regression Models for Count Data in R. *J. Stat. Softw.* **2008**, *27*, 1–25. [CrossRef]
48. Jackman, S. *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*; United States Studies Centre, University of Sydney: Sydney, NSW, Australia, 2020.
49. Zeileis, A.; Köll, S.; Graham, N. Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R. *J. Stat. Softw.* **2020**, *95*, 1–36. [CrossRef]
50. McMurdie, P.J.; Holmes, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **2014**, *10*, e1003531. [CrossRef]
51. Cameron, E.S.; Schmidt, P.J.; Tremblay, B.J.-M.; Emelko, M.B.; Müller, K.M. Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. *Sci. Rep.* **2021**, *11*, 1–13. [CrossRef] [PubMed]
52. Wang, S.; Zhang, Y.-R.; Yu, Y.-B. The important role of fungi in inflammatory bowel diseases. *Scand. J. Gastroenterol.* **2021**, *56*, 1312–1322. [CrossRef] [PubMed]
53. Sokol, H.; Leducq, V.; Aschard, H.; Pham, H.P.; Jegou, S.; Landman, C.; Cohen, D.; Liguori, G.; Bourrier, A.; Nion-Larmurier, I.; et al. Fungal microbiota dysbiosis in IBD. *Gut* **2017**, *66*, 1039–1048. [CrossRef]
54. Liguori, G.; Lamas, B.; Richard, M.L.; Brandi, G.; Da Costa, G.; Hoffmann, T.W.; Di Simone, M.P.; Calabrese, C.; Poggioli, G.; Langella, P.; et al. Fungal Dysbiosis in Mucosa-associated Microbiota of Crohn's Disease Patients. *J. Crohns Colitis* **2016**, *10*, 296–305. [CrossRef]
55. Jain, U.; Heul, A.M.V.; Xiong, S.; Gregory, M.H.; Demers, E.G.; Kern, J.T.; Lai, C.-W.; Muegge, B.D.; Barisas, D.A.G.; Leal-Ekman, J.S.; et al. *Debaryomyces* is enriched in Crohn's disease intestinal tissue and impairs healing in mice. *Science* **2021**, *371*, 1154–1159. [CrossRef]
56. David, L.A.; Maurice, C.F.; Carmody, R.N.; Gootenberg, D.B.; Button, J.E.; Wolfe, B.E.; Ling, A.V.; Devlin, A.S.; Varma, Y.; Fischbach, M.A.; et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **2014**, *505*, 559–563. [CrossRef]
57. Hoffmann, C.; Dollive, S.; Grunberg, S.; Chen, J.; Li, H.; Wu, G.D.; Lewis, J.D.; Bushman, F.D. Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLoS ONE* **2013**, *8*, e66019. [CrossRef] [PubMed]

58. Ali, T.; Bronze, M.S.; Kaitha, S.; Mahmood, S.; Ftaisi, A.; Stone, J. Clinical use of anti-TNF therapy and increased risk of infections. *Drug Health Patient Saf.* **2013**, *5*, 79–99. [CrossRef] [PubMed]
59. Toruner, M.; Loftus, E.; Harmsen, W.S.; Zinsmeister, A.R.; Orenstein, R.; Sandborn, W.J.; Colombel, J.; Egan, L.J. Risk Factors for Opportunistic Infections in Patients With Inflammatory Bowel Disease. *Gastroenterology* **2008**, *134*, 929–936. [CrossRef]
60. Samonis, G.; Gikas, A.; Anaissie, E.J.; Vrenzos, G.; Maraki, S.; Tselentis, Y.; Bodey, G.P. Prospective evaluation of effects of broad-spectrum antibiotics on gastrointestinal yeast colonization of humans. *Antimicrob. Agents Chemother.* **1993**, *37*, 51–53. [CrossRef] [PubMed]
61. Jacobs, C.; Adame, E.C.; Attaluri, A.; Valestin, J.; Rao, S.S.C. Dysmotility and proton pump inhibitor use are independent risk factors for small intestinal bacterial and/or fungal overgrowth. *Aliment. Pharmacol. Ther.* **2013**, *37*, 1103–1111. [CrossRef] [PubMed]
62. Erdogan, A.; Rao, S.S.C. Small Intestinal Fungal Overgrowth. *Curr. Gastroenterol. Rep.* **2015**, *17*, 1–7. [CrossRef] [PubMed]
63. Hartmann, P.; Lang, S.; Zeng, S.; Duan, Y.; Zhang, X.; Wang, Y.; Bondareva, M.; Kruglov, A.; Fouts, D.E.; Stärkel, P.; et al. Dynamic Changes of the Fungal Microbiome in Alcohol Use Disorder. *Front. Physiol.* **2021**, *12*, 699253. [CrossRef]
64. García-Gamboa, R.; Kirchmayr, M.R.; Gradilla-Hernández, M.S.; Pérez-Brocal, V.; Moya, A.; González-Avila, M. The intestinal mycobiota and its relationship with overweight, obesity and nutritional aspects. *J. Hum. Nutr. Diet.* **2021**, *34*, 645–655. [CrossRef]
65. Kelesidis, T.; Pothoulakis, C. Efficacy and safety of the probiotic *Saccharomyces boulardii* for the prevention and therapy of gastrointestinal disorders. *Ther. Adv. Gastroenterol.* **2011**, *5*, 111–125. [CrossRef]
66. Kara, I.; Yıldırım, F.; Özgen, Ö.; Erganiş, S.; Aydoğdu, M.; Dizbay, M.; Gürsel, G.; Kalkancı, A. Saccharomyces cerevisiae fungemia after probiotic treatment in an intensive care unit patient. *J. Mycol. Med.* **2018**, *28*, 218–221. [CrossRef]
67. Martin, I.W.; Tonner, R.; Trivedi, J.; Miller, H.; Lee, R.; Liang, X.; Rotello, L.; Isenbergh, E.; Anderson, J.; Perl, T.; et al. Saccharomyces boulardii probiotic-associated fungemia: Questioning the safety of this preventive probiotic's use. *Diagn. Microbiol. Infect. Dis.* **2017**, *87*, 286–288. [CrossRef] [PubMed]
68. Santino, I.; Alari, A.; Bono, S.; Teti, E.; Marangi, M.; Bernardini, A.; Magrini, L.; Di Somma, S.; Teggi, A. *Saccharomyces Cerevisiae* Fungemia, a Possible Consequence of the Treatment of *Clostridium Difficile* Colitis with a Probioticum. *Int. J. Immunopathol. Pharmacol.* **2014**, *27*, 143–146. [CrossRef] [PubMed]
69. Sivananthan, K.; Petersen, A.M. Review of *Saccharomyces boulardii* as a treatment option in IBD. *Immunopharmacol. Immunotoxicol.* **2018**, *40*, 465–475. [CrossRef]
70. Khatri, I.; Tomar, R.; Ganesan, K.; Prasad, G.S.; Subramanian, S. Complete genome sequence and comparative genomics of the probiotic yeast *Saccharomyces boulardii*. *Sci. Rep.* **2017**, *7*, 1–12. [CrossRef] [PubMed]
71. Petersen, A.M.; Nielsen, H.V.; Stensvold, C.R.; Engberg, J.H.; Friis-Møller, A.; Nordgaard-Lassen, I.; Wildt, S.; Krogfelt, K.A. Blastocystis and Dientamoeba Fragilis in Active and Inactive Inflammatory Bowel Disease. *Gastroenterology* **2011**, *140*, S-329–S-330. [CrossRef]
72. Brands, M.R.; Van de Vijver, E.; Haisma, S.M.; Heida, A.; van Rheenen, P.F. No association between abdominal pain and *Dientamoeba* in Dutch and Belgian children. *Arch. Dis. Child.* **2019**, *104*, 686–689. [CrossRef]
73. Holtman, A.G.; Kranenberg, J.J.; Blanker, M.H.; Ott, A.; Leeuwen, Y.L.-V.; Berger, M.Y. *Di entamoeba fragilis* colonization is not associated with gastrointestinal symptoms in children at primary care level. *Fam. Pract.* **2016**, *34*, 25–29. [CrossRef]
74. Jokelainen, P.; Jensen, B.H.; Andreassen, B.U.; Petersen, A.M.; Röser, D.; Krogfelt, K.A.; Nielsen, H.V.; Stensvold, C.R. Dientamoeba fragilis, a Commensal in Children in Danish Day Care Centers. *J. Clin. Microbiol.* **2017**, *55*, 1707–1713. [CrossRef]
75. Laforest-Lapointe, I.; Arrieta, M.-C. Microbial Eukaryotes: A Missing Link in Gut Microbiome Studies. *mSystems* **2018**, *3*, e00201-17. [CrossRef]
76. Nagata, N.; Marriott, D.; Harkness, J.; Ellis, J.; Stark, D. Current treatment options for Dientamoeba fragilis infections. *Int. J. Parasitol. Drugs Drug Resist.* **2012**, *2*, 204–215. [CrossRef]
77. Roberts, T.; Stark, D.; Harkness, J.; Ellis, J. Update on the pathogenic potential and treatment options for *Blastocystis* sp. *Gut Pathog.* **2014**, *6*, 17. [CrossRef] [PubMed]
78. Haifer, C.; Kelly, C.R.; Paramsothy, S.; Andresen, D.; Papanicolas, E.L.; McKew, G.L.; Borody, T.J.; Kamm, M.; Costello, S.P.; Andrews, J.M.; et al. Australian consensus statements for the regulation, production and use of faecal microbiota transplantation in clinical practice. *Gut* **2020**, *69*, 801–810. [CrossRef] [PubMed]
79. Allaband, C.; McDonald, D.; Vázquez-Baeza, Y.; Minich, J.J.; Tripathi, A.; Brenner, D.A.; Loomba, R.; Smarr, L.; Sandborn, W.J.; Schnabl, B.; et al. Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians. *Clin. Gastroenterol. Hepatol.* **2018**, *17*, 218–230. [CrossRef] [PubMed]
80. Gaulke, C.A.; Sharpton, T.J. The influence of ethnicity and geography on human gut microbiome composition. *Nat. Med.* **2018**, *24*, 1495–1496. [CrossRef]
81. Kabwe, M.H.; Vikram, S.; Mulaudzi, K.; Jansson, J.K.; Makhalanyane, T.P. The gut mycobiota of rural and urban individuals is shaped by geography. *BMC Microbiol.* **2020**, *20*, 1–12. [CrossRef]
82. Martino, C.; Dilmore, A.H.; Burcham, Z.M.; Metcalf, J.L.; Jeste, D.; Knight, R. Microbiota succession throughout life from the cradle to the grave. *Nat. Rev. Genet.* 2022, *Online ahead of print*. [CrossRef]
83. Strati, F.; Di Paola, M.; Stefanini, I.; Albanese, D.; Rizzetto, L.; Lionetti, P.; Calabrò, A.; Jousson, O.; Donati, C.; Cavalieri, D.; et al. Age and Gender Affect the Composition of Fungal Population of the Human Gastrointestinal Tract. *Front. Microbiol.* **2016**, *7*, 1227. [CrossRef]
84. Ahmad, H.F.; Castro Mejia, J.L.; Krych, L.; Khakimov, B.; Kot, W.; Bechshøft, R.L.; Reitelseder, S.; Højfeldt, G.W.; Engelsen, S.B.; Holm, L.; et al. Gut Mycobiome Dysbiosis Is Linked to Hypertriglyceridemia among Home Dwelling Elderly Danes. *bioRxiv* **2020**. [CrossRef]

85. Nagpal, R.; Neth, B.J.; Wang, S.; Mishra, S.P.; Craft, S.; Yadav, H. Gut mycobiome and its interaction with diet, gut bacteria and alzheimer's disease markers in subjects with mild cognitive impairment: A pilot study. *eBioMedicine* **2020**, *59*, 102950. [CrossRef]
86. Eisenhofer, R.; Minich, J.J.; Marotz, C.; Cooper, A.; Knight, R.; Weyrich, L.S. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol.* **2019**, *27*, 105–117. [CrossRef]
87. Hallen-Adams, H.E.; Suhr, M.J. Fungi in the healthy human gastrointestinal tract. *Virulence* **2016**, *8*, 352–358. [CrossRef] [PubMed]
88. Underhill, D.M.; Iliev, I.D. The mycobiota: Interactions between commensal fungi and the host immune system. *Nat. Rev. Immunol.* **2014**, *14*, 405–416. [CrossRef] [PubMed]
89. Olm, M.R.; West, P.T.; Brooks, B.; Firek, B.A.; Baker, R.; Morowitz, M.J.; Banfield, J.F. Genome-resolved metagenomics of eukaryotic populations during early colonization of premature infants and in hospital rooms. *Microbiome* **2019**, *7*, 26. [CrossRef] [PubMed]

Chapter 3

Statement of Authorship

Title of Paper	The utility of dry swabs over stool samples: a case study from a faecal microbiota transplant trial
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Unpublished manuscript formatted for submission to Microbiome.

Principal Author

Name of Principal Author (Candidate)	Gina L. Guzzo		
Contribution to the Paper	Bioinformatic and statistical analyses, visualisation and interpretation of the data, wrote the manuscript.		
Overall percentage (%)	60		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	01/06/2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Provided substantial discussion and feedback to the research and data interpretation, edited all versions of the manuscript.		
Signature		Date	24/06/2022

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Provided discussion and feedback to the research, reviewed the manuscript.		
Signature		Date	08/06/2022

Name of Co-Author	Jane M. Andrews		
Contribution to the Paper	Study conceptualisation, provided funding, provided discussion and feedback to the research, reviewed the manuscript.		
Signature		Date	03/08/2022

Name of Co-Author	Samuel P. Costello		
Contribution to the Paper	Study conceptualisation, provided funding, provided discussion and feedback to the research, reviewed the manuscript.		
Signature		Date	18/07/2022

Name of Co-Author	Chelsea Bickley		
Contribution to the Paper	Study conceptualisation, performed laboratory work and sample processing.		
Signature		Date	07/07/2022

Name of Co-Author	Murthy N. Mittinty		
Contribution to the Paper	Provided discussion and feedback to the research, reviewed the manuscript.		
Signature		Date	28/06/2022

Title

The utility of dry swabs over stool samples: a case study from a faecal microbiota transplant trial

Authors

Gina L. Guzzo[1], Samuel P. Costello[2], Jane M. Andrews[3], Chelsea Bickley[4], Murthy N. Mittinty[5], Bastien Llamas[1,6,7], and Laura S. Weyrich[1,8]

Affiliations

[1]School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia
[2]Inflammatory Bowel Disease Service, Department of Gastroenterology, The Queen Elizabeth Hospital & School of Medicine, Faculty of Health Sciences, University of Adelaide, South Australia, Australia
[3]Inflammatory Bowel Disease Service, Department of Gastroenterology and Hepatology, Royal Adelaide Hospital & School of Medicine, Faculty of Health Sciences, University of Adelaide, South Australia, Australia
[4]CSIRO Health & Biosecurity, Adelaide, South Australia, Australia
[5]School of Public Health, University of Adelaide, South Australia, Australia
[6]Indigenous Genomics Research Group, Telethon Kids Institute, Adelaide, South Australia, Australia
[7]National Centre for Indigenous Genomics, Australian National University, Canberra, Australian Capital Territory, Australia
[8]Department of Anthropology and Huck Institutes of the Life Sciences, Pennsylvania State University, PA, USA

Abstract

The study of the gut microbiome and its relationship to disease has benefited from stool sampling over the invasiveness of mucosal biopsy samples. However, whole stool sampling is difficult to consistently achieve in high-resolution longitudinal studies. Here, we compared whole stool to a sampling alternative, dry swabs of stool, in a longitudinal trial of 25 individuals with ulcerative colitis (UC) who received faecal microbiota transplants (FMT). FMT recipients self-collected dry swabs over 8 weeks post-FMT, resulting in 345 samples across 21 time points, and whole stool samples were also self-collected from the recipients at day 0, and weeks 4 and 8. All samples underwent 16S ribosomal RNA marker gene sequencing to detect bacterial DNA. Microbial diversity of the dry swabs was similar to whole stool, while microbial composition was different and influenced by patient factors, such as age and sex. Additionally, post-FMT remission was associated with significant shifts in the composition and diversity of swabs, and surprisingly, recipients who achieved remission had a lower swab diversity. Given that they captured an informative level of diversity to stool, we conclude that dry swabs are a suitable alternative to whole stool collection in longitudinal studies of the gut microbiome in UC.

Keywords

gut microbiome, swab, stool, ulcerative colitis (UC), faecal microbiota transplant (FMT)

Introduction

Ulcerative colitis (UC) is one of the two major inflammatory bowel diseases and is characterised by chronic inflammation of the large intestine (colon). UC is caused by a combination of immune, environmental, and intestinal microbial factors that lead to aberrant immune function at the colonic mucosa; however, the complexities of these causal mechanisms are not fully understood (Khor, Gardet & Xavier 2011; Loddo & Romano 2015; Ho et al. 2019). Treatments for UC range from pharmacological (e.g., aminosalicylates, steroids and biologics) to surgical and aim to induce and maintain remission (Cai, Wang & Li 2021). While non-surgical treatments for UC have largely centred around immune modulation and suppression, another treatment area of interest is the modification of the intestinal microbiota through faecal microbiota transplantation (FMT) (Knox et al. 2019; Glassner, Abraham & Quigley 2020).

FMT involves the transfer of intestinal microbiota from a healthy donor to a recipient, with the purpose of treating disease. FMT has a high cure rate for the treatment of recurrent and refractory *Clostridioides difficile* infection (Moayyedi et al. 2017; Quraishi et al. 2017). FMT can also induce remission of UC, though with a lower rate of efficacy than for *Clostridioides difficile* infection, that is temporary in most patients without maintenance dosing (Costello, SP et al. 2017; Yalchin et al. 2019; Stojek, Jabłońska & Adrych 2021). Many important questions remain, including which microbes and microbial functions are therapeutic for inducing and maintaining remission in UC, and whether it is important to identify suitable donations with these characteristics and match them to patient factors (Yalchin et al. 2019; Li et al. 2020; Okahara et al. 2020; Olesen & Gerardin 2020; Quraishi, Iqbal & Hart 2020). To address these questions, donor sampling is needed, in addition to high resolution longitudinal sampling of the recipient microbiome post-FMT to understand the initial microbiota changes that may instigate remission. Most FMT trials have used infrequent sampling at the point of clinical assessment (4-12 weeks after FMT has commenced), and this may bias the analysis towards organisms that have benefited from remission rather than organisms that may be inducing remission immediately following FMT.

The intestinal microbiome in IBD is usually sampled via whole stool or, less commonly, via mucosal biopsy (Aldars-García, Chaparro & Gisbert 2021). Stool collection remains the 'gold standard' method for specimen collection due to its “ease” and the invasiveness of biopsies (Biehl et al. 2019), with most FMT trials favouring this approach (Paramsothy et al. 2017). However, stool collection for individuals with UC can be challenging due to frequent loose or bloody stool (Bassis et al. 2017). Additionally, stool collection should be conducted on-site of the storage facility and immediately frozen to reduce the likelihood of bacterial blooms in transit (Amir et al. 2017a; Marotz et al. 2021). This collection method may exclude individuals who are uncomfortable with the ‘yuck factor’ of donating and/or transporting their whole stool. An alternative to stool collection is swab sampling, either of the rectum or toilet paper containing stool (Costello, EK et al. 2009; Vandeputte et al. 2017). Rectal swabs have been shown to be reliable in recapitulating the gut microbiome when compared to stool samples and mucosal biopsies (Araújo-Pérez et al. 2012; Budding et al. 2014; Bassis et al. 2017; Jones et al. 2018; Sun et al. 2021). Toilet paper swabs, or 'dry' swabs, have been particularly popular in citizen science gut microbiome projects (McDonald et al. 2018) because they are easier for the participant to obtain themselves without discomfort, offering the potential for higher participant compliance in clinical trials (Galea & Tracy 2007; Short et

al. 2021). However, few studies have compared dry swabs to stool in great detail (Costello, EK et al. 2009; Sinha et al. 2016; Bokulich et al. 2019), and none have featured self-collected samples from participants.

The objective of our study was to assess the utility of toilet paper swabs over whole stool in a clinical trial of FMT for UC. We compared swab and stool samples from 25 individuals with UC who received donor or autologous FMT at weeks 4 and 8 post-FMT. We also evaluated swabs that were collected at-home by the recipients over 8 weeks post-procedure to describe bacterial microbiome factors associated with remission. In the published literature, our study is the first to compare stool samples and toilet paper swabs collected by participants with IBD undergoing FMT.

Methods

Study design and sample collection

The study design and methods for the collection of stool samples are described previously in Costello et al. (2019). A subset (n = 25) of the FMT recipients in the original study collected toilet paper swabs of their stool (hereafter referred to as ‘swabs') at home daily from days 1 to 7 post-FMT, then for two different days each week from weeks 2 to 8. Each participant was given a -20°C freezer to store their swabs at home before transferring to the on-site storage facility. Whole stool samples were also collected from the recipients at baseline, two days before receiving FMT, and in weeks 4 and 8 post-FMT at the time of clinical follow-up appointments.

Sample extraction and sequencing

Stool from patients and individual donors was frozen without additive at -80°C. Swabs were stored for up to 8 weeks at -20°C prior to transfer to -80°C. Stool from the donor batches was frozen at -80°C with 65% saline and 10% glycerol. Bacterial DNA was extracted from the samples using the MoBio PowerMag Microbial DNA Isolation kit (MoBio Laboratories, Carlsbad, CA, USA) following the manufacturer’s protocol. All stool samples were extracted and processed in duplicate, whereas only single swabs were extracted. The swab extraction was similarly conducted using the PowerMag Isolation kit with an additional step at the beginning: swab tips were removed and placed into lysis buffer, then heated to 56°C for 10 minutes before continuing with the extraction as per the manufacturer's protocol. Amplicon

library preparation was performed using a modified dual-index PCR approach. The first-step primers (515F, 806R) amplified the V4 hypervariable region of the 16S rRNA gene present in bacteria and archaea, and the second set (i5, i7) added the indexed barcodes to enable multiplexing of the samples (Kozich et al. 2013). The library was pooled at equimolar concentrations and run on two independent runs (stool and swab samples) of an Illumina HiSeq2500 Rapid instrument using 2 x 250 bp paired-end chemistry (Ramaciotti Centre for Genomics, UNSW).

To account for contaminant bacterial DNA in the biological samples, negative sample controls were taken for both stool and swabs. Stool negative controls consisted of the lysis buffer only ('stool reagent blank'). There were two types of swab negative controls: the first consisted of the lysis buffer ('swab reagent blank') and the second involved an unused swab tip being extracted in the manner described above ('swab extraction blank'). Controls were prepared according to sample type as described above.

Quality control and bioinformatic analysis

The samples in our study included 507 samples from 25 FMT recipients (431 dry swabs and 76 stool samples), and these were processed and quality controlled within a larger dataset of 599 samples that included 54 donor samples, 14 donor mixes, and 24 negative controls from the original study (Costello, SP et al. 2019). Quality control was performed using the QIIME 2 (v2021.11.0) bioinformatic pipeline for microbiome sequencing data (Bolyen et al. 2019). Stool samples sequenced in lane 1 were selected over those sequenced in lane 2 as more samples successfully sequenced in lane 1. We did not merge samples between lanes to reduce lane-related sequencing bias (Kennedy et al. 2014). Read quality was assessed (qiime quality-filter q-score) and forward reads were chosen to be analysed alone over reverse reads due to a lower overall quality of the reverse reads. Forward reads were then processed with the QIIME 2 plugin Deblur (qiime deblur denoise-16S) to correct low quality sequencing reads, trim the reads to 250 bp (--p-trim-length 250), and produce a count table of unique amplicon sequence variants (ASVs) in each sample and a table of representative sequences (Amir et al. 2017a). A phylogenetic tree was constructed using the SATé-enabled phylogenetic placement (SEPP) (qiime fragment-insertion sepp) (Janssen et al. 2018), which inserted the representative sequences into the SILVA 128 99% identity reference tree (Quast et al. 2012). Taxonomy was assigned to the representative sequences with the QIIME 2 feature-classifier plugin (Bokulich et al. 2018), which uses a classifier (qiime feature-classifier classify-

sklearn) pretrained on the SILVA 138 reference database (Quast et al. 2012). Alpha rarefaction curves (qiime diversity alpha-rarefaction) were generated to inform the ideal depth of rarefaction for the dataset.

Decontamination was performed in R (v4.0.5) (R Core Team 2020) using decontam (v1.14.0) (Davis et al. 2018) by importing the QIIME 2 feature table into phyloseq (v1.38.0) (McMurdie & Holmes 2013) with qiime2R (v0.99.6) (Bisanz 2018). The prevalence method of the isContaminant function was used to identify contaminant amplicon sequence variants (ASVs) in the biological samples from the negative control samples. A stringent threshold score of 0.52 was used to remove the contaminants from the biological samples. Because decontam identifies contaminants based off their prevalence in each sample type, ASVs that were present in the negative controls through suspected cross-contamination from the biological samples were not identified as contaminants. Therefore, these ASVs were not removed during decontamination, resulting in some negative controls remaining in the dataset after this step. These samples were manually removed before downstream analyses (qiime feature-table filter-samples).

The entire dataset was initially rarefied to 1,500 16S rRNA sequences to assess the effects of contamination and other sources of variance with composition and diversity metrics (qiime diversity core-metrics-phylogenetic --p-sampling-depth 1500). This rarefaction depth retained 798,000 (5.11%) sequences in 532 (88.81%) samples from the initial mean sequence count of 26,071 (SD: 69,175) in 599 samples. The dataset was then filtered in two ways: the first for sample type comparisons and the second for swab-only analyses. The dataset for sample type comparisons consisted of the FMT recipients' stool samples from weeks 4 and 8 post-FMT and, as two swabs were collected in each week, the first swab of weeks 4 and 8. For swab-focussed analyses, samples were filtered to FMT recipient swab samples from all time points, and their stool samples from weeks 4 and 8 post-FMT. In both datasets, samples with lower than 1,000 sequences were removed (qiime feature-table filter-samples --p-min-frequency 1000). The dataset for the sample type comparison was rarefied to 1,019 sequences, the size of the smallest sample, for composition and diversity metrics (qiime diversity core-metrics-phylogenetic --p-sampling-depth 1,019), which resulted in 77,444 (5.81%) ASVs from the initial mean sequence count of 17,529 (SD: 12,480) retained in all 76 samples. The rarefaction depth for the swab-only analysis was also 1,019 sequences, which resulted in

351,555 (4.68%) sequences of the initial mean of 21,779 (SD: 71,258) sequences retained in all 345 samples.

Differential abundances were determined at the bacterial genus level using ANCOM-II (v2.1) with unrarefied count tables (Kaul et al. 2017). Data pre-processing was performed using the default values (out_cut = 0.05; zero_cut = 0.90; lib_cut = 1000; neg_lb = FALSE), the model was adjusted for the following covariates: year of FMT and FMT recipient sex, and differentially abundant genera were chosen based on the 70% probability value. Relative abundances for taxa bar plots were generated with phyloseq (v1.38.0) (McMurdie & Holmes 2013), and all plots were made with ggplot2 (v3.3.5) (Wickham 2016), with the help of the tidyverse (v1.3.1) (Wickham et al. 2019), and arranged with cowplot (v1.1.1) (Wilke 2020).

Statistical analysis

For analyses involving participant characteristics, age was defined as a categorical variable according to the Montreal classification for IBD phenotype; the participants under 40 years old were grouped separately to those over 40 (19 to 40 and 41 to 62 years) (Silverberg et al. 2005). Multivariate correlation analyses were run in R by using qiime2R to import abundance unweighted UniFrac distance matrices generated with QIIME 2. Mantel tests were run with the mantel function in the R package vegan (v2.5-7) (Oksanen et al. 2020) using Pearson's correlation and 999 permutations. Procrustes tests were run by converting UniFrac distance matrices to PCoAs with the R package ape (v5.6-2) (Paradis & Schliep 2018) and then using vegan's protest function with the symmetric Procrustes statistic and 999 permutations. Alpha diversity significance tests were calculated in QIIME 2 (qiime diversity alpha-group-significance) by performing Kruskal-Wallis one-way ANOVA tests (Kruskal & Wallis 1952) on Shannon diversity and observed features vectors. Beta diversity significance tests were also calculated in QIIME 2 with adonis multi-factor PERMANOVA (Anderson 2001) set to 999 permutations on unweighted UniFrac distance matrices (qiime diversity adonis --p-permutations 999).

Results

Assessing contamination by sample type

As bacterial contamination influences lower biomass samples in significant ways (Salter et al. 2014; Eisenhofer et al. 2019; Weyrich et al. 2019), we first examined whether contamination had a differential impact on the swabs over the stool samples. Negative controls (n = 24)

were processed and sequenced for each sample type; 10 stool reagent controls, 3 swab extraction controls, and 11 swab reagent controls were collected. Their sequence counts ranged from 8 to 13,782 sequences (mean: 2,422 sequences, SD: 3,872), indicating a high level of contamination in some controls. The controls did not have distinct compositions (beta diversity) based on whether they were processed with swabs or stool (Figure S1A & Figure S1B), and most had a composition more similar to stool than to swabs regardless of the control type (Figure 1A & Figure 1B). The five most abundant genera in the controls were gut taxa: *Blautia*, *Bacteroides*, *Faecalibacterium*, *Prevotella*, and *Collinsella* (Figure S1C).

Decontam was used to identify contaminant ASVs based on their prevalence within the negative controls and biological samples (Figure S2). Decontam identified 2,393 contaminant ASVs in the dataset of 42,072 sequences in 599 biological samples (Table S1). Contaminants were generally not typical of published reagent or environmental contaminants (

Figure S1. Negative control types did not show a distinct microbial signal. The composition of negative controls is shown with unweighted UniFrac distances displayed on PC1 and PC2 in (A) and PC1 and PC3 in (B), and with relative abundances of genera in (C).

) (Salter et al. 2014). Instead, most contaminants (98%) were taxa from the major gut bacterial phyla: Firmicutes (71.5%), Bacteroidota (15.7%), Actinobacteria (5.4%), Proteobacteria (3.9%), Verrumicrobiota (1.1%), and Fusobacteria (0.9%) (

Table **S2**), indicating the contamination may have been faecal in origin. There were more contaminant ASVs identified in the swab sample dataset (2,352) than stool samples (2,110) (Table S3). However, the swab dataset contained more overall samples (n = 445) than the stool dataset (n = 154), so the proportion of contaminants was lower in the swabs (9%) than in stool (13%).

As the contaminant signal did not appear to differ between control types, we did not separate the dataset into each sample type before removing contamination. Out of caution, we removed all 2,393 putative contaminant ASVs from both swab and stool samples. Thus, it is likely that some biological signal at the ASV level was removed in this process (Karstens et al. 2019). Only one stool negative control was removed with decontamination (Table S3); the rest of the controls still remained in the dataset (Figure 1C & Figure 1D). This was to be expected since some of the controls were highly contaminated with gut taxa, making it difficult for decontam to distinguish all true contaminants from gut taxa present in the biological samples. The remaining negative controls were subsequently removed from the dataset. Decontamination did not have a large effect on the composition of the dataset or distribution of swab and stool samples (Figure 1E & Figure 1F). The swabs were not more impacted by decontamination than stool, possibly due to the higher overall number of swabs in the dataset.

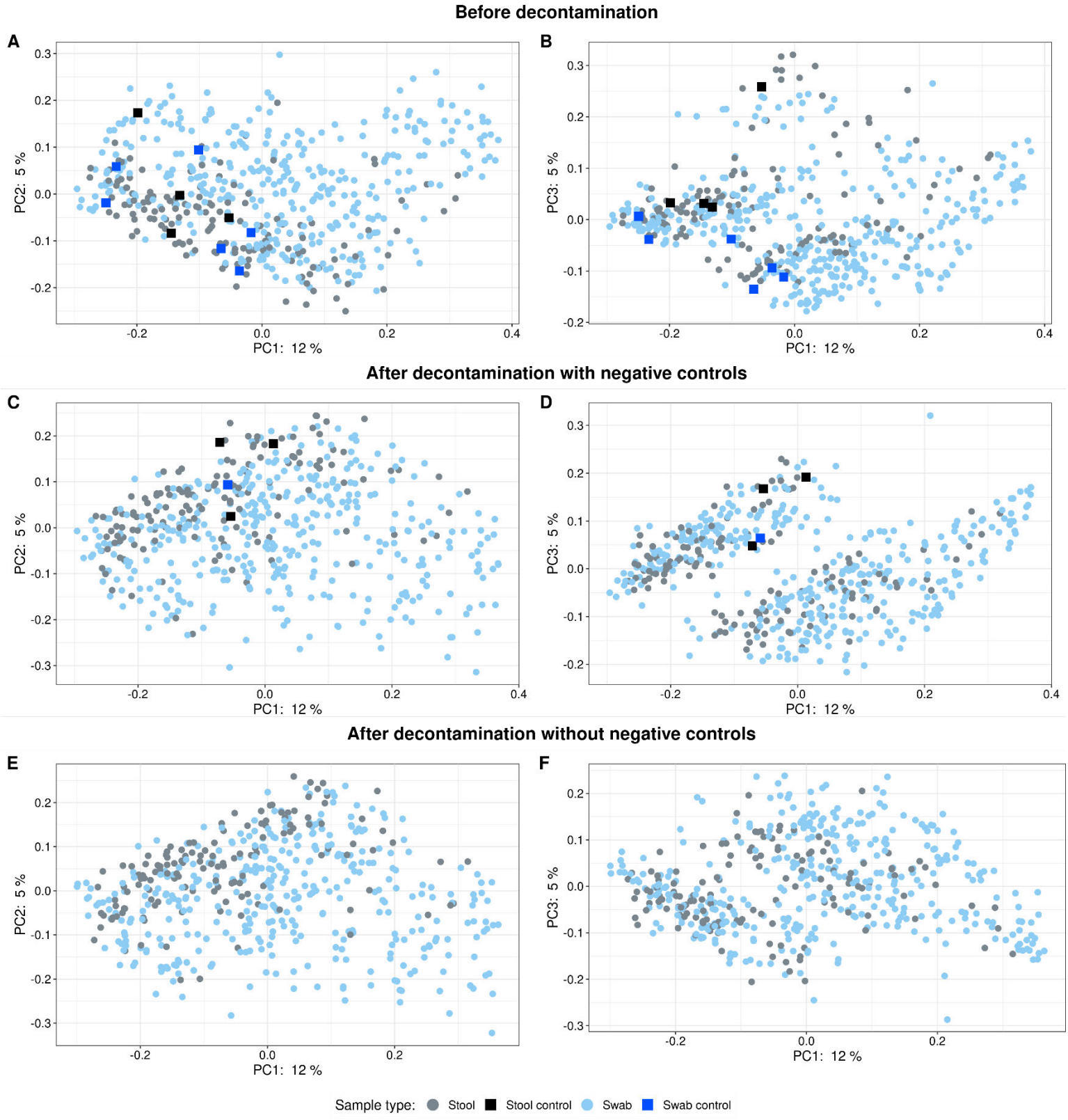

Figure 1. Removing contaminant bacterial DNA did not have a large effect on the compositional differences between swab and stool samples. PCoA plots were generated with unweighted UniFrac distances of samples rarefied to 1,500 16S rRNA sequences. Stool and swab samples from faecal microbiota transplant (FMT) donors and recipients are shown

before decontamination on (A) PC1 and PC2, and (B) PC1 and PC3. All samples after decontamination are shown on (C) PC1 and PC2, and (D) PC1 and PC3. As decontamination did not remove all negative control samples, those that remained were manually removed from the dataset. The remaining samples without negative controls are shown on (E) PC1 and PC2, and (F) PC1 and PC3.

Technical and participant sources of microbiome variance

After filtering out contaminants and negative control samples, the dataset was filtered to FMT recipient samples to test for sources of technical bias that could impact the resulting microbiome signal. One sampling variable—year of FMT administration—had a significant impact on the composition (beta diversity) of the recipients' swabs (Adonis: $R^2 = 0.05$, $p = 0.00$) and stool samples (Adonis: $R^2 = 0.06$, $p = 0.00$) (Figure 2, Figure S3 &

Table **S4**), indicating that the time of sampling influenced the microbiota observed. In particular, samples from 2014 strongly clustered together and were different in alpha diversity from samples taken in 2015 and 2016 (swabs: observed ASVs: H = 41.78, q = 0.00 & Shannon diversity: H = 38.42, q = 0.00; stool: observed ASVs: H = 14.08, q = 0.00 & Shannon diversity: H = 13.95, q = 0.00) (Table S5). There were no differentially abundant genera between the years of FMT in the stool samples, but 117 genera differed in the swab samples, which suggests that the batch effects associated with the year of FMT had a greater impact on the swab samples (Table S6).

We also tested whether participant characteristics influenced the microbiome variation of both sample types. FMT recipient age and sex did not significantly impact the alpha diversity of the swabs (age: observed ASVs: H = 0.89, q = 0.35 & Shannon diversity: H = 0.10, q = 0.76; sex: observed ASVs: H = 0.95, q = 0.33 & Shannon diversity: H = 3.39, q = 0.07), or stool samples (age: observed ASVs: H = 0.08, q = 0.77 & Shannon diversity: H = 0.01, q = 0.92; sex: observed ASVs: H = 2.69, q = 0.10 & Shannon diversity: H = 3.13, q = 0.08) (Table S5). However, these variables had an effect on the composition of each sample types (Figure 2 & Figure S3). Both recipient age and sex equally affected the composition of the swabs (age: Adonis: R^2 = 0.02, p = 0.02; sex: Adonis: R^2 = 0.02, p = 0.00), and recipient sex also affected the composition of the stool samples (Adonis: R^2 = 0.02, p = 0.03) (

Table **S4**).

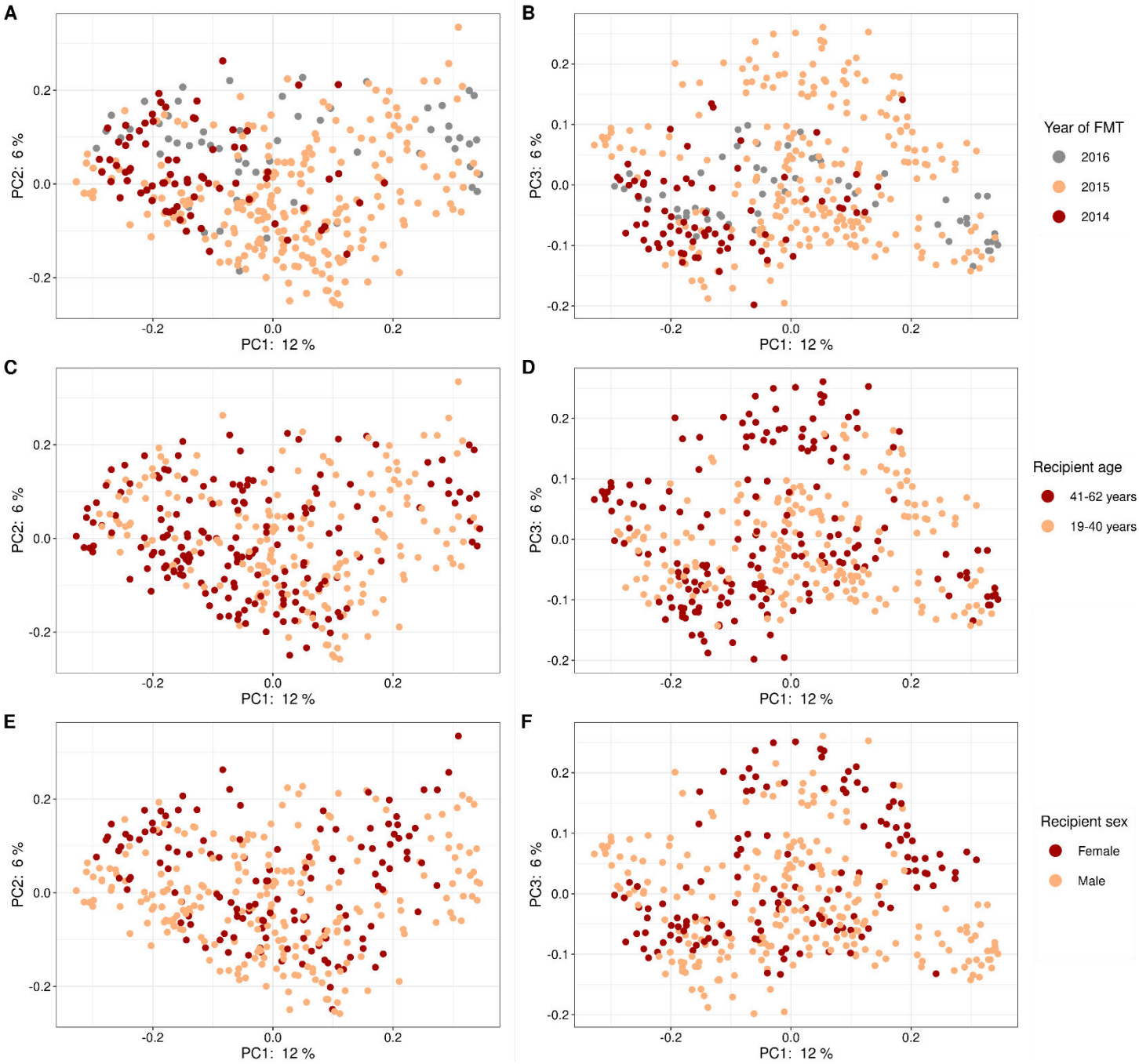

Figure 2. The year the faecal microbiota transplant (FMT) was performed and the FMT recipient age and sex impacted the bacterial composition of the recipient swab samples. PCoA plots were generated with unweighted UniFrac distances of samples rarefied to 1,019 16S rRNA sequences. Composition is shown with PCoA plots of unweighted UniFrac distances and samples are coloured by the year of FMT on (A) PC1 and PC2, and (B) PC1 and PC3. Samples are coloured by FMT recipient sex on (C) PC1 and PC2, and (D) PC1 and PC3.

Swab samples differed from stool in composition but not diversity

We assessed how dry swabs captured the gut microbiome compared to the field standard for specimen collection—whole stool. The dataset was filtered to swab and stool samples collected from FMT recipients in the same weeks—weeks 4 and 8—to mitigate the effect of time variation on the sample comparisons. As swabs were collected twice in each week from weeks 2 to 8, the first swabs collected in weeks 4 and 8 were retained over the second swabs due to a higher amount of missing of the second swabs (Figure S4). Sample filtering resulted in 42 samples from 24 individuals at week 4 post-FMT (20 swab and 22 stool samples), and 34 samples from 23 individuals at week 8 (16 swab and 18 stool samples). The swab dataset contained 36 samples with a mean of 14,239 sequences (SD: 11,863) and the stool samples had a mean of 20,490 (SD: 12,420) sequences in 40 samples. Many of the ASVs across both sample types were assigned to the same bacterial genera, so we collapsed the ASVs to the genus level to test for bacteria that differed between swabs and stool with ANCOM-II (Kaul et al. 2017). Four of the genera that were more abundant in the week 4 swabs over stool—*Finegoldia*, *Anaerococcus*, *Peptoniphilus,* and *Fenollaria*—were also more abundant in the swabs due to the year of FMT. Thus, we attributed these differentially abundant genera to batch effects and removed them from the results. Of the remaining three genera, *Corynebacterium* were more abundant in the swabs, whereas *Bifidobacterium* and *Mitsuokella* were less abundant (Table 1). There were no differentially abundant genera at the second time point of samples, week 8.

We also assessed how the sample types compared in microbiome diversity and composition. Swab and stool samples did not differ in their observed number of ASVs and evenness of ASVs (Shannon diversity) from each other at either time point (week 4 observed ASVs: H = 1.55, q = 0.21 & Shannon diversity: H = 0.01, q = 0.94; week 8 observed ASVs: H = 2.57, q = 0.11 & Shannon diversity: H = 0.58, q = 0.45) (Figure 3A & Figure 3B; Table S7). We tested whether swab and stool differed compositionally using a multi-factor PERMANOVA (adonis) to incorporate the sources of variation (year of FMT and FMT recipient age and sex) that could affect the sample type comparison (Anderson 2001). The year of FMT administration had the greatest impact on the variation in sample compositions (week 4: R^2 = 0.07, p = 0.01; week 8: R^2 = 0.09, p = 0.00), and sample type had the second highest impact (week 4: R^2 = 0.05, p = 0.00; week 8: R^2 = 0.04, p = 0.04) (Table S8).

It is known that gut microbiome samples from the same individual tend to be more similar to each other than to samples from other individuals, and this has been shown in rectal swabs as well (Bassis et al. 2017; Biehl et al. 2019; Reyman et al. 2019). Thus, we examined whether paired swab and stool samples would better illustrate any differences between the sample types. The pairwise dataset consisted of recipients who had both a stool and swab sample at week 4 or a stool and swab sample at week 8. This step reduced the dataset to 36 samples from 18 recipients who had both samples at week 4 (18 swab and 18 stool samples) and 22 samples from 11 recipients who had both samples at week 8 (11 swab and 11 stool samples). The taxonomic makeup of the paired samples showed similarities between some individuals at the bacterial genus level, while others had noticeable differences of particular genera between swabs and stool (Figure S5). Across the paired samples, there were 217 genera identified in the swabs and 189 genera identified in the stool samples. After filtering out the four genera that were likely caused by batch effects, *Finegoldia*, *Anaerococcus*, *Peptoniphilus*, and *Fenollaria*, the results of differential abundance testing on the paired samples revealed that *Anaerostipes* and *Monoglobus* were less abundant in the swabs (Table 1).

Like the unpaired samples, the paired samples did not statistically differ in diversity (week 4 observed ASVs: H = 2.50, q = 0.11 & Shannon diversity: H = 0.14, q = 0.70; week 8 observed ASVs: H = 0.85, q = 0.36 & Shannon diversity: H = 0.05, q = 0.82) (Figure 3C & Figure 3D; Table S7). The microbial compositions of the paired samples were most affected by the year of FMT (week 4: $R^2 = 0.08$, p = 0.01; week 8: $R^2 = 0.16$, p = 0.00), followed by sample type for week 4 only (week4: $R^2 = 0.06$) (Table S8). Week 8 samples were more impacted by recipient age and sex (age: $R^2 = 0.06$, p = 0.07; sex: $R^2 = 0.06$, p = 0.03) than sample type ($R^2 = 0.05$, p = 0.18), though this may have been influenced by the lower sample size at this time point (n = 22). PCoAs of the paired samples showed that samples from the same individuals clustered more closely on PC1 and PC3 than PC1 and PC2 for both time points (Figure S6). This may be explained by the stronger effect of year of FMT on the compositional variation between samples, which may have driven more of the differences seen on PC1 and PC2.

We assessed the microbial similarity between pairwise sample compositions using two multivariate correlation tests—Mantel and Procrustes (Mantel 1967; Peres-Neto & Jackson 2001). The Mantel test reported a significant correlation of 25% (p = 0.03) between the

composition of swab and stool samples, whereas week 8 samples were not significantly similar (p = 0.43) and had no correlation (r = -0.00) (Table S9). Similarly, Procrustes for week 4 showed a significant correlation between swab and stool (p = 0.04) but not for week 8 (p = 0.56). However, week 8 swab and stool samples were more symmetrically correlated (96%) than week 4 samples (92%). These discrepancies are likely due to fewer samples at week 8 (n = 22) than week 4 (n = 36), although this cannot be confirmed from our tests.

Table 1. Bacterial genera that were differentially abundant between swab and stool samples.

Week 4				**Week 8**			
No. samples tested	**Bacterial genus**	**W value**	**Presence in swabs vs stool**	**No. samples tested**	**Bacterial genus**	**W value**	**Presence in swabs vs stool**
Unpaired samples							
n = 42	*Corynebacterium*	153	↑*	n = 34	No differentially abundant taxa		
	Bifidobacterium	152	↓^				
	Mitsuokella	145	↓				
Paired samples							
n = 36	*Bifidobacterium*	156	↓	n = 22	No differentially abundant taxa		
	Corynebacterium	149	↑				
	Anaerostipes	142	↓				
	Monoglobus	138	↓				

*Genus is increased in swabs, decreased in stool. ^Genus is decreased in swabs, increased in stool.

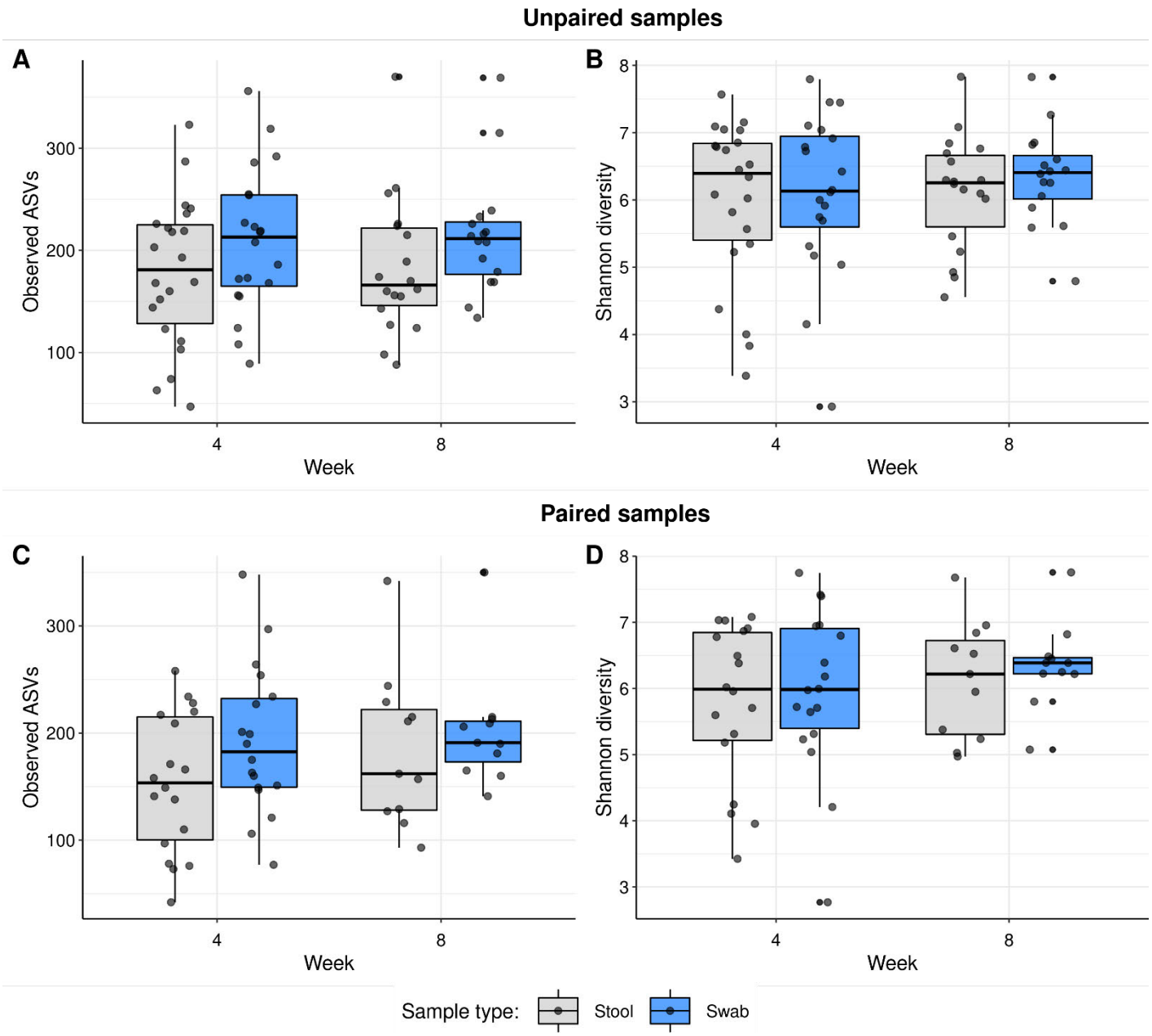

Figure 3. Swab and stool samples from faecal microbiota transplant recipients had comparable levels of diversity across both time points. Alpha diversity (i.e., within-sample diversity) is shown for unpaired samples as the observed of number amplicon sequence variants (ASVs) (A) and evenness of observed ASVs (Shannon diversity) (B). Similarly, the alpha diversity of paired samples is shown as the observed number of ASVs (C) and Shannon diversity (D). All samples were rarefied to 1,019 16S rRNA sequences. The diversities between sample types at each time point were not significant (Kruskal-Wallis, $p > 0.05$).

Diversity and composition of swab samples associated with FMT recipient outcome

After assessing how the swab samples the gut microbiome compared to stool samples, we examined how they associated with recipient outcome which was measured as clinical and endoscopic remission at 8 weeks post-FMT. The swab dataset initially consisted of 386 swab samples collected from 25 individuals across all 8 weeks of the study. Due to the high missingness of some recipients' samples (Figure S4), we removed recipients who had less than 50% of samples across all 21 time points (i.e., less than 11 samples) from the dataset. This filtering step removed 6 individuals and 41 samples, resulting in a dataset of 345 samples from 19 individuals. Microbiome diversity was significantly lower in the post-FMT swabs of individuals who achieved remission over those who did not (observed ASVs: H = 38.38, q = 0.00 & Shannon diversity: H = 59.39, q = 0.00) (Figure 4A & Figure 4B; Table S10). This trend was found in the group of individuals who experienced remission regardless of whether they received donor FMT or autologous FMT (receiving their own stool), and across two-week time points from weeks 1 to 8 (Figure 4C & Figure 4D). The group of individuals who experienced remission did, however, have lower alpha diversity in their pre-FMT stool samples than those who did not respond to FMT, though this difference was not significant (Figure S7).

The association between microbial composition and FMT outcome was tested with an adonis model that incorporated all potential sources of microbial variation (year of FMT, recipient age and sex, and FMT type) and an interaction between FMT type and outcome (Table S11). All factors had a significant effect on composition; the year of FMT explained the most variation (R^2 = 0.05, p = 0.00), followed by FMT type and recipient age (FMT type: R^2 = 0.03, p = 0.00; age: R^2 = 0.03, p = 0.00), recipient sex (R^2 = 0.02, p = 0.00). The multiplicative effect of FMT type and outcome had more of an effect on the composition than outcome alone (FMT type*outcome: R^2 = 0.02, p = 0.00; FMT outcome: R^2 = 0.01, p = 0.00). This association was stronger in the swabs than when tested in the weeks 4 and 8 post-FMT stool samples, which had no significant association with composition and FMT outcome or type (FMT type: R^2 = 0.03, p = 0.14; FMT outcome: R^2 = 0.03, p = 0.10; FMT type*outcome: R^2 = 0.03, p = 0.13) (Table S12). However, the sample size of the stool samples was much smaller than the swab dataset which would have contributed to strength of this association.

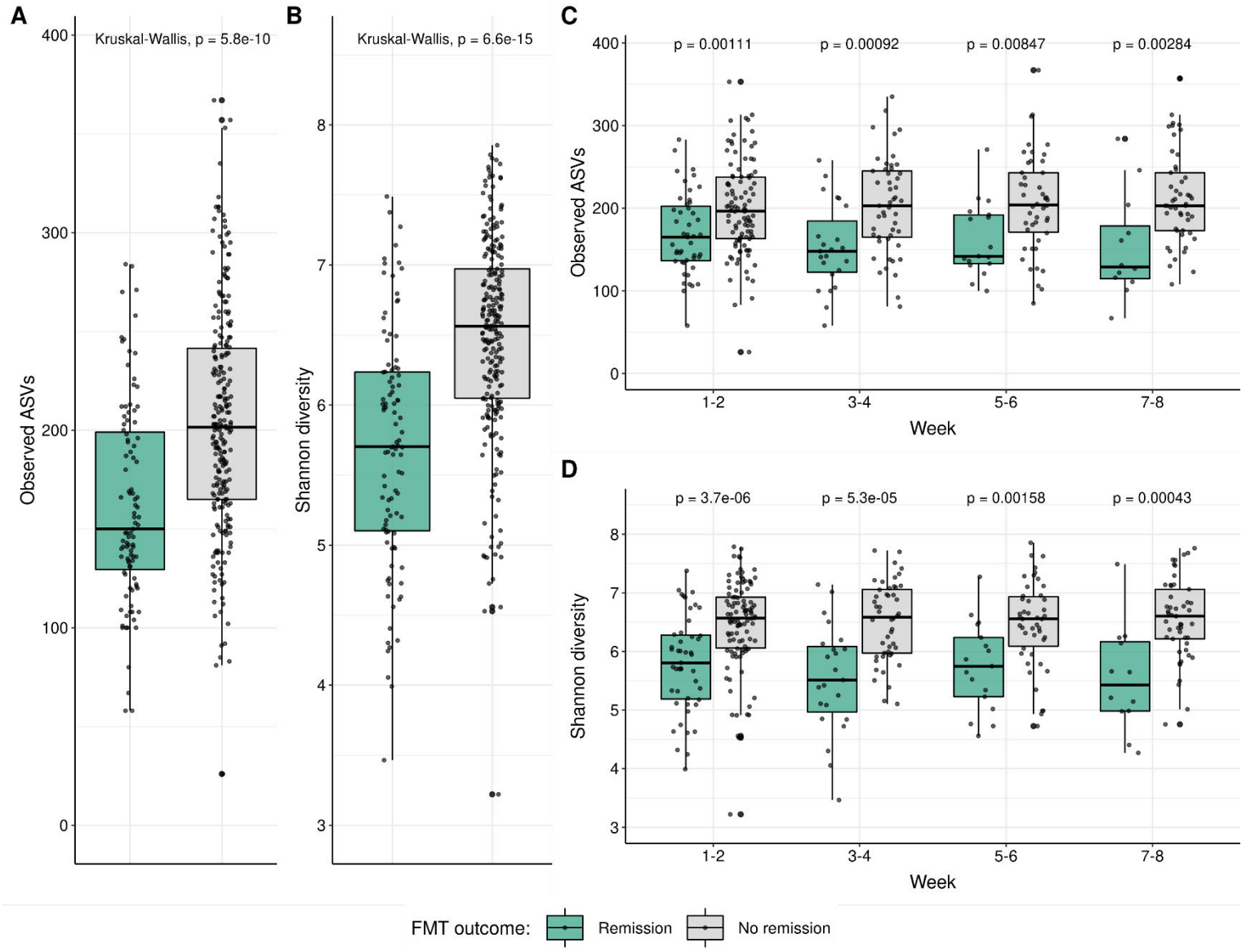

Figure 4. The alpha diversity of dry swabs was decreased in faecal microbiota transplant (FMT) recipients who achieved remission. FMT outcome was defined by clinical and endoscopic remission at week 8 post-FMT. The alpha diversity of swabs is shown for 19 recipients across 8 weeks (n = 345) post-FMT by (A) the observed number of amplicon sequence variants (ASVs) and (B) Shannon diversity, and in two-week intervals showing (C) the observed number of ASVs and (D) Shannon diversity. Significance testing between outcome groups was performed with Kruskal-Wallis tests.

Discussion

Our study featured individuals with UC who self-collected dry swab and whole stool samples following FMT. Swabs were used as an alternative to whole stool samples which could be easily collected at home and feasibly track the microbiome response to FMT on a finer scale over time. Given that stool samples somewhat differ between other sample types such as rectal swabs and mucosal biopsies (Budding et al. 2014; Sun et al. 2021), we expected that dry swabs would likewise show some differences from stool. Indeed, we found that the swabs and stool samples had similar levels of diversity but different compositions.

A limited number of studies have compared dry swabs to whole stool and have found them to be highly similar in composition and diversity (Costello, EK et al. 2009; Sinha et al. 2016; Bokulich et al. 2019). Several studies have shown that rectal swabs also do not differ in diversity from stool (Reyman et al. 2019; Short et al. 2021). Short et al. (2021) found that the composition of rectal and glove tip swabs differed from stool samples, whereas other studies did not share this finding (Bassis et al. 2017; Reyman et al. 2019). Several factors may lead to discrepancies between these studies and ours, such as methodologies and the types of cohorts sampled. For example, Reyman et al. (2019) found that clinical covariates explained more variation in the composition of samples than the sample type. However, their study featured samples collected from infants which would likely be differently impacted by host factors than the samples in our study. Thus, the varying effect sizes of clinical covariates between study cohorts likely play a role in the comparability of sample types, and future studies comparing sample types between different cohorts and health states are needed.

A major concern when working with dry swabs is that, because they have lower biomass than whole stool samples, they will capture a lower, and potentially insufficient, number of microbial taxa. We found that the dry swabs in our study captured a comparable number of gut taxa to whole stool, and this has been similarly observed in rectal swabs (Budding et al. 2014; Bassis et al. 2017; Reyman et al. 2019). Although the swabs had comparable alpha diversity to stool, some of the taxa in the swabs differed from those captured by stool. We hypothesised that this compositional difference could be due to a higher level of skin or mucosal taxa in the swabs. However, we did not detect taxa that are exclusively associated with the skin in our swab samples. Budding et al. (2014) also did not find evidence of contamination by skin taxa in the rectal swabs in their study. They hypothesised that the high

biomass of faeces in swabs likely overwhelms the detection of any skin taxa. We identified one bacterial genus, *Corynebacterium*, that was more abundant in the swabs and is known to colonise both the skin and mucosa (Oliveira et al. 2017). This genus has been previously identified in a study comparing rectal swabs to stool (Biehl et al. 2019). The authors speculated that its presence was likely due to contamination from the skin but also noted that another study found increased levels of *Corynebacterium* in mucosal biopsies (Ringel et al. 2015). Thus, rectal and dry swabs may contain *Corynebacterium* spp. that are mucosal in origin, and further research is needed before categorising these bacteria as contaminants in the context of gut microbiome swabs. To answer this question, negative control swabs of skin from the participants and/or researchers would be invaluable to determine the source of the taxa. Additionally, whole-metagenome sequencing could enable functional analyses to distinguish skin and mucosal taxa.

Other than *Corynebacterium*, we did not detect other aerobic, mucosal-associated bacteria at significant levels in the swabs in our study. Rectal swabs have been shown to represent more of the colonic mucosal microbiome than stool, though both look different to mucosal tissue samples (Araújo-Pérez et al. 2012; Biehl et al. 2019; Sun et al. 2021). We expected the dry swabs to sit somewhere in between rectal swab and stool sample types. An important consideration is that the samples in our study were collected from individuals with colonic inflammation, and this may have reduced the number of mucosal taxa that were present in the individuals. Individuals with UC have a thinner mucus layer in the colon that does not sufficiently support colonisation by commensal microbes, even in the case of some individuals in remission (Johansson et al. 2014; Bankole et al. 2021). Therefore, fewer mucosal bacteria may be present in the swabs of individuals with UC. An important future direction is to compare dry swabs from healthy individuals to swabs from individuals with intestinal inflammation to determine whether they are comprised of different levels of mucosal-associated taxa.

It is also important to note that shotgun metagenomic sequencing would have given a finer taxonomic resolution over the 16S marker gene sequencing used in this study. Other studies have found shotgun sequencing to be more sensitive in detecting phyla (Campanaro et al. 2018) and genera (Durazzi et al. 2021) than 16S sequencing, and the method enables more taxa to be identified at the species and strain levels (Quince et al. 2017). Additionally, shotgun sequencing has successfully been used to profile metagenomes in other swab types

with low microbial biomass such as rectal, vaginal, throat, and skin swabs (Ahannach et al. 2021; Guo et al. 2021; Karpinets et al. 2022). This method could be used to uncover more mucosal taxa in swabs from FMT recipients and distinguish strains associated with skin over colonic ones. This method could also give greater insight into the functional potential of the post-transplant microbiome, and detect other microbial types such as fungi, protozoa, and viruses that may be transplanted or affected by FMT.

After assessing the utility of swabs over whole stool samples, we examined if the swabs would be informative in associating with patient factors such as FMT success. Previous studies have compared the microbial diversity in stool samples from post-FMT individuals with UC and have typically found that the remission groups had higher alpha diversity in weeks 4 to 12 (Rossen et al. 2015; Paramsothy et al. 2017; Chen et al. 2020; Li et al. 2020), though others have not found any correlation between diversity and remission with stool sampling (Costello, SP et al. 2019; Haifer et al. 2022). Our study found the opposite; the swabs of the remission group had lower microbial diversity than the group who did not achieve remission. While this was surprising, it should be noted that the remission group had lower diversity even before receiving FMT, and this may have contributed to the trend seen in the post-FMT weeks. Nonetheless, it is possible that this lower diversity contributed to the success of FMT and/or engraftment of donor bacteria, but we cannot be conclusive as this study is limited by small sample numbers. It is also likely that microbial diversity is not the entire picture, and certain bacterial functions acquired through transplantation of those that enter remission may be more explanatory. For example, acquiring bacteria with the ability to synthesise short-chain fatty acids and secondary bile acids (Paramsothy et al. 2019).

Few studies have examined the differences in beta diversity (composition) between remission and non-remission groups in post-FMT individuals with UC. One study found a difference in the composition of stool samples between these two groups (Moayyedi et al. 2017), whereas another did not (Li et al. 2020). The composition of swab samples in our study strongly associated with the type of FMT and outcome. In contrast, the stool samples did not show such an association, and this was likely affected by the much smaller sample size of the stool dataset, which were only collected in weeks 4 and 8 after FMT. Individuals with UC who have received FMT are typically assessed around weeks 4 and 8 post-procedure for clinical and endoscopic remission, with the recipient's stool microbiome also being sampled around these time points (Paramsothy et al. 2017; Costello, SP et al. 2019; Chen et al. 2020; Li et al.

2020). The use of self-collected swab samples allowed for more flexibility in our study to measure time points in between these weeks, including directly after receiving FMT, and this sample resolution facilitated a greater association with FMT covariates over the standard stool collection time points.

In addition to providing novel insights into the comparability of swab and stool samples, our study highlights the importance of controlling for batch effects and contaminant profiles in microbiome research, regardless of biomass in the biological samples. In particular, batch effects can be differentially associated with sample types and time points, and these should be considered when designing any microbiome study. We collected negative controls in our study to account for contamination during the laboratory processes. Due to the lower biomass of swabs, we expected that contamination might affect these samples more than whole stool. Although some of our controls picked up a considerable number of taxa, it did not appear that the swabs were more affected by contamination than the stool samples. Nevertheless, the high presence of gut taxa in our controls demonstrates the necessity to account for contamination in environments where high biomass samples such as stool are processed. This consideration is crucial when processing lower biomass samples in the same environment, and because of the potential for cross-contamination between high biomass samples. Our diligence in accounting for contamination in our samples may also explain why our results differ from previous studies, as we have not seen such detailed decontamination methods in other publications of FMT in UC.

Another batch effect in our study was the year the FMT was performed, and this covariate significantly impacted downstream analyses. The year of FMT can be considered a proxy for several time-variant factors such as the length of storage time of samples, different donors and recipients during different study years, and other stochastic sources of variation between the years. The older samples differed the most in diversity and composition from the samples from later years in the study. The effect of time spent in frozen storage has not been thoroughly investigated in faecal swabs in the literature, so we were unable to determine how much storage time had an impact on the samples over other sources of variation. Other studies have found that the time spent in transit from collection to storage can cause bacterial blooms in dry swabs (Amir et al. 2017a; McDonald et al. 2018), but this was not a factor in our study as the participants were given -20°C freezers to immediately store their samples.

While our study used the gold standard for specimen preservation, we acknowledge that this may be impractical in many cases. There are several other preservatives available to store samples at room temperature prior to freezing (e.g., OMNIgene GUT, RNAlater, and FTA), but these preservatives have been found to bias diversity results (Dominianni et al. 2014; Choo, Leong & Rogers 2015; Hale et al. 2015; Song et al. 2016; Vandeputte et al. 2017). Therefore, it is recommended when using one of these preservatives at room temperature, the type of preservative should remain consistent throughout an entire study (Nearing, Comeau & Langille 2021). Specifically in faecal swab samples, Marotz et al. (2021) found that preservation in 95% ethanol preservation was the next best alternative to freezing for this sample type.

Our study included the self-collection of dry swabs by participants over the course of eight weeks. Participant adherence is a challenge in clinical research including microbiome studies, and flexible sampling options, such as dry swabs, have been developed to mitigate this issue (Debelius et al. 2016a). Even so, our swab dataset still suffered from the impacts of sample dropout, and we were unable to perform longitudinal regression on the swabs because of missing samples. The sample missingness was random; some days were entirely missing in some participants, whereas other samples were present but did not yield sufficient microbial DNA. Given that our starting sample size was small, any sample missingness made statistical analyses difficult. Our study demonstrates that the effects of missing samples should be considered when designing longitudinal microbiome studies, even when more feasible sample types are used.

Conclusion

In summary, dry faecal swabs are a convenient option for sampling the gut microbiome over whole stool in individuals with ulcerative colitis. While they differ in composition from whole stool, swabs capture an informative level of diversity of colonic bacterial species and can associate with patient factors. Future research is needed to assess how dry swabs differ between healthy individuals from those with a history of intestinal inflammation, and how contamination and other technical batch effects impact dry swabs over other sample types. Finally, a comparison of matched faecal, mucosal, and rectal swab samples to toilet paper swabs from individuals would be of great value in further evaluating the utility of each sample type.

References

Ahannach, S, Delanghe, L, Spacova, I, Wittouck, S, Van Beeck, W, De Boeck, I & Lebeer, S 2021, 'Microbial enrichment and storage for metagenomics of vaginal, skin, and saliva samples', *iScience*, vol. 24, no. 11, 2021/11/19/, p. 103306.

Aldars-García, L, Chaparro, M & Gisbert, JP 2021, 'Systematic Review: The Gut Microbiome and Its Potential Clinical Application in Inflammatory Bowel Disease', *Microorganisms*, vol. 9, no. 5, Apr 30.

Amir, A, McDonald, D, Navas-Molina, JA, Debelius, J, Morton, JT, Hyde, E, Robbins-Pianka, A, Knight, R & Arumugam, M 2017a, 'Correcting for Microbial Blooms in Fecal Samples during Room-Temperature Shipping', *mSystems*, vol. 2, no. 2, pp. e00199-00116.

Anderson, MJ 2001, 'A new method for non-parametric multivariate analysis of variance', *Austral Ecology*, vol. 26, no. 1, 2001/02/01, pp. 32-46.

Araújo-Pérez, F, McCoy, AN, Okechukwu, C, Carroll, IM, Smith, KM, Jeremiah, K, Sandler, RS, Asher, GN & Keku, TO 2012, 'Differences in microbial signatures between rectal mucosal biopsies and rectal swabs', *Gut Microbes*, vol. 3, no. 6, Nov-Dec, pp. 530-535.

Bankole, E, Read, E, Curtis, MA, Neves, JF & Garnett, JA 2021, 'The Relationship between Mucins and Ulcerative Colitis: A Systematic Review', *Journal of clinical medicine*, vol. 10, no. 9, p. 1935.

Bassis, CM, Moore, NM, Lolans, K, Seekatz, AM, Weinstein, RA, Young, VB, Hayden, MK & for the, CDCPEP 2017, 'Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles', *BMC Microbiology*, vol. 17, no. 1, 2017/03/31, p. 78.

Biehl, LM, Garzetti, D, Farowski, F, Ring, D, Koeppel, MB, Rohde, H, Schafhausen, P, Stecher, B & Vehreschild, MJGT 2019, 'Usability of rectal swabs for microbiome sampling in a cohort study of hematological and oncological patients', *PLoS One*, vol. 14, no. 4, pp. e0215428-e0215428.

Bisanz, JE, 2018, *qiime2R: Importing QIIME2 artifacts and associated data into R sessions*, ver. 0.99,

Bokulich, NA, Kaehler, BD, Rideout, JR, Dillon, M, Bolyen, E, Knight, R, Huttley, GA & Gregory Caporaso, J 2018, 'Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin', *Microbiome*, vol. 6, no. 1, 2018/05/17, p. 90.

Bokulich, NA, Maldonado, J, Kang, DW, Krajmalnik-Brown, R & Caporaso, JG 2019, 'Rapidly Processed Stool Swabs Approximate Stool Microbiota Profiles', *mSphere*, vol. 4, no. 2, Apr 10.

Bolyen, E, Rideout, JR, Dillon, MR, Bokulich, NA, Abnet, CC, Al-Ghalith, GA, Alexander, H, Alm, EJ, Arumugam, M, Asnicar, F, Bai, Y, Bisanz, JE, Bittinger, K, Brejnrod, A, Brislawn, CJ, Brown, CT, Callahan, BJ, Caraballo-Rodríguez, AM, Chase, J, Cope, EK, Da Silva, R, Diener, C, Dorrestein, PC, Douglas, GM, Durall, DM, Duvallet, C, Edwardson, CF, Ernst, M, Estaki, M, Fouquier, J, Gauglitz, JM, Gibbons, SM, Gibson, DL, Gonzalez, A, Gorlick, K, Guo, J, Hillmann, B, Holmes, S, Holste, H, Huttenhower, C, Huttley, GA, Janssen, S, Jarmusch, AK, Jiang, L, Kaehler, BD, Kang, KB, Keefe, CR, Keim, P, Kelley, ST, Knights, D, Koester, I, Kosciolek, T, Kreps, J, Langille, MGI, Lee, J, Ley, R, Liu, Y-X, Loftfield, E, Lozupone, C, Maher, M, Marotz, C, Martin, BD, McDonald, D, McIver, LJ, Melnik, AV, Metcalf, JL, Morgan, SC, Morton, JT, Naimey, AT, Navas-Molina, JA, Nothias, LF, Orchanian, SB, Pearson, T, Peoples, SL, Petras, D, Preuss, ML, Pruesse, E, Rasmussen, LB, Rivers, A, Robeson, MS, Rosenthal, P, Segata, N, Shaffer, M, Shiffer, A, Sinha, R, Song, SJ, Spear, JR, Swafford, AD, Thompson, LR, Torres, PJ, Trinh, P, Tripathi, A, Turnbaugh, PJ, Ul-Hasan, S, van der Hooft, JJJ, Vargas, F, Vázquez-Baeza, Y, Vogtmann, E, von Hippel, M, Walters, W, Wan, Y, Wang, M, Warren, J, Weber, KC, Williamson, CHD, Willis, AD, Xu, ZZ, Zaneveld, JR, Zhang, Y, Zhu, Q, Knight, R & Caporaso, JG 2019, 'Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2', *Nature Biotechnology*, vol. 37, no. 8, 2019/08/01, pp. 852-857.

Budding, AE, Grasman, ME, Eck, A, Bogaards, JA, Vandenbroucke-Grauls, CMJE, van Bodegraven, AA & Savelkoul, PHM 2014, 'Rectal swabs for analysis of the intestinal microbiota', *PLoS One*, vol. 9, no. 7, pp. e101344-e101344.

Cai, Z, Wang, S & Li, J 2021, 'Treatment of Inflammatory Bowel Disease: A Comprehensive Review', *Frontiers in Medicine*, vol. 8, 2021-December-20.

Campanaro, S, Treu, L, Kougias, PG, Zhu, X & Angelidaki, I 2018, 'Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies', *Sci Rep*, vol. 8, no. 1, Jan 31, p. 1926.

Chen, HT, Huang, HL, Xu, HM, Luo, QL, He, J, Li, YQ, Zhou, YL, Nie, YQ & Zhou, YJ 2020, 'Fecal microbiota transplantation ameliorates active ulcerative colitis', *Exp Ther Med*, vol. 19, no. 4, Apr, pp. 2650-2660.

Choo, JM, Leong, LEX & Rogers, GB 2015, 'Sample storage conditions significantly influence faecal microbiome profiles', *Sci Rep*, vol. 5, no. 1, 2015/11/17, p. 16350.

Costello, EK, Lauber, CL, Hamady, M, Fierer, N, Gordon, JI & Knight, R 2009, 'Bacterial Community Variation in Human Body Habitats Across Space and Time', *Science*, vol. 326, no. 5960, pp. 1694-1697.

Costello, SP, Hughes, PA, Waters, O, Bryant, RV, Vincent, AD, Blatchford, P, Katsikeros, R, Makanyanga, J, Campaniello, MA, Mavrangelos, C, Rosewarne, CP, Bickley, C, Peters, C, Schoeman, MN, Conlon, MA, Roberts-Thomson, IC & Andrews, JM 2019, 'Effect of Fecal Microbiota Transplantation on 8-Week Remission in Patients With Ulcerative Colitis: A Randomized Clinical Trial', *Jama*, vol. 321, no. 2, Jan 15, pp. 156-164.

Costello, SP, Soo, W, Bryant, RV, Jairath, V, Hart, AL & Andrews, JM 2017, 'Systematic review with meta-analysis: faecal microbiota transplantation for the induction of remission for active ulcerative colitis', *Aliment Pharmacol Ther*, vol. 46, no. 3, Aug, pp. 213-224.

Davis, NM, Proctor, DM, Holmes, SP, Relman, DA & Callahan, BJ 2018, 'Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data', *Microbiome*, vol. 6, no. 1, 2018/12/17, p. 226.

Debelius, JW, Vázquez-Baeza, Y, McDonald, D, Xu, Z, Wolfe, E & Knight, R 2016a, 'Turning Participatory Microbiome Research into Usable Data: Lessons from the American Gut Project', *Journal of microbiology & biology education*, vol. 17, no. 1, pp. 46-50.

Dominianni, C, Wu, J, Hayes, RB & Ahn, J 2014, 'Comparison of methods for fecal microbiome biospecimen collection', *BMC Microbiology*, vol. 14, no. 1, 2014/04/23, p. 103.

Durazzi, F, Sala, C, Castellani, G, Manfreda, G, Remondini, D & De Cesare, A 2021, 'Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota', *Sci Rep*, vol. 11, no. 1, Feb 4, p. 3030.

Eisenhofer, R, Minich, JJ, Marotz, C, Cooper, A, Knight, R & Weyrich, LS 2019, 'Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations', *Trends Microbiol*, vol. 27, no. 2, 2019/02/01/, pp. 105-117.

Galea, S & Tracy, M 2007, 'Participation Rates in Epidemiologic Studies', *Annals of Epidemiology*, vol. 17, no. 9, 2007/09/01/, pp. 643-653.

Glassner, KL, Abraham, BP & Quigley, EMM 2020, 'The microbiome and inflammatory bowel disease', *Journal of Allergy and Clinical Immunology*, vol. 145, no. 1, pp. 16-27.

Guo, H, Rivailler, P, Wang, J, Wang, H, Xu, W, Xu, S, Xu, H & Hu, R 2021, 'Metagenomic analysis of a throat swab sample collected in China on a patient infected with Varicella Zoster Virus', *Sci Rep*, vol. 11, no. 1, 2021/07/06, p. 13874.

Haifer, C, Paramsothy, S, Kaakoush, NO, Saikal, A, Ghaly, S, Yang, T, Luu, LDW, Borody, TJ & Leong, RW 2022, 'Lyophilised oral faecal microbiota transplantation for ulcerative colitis (LOTUS): a randomised, double-blind, placebo-controlled trial', *The Lancet Gastroenterology & Hepatology*, vol. 7, no. 2, pp. 141-151.

Hale, VL, Tan, CL, Knight, R & Amato, KR 2015, 'Effect of preservation method on spider monkey (Ateles geoffroyi) fecal microbiota over 8 weeks', *J Microbiol Methods*, vol. 113, Jun, pp. 16-26.

Ho, SM, Lewis, JD, Mayer, EA, Plevy, SE, Chuang, E, Rappaport, SM, Croitoru, K, Korzenik, JR, Krischer, J, Hyams, JS, Judson, R, Kellis, M, Jerrett, M, Miller, GW, Grant, ML, Shtraizent, N, Honig, G, Hurtado-Lorenzo, A & Wu, GD 2019, 'Challenges in IBD Research: Environmental Triggers', *Inflamm Bowel Dis*, vol. 25, no. Suppl 2, May 16, pp. S13-s23.

Janssen, S, McDonald, D, Gonzalez, A, Navas-Molina, JA, Jiang, L, Xu, ZZ, Winker, K, Kado, DM, Orwoll, E, Manary, M, Mirarab, S, Knight, R & Chia, N 2018, 'Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information', *mSystems*, vol. 3, no. 3, pp. e00021-00018.

Johansson, MEV, Gustafsson, JK, Holmén-Larsson, J, Jabbar, KS, Xia, L, Xu, H, Ghishan, FK, Carvalho, FA, Gewirtz, AT, Sjövall, H & Hansson, GC 2014, 'Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis', *Gut*, vol. 63, no. 2, pp. 281-291.

Jones, RB, Zhu, X, Moan, E, Murff, HJ, Ness, RM, Seidner, DL, Sun, S, Yu, C, Dai, Q, Fodor, AA, Azcarate-Peril, MA & Shrubsole, MJ 2018, 'Inter-niche and inter-individual variation in gut microbial community assessment using stool, rectal swab, and mucosal samples', *Sci Rep*, vol. 8, no. 1, pp. 4139-4139.

Karpinets, TV, Wu, X, Solley, T, El Alam, MB, Sims, TT, Yoshida-Court, K, Lynn, E, Ahmed-Kaddar, M, Biegert, G, Yue, J, Song, X, Sun, H, Petrosino, JF, Mezzari, MP, Okhuysen, P, Eifel, PJ, Jhingran, A, Lin, LL, Schmeler, KM, Ramondetta, L, Ajami, N, Jenq, RR, Futreal, A, Zhang, J, Klopp, AH & Colbert, LE 2022, 'Metagenomes of rectal swabs in larger, advanced stage cervical cancers have enhanced mucus degrading functionalities and distinct taxonomic structure', *BMC Cancer*, vol. 22, no. 1, 2022/09/01, p. 945.

Karstens, L, Asquith, M, Davin, S, Fair, D, Gregory, WT, Wolfe, AJ, Braun, J, McWeeney, S & Gilbert, JA 2019, 'Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments', *mSystems*, vol. 4, no. 4, pp. e00290-00219.

Kaul, A, Mandal, S, Davidov, O & Peddada, SD 2017, 'Analysis of Microbiome Data in the Presence of Excess Zeros', *Front Microbiol*, vol. 8, pp. 2114-2114.

Kennedy, K, Hall, MW, Lynch, MDJ, Moreno-Hagelsieb, G & Neufeld, JD 2014, 'Evaluating bias of illumina-based bacterial 16S rRNA gene profiles', *Applied and Environmental Microbiology*, vol. 80, no. 18, pp. 5717-5722.

Khor, B, Gardet, A & Xavier, RJ 2011, 'Genetics and pathogenesis of inflammatory bowel disease', *Nature*, vol. 474, no. 7351, Jun 15, pp. 307-317.

Knox, NC, Forbes, JD, Peterson, C-L, Van Domselaar, G & Bernstein, CN 2019, 'The Gut Microbiome in Inflammatory Bowel Disease: Lessons Learned From Other Immune-Mediated Inflammatory Diseases', *American Journal of Gastroenterology*, vol. 114, no. 7, Jul, pp. 1051-1070.

Kozich, JJ, Westcott, SL, Baxter, NT, Highlander, SK & Schloss, PD 2013, 'Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform', *Appl Environ Microbiol*, vol. 79, no. 17, Sep, pp. 5112-5120.

Kruskal, WH & Wallis, WA 1952, 'Use of Ranks in One-Criterion Variance Analysis', *Journal of the American Statistical Association*, vol. 47, no. 260, 1952/12/01, pp. 583-621.

Li, Q, Ding, X, Liu, K, Marcella, C, Liu, X, Zhang, T, Liu, Y, Li, P, Xiang, L, Cui, B, Wang, J, Bai, J & Zhang, F 2020, 'Fecal Microbiota Transplantation for Ulcerative Colitis: The Optimum Timing and Gut Microbiota as Predictors for Long-Term Clinical Outcomes', *Clin Transl Gastroenterol*, vol. 11, no. 8, Aug, p. e00224.

Loddo, I & Romano, C 2015, 'Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis', *Frontiers in immunology*, vol. 6, pp. 551-551.

Mantel, N 1967, 'The detection of disease clustering and a generalized regression approach', *Cancer Res*, vol. 27, no. 2, Feb, pp. 209-220.

Marotz, C, Cavagnero Kellen, J, Song Se, J, McDonald, D, Wandro, S, Humphrey, G, Bryant, M, Ackermann, G, Diaz, E, Knight, R & Beiko Robert, G 2021, 'Evaluation of the Effect of Storage Methods on Fecal, Saliva, and Skin Microbiome Composition', *mSystems*, vol. 6, no. 2, pp. e01329-01320.

McDonald, D, Hyde, E, Debelius, JW, Morton, JT, Gonzalez, A, Ackermann, G, Aksenov, AA, Behsaz, B, Brennan, C, Chen, Y, Goldasich, LD, Dorrestein, PC, Dunn, RR, Fahimipour, AK, Gaffney, J, Gilbert, JA, Gogul, G, Green, JL, Hugenholtz, P, Humphrey, G, Huttenhower, C, Jackson, MA, Janssen, S, Jeste, DV, Jiang, L, Kelley, ST, Knights, D, Kosciolek, T, Ladau, J, Leach, J, Marotz, C, Meleshko, D, Melnik, AV, Metcalf, JL, Mohimani, H, Montassier, E, Navas-Molina, J, Nguyen, TT, Peddada, S, Pevzner, P, Pollard, KS, Rahnavard, G, Robbins-Pianka, A, Sangwan, N, Shorenstein, J, Smarr, L, Song, SJ, Spector, T, Swafford, AD, Thackray, VG, Thompson, LR, Tripathi, A, Vázquez-Baeza, Y, Vrbanac, A, Wischmeyer, P, Wolfe, E, Zhu, Q, Knight, R, Greene, CS, Mann, AE, Amir, A, Frazier, A, Martino, C, Lebrilla, C, Lozupone, C, Lewis, CM, Raison, C, Zhang, C, Lauber, CL, Warinner, C, Lowry, CA, Callewaert, C, Bloss, C, Willner, D, Galzerani, DD, Gonzalez, DJ, Mills, DA, Chopra, D, Gevers, D, Berg-Lyons, D, Sears, DD, Wendel, D, Lovelace, E, Pierce, E, TerAvest, E, Bolyen, E, Bushman, FD, Wu, GD, Church, GM, Saxe, G, Holscher, HD, Ugrina, I, German, JB, Caporaso, JG, Wozniak, JM, Kerr, J, Ravel, J, Lewis, JD, Suchodolski, JS, Jansson, JK, Hampton-Marcell, JT, Bobe, J, Raes, J, Chase, JH, Eisen, JA, Monk, J, Clemente, JC, Petrosino, J, Goodrich, J, Gauglitz, J, Jacobs, J, Zengler, K, Swanson, KS, Lewis, K, Mayer, K, Bittinger, K, Dillon, L, Zaramela, LS, Schriml, LM, Dominguez-Bello, MG, Jankowska, MM, Blaser, M, Pirrung, M, Minson, M, Kurisu, M, Ajami, N, Gottel, NR, Chia, N, Fierer, N, White, O, Cani, PD, Gajer, P, Strandwitz, P, Kashyap, P, Dutton, R, Park, RS, Xavier, RJ, Mills, RH, Krajmalnik-Brown, R, Ley, R, Owens, SM, Klemmer, S, Matamoros, S, Mirarab, S, Moorman, S, Holmes, S, Schwartz, T, Eshoo-Anton, TW, Vigers, T, Pandey, V, Treuren, WV, Fang, X, Xu, ZZ, Jarmusch, A, Geier, J, Reeve, N, Silva, R, Kopylova, E, Nguyen, D, Sanders, K, Benitez, RAS, Heale, AC, Abramson, M, Waldispühl, J, Butyaev, A, Drogaris, C, Nazarova, E, Ball, M & Gunderson, B 2018, 'American Gut: an Open Platform for Citizen Science Microbiome Research', *mSystems*, vol. 3, no. 3, pp. e00031-00018.

McMurdie, PJ & Holmes, S 2013, 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data', *PLoS One*, vol. 8, no. 4, p. e61217.

Moayyedi, P, Yuan, Y, Baharith, H & Ford, AC 2017, 'Faecal microbiota transplantation for *Clostridium difficile*-associated diarrhoea: a systematic review of randomised controlled trials', *Med J Aust*, vol. 207, no. 4, Aug 21, pp. 166-172.

Nearing, JT, Comeau, AM & Langille, MGI 2021, 'Identifying biases and their potential solutions in human microbiome studies', *Microbiome*, vol. 9, no. 1, 2021/05/18, p. 113.

Okahara, K, Ishikawa, D, Nomura, K, Ito, S, Haga, K, Takahashi, M, Shibuya, T, Osada, T & Nagahara, A 2020, 'Matching between Donors and Ulcerative Colitis Patients Is Important for Long-Term Maintenance after Fecal Microbiota Transplantation', *Journal of clinical medicine*, vol. 9, no. 6, p. 1650.

Oksanen, J, Blanchet, FG, Friendly, M, Kindt, R, Legendre, P, McGlinn, D, Minchin, PR, O'Hara, RB, Simpson, GL, Solymos, P, Stevens, MHH, Szoecs, E & Wagner, H 2020, 'vegan: Community Ecology Package', 2020.

Olesen, SW & Gerardin, Y 2020, 'Re-Evaluating the Evidence for Faecal Microbiota Transplantation 'Super-Donors' in Inflammatory Bowel Disease', *Journal of Crohn's and Colitis*, vol. 15, no. 3, pp. 453-461.

Oliveira, A, Oliveira, LC, Aburjaile, F, Benevides, L, Tiwari, S, Jamal, SB, Silva, A, Figueiredo, HCP, Ghosh, P, Portela, RW, De Carvalho Azevedo, VA & Wattam, AR 2017, 'Insight of Genus Corynebacterium: Ascertaining the Role of Pathogenic and Non-pathogenic Species', *Front Microbiol*, vol. 8, pp. 1937-1937.

Paradis, E & Schliep, K 2018, 'ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R', *Bioinformatics*, vol. 35, no. 3, pp. 526-528.

Paramsothy, S, Nielsen, S, Kamm, MA, Deshpande, NP, Faith, JJ, Clemente, JC, Paramsothy, R, Walsh, AJ, van den Bogaerde, J, Samuel, D, Leong, RWL, Connor, S, Ng, W, Lin, E, Borody, TJ, Wilkins, MR, Colombel, J-F, Mitchell, HM & Kaakoush, NO 2019, 'Specific Bacteria and Metabolites Associated With Response to Fecal Microbiota Transplantation in Patients With Ulcerative Colitis', *Gastroenterology*, vol. 156, no. 5, pp. 1440-1454.e1442.

Paramsothy, S, Paramsothy, R, Rubin, DT, Kamm, MA, Kaakoush, NO, Mitchell, HM & Castaño-Rodríguez, N 2017, 'Faecal Microbiota Transplantation for Inflammatory Bowel Disease: A Systematic Review and Meta-analysis', *Journal of Crohn's and Colitis*, vol. 11, no. 10, pp. 1180-1199.

Peres-Neto, PR & Jackson, DA 2001, 'How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test', *Oecologia*, vol. 129, no. 2, 2001/10/01, pp. 169-178.

Quast, C, Pruesse, E, Yilmaz, P, Gerken, J, Schweer, T, Yarza, P, Peplies, J & Glöckner, FO 2012, 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic Acids Research*, vol. 41, no. D1, pp. D590-D596.

Quince, C, Walker, AW, Simpson, JT, Loman, NJ & Segata, N 2017, 'Shotgun metagenomics, from sampling to analysis', *Nature Biotechnology*, vol. 35, no. 9, 2017/09/01, pp. 833-844.

Quraishi, MN, Iqbal, TH & Hart, AL 2020, 'Precision Medicine with FMT for Ulcerative Colitis: Are We There Yet?', *Journal of Crohn's and Colitis*, vol. 15, no. 4, pp. 519-520.

Quraishi, MN, Widlak, M, Bhala, N, Moore, D, Price, M, Sharma, N & Iqbal, TH 2017, 'Systematic review with meta-analysis: the efficacy of faecal microbiota transplantation for the treatment of recurrent and refractory Clostridium difficile infection', *Aliment Pharmacol Ther*, vol. 46, no. 5, Sep, pp. 479-493.

R Core Team, 2020, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Reyman, M, van Houten, MA, Arp, K, Sanders, EAM & Bogaert, D 2019, 'Rectal swabs are a reliable proxy for faecal samples in infant gut microbiota research based on 16S-rRNA sequencing', *Sci Rep*, vol. 9, no. 1, 2019/11/05, p. 16072.

Ringel, Y, Maharshak, N, Ringel-Kulka, T, Wolber, EA, Sartor, RB & Carroll, IM 2015, 'High throughput sequencing reveals distinct microbial populations within the mucosal and luminal niches in healthy individuals', *Gut Microbes*, vol. 6, no. 3, pp. 173-181.

Rossen, NG, Fuentes, S, van der Spek, MJ, Tijssen, JG, Hartman, JHA, Duflou, A, Löwenberg, M, van den Brink, GR, Mathus-Vliegen, EMH, de Vos, WM, Zoetendal, EG, D'Haens, GR & Ponsioen, CY 2015, 'Findings From a Randomized Controlled Trial of Fecal Transplantation for Patients With Ulcerative Colitis', *Gastroenterology*, vol. 149, no. 1, 2015/07/01/, pp. 110-118.e114.

Salter, SJ, Cox, MJ, Turek, EM, Calus, ST, Cookson, WO, Moffatt, MF, Turner, P, Parkhill, J, Loman, NJ & Walker, AW 2014, 'Reagent and laboratory contamination can critically impact sequence-based microbiome analyses', *BMC Biology*, vol. 12, no. 1, 2014/11/12, p. 87.

Short, MI, Hudson, R, Besasie, BD, Reveles, KR, Shah, DP, Nicholson, S, Johnson-Pais, TL, Weldon, K, Lai, Z, Leach, RJ, Fongang, B & Liss, MA 2021, 'Comparison of rectal swab, glove tip, and participant-collected stool techniques for gut microbiome sampling', *BMC Microbiology*, vol. 21, no. 1, pp. 26-26.

Silverberg, MS, Satsangi, J, Ahmad, T, Arnott, ID, Bernstein, CN, Brant, SR, Caprilli, R, Colombel, JF, Gasche, C, Geboes, K, Jewell, DP, Karban, A, Loftus, EV, Jr., Peña, AS, Riddell, RH, Sachar, DB, Schreiber, S, Steinhart, AH, Targan, SR, Vermeire, S & Warren, BF 2005, 'Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology', *Can J Gastroenterol*, vol. 19 Suppl A, Sep, pp. 5a-36a.

Sinha, R, Chen, J, Amir, A, Vogtmann, E, Shi, J, Inman, KS, Flores, R, Sampson, J, Knight, R & Chia, N 2016, 'Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies', *Cancer Epidemiology, Biomarkers & Prevention*, vol. 25, no. 2, pp. 407-416.

Song, SJ, Amir, A, Metcalf, JL, Amato, KR, Xu, ZZ, Humphrey, G & Knight, R 2016, 'Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies', *mSystems*, vol. 1, no. 3, pp. e00021-00016.

Stojek, M, Jabłońska, A & Adrych, K 2021, 'The Role of Fecal Microbiota Transplantation in the Treatment of Inflammatory Bowel Disease', *J Clin Med*, vol. 10, no. 18, Sep 8.

Sun, S, Zhu, X, Huang, X, Murff, HJ, Ness, RM, Seidner, DL, Sorgen, AA, Blakley, IC, Yu, C, Dai, Q, Azcarate-Peril, MA, Shrubsole, MJ & Fodor, AA 2021, 'On the robustness of inference of association with the gut microbiota in stool, rectal swab and mucosal tissue samples', *Sci Rep*, vol. 11, no. 1, 2021/07/21, p. 14828.

Vandeputte, D, Tito, RY, Vanleeuwen, R, Falony, G & Raes, J 2017, 'Practical considerations for large-scale gut microbiome studies', *FEMS Microbiol Rev*, vol. 41, no. Supp_1, Aug 1, pp. S154-s167.

Weyrich, LS, Farrer, AG, Eisenhofer, R, Arriola, LA, Young, J, Selway, CA, Handsley-Davis, M, Adler, CJ, Breen, J & Cooper, A 2019, 'Laboratory contamination over time during low-biomass sample analysis', *Molecular Ecology Resources*, vol. 19, no. 4, pp. 982-996.

Wickham, H, 2016, *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York.

Wickham, H, Averick, M, Bryan, J, Chang, W, McGowan, LDa, François, R, Grolemund, G, Hayes, A, Henry, L, Hester, J, Kuhn, M, Pedersen, TL, Miller, E, Bache, SM, Müller, K, Ooms, J, Robinson, D, Seidel, DP, Spinu, V, Takahashi, K, Vaughan, D, Wilke, C, Woo, K & Yutani, H 2019, 'Welcome to the Tidyverse', *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686.

Wilke, CO, 2020, *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*, ver. 1.1.1, R package, CRAN, https://cran.r-project.org/web/packages/cowplot/index.html.

Yalchin, M, Segal, JP, Mullish, BH, Quraishi, MN, Iqbal, TH, Marchesi, JR & Hart, AL 2019, 'Gaps in knowledge and future directions for the use of faecal microbiota transplant in the treatment of inflammatory bowel disease', *Therapeutic Advances in Gastroenterology*, vol. 12, pp. 1756284819891038-1756284819891038.

Chapter 4

Statement of Authorship

Title of Paper	Recommendations and considerations for longitudinal studies of the human microbiome
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Unpublished manuscript formatted for submission to Microbiome.

Principal Author

Name of Principal Author (Candidate)	Gina L. Guzzo		
Contribution to the Paper	Conceptualised the review, conducted the research, wrote the manuscript, and generated the figure and table.		
Overall percentage (%)	75		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	01/06/2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Conceptualised the review, provided substantial discussion and feedback to the research, and edited all versions of the manuscript.		
Signature		Date	24/06/2022

Name of Co-Author	. Bastien Llamas		
Contribution to the Paper	Conceptualised the review, provided substantial discussion and feedback to the research, and reviewed the manuscript.		
Signature		Date	08/06/2022

Name of Co-Author	Murthy N. Mittinty		
Contribution to the Paper	Conceptualised the review, provided substantial discussion and feedback to the research, and edited all versions of the manuscript.		
Signature		Date	28/06/2022

Name of Co-Author	Jane M. Andrews		
Contribution to the Paper	Provided discussion and feedback to the research and reviewed the manuscript.		
Signature		Date	17/8/22

Title

Recommendations and considerations for longitudinal studies of the human microbiome

Authors

Gina L. Guzzo[1], Murthy N. Mittinty[2], Bastien Llamas[1,3,4], Jane M. Andrews[5], and Laura S. Weyrich[1,6]

Affiliations

human microbiome, microbiota, longitudinal, temporal, regression

[1]School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

[2]School of Public Health, University of Adelaide, South Australia, Australia

[3]Indigenous Genomics Research Group, Telethon Kids Institute, Adelaide, South Australia, Australia

[4]National Centre for Indigenous Genomics, Australian National University, Canberra, Australian Capital Territory, Australia

[5]Inflammatory Bowel Disease Service, Department of Gastroenterology and Hepatology, Royal Adelaide Hospital & School of Medicine, Faculty of Health Sciences, University of Adelaide, South Australia, Australia

[6]Department of Anthropology and Huck Institutes of the Life Sciences, Pennsylvania State University, PA, USA

Abstract

The human microbiome is a dynamic ecosystem that plays an integral role in our systemic health. Yet, to date, much of the understanding of this interaction is based on static, cross-sectional, or single-timepoint observations. Accounting for the microbiome's temporal variability offers promising opportunities to characterise its role in diseases and target it for therapies. Therefore, conducting longitudinal microbiome studies is becoming a major focus for clinical research. However, there are several problems unique to longitudinal research that must be considered to effectively design, conduct, and analyse such studies in clinical settings. Accounting for time-dependent sources of variation that can bias inferred associations and the effects of sample missingness are challenging to resolve with existing methods. Here, we discuss the criteria essential for conducting longitudinal microbiome studies in clinical and public health research starting from participant recruitment and sample collection through to statistical analysis. Our discussion also highlights areas where significant advancements are needed to gain higher resolution insights into the microbiome's association with human health.

Keywords

human microbiome, microbiota, longitudinal, temporal, regression

Appreciating temporal dynamics of the human microbiome

Thanks to advances in high-throughput sequencing methods, our understanding of the human body now extends to the trillions of microbes, including bacteria, archaea, fungi, protozoa, viruses, and their genomic and metabolic products, that live on our skin and mucosal surfaces (e.g., gut, mouth, and lungs) (Sender, Fuchs & Milo 2016; Allaband et al. 2019). This ecosystem of microbes, collectively referred to as the 'human microbiome', performs several critical functions and has been linked to numerous health conditions (Berg et al. 2020). Like any ecosystem, the human microbiome is not static, but rather, a dynamic assemblage of microbial communities that are influenced by temporal variation (Stegen, Bottos & Jansson 2018; Uhr, Dohnalová & Thaiss 2019; Berg et al. 2020). Some populations remain relatively stable despite changing conditions, i.e., 'core microbiota', whereas others are transient and dependent on internal, external, and random/stochastic factors (Stegen, Bottos & Jansson 2018). Cross-sectional studies of the microbiome can allow us to draw links between

microbes and conditions, but these studies have a limited ability in determining causality due to the temporal dynamics of microbial communities.

The temporal variability and complexity of the microbiome also makes it difficult to distinguish the effect of noise and confounders from any potential associations (Bharti & Grimm 2019; Wilkinson et al. 2021). These dynamics differ between individuals, meaning that the microbiome has a specific signature in each person that is distinguishable from others, even over time (Costello et al. 2009; Flores et al. 2014). Accounting for this variability with multiple sampling points in each individual can unmask trends associated with the microbiome at a level higher than the individual (e.g., Clooney et al. 2019).

Studies with repeated measures in the same participants, or longitudinal studies, are therefore critically important to advance our understanding of how the microbiome influences human health (Box 1). However, this research is more resource-intensive than studies with cross-sectional designs, and it is crucial for researchers to design and conduct longitudinal studies that can answer their questions in the most efficient and cost-effective manner. To do this, researchers need a solid grasp of the technical and statistical considerations that can confound or limit their analyses. In this review, we provide a guide for researchers intending to embark on a longitudinal high-throughput DNA sequencing (16S or shotgun) study of the microbiome from study design through to analysis by discussing the relevant issues and gaps in knowledge, and recommendations to overcome these.

Box 1. Longitudinal microbiome studies offer many benefits for clinical and public health research.

1. The factors that cause community changes in healthy individuals can be established, which can then be accounted for when studying individuals with diseases or disorders (Cao et al. 2017; Uhr, Dohnalová & Thaiss 2019).
2. Sampling individuals with relapsing-remitting conditions during periods of remission and again during flares can reveal a microbiome signature that precedes and predicts inflammation (Vázquez-Baeza et al. 2018).
3. The microbiome's response to treatments, interventions, and lifestyle changes can be measured, including strain tracking in microbiota-based therapies (Kong et al. 2020).
4. Long-term studies can reveal the microbiome's involvement in aging (Dominguez-Bello et al. 2019) and the development of diseases (e.g., the GEM Project), helping us answer the ultimate cause-or-effect question regarding the microbiome and human health.
5. Characterising the ecological dynamics that make up an individual's microbiome over time can enable its prediction and control for beneficial purposes such as therapies, preventative interventions, and lifestyle changes that favour a healthy state (Stegen, Bottos & Jansson 2018).

Considerations when designing longitudinal studies of the microbiome

The first generation of microbiome studies were mainly exploratory, aimed at describing the immensity of unknown data that exists within human-associated microbes (e.g., Eckburg et al. 2005). From these studies, we now appreciate the importance of microbes in human health and how the microbiome presents the potential to be utilised in multiple facets of medicine: prognostics (Shang & Liu 2018), diagnostics (Raes 2016), and even therapeutics (Jeon et al. 2018). As such, study designs are shifting towards more hypothesis-driven research (Tripathi et al. 2018), including both observational and experimental/interventional designs (Thiese 2014). Across these study designs, factors related to participants and sampling should be considered during the planning phase to minimise attrition and bias (Figure 1).

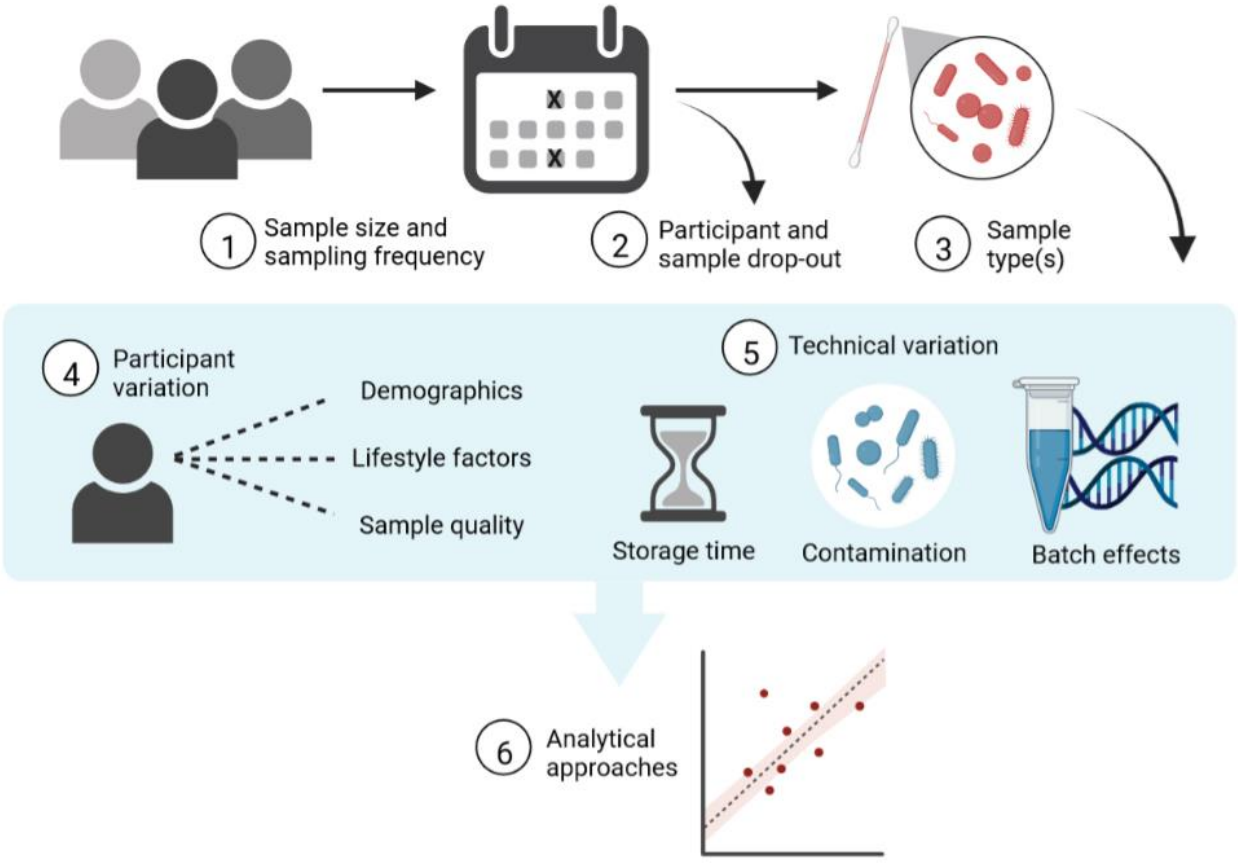

Figure 1. Diagram of the main factors that influence longitudinal microbiome study design and analysis. Figure created with BioRender.com.

i. Sample size and sampling frequency

Designing a longitudinal study requires researchers to consider (1) the sample size, including the cohort size and number of time points to be sampled, and (2) the amount of time between each sample. The minimum number of samples needed is normally determined by available funding and statistical power calculations. However, power calculations for microbiome studies are not straightforward and largely understudied (Debelius et al. 2016b; Silverman et al. 2018a). Longitudinal sampling adds a layer of complexity to an already-challenging exercise due to the increased parameter estimation for repeated measurements and correlations between non-independent measurements (Guo & Pandis 2015). The standards for calculating statistical power can be used (Debelius et al. 2016b), but it is recommended to supplement these with simulation studies (e.g., Fukuyama et al. 2017) and/or examining existing empirical data (e.g., Vázquez-Baeza et al. 2018 and Casals-Pascual et al. 2020).

The sampling frequency will largely depend on the study design and questions. In experimental studies where the exposure is predicted to initially cause high variability in the microbiome (e.g., medication, faecal microbiota transplantation, and surgery), it is recommended to concentrate sampling around the time points when variation is predicted to be higher, and trying to capture equal variation between subsequent time points to conserve resources (Silverman et al. 2018a). In observational studies where the microbial dynamics are predicted to be relatively homogeneous within each participant, longer time between each sample collection is appropriate, thereby allowing resources to sample a larger cohort size. Prediction models can also assist in determining the number of time points needed if the volatility of the microbiome is unknown in a particular disease. For example, Vázquez-Baeza et al. (2018) used a Random Forest model and found that sampling fewer individuals with Crohn's disease at more time points was more robust than sampling more individuals at fewer time points. While their model was applied to individuals with Crohn's disease, this approach could be extended to other conditions and outcomes. However, this strategy requires knowledge of the microbiome dynamics which may be difficult to hypothesise if no prior similar study has been conducted. Thus, further research into statistically predicting the appropriate sample size and frequency in longitudinal microbiome studies is needed (Box 2).

Box 2. Opportunities for future research and development to improve longitudinal microbiome research.

Study design

- Expanding the accessibility of longitudinal microbiome studies to underrepresented demographics.
- Benchmarking new, flexible sampling types and improving existing ones.

Technical effects

- Determining the effect of storage time on different sample types.
- Determining the effect of contamination on temporal sampling.

Statistical methods

- Developing power calculations for repeated measures microbiome sampling.
- Developing imputation methods for missing microbiome samples.
- Consolidating the theoretical assumptions of missing data for microbiome research.
- Computationally efficient models that include random slopes and intercepts, time-varying covariates, and can handle missing data.

ii. Participant recruitment, retention, and compliance

Recruiting and retaining a study group over any period of time can be challenging. Individuals are less likely to participate in studies that have longer time commitments (Galea & Tracy 2007), and the length of study period has been identified as a key factor affecting retention rates (Chaudhari et al. 2020). Several strategies have been developed to target common barriers preventing participation and retention in clinical and epidemiological studies that are discussed elsewhere (Galea & Tracy 2007; Gul & Ali 2010; Abshire et al. 2017; Chaudhari et al. 2020).

When recruiting participants for longitudinal microbiome studies, researchers should be aware of selection biases that may occur (Debelius et al. 2016a). Factors such as geographical location, gender, income, and cultural beliefs can contribute to an individual's ability and

willingness to enter and remain in a particular study. These factors have contributed to the unequal representation of demographics in microbiome research (Nath et al. 2021; Abdill, Adamowicz & Blekhman 2022). For example, Carson, Little and Townsend (2019) emphasised that black American women are underrepresented in microbiome studies, even though they are disproportionately affected by chronic diseases. By using strategies such as open communication about the study purpose, sampling process, and data dissemination, as well as displaying diversity amongst their research team, they were able to recruit and retain an adequate proportion of black women in a longitudinal microbiome study with blood, stool, and saliva samples. (Carson, Little & Townsend 2019; Abdill, Adamowicz & Blekhman 2022).

Participant retention and compliance are particularly important in microbiome studies because it is very difficult to statistically compensate for missing or low-quality samples (discussed below). In the case of self-collected samples, sample quality can be improved by providing simple and clear instructions with visuals such as images and/or videos of how to collect samples and what constitutes a good sample versus a bad one (Debelius et al. 2016a; Carson, Little & Townsend 2019). Sampling invasiveness can also influence how likely a participant is willing to continue collection (i.e., the easier the sample is to collect, the more likely it is that participants will follow through) (Chaudhari et al. 2020). Thus, the sampling method should be less invasive where possible. As with recruitment strategies, regular contact with participants can also help overcome issues such as hesitation over sampling invasiveness that may normally lead to attrition (Carson, Little & Townsend 2019).

iii. Sample type(s)

A variety of sample types are now available to sample the microbiome of various body sites. These include less invasive and more flexible alternatives that can promote participant recruitment by supporting self-collection outside of the clinic (Allaband et al. 2019). Examples of less invasive alternatives include sampling the lung microbiome with breath condensate instead of lung brushings (Glendinning et al. 2017), sampling the oral microbiome with saliva instead of dental calculus (Zaura et al. 2021), and sampling the gut microbiome with rectal swabs (Budding et al. 2014; Reyman et al. 2019), glove tip swabs (Short et al. 2021), or whole stool instead of mucosal biopsies. Dry swabs of used toilet paper are an example of gut microbiome sampling that can be self-collected at home (Sinha et al.

2016; McDonald et al. 2018; Bokulich et al. 2019). Although, it should be noted that sample proxies do not yield the same microbial composition as the reference sample, and the study questions should be balanced against the potential invasiveness of sampling (Glendinning et al. 2017; Sun et al. 2021). If an alternative sample type is used, researchers need to ensure that adequate comparative analyses have been previously documented, or conduct these themselves with detailed diversity, composition, and differential abundance analyses to determine how representative the alternative sample is compared to the reference sample. Another drawback of flexible sampling types is that they are generally lower in biomass, which may hinder the ability to divide them for storage and analysis (e.g., biobanking, culturing, and various '-omics') (Wilkinson et al. 2021).

Data collection to account for sources of microbiome variation

One of the primary goals of clinical and public health research is to determine if a treatment or exposure (e.g., medication, therapy, diet, lifestyle factor) affects an outcome of interest. In microbiome research, the outcome can be a change to the microbiome's overall composition, diversity, or function, or to individual taxa. Occasionally, the microbiome may be used as an exposure or covariate as well such as to study how the microbiome affects the bioactivity of drugs (Haiser et al. 2013; Maier et al. 2018; Wilkinson et al. 2021). The relationship between the microbiome and an exposure/treatment can only be established when confounding sources of noise are accounted for in the study. Noise can originate from participant characteristics, technical effects, stochasticity, and other unknown factors. Participant and technical factors (i.e., covariates) in microbiome studies can vary in their effect sizes (Debelius et al. 2016b) and can vary over time in longitudinal studies (i.e., time-varying covariates) (Mirzayi et al. 2021), adding an additional layer of consideration for metadata collection.

i. Participant sources of microbiome variation

Determining which participant covariates to record in a microbiome study should be guided by expert knowledge of the area (e.g., standard demographic data, clinical biomarkers, and clinical phenotype classifications), as well as a literature search of similar studies and potential confounders (e.g., Vujkovic-Cvijin et al. 2020). In longitudinal microbiome studies, determining which time-varying covariates to consider is dependent on the duration of study

and research questions. For example, studying the gut microbiome over a short timespan to determine the effect of a probiotic does not require adjustments to age, but it would likely include records of weight or diet. There are also sampling factors to consider that can affect the quality and variability of the microbial DNA captured. For example, stool quality and moisture content can determine which microbes are likely to be present in a sample and is a significant contributor to inter-individual variation (Vandeputte et al. 2016; Vandeputte et al. 2017).

ii. Technical sources of microbiome variation

Technical aspects of a microbiome study can cause variation that obscures the underlying biological signal (Allaband et al. 2019; Wang & LêCao 2020). Discussions of general technical biases are available and can provide valuable information to consider when planning longitudinal studies (Pollock et al. 2018; Allaband et al. 2019; Bharti & Grimm 2019; Nearing, Comeau & Langille 2021). There are several time-dependent factors that can affect the microbial makeup of a sample: the time spent in storage, contamination, and stochasticity of laboratory processes (Figure 1) (Shaw et al. 2016; Weyrich et al. 2019; Wang & LêCao 2020). Typically, once a sample is collected, it is immediately preserved with a chemical fixative or, ideally, frozen and stored (Nearing, Comeau & Langille 2021). Short-term storage in the manner of days or weeks does not appear to affect sample composition, but longer storage over months and years may have an effect (Lauber et al. 2010; Shaw et al. 2016). Out of caution, researchers should take note of sample storage dates and test for their association with the data.

The gold standard for microbiome research is to minimise contamination throughout the sample processing steps by using sterile equipment and practises (Eisenhofer et al. 2019; Hornung, Zwittink & Kuijper 2019). Microbial contamination (i.e., foreign microbial material that is introduced to a sample) can originate from many sources including the sample collection kit, reagents, environment of sampling, laboratory, and technicians (Hornung, Zwittink & Kuijper 2019; Jurasz, Pawłowski & Perlejewski 2021). In longitudinal studies, researchers risk attributing shifts in the microbiome to a change in time or clinical variable when they are instead caused by contamination (Salter et al. 2014; Weiss et al. 2014). Contamination is particularly important for low-biomass specimens, such as skin or placental swabs, tissue biopsies, blood, urine, and breath condensate (Weiss et al. 2014). However, the

effect of contamination on higher biomass samples such as whole stool should also be considered, especially when conducting more sensitive analyses of stool such as donor-to-recipient strain tracking in faecal microbiota transplantation (Kong et al. 2020). Additionally, the magnitude of this effect is not known for lower biomass proxies for stool such as rectal and dry swabs. Controls—both negative and positive—should be taken during sample collection, extraction, and/or sequencing to allow identification of contamination from external sources and other samples (cross-contamination). As contaminants and the sources of contaminants can change over time (e.g., new researchers join the study, the laboratory location changes, or the reagents change), negative controls should be collected repeatedly throughout a study (Weyrich et al. 2019). The study protocol should remain as uniform as possible throughout, but any changes should be noted as covariates for later analyses (discussed below).

Lastly, stochastic factors in the laboratory can influence the microbial output of a sample. Technicians, machinery, sequencing primers, and reagents do not behave identically for every extraction and sequencing batch, and this randomness of output combined with contamination leads to batch variation (Salter et al. 2014; McDonald et al. 2018; Oh & Li 2022). As this stochasticity may behave differently over time, samples should be processed as close together as possible and/or randomised during processing to reduce technical batch effects (Eisenhofer et al. 2019; Holmes 2019). Like other sources of known variation, batch numbers should also be recorded as covariates so the resulting variation can be accounted for in downstream analyses.

Data cleaning

Once sequencing data have been generated, an important first step is to ensure that the data are sufficiently quality tested to minimise the effects of technical variation before downstream analysis. This step can be achieved using tools specific for sequencing quality control (e.g., Trimmomatic and fastp), or built in R packages (e.g., phyloseq and DADA2) and standalone programs (e.g., QIIME2, MEGAN-CE, and MG-RAST) (McMurdie & Holmes 2013; Bolger, Lohse & Usadel 2014; Callahan et al. 2016; Huson et al. 2016; Keegan, Glass & Meyer 2016; Chen, S et al. 2018; Bolyen et al. 2019). Bioinformatic tools can also be optionally used to deal with microbial contamination. . Decontamination can be performed by the open-source program *decontam*, which identifies contaminants in biological samples based on their prevalence or frequency in negative control samples (Davis et al.

2018). This step can be performed on a per-batch basis if the samples were processed at different times over a considerable period. Other bioinformatic tools and methods exist to remove unwanted technical variation and are discussed in detail by Wang and LêCao (2020). Alternatively, technical covariates can be included in the statistical models discussed below, but the drawback to this approach is that it is restricted to known/observed covariates (Wang & LêCao 2020).

Analysing longitudinal microbiome data

Longitudinal models are more complex than standard models as they need to integrate the multi-level structure of the data (i.e., participants, time, and microbial taxa) and the repeated measurements between participants that are correlated. In clinical research, these models need to accept covariates to account for the technical and biological variation discussed previously and, ideally, handle missing data which commonly occurs in human studies. Microbiome data also have specific features that complicate the analysis—the data are sparse and compositional (Gloor et al. 2017; Mallick et al. 2021; Pan 2021), and statistical models for regression and prediction should be carefully selected (Silverman et al. 2018b). Models that accommodate these issues must be optimised to interpret microbiome data (e.g., microbial abundances), or else it will be computationally expensive and cumbersome to navigate results (Zhang & Yi 2020). Given the complexities of microbiome data, a limited number of open-source longitudinal models are currently available (Table 1). While we provide an overview of the notable statistical considerations, research teams should include expert biostatisticians and bioinformaticians to manage these issues and determine how best to design a study that is in agreement with available resources.

i. Zero-inflated microbiome data

Microbiome data are notoriously sparse; many taxa have very low abundances and few taxa have very high abundances (Kaul et al. 2017; Pan 2021). Across a group of participants, there will likely be many taxa that are present in only one or few individuals and absent in the remainder. Taxa can be missing because of technical reasons including insufficient sampling (e.g., low-quality sample or old/expired sample) or sequencing depth (false negatives), or because they are actually absent in an individual at a particular sampling time (true negatives) (Zhang & Yi 2020; Han et al. 2021).

Due to this missingness, abundance counts of taxa are over-populated with zeros (zero-inflated) and form right-skewed data distributions. Two common methods to deal with zero-inflation are to add a pseudo-count of one to replace zeros before an analysis (which is mathematically convenient but shifts the model away from being biologically meaningful/true), or to incorporate a zero-inflated probability distribution into the model (Chen & Li 2016; Zhang, Guo & Yi 2020; Han et al. 2021). Several available longitudinal methods incorporate the zero-inflated probability distribution (Chen & Li 2016; Zhang, Guo & Yi 2020; Han et al. 2021). However, Kaul et al. (2017) argue that these solutions are too simplistic as they do not account for the different types of zero counts that occur (i.e., true and false negatives are incorrectly grouped together). Their method, ANCOM-II, models three types of zero counts instead of using a single probability distribution. Another newer option is *MaAsLin2*, which has a default log-transformed linear model that shows a low false discovery rate while maintaining a higher level of sensitivity over other methods (Mallick et al. 2021). Given the variation in methods, it is sensible to try multiple methods to compare and collate results, and explore notable limitations of each method pertinent to each study.

While zero-inflated modelling helps mitigate the effect of excess zeroes in a dataset, some taxa are present at levels too low to be included in modelling. Extremely rare taxa should be excluded from the analysis as they can skew results when included alongside more prevalent taxa. The exclusion of rare taxa can also reduce technical variation in the dataset while preserving diversity analyses (Cao, Q et al. 2021). Filtering rare taxa can be performed based on their prevalence and/or abundance, either prior to analysis (e.g., QIIME 2's plugin 'qiime feature-table filter-features') or within a multivariate pipeline (e.g., MaAsLin2) (Bolyen et al. 2019; Mallick et al. 2021). It is recommended to use at least 10% prevalence to exclude rare taxa (Cao, Q et al. 2021). However, depending on the dataset and research questions, a more stringent filter can be used. Another filtering option is PERFect, which incorporates a permutation test to determine rare taxa that are likely a result of technical noise (Smirnova, Huzurbazar & Jafari 2018)

ii. Missing samples

Missing microbiome data can arise at multiple levels in a study and cause missing taxa or the loss of entire samples. Data missingness in longitudinal studies can also arise from sample dropout caused by a lack of sampling at a particular time point or due to low-quality sampling that does not yield substantive data. Missing data is a common occurrence in all clinical research, and single or multiple imputation is normally used to replace missing values (Little

& Rubin 2020). Missing microbiome samples, however, are difficult to deal with because each sample is composed of several hundreds or thousands of taxa and imputation would require replacing all of these taxa counts (Gloor et al. 2017). It is also difficult to determine whether the assumptions for imputation are met. Imputation requires that missing data are missing-at-random, such that the observed values can be used to explain the missing values (Sterne et al. 2009). This assumption may not necessarily be true with missing microbial taxa (e.g., can the earlier or later measurements of an individual's microbiome inform on how to impute the abundance and composition of their missing sample?). Further research is needed to understand microbial dynamics before imputation can be accepted as a robust approach. Unless the analysis is restricted to single values such as diversity metrics, we do not recommend imputing missing microbiome samples given the aforementioned biases, and because of the current lack of appropriate microbiome-specific analytical tools to do so. Thus, where sample dropout is an issue, the statistical model should handle missing values (Table 1). Although, if there are many missing samples and the sample size is small, most models will likely fail to converge. Therefore, it is important to minimise sample dropout through the above strategies during the experimental/observational design and data collection phases of the study.

iii. Model parameters

Statistical models more closely reflect the real world when they incorporate covariates, including those that change over time, and the model parameters reflect the dynamics of the data. Longitudinal models that implement covariates are most often fixed, random, or mixed (both fixed and random) effects models (Bell, Fairbrother & Jones 2019) (Table 1). Time-varying covariates are modelled with random effects models and are therefore included in mixed effects models as well. These are the most common models available for longitudinal microbiome regression. When choosing other parameters such as the slopes and intercepts, these should reflect the known and/or hypothesised dynamics of the microbial communities being sampled. For example, if it is known that all participants start at the same level of diversity or abundance, then a fixed intercept can be chosen. Similarly for the slope, if the participants will have the same direction and rate of trajectory of their microbiome's properties (e.g., abundance or diversity increase or decrease), then a fixed slope can be chosen. However, we argue that these scenarios are uncommon as most people have microbiomes that are unique to themselves, even over time, and may respond differently to

exposure/treatments (Gilbert et al. 2018). Therefore, the model that is most likely to reflect empirical data is the random slopes and random intercepts model.

A study can also have exposures that impact time-varying confounders which can in turn impact the exposures measured at later times (e.g., altering the drug dosage throughout a study). Such mechanisms cause more complex mediation effects which are not properly handled by standard random effect models. These require special models such as G-estimation (Robins, Hernán & Brumback 2000) which are currently not used in microbiome analysis.

iv. Interpretation of results

Lastly, a statistical model is useful insofar as its results are interpretable to the questions at hand. Researchers need to determine which microbial metrics they are interested in testing to inform which model(s) to choose. Most available models use differential abundance as the unit of measurement (Table 1), which can be tested at any taxonomic level the dataset has been classified to (e.g., phylum through to strain level). Though, caution should be taken when testing large, complex datasets at the lowest taxonomic levels (e.g., at the level of amplicon sequence variant (ASV) or operational taxonomic unit (OTU)) as mixed effects regression on such large matrices generally requires a large amount of computational memory unless it has been highly optimised for this process (Zhang & Yi 2020). Only one program, q2-longitudinal, allows the user to regress on diversity metrics with covariates over time (Bokulich et al. 2018), but it does not include an option for random slopes. New and improved programs are greatly needed to model realistic dynamics of the human microbiome as longitudinal research continues to increase.

Concluding remarks

Longitudinal research has the potential to uncover key aspects of the microbiome's role in health and disease, but studies must be carefully planned and executed to minimise bias and difficulties with downstream analyses. Improvements to current statistical methods are needed to propel the field towards more robust hypothesis-driven research. As the scale of studies and volume of data increases, these methods need to be optimised for extensive datasets that include many informative covariates, including time-varying covariates, and can account for missing data at multiple levels of the microbiome samples. We also urge

researchers to make their analysis code publicly available on data repositories to allow for faster standardisation of methods across the field (e.g., Github, Bitbucket, or Zenodo).

Table 1. Open-source statistical programs available for longitudinal analysis of the human microbiome high-throughput sequencing data.

Program name	**Model**	**Unit(s) of measurement**	**Advantages**	**Assumptions and limitations**
MaAsLin 2 (Mallick et al. 2021)	Multivariable linear mixed effects models	Differential abundances	• Available as R package. • High stringency reduces likelihood of false positives. • Includes options for random slopes and intercepts. • Several normalisation methods available. • Several mixed effects models available. • Handles missing samples.	• High stringency may reduce sensitivity.
TPLMM (Han et al. 2021)	Two-part linear mixed model with shared random effects	Differential abundances	• Log-transformed normalisation for relative abundances. • Option for random intercepts. • Handles missing samples.	• Code only available in manuscript. • Does not include random slopes.
ZGIMM (Zhang, Guo & Yi 2020)	Zero-inflated gaussian mixed models	Differential abundances	• Available as R package. • Runs with either proportional or count data. • Option for random intercepts. • Handles missing samples.	• Does not include random slopes.
Q2-longitudinal (Bokulich et al. 2018)	Linear mixed effects model	Diversity metrics and temporal dynamics	• Available in QIIME2. • Option for random intercepts. • Handles missing samples.	• Requires rarefied input table which may lead to loss of information (McMurdie & Holmes 2014). • Does not include options for random slopes.
ANCOM-II (Kaul et al. 2017)	Linear mixed effects model	Differential abundances	• Available as R code. • High stringency reduces likelihood of false positives. • Includes options for random slopes and intercepts. • Handles missing samples.	• Preferable to run on count data.

			• Centred-log ratio normalisation for relative abundances.	
ZIBR (Chen & Li 2016)	Two-part zero-inflated beta regression model with random effects	Differential abundances	• Available as R package. • Option for random intercepts. • Two-part model to account for presence/absence of taxa and non-zero abundances of taxa.	• Cannot handle missing samples. • Input needs to be relative abundances, not count data. • Does not include random slopes, but model can be extended to include them.

References

Abdill, RJ, Adamowicz, EM & Blekhman, R 2022, 'Public human microbiome data are dominated by highly developed countries', *PLoS Biol*, vol. 20, no. 2, p. e3001536.

Abshire, M, Dinglas, VD, Cajita, MIA, Eakin, MN, Needham, DM & Himmelfarb, CD 2017, 'Participant retention practices in longitudinal clinical research studies with high retention rates', *BMC Medical Research Methodology*, vol. 17, no. 1, 2017/02/20, p. 30.

Allaband, C, McDonald, D, Vazquez-Baeza, Y, Minich, JJ, Tripathi, A, Brenner, DA, Loomba, R, Smarr, L, Sandborn, WJ, Schnabl, B, Dorrestein, P, Zarrinpar, A & Knight, R 2019, 'Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians', *Clin Gastroenterol Hepatol*, vol. 17, no. 2, Jan, pp. 218-230.

Bell, A, Fairbrother, M & Jones, K 2019, 'Fixed and random effects models: making an informed choice', *Quality & Quantity*, vol. 53, no. 2, 2019/03/01, pp. 1051-1074.

Berg, G, Rybakova, D, Fischer, D, Cernava, T, Vergès, M-CC, Charles, T, Chen, X, Cocolin, L, Eversole, K, Corral, GH, Kazou, M, Kinkel, L, Lange, L, Lima, N, Loy, A, Macklin, JA, Maguin, E, Mauchline, T, McClure, R, Mitter, B, Ryan, M, Sarand, I, Smidt, H, Schelkle, B, Roume, H, Kiran, GS, Selvin, J, Souza, RSCd, van Overbeek, L, Singh, BK, Wagner, M, Walsh, A, Sessitsch, A & Schloter, M 2020, 'Microbiome definition re-visited: old concepts and new challenges', *Microbiome*, vol. 8, no. 1, 2020/06/30, p. 103.

Bharti, R & Grimm, DG 2019, 'Current challenges and best-practice protocols for microbiome analysis', *Briefings in Bioinformatics*, vol. 22, no. 1, pp. 178-193.

Bokulich, NA, Dillon, MR, Zhang, Y, Rideout, JR, Bolyen, E, Li, H, Albert, PS, Caporaso, JG & Arumugam, M 2018, 'q2-longitudinal: Longitudinal and Paired-Sample Analyses of Microbiome Data', *mSystems*, vol. 3, no. 6, pp. e00219-00218.

Bokulich, NA, Maldonado, J, Kang, DW, Krajmalnik-Brown, R & Caporaso, JG 2019, 'Rapidly Processed Stool Swabs Approximate Stool Microbiota Profiles', *mSphere*, vol. 4, no. 2, Apr 10.

Bolger, AM, Lohse, M & Usadel, B 2014, 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, vol. 30, no. 15, Aug 1, pp. 2114-2120.

Bolyen, E, Rideout, JR, Dillon, MR, Bokulich, NA, Abnet, CC, Al-Ghalith, GA, Alexander, H, Alm, EJ, Arumugam, M, Asnicar, F, Bai, Y, Bisanz, JE, Bittinger, K, Brejnrod, A, Brislawn, CJ, Brown, CT, Callahan, BJ, Caraballo-Rodríguez, AM, Chase, J, Cope, EK, Da Silva, R, Diener, C, Dorrestein, PC, Douglas, GM, Durall, DM, Duvallet, C, Edwardson, CF, Ernst, M, Estaki, M, Fouquier, J, Gauglitz, JM, Gibbons, SM, Gibson, DL, Gonzalez, A, Gorlick, K, Guo, J, Hillmann, B, Holmes, S, Holste, H, Huttenhower, C, Huttley, GA, Janssen, S, Jarmusch, AK, Jiang, L, Kaehler, BD, Kang, KB, Keefe, CR, Keim, P, Kelley, ST, Knights, D, Koester, I, Kosciolek, T, Kreps, J, Langille, MGI, Lee, J, Ley, R, Liu, Y-X, Loftfield, E, Lozupone, C, Maher, M, Marotz, C, Martin, BD, McDonald, D, McIver, LJ, Melnik, AV, Metcalf, JL, Morgan, SC, Morton, JT, Naimey, AT, Navas-Molina, JA, Nothias,

LF, Orchanian, SB, Pearson, T, Peoples, SL, Petras, D, Preuss, ML, Pruesse, E, Rasmussen, LB, Rivers, A, Robeson, MS, Rosenthal, P, Segata, N, Shaffer, M, Shiffer, A, Sinha, R, Song, SJ, Spear, JR, Swafford, AD, Thompson, LR, Torres, PJ, Trinh, P, Tripathi, A, Turnbaugh, PJ, Ul-Hasan, S, van der Hooft, JJJ, Vargas, F, Vázquez-Baeza, Y, Vogtmann, E, von Hippel, M, Walters, W, Wan, Y, Wang, M, Warren, J, Weber, KC, Williamson, CHD, Willis, AD, Xu, ZZ, Zaneveld, JR, Zhang, Y, Zhu, Q, Knight, R & Caporaso, JG 2019, 'Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2', *Nature Biotechnology*, vol. 37, no. 8, 2019/08/01, pp. 852-857.

Budding, AE, Grasman, ME, Eck, A, Bogaards, JA, Vandenbroucke-Grauls, CMJE, van Bodegraven, AA & Savelkoul, PHM 2014, 'Rectal swabs for analysis of the intestinal microbiota', *PLoS One*, vol. 9, no. 7, pp. e101344-e101344.

Callahan, BJ, McMurdie, PJ, Rosen, MJ, Han, AW, Johnson, AJA & Holmes, SP 2016, 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*, vol. 13, no. 7, 2016/07/01, pp. 581-583.

Cao, H-T, Gibson, TE, Bashan, A & Liu, Y-Y 2017, 'Inferring human microbial dynamics from temporal metagenomics data: Pitfalls and lessons', *BioEssays*, vol. 39, no. 2, p. 1600188.

Cao, Q, Sun, X, Rajesh, K, Chalasani, N, Gelow, K, Katz, B, Shah, VH, Sanyal, AJ & Smirnova, E 2021, 'Effects of Rare Microbiome Taxa Filtering on Statistical Analysis', *Front Microbiol*, vol. 11, 2021-January-12.

Carson, TL, Little, RB & Townsend, SN 2019, 'Preliminary feasibility for recruiting and retaining black and white females to provide fecal samples for longitudinal research', *Gut Pathogens*, vol. 11, no. 1, 2019/08/26, p. 43.

Casals-Pascual, C, González, A, Vázquez-Baeza, Y, Song, SJ, Jiang, L & Knight, R 2020, 'Microbial Diversity in Clinical Microbiome Studies: Sample Size and Statistical Power Considerations', *Gastroenterology*, vol. 158, no. 6, 2020/05/01/, pp. 1524-1528.

Chaudhari, N, Ravi, R, Gogtay, N & Thatte, U 2020, 'Recruitment and retention of the participants in clinical trials: Challenges and solutions', *Perspectives in Clinical Research*, vol. 11, no. 2, April 1, 2020, pp. 64-69.

Chen, EZ & Li, H 2016, 'A two-part mixed-effects model for analyzing longitudinal microbiome compositional data', *Bioinformatics*, vol. 32, no. 17, Sep 1, pp. 2611-2617.

Chen, S, Zhou, Y, Chen, Y & Gu, J 2018, 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics*, vol. 34, no. 17, pp. i884-i890.

Clooney, AG, Sutton, TDS, Shkoporov, AN, Holohan, RK, Daly, KM, O'Regan, O, Ryan, FJ, Draper, LA, Plevy, SE, Ross, RP & Hill, C 2019, 'Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease', *Cell Host Microbe*, vol. 26, no. 6, Dec 11, pp. 764-778.e765.

Costello, EK, Lauber, CL, Hamady, M, Fierer, N, Gordon, JI & Knight, R 2009, 'Bacterial Community Variation in Human Body Habitats Across Space and Time', *Science*, vol. 326, no. 5960, pp. 1694-1697.

Davis, NM, Proctor, DM, Holmes, SP, Relman, DA & Callahan, BJ 2018, 'Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data', *Microbiome*, vol. 6, no. 1, 2018/12/17, p. 226.

Debelius, JW, Song, SJ, Vazquez-Baeza, Y, Xu, ZZ, Gonzalez, A & Knight, R 2016b, 'Tiny microbes, enormous impacts: what matters in gut microbiome studies?', *Genome Biol*, vol. 17, no. 1, 2016/10/19, p. 217.

Debelius, JW, Vázquez-Baeza, Y, McDonald, D, Xu, Z, Wolfe, E & Knight, R 2016a, 'Turning Participatory Microbiome Research into Usable Data: Lessons from the American Gut Project', *Journal of microbiology & biology education*, vol. 17, no. 1, pp. 46-50.

Dominguez-Bello, MG, Godoy-Vitorino, F, Knight, R & Blaser, MJ 2019, 'Role of the microbiome in human development', *Gut*, vol. 68, no. 6, pp. 1108-1114.

Eckburg, PB, Bik, EM, Bernstein, CN, Purdom, E, Dethlefsen, L, Sargent, M, Gill, SR, Nelson, KE & Relman, DA 2005, 'Diversity of the human intestinal microbial flora', *Science*, vol. 308, no. 5728, Jun 10, pp. 1635-1638.

Eisenhofer, R, Minich, JJ, Marotz, C, Cooper, A, Knight, R & Weyrich, LS 2019, 'Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations', *Trends Microbiol*, vol. 27, no. 2, 2019/02/01/, pp. 105-117.

Flores, GE, Caporaso, JG, Henley, JB, Rideout, JR, Domogala, D, Chase, J, Leff, JW, Vázquez-Baeza, Y, Gonzalez, A, Knight, R, Dunn, RR & Fierer, N 2014, 'Temporal variability is a personalized feature of the human microbiome', *Genome Biol*, vol. 15, no. 12, 2014/12/03, p. 531.

Fukuyama, J, Rumker, L, Sankaran, K, Jeganathan, P, Dethlefsen, L, Relman, DA & Holmes, SP 2017, 'Multidomain analyses of a longitudinal human microbiome intestinal cleanout perturbation experiment', *PLOS Computational Biology*, vol. 13, no. 8, p. e1005706.

Galea, S & Tracy, M 2007, 'Participation Rates in Epidemiologic Studies', *Annals of Epidemiology*, vol. 17, no. 9, 2007/09/01/, pp. 643-653.

GEM Project 2020, *Crohn's and Colitis Canada Study on the Genetic, Environmental and Microbial Interactions that cause IBD*, The Crohn's and Colitis Canada Genetic, Environmental, Microbial (GEM) Project, Crohn's and Colitis Canada, <https://crohnsandcolitis.ca/gem>.

Gilbert, JA, Blaser, MJ, Caporaso, JG, Jansson, JK, Lynch, SV & Knight, R 2018, 'Current understanding of the human microbiome', *Nat Med*, vol. 24, no. 4, Apr 10, pp. 392-400.

Glendinning, L, Wright, S, Tennant, P, Gill, AC, Collie, D & McLachlan, G 2017, 'Microbiota in Exhaled Breath Condensate and the Lung', *Applied and Environmental Microbiology*, vol. 83, no. 12, pp. e00515-00517.

Gloor, GB, Macklaim, JM, Pawlowsky-Glahn, V & Egozcue, JJ 2017, 'Microbiome Datasets Are Compositional: And This Is Not Optional', *Front Microbiol*, vol. 8, 2017-November-15.

Gul, RB & Ali, PA 2010, 'Clinical trials: the challenge of recruitment and retention of participants', *Journal of Clinical Nursing*, vol. 19, no. 1-2, pp. 227-233.

Guo, Y & Pandis, N 2015, 'Sample-size calculation for repeated-measures and longitudinal studies', *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 147, no. 1, 2015/01/01/, pp. 146-149.

Haiser, HJ, Gootenberg, DB, Chatman, K, Sirasani, G, Balskus, EP & Turnbaugh, PJ 2013, 'Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta', *Science (New York, N.Y.)*, vol. 341, no. 6143, pp. 295-298.

Han, Y, Baker, C, Vogtmann, E, Hua, X, Shi, J & Liu, D 2021, 'Modeling Longitudinal Microbiome Compositional Data: A Two-Part Linear Mixed Model with Shared Random Effects', *Statistics in Biosciences*, vol. 13, no. 2, 2021/07/01, pp. 243-266.

Holmes, S 2019, 'Successful strategies for human microbiome data generation, storage and analyses', *Journal of Biosciences*, vol. 44, no. 5, 2019/09/20, p. 111.

Hornung, BVH, Zwittink, RD & Kuijper, EJ 2019, 'Issues and current standards of controls in microbiome research', *FEMS Microbiology Ecology*, vol. 95, no. 5.

Huson, DH, Beier, S, Flade, I, Gorska, A, El-Hadidi, M, Mitra, S, Ruscheweyh, HJ & Tappu, R 2016, 'MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data', *PLoS Comput Biol*, vol. 12, no. 6, Jun, p. e1004957.

Jeon, SR, Chai, J, Kim, C & Lee, CH 2018, 'Current Evidence for the Management of Inflammatory Bowel Diseases Using Fecal Microbiota Transplantation', *Current Infectious Disease Reports*, vol. 20, no. 8, Aug.

Jurasz, H, Pawłowski, T & Perlejewski, K 2021, 'Contamination Issue in Viral Metagenomics: Problems, Solutions, and Clinical Perspectives', *Front Microbiol*, vol. 12, 2021-October-20.

Kaul, A, Mandal, S, Davidov, O & Peddada, SD 2017, 'Analysis of Microbiome Data in the Presence of Excess Zeros', *Front Microbiol*, vol. 8, pp. 2114-2114.

Keegan, KP, Glass, EM & Meyer, F 2016, 'MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function', *Methods Mol Biol*, vol. 1399, pp. 207-233.

Kong, L, Lloyd-Price, J, Vatanen, T, Seksik, P, Beaugerie, L, Simon, T, Vlamakis, H, Sokol, H & Xavier, RJ 2020, 'Linking Strain Engraftment in Fecal Microbiota Transplantation With Maintenance of Remission in Crohn's Disease', *Gastroenterology*, vol. 159, no. 6, Dec, pp. 2193-2202.e2195.

Lauber, CL, Zhou, N, Gordon, JI, Knight, R & Fierer, N 2010, 'Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples', *FEMS Microbiology Letters*, vol. 307, no. 1, pp. 80-86.

Little, RJA & Rubin, DB 2020, *Statistical analysis with missing data*, Wiley series in probability and statistics, Third edition. edn, Wiley, Hoboken, NJ.

Maier, L, Pruteanu, M, Kuhn, M, Zeller, G, Telzerow, A, Anderson, EE, Brochado, AR, Fernandez, KC, Dose, H, Mori, H, Patil, KR, Bork, P & Typas, A 2018, 'Extensive impact of non-antibiotic drugs on human gut bacteria', *Nature*, vol. 555, no. 7698, pp. 623-628.

Mallick, H, Rahnavard, A, McIver, LJ, Ma, S, Zhang, Y, Nguyen, LH, Tickle, TL, Weingart, G, Ren, B, Schwager, EH, Chatterjee, S, Thompson, KN, Wilkinson, JE, Subramanian, A, Lu, Y, Waldron, L, Paulson, JN, Franzosa, EA, Bravo, HC & Huttenhower, C 2021, 'Multivariable association discovery in population-scale meta-omics studies', *PLOS Computational Biology*, vol. 17, no. 11, p. e1009442.

McDonald, D, Hyde, E, Debelius, JW, Morton, JT, Gonzalez, A, Ackermann, G, Aksenov, AA, Behsaz, B, Brennan, C, Chen, Y, Goldasich, LD, Dorrestein, PC, Dunn, RR, Fahimipour, AK, Gaffney, J, Gilbert, JA, Gogul, G, Green, JL, Hugenholtz, P, Humphrey, G, Huttenhower, C, Jackson, MA, Janssen, S, Jeste, DV, Jiang, L, Kelley, ST, Knights, D, Kosciolek, T, Ladau, J, Leach, J, Marotz, C, Meleshko, D, Melnik, AV, Metcalf, JL, Mohimani, H, Montassier, E, Navas-Molina, J, Nguyen, TT, Peddada, S, Pevzner, P, Pollard, KS, Rahnavard, G, Robbins-Pianka, A, Sangwan, N, Shorenstein, J, Smarr, L, Song, SJ, Spector, T, Swafford, AD, Thackray, VG, Thompson, LR, Tripathi, A, Vázquez-Baeza, Y, Vrbanac, A, Wischmeyer, P, Wolfe, E, Zhu, Q, Knight, R, Greene, CS, Mann, AE, Amir, A, Frazier, A, Martino, C, Lebrilla, C, Lozupone, C, Lewis, CM, Raison, C, Zhang, C, Lauber, CL, Warinner, C, Lowry, CA, Callewaert, C, Bloss, C, Willner, D, Galzerani, DD, Gonzalez, DJ, Mills, DA, Chopra, D, Gevers, D, Berg-Lyons, D, Sears, DD, Wendel, D, Lovelace, E, Pierce, E, TerAvest, E, Bolyen, E, Bushman, FD, Wu, GD, Church, GM, Saxe, G, Holscher, HD, Ugrina, I, German, JB, Caporaso, JG, Wozniak, JM, Kerr, J, Ravel, J, Lewis, JD, Suchodolski, JS, Jansson, JK, Hampton-Marcell, JT, Bobe, J, Raes, J, Chase, JH, Eisen, JA, Monk, J, Clemente, JC, Petrosino, J, Goodrich, J, Gauglitz, J, Jacobs, J, Zengler, K, Swanson, KS, Lewis, K, Mayer, K, Bittinger, K, Dillon, L, Zaramela, LS, Schriml, LM, Dominguez-Bello, MG, Jankowska, MM, Blaser, M, Pirrung, M, Minson, M, Kurisu, M, Ajami, N, Gottel, NR, Chia, N, Fierer, N, White, O, Cani, PD, Gajer, P, Strandwitz, P, Kashyap, P, Dutton, R, Park, RS, Xavier, RJ, Mills, RH, Krajmalnik-Brown, R, Ley, R, Owens, SM, Klemmer, S, Matamoros, S, Mirarab, S, Moorman, S, Holmes, S, Schwartz, T, Eshoo-Anton, TW, Vigers, T, Pandey, V, Treuren, WV, Fang, X, Xu, ZZ, Jarmusch, A, Geier, J, Reeve, N, Silva, R, Kopylova, E, Nguyen, D, Sanders, K, Benitez, RAS, Heale, AC, Abramson, M, Waldispühl, J, Butyaev, A, Drogaris, C, Nazarova, E, Ball, M & Gunderson, B 2018,

'American Gut: an Open Platform for Citizen Science Microbiome Research', *mSystems*, vol. 3, no. 3, pp. e00031-00018.

McMurdie, PJ & Holmes, S 2013, 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data', *PLoS One*, vol. 8, no. 4, p. e61217.

McMurdie, PJ & Holmes, S 2014, 'Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible', *PLOS Computational Biology*, vol. 10, no. 4, p. e1003531.

Mirzayi, C, Renson, A, Furlanello, C, Sansone, S-A, Zohra, F, Elsafoury, S, Geistlinger, L, Kasselman, LJ, Eckenrode, K, van de Wijgert, J, Loughman, A, Marques, FZ, MacIntyre, DA, Arumugam, M, Azhar, R, Beghini, F, Bergstrom, K, Bhatt, A, Bisanz, JE, Braun, J, Bravo, HC, Buck, GA, Bushman, F, Casero, D, Clarke, G, Collado, MC, Cotter, PD, Cryan, JF, Demmer, RT, Devkota, S, Elinav, E, Escobar, JS, Fettweis, J, Finn, RD, Fodor, AA, Forslund, S, Franke, A, Furlanello, C, Gilbert, J, Grice, E, Haibe-Kains, B, Handley, S, Herd, P, Holmes, S, Jacobs, JP, Karstens, L, Knight, R, Knights, D, Koren, O, Kwon, DS, Langille, M, Lindsay, B, McGovern, D, McHardy, AC, McWeeney, S, Mueller, NT, Nezi, L, Olm, M, Palm, N, Pasolli, E, Raes, J, Redinbo, MR, Rühlemann, M, Balfour Sartor, R, Schloss, PD, Schriml, L, Segal, E, Shardell, M, Sharpton, T, Smirnova, E, Sokol, H, Sonnenburg, JL, Srinivasan, S, Thingholm, LB, Turnbaugh, PJ, Upadhyay, V, Walls, RL, Wilmes, P, Yamada, T, Zeller, G, Zhang, M, Zhao, N, Zhao, L, Bao, W, Culhane, A, Devanarayan, V, Dopazo, J, Fan, X, Fischer, M, Jones, W, Kusko, R, Mason, CE, Mercer, TR, Sansone, S-A, Scherer, A, Shi, L, Thakkar, S, Tong, W, Wolfinger, R, Hunter, C, Segata, N, Huttenhower, C, Dowd, JB, Jones, HE, Waldron, L, Genomic Standards, C, Massive, A & Quality Control, S 2021, 'Reporting guidelines for human microbiome research: the STORMS checklist', *Nature Medicine*, vol. 27, no. 11, 2021/11/01, pp. 1885-1892.

Nath, S, Handsley-Davis, M, Weyrich, LS & Jamieson, LM 2021, 'Diversity and bias in oral microbiome research: A commentary', *eClinicalMedicine*, vol. 36.

Nearing, JT, Comeau, AM & Langille, MGI 2021, 'Identifying biases and their potential solutions in human microbiome studies', *Microbiome*, vol. 9, no. 1, 2021/05/18, p. 113.

Oh, V-KS & Li, RW 2022, 'Large-Scale Meta-Longitudinal Microbiome Data with a Known Batch Factor', *Genes*, vol. 13, no. 3, p. 392.

Pan, AY 2021, 'Statistical analysis of microbiome data: The challenge of sparsity', *Current Opinion in Endocrine and Metabolic Research*, vol. 19, 2021/08/01/, pp. 35-40.

Pollock, J, Glendinning, L, Wisedchanwet, T, Watson, M & Liu, S-J 2018, 'The Madness of Microbiome: Attempting To Find Consensus “Best Practice” for 16S Microbiome Studies', *Applied and Environmental Microbiology*, vol. 84, no. 7, pp. e02627-02617.

Raes, J 2016, 'Microbiome-based companion diagnostics: no longer science fiction?', *Gut*, vol. 65, no. 6, pp. 896-897.

Reyman, M, van Houten, MA, Arp, K, Sanders, EAM & Bogaert, D 2019, 'Rectal swabs are a reliable proxy for faecal samples in infant gut microbiota research based on 16S-rRNA sequencing', *Sci Rep*, vol. 9, no. 1, 2019/11/05, p. 16072.

Robins, JM, Hernán, MA & Brumback, B 2000, 'Marginal structural models and causal inference in epidemiology', *Epidemiology*, vol. 11, no. 5, Sep, pp. 550-560.

Salter, SJ, Cox, MJ, Turek, EM, Calus, ST, Cookson, WO, Moffatt, MF, Turner, P, Parkhill, J, Loman, NJ & Walker, AW 2014, 'Reagent and laboratory contamination can critically impact sequence-based microbiome analyses', *BMC Biology*, vol. 12, no. 1, 2014/11/12, p. 87.

Sender, R, Fuchs, S & Milo, R 2016, 'Revised Estimates for the Number of Human and Bacteria Cells in the Body', *PLoS Biol*, vol. 14, no. 8, Aug, p. e1002533.

Shang, FM & Liu, HL 2018, 'Fusobacterium nucleatum and colorectal cancer: A review', *World J Gastrointest Oncol*, vol. 10, no. 3, Mar 15, pp. 71-81.

Shaw, AG, Sim, K, Powell, E, Cornwell, E, Cramer, T, McClure, ZE, Li, M-S & Kroll, JS 2016, 'Latitude in sample handling and storage for infant faecal microbiota studies: the elephant in the room?', *Microbiome*, vol. 4, no. 1, 2016/07/30, p. 40.

Short, MI, Hudson, R, Besasie, BD, Reveles, KR, Shah, DP, Nicholson, S, Johnson-Pais, TL, Weldon, K, Lai, Z, Leach, RJ, Fongang, B & Liss, MA 2021, 'Comparison of rectal swab, glove tip, and participant-collected stool techniques for gut microbiome sampling', *BMC Microbiology*, vol. 21, no. 1, pp. 26-26.

Silverman, JD, Durand, HK, Bloom, RJ, Mukherjee, S & David, LA 2018b, 'Dynamic linear models guide design and analysis of microbiota studies within artificial human guts', *Microbiome*, vol. 6, no. 1, 2018/11/12, p. 202.

Silverman, JD, Shenhav, L, Halperin, E, Mukherjee, S & David, LA 2018a, 'Statistical Considerations in the Design and Analysis of Longitudinal Microbiome Studies', *bioRxiv*, p. 448332.

Sinha, R, Chen, J, Amir, A, Vogtmann, E, Shi, J, Inman, KS, Flores, R, Sampson, J, Knight, R & Chia, N 2016, 'Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies', *Cancer Epidemiology, Biomarkers & Prevention*, vol. 25, no. 2, pp. 407-416.

Smirnova, E, Huzurbazar, S & Jafari, F 2018, 'PERFect: PERmutation Filtering test for microbiome data', *Biostatistics*, vol. 20, no. 4, pp. 615-631.

Stegen, JC, Bottos, EM & Jansson, JK 2018, 'A unified conceptual framework for prediction and control of microbiomes', *Current Opinion in Microbiology*, vol. 44, 2018/08/01/, pp. 20-27.

Sterne, JA, White, IR, Carlin, JB, Spratt, M, Royston, P, Kenward, MG, Wood, AM & Carpenter, JR 2009, 'Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls', *BMJ (Clinical research ed.)*, vol. 338, Jun 29, p. b2393.

Sun, S, Zhu, X, Huang, X, Murff, HJ, Ness, RM, Seidner, DL, Sorgen, AA, Blakley, IC, Yu, C, Dai, Q, Azcarate-Peril, MA, Shrubsole, MJ & Fodor, AA 2021, 'On the robustness of inference of association with the gut microbiota in stool, rectal swab and mucosal tissue samples', *Sci Rep*, vol. 11, no. 1, 2021/07/21, p. 14828.

Thiese, MS 2014, 'Observational and interventional study design types; an overview', *Biochemia medica*, vol. 24, no. 2, pp. 199-210.

Tripathi, A, Marotz, C, Gonzalez, A, Vázquez-Baeza, Y, Song, SJ, Bouslimani, A, McDonald, D, Zhu, Q, Sanders, JG, Smarr, L, Dorrestein, PC & Knight, R 2018, 'Are microbiome studies ready for hypothesis-driven research?', *Curr Opin Microbiol*, vol. 44, Aug, pp. 61-69.

Uhr, GT, Dohnalová, L & Thaiss, CA 2019, 'The Dimension of Time in Host-Microbiome Interactions', *mSystems*, vol. 4, no. 1, pp. e00216-00218.

Vandeputte, D, Falony, G, Vieira-Silva, S, Tito, RY, Joossens, M & Raes, J 2016, 'Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates', *Gut*, vol. 65, no. 1, pp. 57-62.

Vandeputte, D, Kathagen, G, D'hoe, K, Vieira-Silva, S, Valles-Colomer, M, Sabino, J, Wang, J, Tito, RY, De Commer, L, Darzi, Y, Vermeire, S, Falony, G & Raes, J 2017, 'Quantitative microbiome profiling links gut community variation to microbial load', *Nature*, vol. 551, no. 7681, 2017/11/01, pp. 507-511.

Vázquez-Baeza, Y, Gonzalez, A, Xu, ZZ, Washburne, A, Herfarth, HH, Sartor, RB & Knight, R 2018, 'Guiding longitudinal sampling in IBD cohorts', *Gut*, vol. 67, no. 9, pp. 1743-1745.

Vujkovic-Cvijin, I, Sklar, J, Jiang, L, Natarajan, L, Knight, R & Belkaid, Y 2020, 'Host variables confound gut microbiota studies of human disease', *Nature*, vol. 587, no. 7834, 2020/11/01, pp. 448-454.

Wang, Y & LêCao, KA 2020, 'Managing batch effects in microbiome data', *Brief Bioinform*, vol. 21, no. 6, Dec 1, pp. 1954-1970.

Weiss, S, Amir, A, Hyde, ER, Metcalf, JL, Song, SJ & Knight, R 2014, 'Tracking down the sources of experimental contamination in microbiome studies', *Genome Biol*, vol. 15, no. 12, 2014/12/17, p. 564.

Weyrich, LS, Farrer, AG, Eisenhofer, R, Arriola, LA, Young, J, Selway, CA, Handsley-Davis, M, Adler, CJ, Breen, J & Cooper, A 2019, 'Laboratory contamination over time during low-biomass sample analysis', *Molecular Ecology Resources*, vol. 19, no. 4, pp. 982-996.

Wilkinson, JE, Franzosa, EA, Everett, C, Li, C, Hu, FB, Wirth, DF, Song, M, Chan, AT, Rimm, E, Garrett, WS & Huttenhower, C 2021, 'A framework for microbiome science in public health', *Nat Med*, vol. 27, no. 5, May, pp. 766-774.

Zaura, E, Pappalardo, VY, Buijs, MJ, Volgenant, CMC & Brandt, BW 2021, 'Optimizing the quality of clinical studies on oral microbiome: A practical guide for planning, performing, and reporting', *Periodontol 2000*, vol. 85, no. 1, Feb, pp. 210-236.

Zhang, X, Guo, B & Yi, N 2020, 'Zero-Inflated gaussian mixed models for analyzing longitudinal microbiome data', *PLoS One*, vol. 15, no. 11, pp. e0242073-e0242073.

Zhang, X & Yi, N 2020, 'NBZIMM: negative binomial and zero-inflated mixed models, with application to microbiome/metagenomics data analysis', *BMC Bioinformatics*, vol. 21, no. 1, 2020/10/30, p. 488.

Thesis Discussion

Overview

The prevalence of inflammatory bowel disease (IBD) is continuing to rise on a global scale, and there is an urgent need for effective therapies and treatments for the disease (Alatab et al. 2020; Cai, Wang & Li 2021). The discovery of the gut microbiome and its link to intestinal diseases represented a major milestone for human health research, and the ability to study it through high-throughput sequencing methods has dramatically improved our understanding of IBD (Pittayanon et al. 2020). Through this research, we have gained valuable insights into the etiopathology of the disease, and, importantly, how it can be managed (Glassner, Abraham & Quigley 2020; Aldars-García, Chaparro & Gisbert 2021). However, the field of gut microbiome research is still relatively young—there are many knowledge gaps yet to be filled, and a continuing need for improvements to the existing methods, approaches, and tools (Proctor 2019). The goal of this thesis was to advance our understanding of IBD through knowledge gained from studying several facets of the gut microbiome. The body of work presented in this thesis can be summarised into three main themes:

1. Improve our understanding of the relationship between the gut microbiome and IBD.
2. Refine our ability to treat IBD through research of the gut microbiome.
3. Gain insights and stimulate future research on IBD through an interdisciplinary approach to microbiome science.

In this thesis discussion, I highlight how my research within each chapter relates to each of these themes. I also discuss the significance of my research to the field, the current limitations of available approaches, and the gaps to be addressed with future of research on the gut microbiome and IBD.

Improving our understanding of the link between the gut microbiome and IBD

Nonbacterial gut microbes help explain 'dysbiosis' in IBD

The term 'dysbiosis' is commonly used to describe how the gut microbiomes of individuals with IBD differ in some way from unaffected, healthy individuals. It is an umbrella term to describe several patterns, such as microbiomes that are altered in diversity, composition, function, or specific taxa from a healthy state (Hooks & O'Malley 2017). While broad terms such as 'dysbiosis' can be convenient to describe complex phenomena, they can also be

misleading due to their generalised usage. It is for this reason that I refrained from using the term ‘dysbiosis’ throughout this thesis except when referring to its contextual association. Instead, I opted to explicitly describe how the IBD gut microbiome differs from a healthy one, which is an approach that I and other scientists believe is more beneficial for the field (Hooks & O'Malley 2017; Aldars-García, Chaparro & Gisbert 2021). In **Chapter 1**, I highlighted that ‘dysbiosis’ has become misleading in IBD research for the aforementioned reason. The example I gave is how most gut microbiome studies to date have only included bacteria but describe the entirety of the microbiome as being ‘dysbiotic’. Including nonbacterial gut microbes in more studies of IBD will therefore expand our knowledge of what ‘dysbiosis’ actually means for individuals with the disease.

In **Chapter 2**, I supported the assertion that nonbacterial microbes are important in IBD by identifying intestinal eukaryotes in a large shotgun metagenomic dataset of individuals with the disease. This research showed that there are indeed differences in the distributions of fungal and protozoan species in IBD. Fungi have generally been described as having a pathogenic role in IBD (Sartor & Wu 2017; Sokol et al. 2017), though mouse models have identified a potentially therapeutic effect of *Saccharomyces* species (Sivignon et al. 2015; Wang, C et al. 2019). While this study did not ascertain the mechanistic involvement of fungi in the disease, it showed how IBD patients harbour a higher abundance and diversity of fungi with greater resolution than previous studies. This result can be used to justify more resource-intensive studies that investigate the molecular involvement of fungi in the pathogenesis of IBD. For example, more studies with ex-vivo tissue samples are needed to examine the interactions between fungi, other microbial cells, and immune cells at the mucosa (e.g., Jain et al. 2021).

It has been suggested that the relative abundance of fungi differs along the gastrointestinal tract (Sender, Fuchs & Milo 2016; Richard & Sokol 2019), wherein fungi have a higher abundance relative to bacteria in the stomach and duodenum than the colon. Investigating fungal interactions at non-colonic mucosal sites may therefore explain the anatomical differences seen in subtypes of CD. Sampling along the gastrointestinal tract would also be especially helpful in CD where inflammation can occur in regions other than the colon (e.g., ileitis, gastroduodenal CD, and jejunoileitis) such that stool samples may be less representative for patients with these phenotypes (Tang et al. 2020). An additional layer of critical information will be to add host genetics to these studies (Glassner, Abraham & Quigley 2020). Mutations for antifungal immune genes have been found in only some

individuals with IBD (Richard et al. 2015), and it will be meaningful to know if the distribution of fungi differs in individuals with these mutations. Lastly, using methods such as metatranscriptomics to capture the gene expression of fungal cells will allow us to determine their functional activity in the gut (Lavelle & Sokol 2018).

These avenues for future research of fungi can likewise be applied to protozoa, where even less is currently known about how and where they colonise along the gastrointestinal tract, or the microbial cells they interact with (Garcia 2016; Chabé, Lokmer & Ségurel 2017; Deng et al. 2021). Common intestinal protozoa such as *Blastocystis* and *Dientamoeba fragilis* were historically known as pathogens as they were often identified in the faeces of individuals with gastroenteritis (Barratt et al. 2011; Coyle et al. 2011). However, **Chapter 1** discussed the recent discoveries made by high-throughput sequencing studies that have found *Blastocystis* and *D. fragilis* in healthy, asymptomatic adults and children, and how their absence is associated with lower bacterial diversity (Nourrisson et al. 2014; Holtman et al. 2016; Jokelainen et al. 2017; Nieves-Ramírez et al. 2018; Brands et al. 2019; Dullaert-de Boer et al. 2019; Tito et al. 2019). It has been asserted that these protozoa may in fact feed on intestinal pathogens, and the loss or absence of these genera leads to an overgrowth of their prey and a loss of bacterial diversity (Laforest-Lapointe & Arrieta 2018; Deng et al. 2021). In **Chapter 2**, the high prevalence of protozoa, particularly *Blastocystis* spp. and *D. fragilis,* in healthy individuals and the notable absence of these species in the IBD cohort was one of the most dramatic results from this study. This result indicates that these species should no longer be overlooked in microbiome research given that there appears to be a strong link between their prevalence and health.

Understanding why and how protozoa are absent in IBD can provide information on the mechanisms that underpin IBD. While recent hypotheses can explain the potential role for these microbes in a healthy state, they do not explain how protozoa, such as *Blastocystis*, are lost from the gut microbiome in the first place. One explanation is that the use of antibiotics in patients causes their depletion, both by direct depletion of *Blastocystis* and indirectly through depletion of bacteria they interact with, i.e., their absence is a consequence of the disease and its treatment (Nagata et al. 2012; Roberts et al. 2014; Cai, Wang & Li 2021). Similarly, antibiotic use during childhood is a risk factor for developing IBD (Shaw, Blanchard & Bernstein 2010; Kronman et al. 2012), so it is possible that a depletion in protozoa is a causal factor in the development of IBD. Looking at the distribution of intestinal protozoa across the world shows that there is a link between lifestyle factors and their

prevalence (Speich et al. 2016; Gizaw et al. 2018; Omarova et al. 2018). Individuals living in rural, non-industrialised regions have a markedly higher abundance and diversity of protozoa (El Safadi et al. 2014; Morton et al. 2015; Lokmer et al. 2019). These species are generally considered rare in the gut microbiomes of industrialised populations; even the levels seen in the healthy individuals in **Chapter 2** are much lower than what is seen in individuals who live non-industrialised lifestyles (Lokmer et al. 2019). Interestingly, an industrialised lifestyle is also a risk factor for IBD (Barreiro-de Acosta et al. 2011; Ho et al. 2019). The incidence of IBD is considerably higher in 'Western' nations, and regions of Africa and Asia that are becoming more industrialised are associated with an increased incidence of IBD that outpaces a genetic explanation (Ng et al. 2018). With this lifestyle comes several factors that are linked to the loss of microbial diversity such as increased antibiotic use during childhood, increased hygiene practices, exposure to pollution, reduced diversity of food sources, and a decreased exposure to green spaces (López-Serrano et al. 2010; M'Koma 2013). The link between industrialisation and immune dysfunction has similarly been observed in other immune-mediated conditions, including allergic rhinitis, atopic dermatitis, asthma, and type 1 diabetes (Kondrashova et al. 2005; Ege et al. 2011; Laatikainen et al. 2011). It has been posited that a lack of exposure to and colonisation by a diversity of microbes during childhood confers immune dysfunction and may thus prevent the development of immune-mediated conditions such as IBD (Bach 2018). Therefore, studying protozoa alongside bacteria, and their interactions with bacteria, in different populations and health conditions will give broader insights into the development of IBD and other immune-mediated conditions. Even if the loss of protozoa does not directly cause IBD, it may at least be an important warning sign (i.e., 'biomarker') that can be used in newly emerging microbiome-based diagnostics to recognise those at higher risk for developing it.

One critical area that was not explored in **Chapter 2** is the distribution of intestinal viruses, particularly bacteriophages, in IBD and its subtypes. **Chapter 1** discussed a handful of studies that have investigated the gut virome in paediatric and adult patients with IBD, showing mixed results so far. While incredibly small, viruses outnumber bacteria by up to 10:1 in the gut, and they have the capacity to regulate microbial communities (Minot et al. 2013; Mukhopadhya et al. 2019). Therefore, studying these tiny microbes offers an additional opportunity to describe the microbiome in IBD, and to understand the trans-kingdom interactions that may underpin the compositional changes seen in bacteria. There is also evidence to suggest that intestinal viruses contribute directly to inflammation (Adiliaghdam

et al. 2022), and it has been hypothesised that viral infection of the mucosa may be responsible for triggering IBD in susceptible individuals (Ungaro et al. 2019).

Studying viruses through metagenomics remains challenging, and I personally decided against conducting this analysis for several reasons. Firstly, the primary studies in **Chapter 2** did not use appropriate methods to enrich viral DNA in their laboratory protocols, such as viral protein isolation (Kleiner, Hooper & Duerkop 2015), and it is unlikely that the standard methods used to enrich bacterial DNA would capture sufficient viral DNA (Dutilh et al. 2017). Secondly, the study design was a cross-cohort comparison, which meant that the samples came from different labs. Previous research on the gut virome has shown that laboratory-specific effects, such as contamination, can overshadow biological effects, and this could have significantly confounded our results (Gregory et al. 2020; Jurasz, Pawłowski & Perlejewski 2021). The current research suggests that it is best to study the gut virome by specifically designing this intention into the study from the beginning, and not with a retrospective, observational design (Zhang, YZ, Shi & Holmes 2018). Lastly, viruses are more bioinformatically challenging to study than eukaryotes given they have no universal marker genes and are vastly under-described (Khan Mirzaei et al. 2021). While there were a small number of tools to help with analysis (see **Chapter 1**), there were no open-source, user-friendly, and comprehensive end-to-end bioinformatic pipelines to identify and analyse viral DNA from raw sequencing data during the analysis phase of **Chapter 2** (Mukhopadhya et al. 2019), and I did not have the resources to develop and perform such an intensive analysis.

Fortunately, more recent advancements in this space have led to available end-to-end pipelines such as VirSorter2 (Guo et al. 2021), VIBRANT (Kieft, Zhou & Anantharaman 2020), and DeepVirFinder (Ren et al. 2020). Tools like these can identify viruses with greater sensitivity thanks to improved databases and machine learning algorithms that include reference-independent methods (Kieft, Zhou & Anantharaman 2020; Ren et al. 2020). For example, Clooney et al. (2019) created a reference-independent viral metagenomics pipeline that allowed them to identify more viruses in a dataset of individuals with IBD than had previously been published (Norman et al. 2015). They also grouped viral strains into higher taxonomic rankings using protein homology to overcome the high inter-individual variability of gut viromes. This step allowed them to see broader, cohort-level differences between the individuals with UC, CD, and without IBD and they found that healthy individuals had a core gut virome that was absent in individuals with CD. They were also able to show that both gut bacteria and viruses had parallel shifts in individuals with IBD. This interkingdom connection

provides a powerful tool for understanding the gut microbiome's role in the pathophysiology of IBD. Understanding connections like these will only continue to expand given that the increasing interest in viromics suggests we are on the precipice of an explosion of clinical virome research (Khan Mirzaei et al. 2021).

Longitudinal sampling can better define the gut microbiome in IBD

The cohort comparison in **Chapter 2** consisted of samples from two cross-sectional studies. These samples were collected from each participant at only a single timepoint, which was a notable limitation of this study. This is because cross-sectional sampling of the microbiome generally cannot delineate causal roles of microbes in IBD, only associations (Savitz & Wellenius 2022). The strength of an association can be stronger when a large, well-defined cohort is studied, but it is difficult to attribute mechanistic roles without observing the temporal distribution of the gut microbiome in association with the development of clinical symptoms (Brüssow 2016). It is for this reason that, in **Chapter 4**, I stressed the importance of capturing the dynamics of the microbiome with longitudinal sampling. Longitudinal research is a critical focus for future microbiome research in IBD, as well as more broadly for other microbiome-associated diseases and conditions.

Resources, such as those discussed in **Chapter 4,** are invaluable in arming researchers with the necessary tools and knowledge to conduct these studies. Longitudinal research is generally more demanding and expensive than cross-sectional, and **Chapter 4** provides considerations for researchers to carry out the most resource-efficient studies (Allaband et al. 2019). This chapter was written beyond the scope of IBD and the gut microbiome, and instead discussed general principles for all clinical microbiome research. I took this approach because the considerations and available statistical methods broadly remain the same for all body sites, thereby widening the target audience to deliver this important message to more individuals in the field and adjacent fields. I also discussed limitations of the available statistical tools for longitudinal microbiome studies, such as determining statistical power, adequate sampling size and frequency, imputation for missing microbiome data, and regression methods that are optimised for larger datasets. Therefore, this chapter is a resource for statistical gaps in the field that can be addressed with future research, thereby leading to improvements in methods that can benefit our understanding of IBD too.

IBD research specifically can benefit from longitudinal microbiome sampling because it is a multifactorial disease with symptoms that change over time (Loddo & Romano 2015). As the microbiome itself is a temporally fluctuating ecosystem, its variability must be accounted for to determine correlation and causality with the disease itself (Stegen, Bottos & Jansson 2018; Uhr, Dohnalová & Thaiss 2019). Longitudinal studies have found that individuals with IBD have gut microbiomes that are more temporally unstable than unaffected individuals (Halfvarson et al. 2017; Clooney et al. 2021), and this instability can be further affected by medical interventions related to the disease. For example, Halfvarson et al. (2017) sampled individuals with CD and UC and healthy controls every three months over a 2-year period and found that the individuals with CD who had undergone surgery (ileocaecal resection) were more likely to have unstable gut microbiota. Moreover, Clooney et al. (2021) discovered that fluctuations in the gut microbiomes of CD and UC patients occurred as a result of changes to medication (corticosteroids) across the study period of 4-month sampling intervals. These changes were able to be observed because longitudinal sampling was conducted, and a cross-sectional study at a single time point likely would not show this.

Whether a patient is experiencing a flare or is in remission is another important factor to consider when studying the microbiome in IBD because the level of inflammation along the gastrointestinal tract affects how and which microbes can colonise the region (Mirsepasi-Lauridsen et al. 2018; Clooney et al. 2021; Jain et al. 2021). As the disease is a relapsing-remitting condition, sampling across time can allow us to track the effects of disease activity on the microbiota. Flares can also affect other lifestyle factors of a patient—for example, flares can exacerbate stress and affect dietary choices (Vagianos et al. 2022)—which can in turn affect the microbial composition of the gut. Clooney et al. (2021) found that the largest microbiome variance occurred in patients who transitioned between active and inactive IBD. While these studies looked at bacteria, future studies could include eukaryotes to see, for example, if fungal abundance increases in response to inflammation. This relationship is better defined when the microbiome can be measured before disease transitions occur. Additionally, this study design can enable us to determine the directionality of the microbiome-disease relationship. Nishihara et al. (2021) sampled colonic mucosa in individuals with UC over five years and found that lower alpha diversity was a precursor to flares, suggesting that alpha diversity could be used as an indicator for risk of relapse. However, this study did not measure faecal calprotectin levels, an indicator of intestinal inflammation (Bjarnason 2017), and it would be interesting to know whether these levels

increased after bacterial diversity decreased or if the opposite occurred. If the former occurred, this may indicate that changes to bacterial diversity had a causal role in immune inflammation leading to flare.

While these studies tracked participants who had already been diagnosed with IBD, longitudinal epidemiological studies of undiagnosed individuals are also needed to understand the earlier microbiological developments that eventuate in the disease. These studies will help us understand the microbial indicators of IBD risk (Agrawal et al. 2022), as well as understanding the causal mechanisms related to the microbiota that can be exploited in microbiota-based therapies (Stegen, Bottos & Jansson 2018; Uhr, Dohnalová & Thaiss 2019). Epidemiological microbiome studies that span across many years are historically uncommon due to the resources required to undertake such studies, but are becoming increasingly popular as public interest in the microbiome has led to more funding from the public and private sectors, engagement from the public, and international collaborative efforts between medical collection points (Wilkinson et al. 2021). Largescale IBD studies are currently underway, such as the Crohn's and Colitis Canada Study on the Genetic, Environmental and Microbial Interactions that cause IBD (GEM Project) that has longitudinally-collected samples, including stool, from individuals with CD and their first-degree unaffected relatives across seven countries (GEM Project 2020). Some of these first-degree relatives were diagnosed with CD within the study period, and their present-day samples can be compared to their pre-diagnosis samples to identify changes to their microbiota. Additionally, large citizen science projects that have collected gut microbiome samples from thousands of individuals (e.g., The American Gut Project, Human Microbiome Project, The Human Diets & Microbiome Initiative) may be able to retrospectively sample individuals who have been diagnosed with IBD since they donated samples to the initial projects (Turnbaugh et al. 2007; McDonald et al. 2018; Taylor et al. 2020). As most studies of the microbiome in IBD are conducted after the disease has developed, these longitudinal efforts are ideal to characterise the microbial factors involved in its manifestation.

Longitudinal studies that span from childhood can also help determine which environmental triggers lead to the disease. It is beneficial to know when the critical exposure window occurs for factors that deplete the gut microbiome such as antibiotic use during childhood and a lack of exposure to microbially-rich environments contribute to the development of IBD (Ho et al. 2019; Kellermayer & Zilbauer 2020). For example, what is the age and exposure to antibiotics that triggers the development of IBD? How can we safely supplement ourselves

and our microbially-depleted environments with more beneficial microbes (i.e., probiotics)? What other lifestyle choices can we encourage to maintain healthy microbiota and reduce the development of IBD? These questions are critical to answer as more countries are becoming industrialised and urbanised (M'Koma 2013). Additionally, the hygiene practices adopted during the COVID-19 pandemic may have lasting impacts on children who were born and developing during this time, and these questions need to be addressed so we are better able to handle, and hopefully reverse, the potential immune consequences of highly sanitised lifestyles (Burchill et al. 2021).

Chapter 4 also had a strong focus on standardising technical aspects of longitudinal sampling in IBD research. Batch effects are a large source of variation in microbiome sampling, and it is important to account for these by trying to minimise them where possible and track them as covariates for downstream analyses (Wang, Y & LêCao 2020; Wilkinson et al. 2021). Longer study periods have an increased likelihood that technical aspects will change throughout the course of the study (e.g., reagents, sequencing protocol, sources of contamination), therefore increasing the need for standardising the technical efforts. Using flexible sampling types can also increase the accessibility of longitudinal studies by making it easier for participants to self-collect their samples (Debelius et al. 2016a). This is exemplified in **Chapter 3**, where dry swabs were implemented in place of whole stool. Of course, not all samples are created equal, and there will be variation in the efficiency and accuracy of how each sample type captures the gut microbiome. Therefore, a critical first step in using a sample type that is different from the reference sample is to conduct a comparative analysis.

In **Chapter 3**, dry swabs were compared to their reference sample type, whole stool, to determine whether they are a valid proxy. Dry swabs have been previously used in several publications, including the large-scale citizen science project, the American Gut Project (McDonald et al. 2018). The use of this sample type in the literature dates to 2009 (Costello, EK et al. 2009), and two subsequent studies included comparisons to stool (Sinha et al. 2016; Bokulich et al. 2019). However, **Chapter 3** was the first to conduct such a detailed comparison. Additionally, the study showed that dry swabs still maintain their utility in a real-world setting where participants collected their samples, instead of in an ideal laboratory setting. The chapter also provides a template for researchers wanting to benchmark new sample types, and these sample proxies do not necessarily need to be gut samples. Alternatives for sampling microbiota of other body sites are discussed in **Chapter 4**, where it is recommended to conduct a comparative analysis before undertaking an entire study with a

potentially low-quality sample type. Expanding and improving sampling methods is a crucial step towards making microbiome science accessible to more individuals, particularly in long-term studies where participant attrition is a larger issue.

Refining our ability to treat IBD

The primary goal of studying the microbiome in IBD is to understand the microbial and immune mechanisms that lead to its development and persistence. This is so that we can understand how to prevent, treat, and one day find a cure for the disease. The strong links observed between the gut microbiome and IBD have led clinicians and researchers to understand the effects of existing therapeutics on the gut microbiome and to trial several microbiota-altering therapies (Graham, Subramanian & Xavier 2019; Knox et al. 2019; Plichta et al. 2019). My thesis provides several contributions to the advancements of these therapies for IBD.

Studying nonbacterial gut microbes can improve the efficacy of therapies for IBD

Therapies for IBD range from immunomodulators and immunosuppressants (e.g., aminosalicylates, corticosteroids, and antibody therapies) to therapies that work by altering the gut microbiota (Cai, Wang & Li 2021). All of these options have moderate success rates, and improvements to therapies are desperately needed. **Chapter 1** and **Chapter 2** demonstrate the necessity to study all components of the gut microbiota to better understand IBD. This understanding goes beyond characterising the 'IBD gut microbiome' for prognostics and diagnostics and extends to improving therapies for IBD. While I discussed these advantages briefly in **Chapter 1**, I will expand on those ideas here.

Firstly, nonbacterial microbes are an additional avenue to understand how existing treatments work, and why they are not be effective in every patient. The gut microbiome plays a role in the bioavailability of drugs which may be affected by the composition and diversity of nonbacterial microbes that are present upon administration (Weersma, Zhernakova & Fu 2020; Zhang, F et al. 2020). Therefore, identifying which microbes are predictive of a better response will help target treatments to patients. Additionally, the interaction between drugs and the nonbacterial microbiome may provide insights into how and if a treatment will work.

For example, Jun et al. (2019) examined the effect of 5-aminosalicylic acid (5-ASA), an aminosalicylate first-line treatment for colonic IBD, on the fungal microbiome of the colonic mucosa in individuals with UC. The inflamed mucosa in the treatment-naïve group were enriched for fungi, such as *Wickerhamomyces*. Conversely, fungi were less diverse in patients who were treated with 5-ASA, and *Wickerhamomyces* was less abundant. This study suggests that there is a link between immune modulation and the distribution of mucosal fungi. How and why the distribution of fungi changes due to therapy remains to be answered. Fungi may respond to the therapy directly, changes in the immune system, or changes in other microbes, such as bacteria that are affected by 5-ASA as well. Furthermore, it is unclear whether a change to the distribution of fungi augments the effect of 5-ASA. Studies conducted *in vivo* may help uncover the chain of molecular events that take place in response to 5-ASA administration.

Nonbacterial bacteria can also inform on much-needed refinements to microbiota-based therapies for IBD (Ungaro et al. 2019; Houshyar et al. 2021; Underhill & Braun 2022). Only two types of microbiota-based therapies are currently approved for clinical use in Australia: probiotics (*Escherichia coli* Nissle 1917) and faecal microbiota transplantation (FMT) (GESA 2018). The efficacy of these therapies sits at a moderate level, on par with other interventions for IBD (Costello, SP et al. 2017; Caldeira et al. 2020). There is currently little evidence to support the efficacy of probiotics in IBD, and given the difficulties in standardisation, they are generally not recommended to treat or prevent IBD unless the administration follows antibiotic treatment (Gulliver et al. 2022). A hypothesis for their lack of effectiveness is that probiotics interact with the luminal microbiota more than mucosal microbiota which are at the interface of inflammation (Jonkers & Stockbrügger 2003). Furthermore, probiotics rely on identifying key strains that deliver anti-inflammatory effects, which can be difficult given the polymicrobial nature of the disease. Unlike probiotics, FMT is a complex mixture of microbes and nutrients that proves difficult to describe (Gulliver et al. 2022). However, characterising the constitution of faeces that is most beneficial for recipients can lead to breakthroughs in identifying probiotic strains, and these two therapies are not mutually exclusive.

Including nonbacterial microbes in studies of FMT can help us identify favourable and unfavourable taxa to transplant from donor to recipient. In **Chapter 2**, I identified eukaryotes in recipients of FMT to show that fungi and protozoa were present in these individuals before and after receiving FMT. While the sample size was small and it was difficult to discern

species-level trends, a main takeaway from this analysis was that some eukaryotes appeared to persist even after FMT. This finding warrants further investigation, to identify which taxa are recalcitrant to FMT, and to determine whether these resistant taxa assist or hinder transplant success. Another important observation was that *Blastocystis* sp. subtype 2 were present in an individual who achieved remission both pre- and post-FMT, as well as in a donor whose patient achieved remission. This finding, along with the other findings in **Chapter 2**, contributes to the reconsideration of *Blastocystis* as pathogens in the disease. A similar study has shown that *Blastocystis* were transferred to recipients of FMT for *Clostridioides difficile* with no adverse effects (Terveer et al. 2019). Nevertheless, strain-level analyses would be most beneficial to determine whether only specific *Blastocystis* strains are not pathogenic.

Along with eukaryotes, the virome could be a 'missing link' to refine FMT (Plichta et al. 2019; Ungaro et al. 2019). Viruses are generally regarded as unfavourable microorganisms for their associations with host diseases, but most of the viruses found in the human gut are bacteriophages that contribute to the distribution of intestinal bacteria and archaea (Carding, Davis & Hoyles 2017). Thus, certain bacteriophages may be beneficial to control pathogenic bacteria in individuals with IBD. I touched on this topic in **Chapter 1** to raise awareness of how future FMT studies should investigate the viral composition of donor and patient samples. For brevity, the scope of this chapter was narrowed to discussing findings of the virome in IBD. However, insights can be gained from findings of FMT for *C. difficile* infections. In particular, one study filtered out small particles and bacteria from donor stool and used the remaining filtrate containing molecules such as DNA, metabolites, proteins, and viruses, rather than whole stool, to treat *C. difficile* infections in five patients (Ott et al. 2017). Donor-derived phages, particularly *Lactococcus* phages, were present in one of the recipients up to six weeks post-transfer of the filtrate. All five were cured as a result of receiving the filtrate and, while the sample size was small, suggests that bacteriophages could have been involved in resolving inflammation. While this study looked at patients of *C. difficile* infections, it would be beneficial to know if a similar phenomenon occurs in IBD patients. If phages alone can have therapeutic effects, it will be beneficial to select donors who have an abundance of favourable phages. This could act as a quasi-phage therapy until further advancements to treatments are made. Identifying the groups of phages that lead to better responses can also tell us more about which bacteria are the cause of inflammation in patients. While we are still a way off from this, the initial steps of characterising the

distribution of viruses in IBD and healthy individuals have begun (Norman et al. 2015; Pérez-Brocal et al. 2015; Zuo et al. 2019).

Longitudinal microbiome sampling is necessary in clinical trials for IBD

Like the above study by Ott et al. (2017), several longitudinal studies of FMT have illustrated how a repeated measures study design is essential to elucidate the nonbacterial microbiome's response to therapy in IBD (Rossen et al. 2015; Moayyedi et al. 2017; Paramsothy et al. 2019; Chen, HT et al. 2020; Li et al. 2020). Experimental study designs for clinical trials can vary in sampling frequency and duration, depending on the study questions. Weekly or biweekly sampling is common to track general compositional trends in response to treatment. High resolution sampling has also been used to track the success of donor FMT engraftment in individuals with CD, including tracking donor-to-strain reception (Kong et al. 2020; Aggarwala et al. 2021). Longer term sampling can be used to track the duration of efficacy of treatment, for example how long it takes for recurrence of symptoms following FMT. There are also irregularly spaced sampling strategies to conserve resources while capturing the expected spike in variation following treatment (Silverman et al. 2018). This is exemplified in **Chapter 3**, where the first week post-FMT was sampled by the participants with daily dry stool swabs then twice weekly for the next seven weeks. The purpose of this study design was primarily to define two features of FMT: (1) to observe the foundational bacteria in the first week following FMT that contributed to remission observed in week 4, and (2) to observe the broader compositional and diversity changes that differed between the responder and non-responder groups. However, obtaining these results proved more difficult than expected.

The swab dataset in **Chapter 3** suffered from significant sample missingness that impacted the ability of available statistical methods to produce informative results. I explored several different regression models that were designed for microbiome data and could be used for repeated measures study designs (i.e., the subjects could be included as random effects). All of these either failed to converge or were not permissive due to study design limitations (e.g., ZIBR which did not allow missing samples) (Chen, EZ & Li 2016). This problem was additionally compounded by the overall lack of appropriate regression models for longitudinal microbiome data. For example, most available models do not have an option for both random slopes and intercepts (see **Chapter 4**), even though this option is expected to

reflect the dynamics of the post-FMT microbiome most accurately. The one model that allowed both random slopes and intercepts (ANCOM-II from Kaul et al. 2017) was computationally expensive and did not finish running, even on a high-performance computing cluster with over a TB of memory.

It was from these limitations that **Chapter 4** was conceived. This chapter was designed to encourage researchers who are planning to conduct longitudinal microbiome studies to consider the many technical and statistical aspects of the study design. The chapter was written in a manner that included accessible information on biostatistics for researchers and clinicians who may be unfamiliar with some of these technical aspects but could greatly benefit from awareness of them. Once published, this chapter is expected to contribute to the improvement in quality and standards for longitudinal microbiome studies, while also stimulating more research into improved statistical methods for microbiome research.

In addition to the methodological recommendations in **Chapter 4**, there are also recommendations for longitudinal studies of FMT based off previous studies in the literature. Several placebo-controlled trials have grouped recipients into treatment type (e.g., placebo and donor FMT) for the microbiome analysis portion of the study (Paramsothy et al. 2017; Chen, HT et al. 2020; Li et al. 2020). This grouping was likely to establish the efficacy of donor FMT through observing whether the recipient microbiome more closely resembled the donor. Future publications of FMT need to group recipients by their response (remission versus no remission) as well to investigate whether there are microbiome differences between the responder and non-responder groups. This grouping can be performed regardless of FMT type, as factors that lead to remission in the placebo group are still important to consider and will improve our ability to isolate therapeutic microbes. This was the strategy that was implemented in **Chapter 3**, because I was interested in the microbial instigators of remission, regardless of the treatment type. I also took this approach because it increased the statistical power over splitting the already-small sample size into four sub-groups (i.e., donor FMT and remission, donor FMT and no remission, autologous FMT and remission, and autologous FMT, and no remission). Clinical trials of new therapies often have small sample sizes, so strategies such as this should be taken to increase statistical power where possible. The ability to compare the microbiome samples of recipients before and after receiving FMT allows us to answer important question about how the recipient microbiome changes as a result of FMT. Yet, less obvious but meaningful questions can also be addressed. Do the microbiomes of patients who achieve remission look different to non-responders even before receiving FMT?

Answering this question can help us understand if certain patients are more likely to be responsive to FMT. Other questions relate to the therapeutic potential of specific microbial strains. Do non-responders possess certain microbes that are resistant to FMT and prevent its mechanism of action? Conversely, which microbes can successfully engraft and restore the microbiome?

Advancing IBD research through an interdisciplinary approach to microbiome science

Synthesising knowledge from several fields through collaboration is one of the best ways to advance a research area (Smye & Frangi 2021). The field of IBD research and, importantly, people with IBD benefit the most when experts from several disciplines come together to conduct thorough and well-designed studies (Wilkinson et al. 2021). The work presented in this thesis is a result of fruitful collaborations with professionals from different research areas including gastroenterology and biostatistics, and I will herein discuss the merits of this experience, including some examples of how these complementary fields can learn from each other to support IBD research.

Collaborating with clinicians

Bridging the gap between basic and clinical (applied) microbiome research was a major priority of this thesis. Interpreting basic microbiological science through a clinical lens improves the likelihood of the research being incorporated into future translational research (Shankar 2017). There are several examples of this approach throughout the thesis. Firstly, **Chapter 1** was written to be accessible to clinicians and was published in a clinical journal (*Inflammatory Bowel Diseases*). These insights of the nonbacterial microbiome now have a better chance to be implemented in a clinical context. **Chapter 3** was developed in close collaboration with a team of gastroenterologists who were essential to the clinical interpretations of the results, while they also benefited from a technical investigation of the data. Lastly, **Chapter 4** presented information from basic microbiology and biostatistics to be applied in health and medical research of the microbiome and is an important resource to help accelerate discoveries of the microbiome in IBD.

From my experience in producing this thesis, the practical experience of clinicians greatly improved the interpretation of the data and results in a physiological and patient-centred context. In IBD, being familiar with aspects such as patient symptoms, the effects of medications, and lifestyle factors common in patients can supplement microbiome insights. An example of these insights was apparent when I was interpreting the reasons for sample missingness in **Chapter 3**, particularly in samples that yielded very low amounts of microbial DNA. When discussing with the clinicians of the project, I was reminded that the patients recruited for this study had active UC, which meant that many of them had frequent loose, bloody stool. This stool quality could explain why sampling on some days was unsuccessful, as the presence of blood in stool could overwhelm any microbial DNA. Variable stool quality may also be one of the reasons for the higher temporal variability generally observed in IBD patients (Halfvarson et al. 2017), as stool quality is linked to gut microbiome composition in healthy individuals (Vandeputte et al. 2016). This temporal variability would also compound the difficulty of convergence of regression models, particularly because many of the models did not offer options for random slopes (see **Chapter 4**). To improve longitudinal sampling, studies should attach a questionnaire with every sample collection to record critical information that could affect sample quality such as the Bristol Stool Scale to classify the quality of stool, and to record whether there is blood present (Vandeputte et al. 2016).

It is also important to incorporate clinical knowledge of the physiology and behaviour of flares in individuals with IBD. For example, in **Chapter 2**, the primary study of the IBD cohort did not record whether each individual was currently in an active or inactive state of disease at the time of sample collection (Imhann et al. 2019). This lack of metadata impacted my ability to conclude on the association between flare and the nonbacterial microbiome, and how the results of this study relate to other studies. While one might think that the consequences of flare on the microbiome would be substantial due to the elevated immune activity and changes to mucosa, recent longitudinal studies how not found a significant intra-individual difference as a consequence of flare disease activity in either CD (Galazzo et al. 2019; Strömbeck et al. 2020) or UC (Öhman et al. 2021). However, this sample size is still small to be conclusive and these studies also did not characterise the nonbacterial microbiome in response to flare. Additionally, an increase in disease activity affects patients in other ways such as diet or medication changes to manage flare, and increased levels of stress/anxiety/depression as a consequence of a flare (Carter, Lobo & Travis 2004; Mikocka-Walus et al. 2007; Limdi 2018). Disease activity should therefore be an essential criterion

when recording participant information. Clinicians will be intimately familiar with these associated changes and this knowledge can help inform how covariates may interact with each other in a study.

Health and medical science can also benefit from technical and theoretical insights from basic microbiology. The extensive impact of technical effects, including contamination, on the microbiome are crucial to acknowledge and account for in clinical studies (Allaband et al. 2019). Many studies of the microbiome do not report controls such as negative and positive extraction and sequencing controls, such as all of the studies analysed in **Chapter 2** (Schirmer et al. 2016; Imhann et al. 2019; Kong et al. 2020). This was an important aspect to highlight in **Chapter 4**, but these requirements go beyond longitudinal sampling and extend to all study designs. There remains an implicit assumption that higher biomass sampling, such as gut microbiome sampling, does not require negative controls because the biomass of the sample will overwhelm any introduced DNA from external sources (Karstens et al. 2019). However, this simplistic assumption may bias results for several reasons. Firstly, this assumption does not account for cross-contamination that can occur between samples in sequencing wells that can in turn alter the diversity of neighbouring samples (Minich et al. 2019). Secondly, higher biomass studies are not exclusively interested in the most abundant microbes, and this thesis included several examples of gut microbiome studies focussed on rarer microbes. For example, colonic mucosal taxa are found in much lower abundance than luminal taxa, but are of interest when investigating a sample type that may be presumed to capture more mucosal taxa such as dry or rectal swabs (Araújo-Pérez et al. 2012; Budding et al. 2014; Reyman et al. 2019). More broadly, gut microbiome studies of IBD with whole stool may still be interested in identifying mucosal taxa because they are situated at the interface of inflammation and interact directly with immune cells (Priya et al. 2022). Given the higher oxygen availability at the mucosa, mucosal species are generally anaerobes or facultative anaerobes, and several may be able to colonise the mucosa, rectum, and anus, such as *Staphylococcus* (Priya et al. 2022) and *Corynebacterium* identified in **Chapter 3** and by Biehl et al. (2019). Therefore, it would be ideal to sample the skin of participants and technicians to account for the possibility of contamination of swabs or biopsies by skin taxa.

Another example of rare microbes being of interest in IBD is through studying nonbacterial microbes. In **Chapter 2**, some of the identified fungi are also able to colonise the skin (Seed 2014), and I was unable to discount the possibility of their presence being caused by contamination. None of the primary publications included negative controls in their published

datasets, which is a major issue when trying to make use of published data. Fortunately, more publications are adopting standards for research articles such as reporting controls thanks to published recommendations (Eisenhofer et al. 2019; Hornung, Zwittink & Kuijper 2019). Finally, the term 'high biomass' is not a static definition, and sample types can range in their abilities to capture biospecimen. For example, mucosal biopsies are higher biomass when compared to skin swabs, but much lower when compared to whole stool. Contamination may differently affect sample types and should always be a precautionary step during laboratory procedures.

Insights from biostatisticians

Robust scientific findings are built upon a foundation of well-designed studies, and these require a considered approach to statistical analysis. Due to the immense scale of microbiome data, the field is reliant on complex statistical models and methods to clean data and interpret results (Mallick et al. 2021). The mathematisation of fields, such as microbiome science, is increasingly common as the necessity and ability to obtain larger datasets continues to rise, and it is beneficial for all non-statistician researchers and clinicians to endeavour to understand and implement recommendations from statisticians. This approach only succeeds when thorough discussions are held between both types of experts. Statisticians need to understand the biological dynamics, biases, and assumptions that pertain to microbiome data, and microbiome scientists and clinicians need to understand the assumptions and limitations of statistical models for study design and analysis.

Sitting at the junction between microbiology, gastroenterology, and biostatistics can be challenging, but it is worth persevering to stimulate innovations and improve outcomes for patients (Kivits, Ricci & Minary 2019). **Chapter 4** was written while working closely with an epidemiological statistician to bring insights from public health statistics to the forefront of clinical microbiome research. The chapter emphasises statistical considerations that significantly impact longitudinal microbiome studies. One example is how I emphasised the impact of missing data, which may not be fully appreciated by clinicians who are accustomed to imputing single data values such as participant anthropometrics and demographics, and microbiome specialists who have not used regression methods for longitudinal microbiome data.

Another example is how time-varying covariates are important to consider in longitudinal research and incorporate into the available regression methods. This point exemplifies the advantages of working directly with statisticians, as they can examine which covariates to collect and how to collect them, and how these will impact the models used. This is essential as we strive to characterise more of the sources of variation in the microbiome. For example, Clooney et al. (2021) found that even after recording seemingly extensive metadata, 90% of the microbiome variance of individuals with IBD was due to stochasticity and unmeasured covariates. This unaccounted variance will ultimately affect the reliability of conclusions, and improved statistical methods are warranted to understand the temporal dynamics of the microbiome and covariate interactions. In summary, microbiome projects should include at least one statistician as a team member or consultant to mitigate the potential for sources of bias and to identify methodological constraints which can ultimately be framed as opportunities for future advancements.

Conclusion

The gut microbiome is a recently tapped resource to further our understanding of multifactorial immune-mediated diseases such as IBD. This thesis presents a multifaceted approach to studying the microbiome and how it can be used to assess and treat IBD. By spotlighting nonbacterial intestinal microbes, the work in this thesis gives a comprehensive description of the microbiome in IBD that has been neglected in previous studies. It is now clear that nonbacterial microbes hold insights into how IBD arises and how it can be treated, and the findings in this thesis may lead to future research on these topics. This thesis also provides a framework to improve multiple aspects of microbiome study design, including benchmarking microbiome samples for clinical trials and designing longitudinal microbiome studies. Through this interdisciplinary work, the field of clinical microbiome research will be aware of crucial technical and statistical considerations to properly conduct studies. Ultimately, the work contained in this thesis may inspire translational research leading to key medical advancements for individuals with IBD.

References

Adiliaghdam, F, Amatullah, H, Digumarthi, S, Saunders, TL, Rahman, RU, Wong, LP, Sadreyev, R, Droit, L, Paquette, J, Goyette, P, Rioux, JD, Hodin, R, Mihindukulasuriya, KA, Handley, SA & Jeffrey, KL 2022, 'Human enteric viruses autonomously shape inflammatory bowel disease phenotype through divergent innate immunomodulation', *Sci Immunol*, vol. 7, no. 70, Apr 8, p. eabn6660.

Aggarwala, V, Mogno, I, Li, Z, Yang, C, Britton, GJ, Chen-Liaw, A, Mitcham, J, Bongers, G, Gevers, D, Clemente, JC, Colombel, J-F, Grinspan, A & Faith, J 2021, 'Precise quantification of bacterial strains after fecal microbiota transplantation delineates long-term engraftment and explains outcomes', *Nat Microbiol*, vol. 6, no. 10, 2021/10/01, pp. 1309-1318.

Agrawal, M, Allin, KH, Petralia, F, Colombel, J-F & Jess, T 2022, 'Multiomics to elucidate inflammatory bowel disease risk factors and pathways', *Nature Reviews Gastroenterology & Hepatology*, vol. 19, no. 6, 2022/06/01, pp. 399-409.

Alatab, S, Sepanlou, SG, Ikuta, K, Vahedi, H, Bisignano, C, Safiri, S, Sadeghi, A, Nixon, MR, Abdoli, A, Abolhassani, H, Alipour, V, Almadi, MAH, Almasi-Hashiani, A, Anushiravani, A, Arabloo, J, Atique, S, Awasthi, A, Badawi, A, Baig, AAA, Bhala, N, Bijani, A, Biondi, A, Borzì, AM, Burke, KE, Carvalho, F, Daryani, A, Dubey, M, Eftekhari, A, Fernandes, E, Fernandes, JC, Fischer, F, Haj-Mirzaian, A, Haj-Mirzaian, A, Hasanzadeh, A, Hashemian, M, Hay, SI, Hoang, CL, Househ, M, Ilesanmi, OS, Jafari Balalami, N, James, SL, Kengne, AP, Malekzadeh, MM, Merat, S, Meretoja, TJ, Mestrovic, T, Mirrakhimov, EM, Mirzaei, H, Mohammad, KA, Mokdad, AH, Monasta, L, Negoi, I, Nguyen, TH, Nguyen, CT, Pourshams, A, Poustchi, H, Rabiee, M, Rabiee, N, Ramezanzadeh, K, Rawaf, DL, Rawaf, S, Rezaei, N, Robinson, SR, Ronfani, L, Saxena, S, Sepehrimanesh, M, Shaikh, MA, Sharafi, Z, Sharif, M, Siabani, S, Sima, AR, Singh, JA, Soheili, A, Sotoudehmanesh, R, Suleria, HAR, Tesfay, BE, Tran, B, Tsoi, D, Vacante, M, Wondmieneh, AB, Zarghi, A, Zhang, Z-J, Dirac, M, Malekzadeh, R & Naghavi, M 2020, 'The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017', *The Lancet Gastroenterology & Hepatology*, vol. 5, no. 1, pp. 17-30.

Aldars-García, L, Chaparro, M & Gisbert, JP 2021, 'Systematic Review: The Gut Microbiome and Its Potential Clinical Application in Inflammatory Bowel Disease', *Microorganisms*, vol. 9, no. 5, Apr 30.

Allaband, C, McDonald, D, Vazquez-Baeza, Y, Minich, JJ, Tripathi, A, Brenner, DA, Loomba, R, Smarr, L, Sandborn, WJ, Schnabl, B, Dorrestein, P, Zarrinpar, A & Knight, R 2019, 'Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians', *Clin Gastroenterol Hepatol*, vol. 17, no. 2, Jan, pp. 218-230.

Araújo-Pérez, F, McCoy, AN, Okechukwu, C, Carroll, IM, Smith, KM, Jeremiah, K, Sandler, RS, Asher, GN & Keku, TO 2012, 'Differences in microbial signatures between rectal mucosal biopsies and rectal swabs', *Gut Microbes*, vol. 3, no. 6, Nov-Dec, pp. 530-535.

Bach, JF 2018, 'The hygiene hypothesis in autoimmunity: the role of pathogens and commensals', *Nat Rev Immunol*, vol. 18, no. 2, Feb, pp. 105-120.

Barratt, JLN, Harkness, J, Marriott, D, Ellis, JT & Stark, D 2011, 'A review of Dientamoeba fragilis carriage in humans: Several reasons why this organism should be considered in the diagnosis of gastrointestinal illness', *Gut Microbes*, vol. 2, no. 1, 2011/01/01, pp. 3-12.

Barreiro-de Acosta, M, Alvarez Castro, A, Souto, R, Iglesias, M, Lorenzo, A & Dominguez-Muñoz, JE 2011, 'Emigration to western industrialized countries: A risk factor for developing inflammatory bowel disease', *Journal of Crohn's and Colitis*, vol. 5, no. 6, pp. 566-569.

Biehl, LM, Garzetti, D, Farowski, F, Ring, D, Koeppel, MB, Rohde, H, Schafhausen, P, Stecher, B & Vehreschild, MJGT 2019, 'Usability of rectal swabs for microbiome sampling in a cohort study of hematological and oncological patients', *PLoS One*, vol. 14, no. 4, pp. e0215428-e0215428.

Bjarnason, I 2017, 'The Use of Fecal Calprotectin in Inflammatory Bowel Disease', *Gastroenterol Hepatol (N Y)*, vol. 13, no. 1, pp. 53-56.

Bokulich, NA, Maldonado, J, Kang, DW, Krajmalnik-Brown, R & Caporaso, JG 2019, 'Rapidly Processed Stool Swabs Approximate Stool Microbiota Profiles', *mSphere*, vol. 4, no. 2, Apr 10.

Brands, MR, Van de Vijver, E, Haisma, SM, Heida, A & van Rheenen, PF 2019, 'No association between abdominal pain and *Dientamoeba*; in Dutch and Belgian children', *Archives of Disease in Childhood*, vol. 104, no. 7, p. 686.

Brüssow, H 2016, 'How stable is the human gut microbiota? And why this question matters', *Environ Microbiol*, vol. 18, no. 9, Sep, pp. 2779-2783.

Budding, AE, Grasman, ME, Eck, A, Bogaards, JA, Vandenbroucke-Grauls, CMJE, van Bodegraven, AA & Savelkoul, PHM 2014, 'Rectal swabs for analysis of the intestinal microbiota', *PLoS One*, vol. 9, no. 7, pp. e101344-e101344.

Burchill, E, Lymberopoulos, E, Menozzi, E, Budhdeo, S, McIlroy, JR, Macnaughtan, J & Sharma, N 2021, 'The Unique Impact of COVID-19 on Human Gut Microbiome Research', *Frontiers in Medicine*, vol. 8, 2021-March-16.

Cai, Z, Wang, S & Li, J 2021, 'Treatment of Inflammatory Bowel Disease: A Comprehensive Review', *Frontiers in Medicine*, vol. 8, 2021-December-20.

Caldeira, LdF, Borba, HH, Tonin, FS, Wiens, A, Fernandez-Llimos, F & Pontarolo, R 2020, 'Fecal microbiota transplantation in inflammatory bowel disease patients: A systematic review and meta-analysis', *PLoS One*, vol. 15, no. 9, pp. e0238910-e0238910.

Carding, SR, Davis, N & Hoyles, L 2017, 'Review article: the human intestinal virome in health and disease', *Alimentary Pharmacology & Therapeutics*, vol. 46, no. 9, Nov, pp. 800-815.

Carter, MJ, Lobo, AJ & Travis, SPL 2004, 'Guidelines for the management of inflammatory bowel disease in adults', *Gut*, vol. 53, no. suppl 5, pp. v1-v16.

Chabé, M, Lokmer, A & Ségurel, L 2017, 'Gut Protozoa: Friends or Foes of the Human Gut Microbiota?', *Trends Parasitol*, vol. 33, no. 12, 2017/12/01/, pp. 925-934.

Chen, EZ & Li, H 2016, 'A two-part mixed-effects model for analyzing longitudinal microbiome compositional data', *Bioinformatics*, vol. 32, no. 17, Sep 1, pp. 2611-2617.

Chen, HT, Huang, HL, Xu, HM, Luo, QL, He, J, Li, YQ, Zhou, YL, Nie, YQ & Zhou, YJ 2020, 'Fecal microbiota transplantation ameliorates active ulcerative colitis', *Exp Ther Med*, vol. 19, no. 4, Apr, pp. 2650-2660.

Clooney AG, Sutton, TDS, Shkoporov, AN, Holohan, RK, Daly, KM, O'Regan, O, Ryan, FJ, Draper, LA, Plevy, SE, Ross, RP & Hill, C 2019, 'Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease', *Cell Host Microbe*, vol. 26, no. 6, Dec 11, pp. 764-778.e765.

Clooney, AG, Eckenberger, J, Laserna-Mendieta, E, Sexton, KA, Bernstein, MT, Vagianos, K, Sargent, M, Ryan, FJ, Moran, C, Sheehan, D, Sleator, RD, Targownik, LE, Bernstein, CN, Shanahan, F & Claesson, MJ 2021, 'Ranking microbiome variance in inflammatory bowel disease: a large longitudinal intercontinental study', *Gut*, vol. 70, no. 3, p. 499.

Costello, EK, Lauber, CL, Hamady, M, Fierer, N, Gordon, JI & Knight, R 2009, 'Bacterial Community Variation in Human Body Habitats Across Space and Time', *Science*, vol. 326, no. 5960, pp. 1694-1697.

Costello, SP, Soo, W, Bryant, RV, Jairath, V, Hart, AL & Andrews, JM 2017, 'Systematic review with meta-analysis: faecal microbiota transplantation for the induction of remission for active ulcerative colitis', *Aliment Pharmacol Ther*, vol. 46, no. 3, Aug, pp. 213-224.

Coyle, CM, Varughese, J, Weiss, LM & Tanowitz, HB 2011, 'Blastocystis: To Treat or Not to Treat…', *Clinical Infectious Diseases*, vol. 54, no. 1, pp. 105-110.

Debelius, JW, Vázquez-Baeza, Y, McDonald, D, Xu, Z, Wolfe, E & Knight, R 2016a, 'Turning Participatory Microbiome Research into Usable Data: Lessons from the American Gut Project', *Journal of microbiology & biology education*, vol. 17, no. 1, pp. 46-50.

Deng, L, Wojciech, L, Gascoigne, NRJ, Peng, G & Tan, KSW 2021, 'New insights into the interactions between Blastocystis, the gut microbiota, and host immunity', *PLoS pathogens*, vol. 17, no. 2, p. e1009253.

Dullaert-de Boer, M, Schuurs, TA, Vermeer, M, Ruijs, G, van der Zanden, AGM, Weel, JF & van Coppenraet, L 2019, 'Distribution and relevance of Dientamoeba fragilis and Blastocystis species in gastroenteritis: results from a case-control study', *European Journal of Clinical Microbiology & Infectious Diseases*, p. 7.

Dutilh, BE, Reyes, A, Hall, RJ & Whiteson, KL 2017, 'Editorial: Virus Discovery by Metagenomics: The (Im)possibilities', *Front Microbiol*, vol. 8, p. 1710.

Ege, MJ, Mayer, M, Normand, AC, Genuneit, J, Cookson, WO, Braun-Fahrländer, C, Heederik, D, Piarroux, R & von Mutius, E 2011, 'Exposure to environmental microorganisms and childhood asthma', *N Engl J Med*, vol. 364, no. 8, Feb 24, pp. 701-709.

Eisenhofer, R, Minich, JJ, Marotz, C, Cooper, A, Knight, R & Weyrich, LS 2019, 'Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations', *Trends Microbiol*, vol. 27, no. 2, 2019/02/01/, pp. 105-117.

El Safadi, D, Gaayeb, L, Meloni, D, Cian, A, Poirier, P, Wawrzyniak, I, Delbac, F, Dabboussi, F, Delhaes, L, Seck, M, Hamze, M, Riveau, G & Viscogliosi, E 2014, 'Children of Senegal River Basin show the highest prevalence of Blastocystis sp. ever observed worldwide', *BMC infectious diseases*, vol. 14, pp. 164-164.

Galazzo, G, Tedjo, DI, Wintjens, DSJ, Savelkoul, PHM, Masclee, AAM, Bodelier, AGL, Pierik, MJ, Jonkers, D & Penders, J 2019, 'Faecal Microbiota Dynamics and their Relation to Disease Course in Crohn's Disease', *Journal of Crohn's & colitis*, vol. 13, no. 10, Sep 27, pp. 1273-1282.

Garcia, LS 2016, 'Dientamoeba fragilis, One of the Neglected Intestinal Protozoa', *Journal of Clinical Microbiology*, vol. 54, no. 9, p. 2243.

GEM Project 2020, *Crohn's and Colitis Canada Study on the Genetic, Environmental and Microbial Interactions that cause IBD*, The Crohn's and Colitis Canada Genetic, Environmental, Microbial (GEM) Project, Crohn's and Colitis Canada, <https://crohnsandcolitis.ca/gem>.

GESA 2018, *Australian Guidelines for General Practitioners and Physicians: Inflammatory Bowel Disease*, The Gastroenterological Society of Australia (GESA), Melbourne, Victoria, Australia.

Gizaw, Z, Adane, T, Azanaw, J, Addisu, A & Haile, D 2018, 'Childhood intestinal parasitic infection and sanitation predictors in rural Dembiya, northwest Ethiopia', *Environmental health and preventive medicine*, vol. 23, no. 1, pp. 26-26.

Glassner, KL, Abraham, BP & Quigley, EMM 2020, 'The microbiome and inflammatory bowel disease', *Journal of Allergy and Clinical Immunology*, vol. 145, no. 1, pp. 16-27.

Graham, DB, Subramanian, S & Xavier, RJ 2019, 'Therapeutic Opportunities in Inflammatory Bowel Disease: Mechanistic Dissection of Host-Microbiome Relationships', *Cell.*, vol. 178, no. 5, pp. 1041-1056.

Gregory, AC, Zablocki, O, Zayed, AA, Howell, A, Bolduc, B & Sullivan, MB 2020, 'The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut', *Cell Host Microbe*, 2020/08/24/.

Gulliver, EL, Young, RB, Chonwerawong, M, D'Adamo, GL, Thomason, T, Widdop, JT, Rutten, EL, Rossetto Marcelino, V, Bryant, RV, Costello, SP, O'Brien, CL, Hold, GL, Giles, EM & Forster, SC 2022, 'Review article: the future of microbiome-based therapeutics', *Aliment Pharmacol Ther*, May 24.

Guo, J, Bolduc, B, Zayed, AA, Varsani, A, Dominguez-Huerta, G, Delmont, TO, Pratama, AA, Gazitúa, MC, Vik, D, Sullivan, MB & Roux, S 2021, 'VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses', *Microbiome*, vol. 9, no. 1, 2021/02/01, p. 37.

Halfvarson, J, Brislawn, CJ, Lamendella, R, Vázquez-Baeza, Y, Walters, WA, Bramer, LM, D'Amato, M, Bonfiglio, F, McDonald, D, Gonzalez, A, McClure, EE, Dunklebarger, MF, Knight, R & Jansson, JK 2017, 'Dynamics of the human gut microbiome in inflammatory bowel disease', *Nat Microbiol*, vol. 2, Feb 13, p. 17004.

Ho, SM, Lewis, JD, Mayer, EA, Plevy, SE, Chuang, E, Rappaport, SM, Croitoru, K, Korzenik, JR, Krischer, J, Hyams, JS, Judson, R, Kellis, M, Jerrett, M, Miller, GW, Grant, ML, Shtraizent, N, Honig, G, Hurtado-Lorenzo, A & Wu, GD 2019, 'Challenges in IBD Research: Environmental Triggers', *Inflamm Bowel Dis*, vol. 25, no. Suppl 2, May 16, pp. S13-s23.

Holtman, GA, Kranenberg, JJ, Blanker, MH, Ott, A, Lisman-van Leeuwen, Y & Berger, MY 2016, 'Dientamoeba fragilis colonization is not associated with gastrointestinal symptoms in children at primary care level', *Family Practice*, vol. 34, no. 1, pp. 25-29.

Hooks, KB & O'Malley, MA 2017, 'Dysbiosis and Its Discontents', *MBio*, vol. 8, no. 5, Oct 10.

Hornung, BVH, Zwittink, RD & Kuijper, EJ 2019, 'Issues and current standards of controls in microbiome research', *FEMS Microbiology Ecology*, vol. 95, no. 5.

Houshyar, Y, Massimino, L, Lamparelli, LA, Danese, S & Ungaro, F 2021, 'Going Beyond Bacteria: Uncovering the Role of Archaeome and Mycobiome in Inflammatory Bowel Disease', *Frontiers in Physiology*, vol. 12, 2021-December-06.

Imhann, F, Van der Velde, KJ, Barbieri, R, Alberts, R, Voskuil, MD, Vich Vila, A, Collij, V, Spekhorst, LM, Van der Sloot, KWJ, Peters, V, Van Dullemen, HM, Visschedijk, MC, Festen, EAM, Swertz, MA, Dijkstra, G & Weersma, RK 2019, 'The 1000IBD project: multi-omics data of 1000 inflammatory bowel disease patients; data release 1', *BMC Gastroenterol*, vol. 19, no. 1, 2019/01/08, p. 5.

Jain, U, Ver Heul, AM, Xiong, S, Gregory, MH, Demers, EG, Kern, JT, Lai, CW, Muegge, BD, Barisas, DAG, Leal-Ekman, JS, Deepak, P, Ciorba, MA, Liu, TC, Hogan, DA, Debbas, P, Braun, J, McGovern, DPB, Underhill, DM & Stappenbeck, TS 2021, 'Debaryomyces is enriched in Crohn's disease intestinal tissue and impairs healing in mice', *Science*, vol. 371, no. 6534, Mar 12, pp. 1154-1159.

Jokelainen, P, Hebbelstrup Jensen, B, Andreassen, BU, Petersen, AM, Röser, D, Krogfelt, KA, Nielsen, HV & Stensvold, CR 2017, 'Dientamoeba fragilis, a Commensal in Children in Danish Day Care Centers', *Journal of Clinical Microbiology*, vol. 55, no. 6, p. 1707.

Jonkers, D & Stockbrügger, R 2003, 'Probiotics and inflammatory bowel disease', *Journal of the Royal Society of Medicine*, vol. 96, no. 4, pp. 167-171.

Jun, X, Ning, C, Yang, S, Zhe, W, Na, W, Yifan, Z, Xinhua, R & Yulan, L 2019, 'Alteration of Fungal Microbiota After 5-ASA Treatment in UC Patients', *Inflamm Bowel Dis*, vol. 26, no. 3, pp. 380-390.

Jurasz, H, Pawłowski, T & Perlejewski, K 2021, 'Contamination Issue in Viral Metagenomics: Problems, Solutions, and Clinical Perspectives', *Front Microbiol*, vol. 12, 2021-October-20.

Karstens, L, Asquith, M, Davin, S, Fair, D, Gregory, WT, Wolfe, AJ, Braun, J, McWeeney, S & Gilbert, JA 2019, 'Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments', *mSystems*, vol. 4, no. 4, pp. e00290-00219.

Kaul, A, Mandal, S, Davidov, O & Peddada, SD 2017, 'Analysis of Microbiome Data in the Presence of Excess Zeros', *Front Microbiol*, vol. 8, pp. 2114-2114.

Kellermayer, R & Zilbauer, M 2020, 'The Gut Microbiome and the Triple Environmental Hit Concept of Inflammatory Bowel Disease Pathogenesis', *J Pediatr Gastroenterol Nutr*, vol. 71, no. 5, Nov, pp. 589-595.

Khan Mirzaei, M, Xue, J, Costa, R, Ru, J, Schulz, S, Taranu, ZE & Deng, L 2021, 'Challenges of Studying the Human Virome – Relevant Emerging Technologies', *Trends Microbiol*, vol. 29, no. 2, pp. 171-181.

Kieft, K, Zhou, Z & Anantharaman, K 2020, 'VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences', *Microbiome*, vol. 8, no. 1, 2020/06/10, p. 90.

Kivits, J, Ricci, L & Minary, L 2019, 'Interdisciplinary research in public health: the 'why' and the 'how'', *Journal of Epidemiology and Community Health*, vol. 73, no. 12, pp. 1061-1062.

Kleiner, M, Hooper, LV & Duerkop, BA 2015, 'Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes', *Bmc Genomics*, vol. 16, no. 1, 2015/01/22, p. 7.

Knox, NC, Forbes, JD, Peterson, C-L, Van Domselaar, G & Bernstein, CN 2019, 'The Gut Microbiome in Inflammatory Bowel Disease: Lessons Learned From Other Immune-Mediated Inflammatory Diseases', *American Journal of Gastroenterology*, vol. 114, no. 7, Jul, pp. 1051-1070.

Kondrashova, A, Reunanen, A, Romanov, A, Karvonen, A, Viskari, H, Vesikari, T, Ilonen, J, Knip, M & Hyöty, H 2005, 'A six-fold gradient in the incidence of type 1 diabetes at the eastern border of Finland', *Ann Med*, vol. 37, no. 1, pp. 67-72.

Kong, L, Lloyd-Price, J, Vatanen, T, Seksik, P, Beaugerie, L, Simon, T, Vlamakis, H, Sokol, H & Xavier, RJ 2020, 'Linking Strain Engraftment in Fecal Microbiota Transplantation With Maintenance of Remission in Crohn's Disease', *Gastroenterology*, vol. 159, no. 6, Dec, pp. 2193-2202.e2195.

Kronman, MP, Zaoutis, TE, Haynes, K, Feng, R & Coffin, SE 2012, 'Antibiotic exposure and IBD development among children: a population-based cohort study', *Pediatrics*, vol. 130, no. 4, Oct, pp. e794-803.

Laatikainen, T, von Hertzen, L, Koskinen, JP, Mäkelä, MJ, Jousilahti, P, Kosunen, TU, Vlasoff, T, Ahlström, M, Vartiainen, E & Haahtela, T 2011, 'Allergy gap between Finnish and Russian Karelia on increase', *Allergy*, vol. 66, no. 7, Jul, pp. 886-892.

Laforest-Lapointe, I & Arrieta, MC 2018, 'Microbial Eukaryotes: a Missing Link in Gut Microbiome Studies', *mSystems*, vol. 3, no. 2, Mar-Apr.

Lavelle, A & Sokol, H 2018, 'Gut microbiota: Beyond metagenomics, metatranscriptomics illuminates microbiome functionality in IBD', *Nat Rev Gastroenterol Hepatol*, vol. 15, no. 4, Apr, pp. 193-194.

Li, Q, Ding, X, Liu, K, Marcella, C, Liu, X, Zhang, T, Liu, Y, Li, P, Xiang, L, Cui, B, Wang, J, Bai, J & Zhang, F 2020, 'Fecal Microbiota Transplantation for Ulcerative Colitis: The Optimum Timing and Gut Microbiota as Predictors for Long-Term Clinical Outcomes', *Clin Transl Gastroenterol*, vol. 11, no. 8, Aug, p. e00224.

Limdi, JK 2018, 'Dietary practices and inflammatory bowel disease', *Indian journal of gastroenterology : official journal of the Indian Society of Gastroenterology*, vol. 37, no. 4, pp. 284-292.

Loddo, I & Romano, C 2015, 'Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis', *Frontiers in immunology*, vol. 6, pp. 551-551.

Lokmer, A, Cian, A, Froment, A, Gantois, N, Viscogliosi, E, Chabé, M & Ségurel, L 2019, 'Use of shotgun metagenomics for the identification of protozoa in the gut microbiota of healthy individuals from worldwide populations with various industrialization levels', *PLoS One*, vol. 14, no. 2, p. e0211139.

López-Serrano, P, Pérez-Calle, JL, Pérez-Fernández, MT, Fernández-Font, JM, Boixeda de Miguel, D & Fernández-Rodríguez, CM 2010, 'Environmental risk factors in inflammatory bowel diseases. Investigating the hygiene hypothesis: a Spanish case-control study', *Scand J Gastroenterol*, vol. 45, no. 12, Dec, pp. 1464-1471.

M'Koma, AE 2013, 'Inflammatory bowel disease: an expanding global health problem', *Clin Med Insights Gastroenterol*, vol. 6, pp. 33-47.

Mallick, H, Rahnavard, A, McIver, LJ, Ma, S, Zhang, Y, Nguyen, LH, Tickle, TL, Weingart, G, Ren, B, Schwager, EH, Chatterjee, S, Thompson, KN, Wilkinson, JE, Subramanian, A, Lu, Y, Waldron, L, Paulson, JN, Franzosa, EA, Bravo, HC & Huttenhower, C 2021, 'Multivariable association discovery in population-scale meta-omics studies', *PLOS Computational Biology*, vol. 17, no. 11, p. e1009442.

McDonald, D, Hyde, E, Debelius, JW, Morton, JT, Gonzalez, A, Ackermann, G, Aksenov, AA, Behsaz, B, Brennan, C, Chen, Y, Goldasich, LD, Dorrestein, PC, Dunn, RR, Fahimipour, AK, Gaffney, J, Gilbert, JA, Gogul, G, Green, JL, Hugenholtz, P, Humphrey, G, Huttenhower, C, Jackson, MA, Janssen, S, Jeste, DV, Jiang, L, Kelley, ST, Knights, D, Kosciolek, T, Ladau, J, Leach, J, Marotz, C, Meleshko, D, Melnik, AV, Metcalf, JL, Mohimani, H, Montassier, E, Navas-Molina, J, Nguyen, TT, Peddada, S, Pevzner, P, Pollard, KS, Rahnavard, G, Robbins-Pianka, A, Sangwan, N, Shorenstein, J, Smarr, L, Song, SJ, Spector, T, Swafford, AD, Thackray, VG, Thompson, LR, Tripathi, A, Vázquez-Baeza, Y, Vrbanac, A, Wischmeyer, P, Wolfe, E, Zhu, Q, Knight, R, Greene, CS, Mann, AE, Amir, A, Frazier, A, Martino, C, Lebrilla, C, Lozupone, C, Lewis, CM, Raison, C, Zhang, C, Lauber, CL, Warinner, C, Lowry, CA, Callewaert, C, Bloss, C, Willner, D, Galzerani, DD, Gonzalez, DJ, Mills, DA, Chopra, D, Gevers, D, Berg-Lyons, D, Sears, DD, Wendel, D, Lovelace, E, Pierce, E, TerAvest, E, Bolyen, E, Bushman, FD, Wu, GD, Church, GM, Saxe, G, Holscher, HD, Ugrina, I, German, JB, Caporaso, JG, Wozniak, JM, Kerr, J, Ravel, J, Lewis, JD, Suchodolski, JS, Jansson, JK, Hampton-Marcell, JT, Bobe, J, Raes, J, Chase, JH, Eisen, JA, Monk, J, Clemente, JC, Petrosino, J, Goodrich, J, Gauglitz, J, Jacobs, J, Zengler, K, Swanson, KS, Lewis, K, Mayer, K, Bittinger, K, Dillon, L, Zaramela, LS, Schriml, LM, Dominguez-Bello, MG, Jankowska, MM, Blaser, M, Pirrung, M, Minson, M, Kurisu, M, Ajami, N, Gottel, NR, Chia, N, Fierer, N, White, O, Cani, PD, Gajer, P, Strandwitz, P, Kashyap, P, Dutton, R, Park, RS, Xavier, RJ, Mills, RH, Krajmalnik-Brown, R, Ley, R, Owens, SM,

Klemmer, S, Matamoros, S, Mirarab, S, Moorman, S, Holmes, S, Schwartz, T, Eshoo-Anton, TW, Vigers, T, Pandey, V, Treuren, WV, Fang, X, Xu, ZZ, Jarmusch, A, Geier, J, Reeve, N, Silva, R, Kopylova, E, Nguyen, D, Sanders, K, Benitez, RAS, Heale, AC, Abramson, M, Waldispühl, J, Butyaev, A, Drogaris, C, Nazarova, E, Ball, M & Gunderson, B 2018, 'American Gut: an Open Platform for Citizen Science Microbiome Research', *mSystems*, vol. 3, no. 3, pp. e00031-00018.

Mikocka-Walus, AA, Turnbull, DA, Moulding, NT, Wilson, IG, Andrews, JM & Holtmann, GJ 2007, 'Controversies surrounding the comorbidity of depression and anxiety in inflammatory bowel disease patients: a literature review', *Inflamm Bowel Dis*, vol. 13, no. 2, Feb, pp. 225-234.

Minich, JJ, Sanders, JG, Amir, A, Humphrey, G, Gilbert, JA & Knight, R 2019, 'Quantifying and Understanding Well-to-Well Contamination in Microbiome Research', *mSystems*, vol. 4, no. 4, Jun 25.

Minot, S, Bryson, A, Chehoud, C, Wu, GD, Lewis, JD & Bushman, FD 2013, 'Rapid evolution of the human gut virome', *Proceedings of the National Academy of Sciences*, vol. 110, no. 30, pp. 12450-12455.

Mirsepasi-Lauridsen, HC, Vrankx, K, Engberg, J, Friis-Moller, A, Brynskov, J, Nordgaard-Lassen, I, Petersen, AM & Krogfelt, KA 2018, 'Disease-Specific Enteric Microbiome Dysbiosis in Inflammatory Bowel Disease', *Front Med (Lausanne)*, vol. 5, no. 304, 2018-November-20, p. 304.

Moayyedi, P, Yuan, Y, Baharith, H & Ford, AC 2017, 'Faecal microbiota transplantation for *Clostridium difficile*-associated diarrhoea: a systematic review of randomised controlled trials', *Med J Aust*, vol. 207, no. 4, Aug 21, pp. 166-172.

Morton, ER, Lynch, J, Froment, A, Lafosse, S, Heyer, E, Przeworski, M, Blekhman, R & Segurel, L 2015, 'Variation in Rural African Gut Microbiota Is Strongly Correlated with Colonization by Entamoeba and Subsistence', *PLoS Genet*, vol. 11, no. 11, Nov, p. e1005658.

Mukhopadhya, I, Segal, JP, Carding, SR, Hart, AL & Hold, GL 2019, 'The gut virome: the 'missing link' between gut bacteria and host immunity?', *Therapeutic Advances in Gastroenterology*, vol. 12, 2019/01/01, p. 1756284819836620.

Nagata, N, Marriott, D, Harkness, J, Ellis, JT & Stark, D 2012, 'Current treatment options for Dientamoeba fragilis infections', *International journal for parasitology. Drugs and drug resistance*, vol. 2, pp. 204-215.

Ng, SC, Shi, HY, Hamidi, N, Underwood, FE, Tang, W, Benchimol, EI, Panaccione, R, Ghosh, S, Wu, JCY, Chan, FKL, Sung, JJY & Kaplan, GG 2018, 'Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies', *Lancet*, vol. 390, no. 10114, Dec 23, pp. 2769-2778.

Nieves-Ramírez, ME, Partida-Rodríguez, O, Laforest-Lapointe, I, Reynolds, LA, Brown, EM, Valdez-Salazar, A, Morán-Silva, P, Rojas-Velázquez, L, Morien, E, Parfrey, LW, Jin, M, Walter, J, Torres, J, Arrieta, MC, Ximénez-García, C & Finlay, BB 2018, 'Asymptomatic Intestinal Colonization with Protist Blastocystis Is Strongly Associated with Distinct Microbiome Ecological Patterns', *mSystems*, vol. 3, no. 3, pp. e00007-00018.

Nishihara, Y, Ogino, H, Tanaka, M, Ihara, E, Fukaura, K, Nishioka, K, Chinen, T, Tanaka, Y, Nakayama, J, Kang, D & Ogawa, Y 2021, 'Mucosa-associated gut microbiota reflects clinical course of ulcerative colitis', *Sci Rep*, vol. 11, no. 1, 2021/07/02, p. 13743.

Norman, JM, Handley, SA, Baldridge, MT, Droit, L, Liu, CY, Keller, BC, Kambal, A, Monaco, CL, Zhao, G, Fleshner, P, Stappenbeck, TS, McGovern, DP, Keshavarzian, A, Mutlu, EA, Sauk, J, Gevers, D, Xavier, RJ, Wang, D, Parkes, M & Virgin, HW 2015, 'Disease-specific alterations in the enteric virome in inflammatory bowel disease', *Cell*, vol. 160, no. 3, Jan 29, pp. 447-460.

Nourrisson, C, Scanzi, J, Pereira, B, NkoudMongo, C, Wawrzyniak, I, Cian, A, Viscogliosi, E, Livrelli, V, Delbac, F, Dapoigny, M & Poirier, P 2014, 'Blastocystis is associated with decrease of fecal microbiota protective bacteria: comparative analysis between patients with irritable bowel syndrome and control subjects', *PLoS One*, vol. 9, no. 11, p. e111868.

Öhman, L, Lasson, A, Strömbeck, A, Isaksson, S, Hesselmar, M, Simrén, M, Strid, H & Magnusson, MK 2021, 'Fecal microbiota dynamics during disease activity and remission in newly diagnosed and established ulcerative colitis', *Sci Rep*, vol. 11, no. 1, 2021/04/21, p. 8641.

Omarova, A, Tussupova, K, Berndtsson, R, Kalishev, M & Sharapatova, K 2018, 'Protozoan Parasites in Drinking Water: A System Approach for Improved Water, Sanitation and Hygiene in Developing Countries', *International Journal of Environmental Research and Public Health*, vol. 15, no. 3, p. 495.

Ott, SJ, Waetzig, GH, Rehman, A, Moltzau-Anderson, J, Bharti, R, Grasis, JA, Cassidy, L, Tholey, A, Fickenscher, H, Seegert, D, Rosenstiel, P & Schreiber, S 2017, 'Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With *Clostridium difficile* Infection', *Gastroenterology*, vol. 152, no. 4, pp. 799-811.e797.

Paramsothy, S, Kamm, MA, Kaakoush, NO, Walsh, AJ, van den Bogaerde, J, Samuel, D, Leong, RWL, Connor, S, Ng, W, Paramsothy, R, Xuan, W, Lin, E, Mitchell, HM & Borody, TJ 2017, 'Multidonor intensive faecal microbiota transplantation for active ulcerative colitis: a randomised placebo-controlled trial', *The Lancet*, vol. 389, no. 10075, 2017/03/25/, pp. 1218-1228.

Paramsothy, S, Nielsen, S, Kamm, MA, Deshpande, NP, Faith, JJ, Clemente, JC, Paramsothy, R, Walsh, AJ, van den Bogaerde, J, Samuel, D, Leong, RWL, Connor, S, Ng, W, Lin, E, Borody, TJ, Wilkins, MR, Colombel, J-F, Mitchell, HM & Kaakoush, NO 2019, 'Specific Bacteria and Metabolites Associated With Response to Fecal Microbiota Transplantation in Patients With Ulcerative Colitis', *Gastroenterology*, vol. 156, no. 5, pp. 1440-1454.e1442.

Pérez-Brocal, V, García-López, R, Nos, P, Beltrán, B, Moret, I & Moya, A 2015, 'Metagenomic Analysis of Crohn's Disease Patients Identifies Changes in the Virome and Microbiome Related to Disease Status and Therapy, and Detects Potential Interactions and Biomarkers', *Inflamm Bowel Dis*, vol. 21, no. 11, pp. 2515-2532.

Pittayanon, R, Lau, JT, Leontiadis, GI, Tse, F, Yuan, Y, Surette, M & Moayyedi, P 2020, 'Differences in Gut Microbiota in Patients With vs Without Inflammatory Bowel Diseases: A Systematic Review', *Gastroenterology*, vol. 158, no. 4, Mar, pp. 930-946.e931.

Plichta, DR, Graham, DB, Subramanian, S & Xavier, RJ 2019, 'Therapeutic Opportunities in Inflammatory Bowel Disease: Mechanistic Dissection of Host-Microbiome Relationships', *Cell*, vol. 178, no. 5, pp. 1041-1056.

Priya, S, Burns, MB, Ward, T, Mars, RAT, Adamowicz, B, Lock, EF, Kashyap, PC, Knights, D & Blekhman, R 2022, 'Identification of shared and disease-specific host gene–microbiome associations across human diseases using multi-omic integration', *Nat Microbiol*, vol. 7, no. 6, 2022/06/01, pp. 780-795.

Proctor, L 2019, 'Priorities for the next 10 years of human microbiome research', *Nature*, vol. 569, no. 7758, May, pp. 623-625.

Ren, J, Song, K, Deng, C, Ahlgren, NA, Fuhrman, JA, Li, Y, Xie, X, Poplin, R & Sun, F 2020, 'Identifying viruses from metagenomic data using deep learning', *Quant Biol*, vol. 8, no. 1, Mar, pp. 64-77.

Reyman, M, van Houten, MA, Arp, K, Sanders, EAM & Bogaert, D 2019, 'Rectal swabs are a reliable proxy for faecal samples in infant gut microbiota research based on 16S-rRNA sequencing', *Sci Rep*, vol. 9, no. 1, 2019/11/05, p. 16072.

Richard, ML, Lamas, B, Liguori, G, Hoffmann, TW & Sokol, H 2015, 'Gut fungal microbiota: the Yin and Yang of inflammatory bowel disease', *Inflamm Bowel Dis*, vol. 21, no. 3, Mar, pp. 656-665.

Richard, ML & Sokol, H 2019, 'The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases', *Nature Reviews Gastroenterology & Hepatology*, vol. 16, no. 6, Jun, pp. 331-345.

Roberts, T, Stark, D, Harkness, J & Ellis, J 2014, 'Update on the pathogenic potential and treatment options for Blastocystis sp', *Gut Pathogens*, vol. 6, no. 1, 2014/05/28, p. 17.

Rossen, NG, Fuentes, S, van der Spek, MJ, Tijssen, JG, Hartman, JHA, Duflou, A, Löwenberg, M, van den Brink, GR, Mathus-Vliegen, EMH, de Vos, WM, Zoetendal, EG, D'Haens, GR & Ponsioen, CY 2015, 'Findings From a Randomized Controlled Trial of Fecal Transplantation for Patients With Ulcerative Colitis', *Gastroenterology*, vol. 149, no. 1, 2015/07/01/, pp. 110-118.e114.

Sartor, RB & Wu, GD 2017, 'Roles for Intestinal Bacteria, Viruses, and Fungi in Pathogenesis of Inflammatory Bowel Diseases and Therapeutic Approaches', *Gastroenterology*, vol. 152, no. 2, Feb, pp. 327-339.e324.

Savitz, DA & Wellenius, GA 2022, 'Can Cross-Sectional Studies Contribute to Causal Inference? It Depends', *Am J Epidemiol*, Mar 1.

Schirmer, M, Smeekens, SP, Vlamakis, H, Jaeger, M, Oosting, M, Franzosa, EA, Ter Horst, R, Jansen, T, Jacobs, L, Bonder, MJ, Kurilshikov, A, Fu, J, Joosten, LAB, Zhernakova, A, Huttenhower, C, Wijmenga, C, Netea, MG & Xavier, RJ 2016, 'Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity', *Cell*, vol. 167, no. 4, Nov 3, pp. 1125-1136.e1128.

Seed, PC 2014, 'The human mycobiome', *Cold Spring Harbor perspectives in medicine*, vol. 5, no. 5, pp. a019810-a019810.

Sender, R, Fuchs, S & Milo, R 2016, 'Revised Estimates for the Number of Human and Bacteria Cells in the Body', *PLoS Biol*, vol. 14, no. 8, Aug, p. e1002533.

Shankar, J 2017, 'Insights into study design and statistical analyses in translational microbiome studies', *Ann Transl Med*, vol. 5, no. 12, Jun, p. 249.

Shaw, SY, Blanchard, JF & Bernstein, CN 2010, 'Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease', *Am J Gastroenterol*, vol. 105, no. 12, Dec, pp. 2687-2692.

Silverman, JD, Shenhav, L, Halperin, E, Mukherjee, S & David, LA 2018, 'Statistical Considerations in the Design and Analysis of Longitudinal Microbiome Studies', *bioRxiv*, p. 448332.

Sinha, R, Chen, J, Amir, A, Vogtmann, E, Shi, J, Inman, KS, Flores, R, Sampson, J, Knight, R & Chia, N 2016, 'Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies', *Cancer Epidemiology, Biomarkers & Prevention*, vol. 25, no. 2, pp. 407-416.

Sivignon, A, de Vallee, A, Barnich, N, Denizot, J, Darcha, C, Pignede, G, Vandekerckove, P & Darfeuille-Michaud, A 2015, 'Saccharomyces cerevisiae CNCM I-3856 prevents colitis induced by AIEC bacteria in the transgenic mouse model mimicking Crohn's disease', *Inflamm Bowel Dis*, vol. 21, no. 2, Feb, pp. 276-286.

Smye, SW & Frangi, AF 2021, 'Interdisciplinary research: shaping the healthcare of the future', *Future healthcare journal*, vol. 8, no. 2, pp. e218-e223.

Sokol, H, Leducq, V, Aschard, H, Pham, HP, Jegou, S, Landman, C, Cohen, D, Liguori, G, Bourrier, A, Nion-Larmurier, I, Cosnes, J, Seksik, P, Langella, P, Skurnik, D, Richard, ML & Beaugerie, L 2017, 'Fungal microbiota dysbiosis in IBD', *Gut*, vol. 66, no. 6, Jun, pp. 1039-1048.

Speich, B, Croll, D, Fürst, T, Utzinger, J & Keiser, J 2016, 'Effect of sanitation and water treatment on intestinal protozoa infection: a systematic review and meta-analysis', *The Lancet Infectious Diseases*, vol. 16, no. 1, 2016/01/01/, pp. 87-99.

Stegen, JC, Bottos, EM & Jansson, JK 2018, 'A unified conceptual framework for prediction and control of microbiomes', *Current Opinion in Microbiology*, vol. 44, 2018/08/01/, pp. 20-27.

Strömbeck, A, Lasson, A, Strid, H, Sundin, J, Stotzer, P-O, Simrén, M, Magnusson, MK & Öhman, L 2020, 'Fecal microbiota composition is linked to the postoperative disease course in patients with Crohn's disease', *BMC Gastroenterol*, vol. 20, no. 1, 2020/05/04, p. 130.

Tang, Q, Jin, G, Wang, G, Liu, T, Liu, X, Wang, B & Cao, H 2020, 'Current Sampling Methods for Gut Microbiota: A Call for More Precise Devices', *Frontiers in cellular and infection microbiology*, vol. 10, pp. 151-151.

Taylor, BC, Lejzerowicz, F, Poirel, M, Shaffer, JP, Jiang, L, Aksenov, A, Litwin, N, Humphrey, G, Martino, C, Miller-Montgomery, S, Dorrestein, PC, Veiga, P, Song, SJ, McDonald, D, Derrien, M & Knight, R 2020, 'Consumption of Fermented Foods Is Associated with Systematic Differences in the Gut Microbiome and Metabolome', *mSystems*, vol. 5, no. 2, Mar 17.

Terveer, EM, van Gool, T, Ooijevaar, RE, Sanders, IMJG, Boeije-Koppenol, E, Keller, JJ, Bart, A, Kuijper, EJ & Group, NDFBS 2019, 'Human Transmission of Blastocystis by Fecal Microbiota Transplantation Without Development of Gastrointestinal Symptoms in Recipients', *Clinical Infectious Diseases*.

Tito, RY, Chaffron, S, Caenepeel, C, Lima-Mendez, G, Wang, J, Vieira-Silva, S, Falony, G, Hildebrand, F, Darzi, Y, Rymenans, L, Verspecht, C, Bork, P, Vermeire, S, Joossens, M & Raes, J 2019, 'Population-level analysis of *Blastocystis* subtype prevalence and variation in the human gut microbiota', *Gut*, vol. 68, no. 7, pp. 1180-1189.

Turnbaugh, PJ, Ley, RE, Hamady, M, Fraser-Liggett, CM, Knight, R & Gordon, JI 2007, 'The Human Microbiome Project', *Nature*, vol. 449, no. 7164, 2007/10/01, pp. 804-810.

Uhr, GT, Dohnalová, L & Thaiss, CA 2019, 'The Dimension of Time in Host-Microbiome Interactions', *mSystems*, vol. 4, no. 1, pp. e00216-00218.

Underhill, DM & Braun, J 2022, 'Fungal microbiome in inflammatory bowel disease: a critical assessment', *The Journal of clinical investigation*, vol. 132, no. 5, p. e155786.

Ungaro, F, Massimino, L, D'Alessio, S & Danese, S 2019, 'The gut virome in inflammatory bowel disease pathogenesis: From metagenomics to novel therapeutic approaches', *United European Gastroenterol J*, vol. 7, no. 8, Oct, pp. 999-1007.

Vagianos, K, Shafer, LA, Witges, K, Graff, LA, Targownik, LE & Bernstein, CN 2022, 'Self-reported flares among people living with inflammatory bowel disease are associated with stress and worry but not associated with recent diet changes: The Manitoba Living with IBD Study', *JPEN J Parenter Enteral Nutr*, Feb 11.

Vandeputte, D, Falony, G, Vieira-Silva, S, Tito, RY, Joossens, M & Raes, J 2016, 'Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates', *Gut*, vol. 65, no. 1, pp. 57-62.

Wang, C, Li, W, Wang, H, Ma, Y, Zhao, X, Zhang, X, Yang, H, Qian, J & Li, J 2019, 'Saccharomyces boulardii alleviates ulcerative colitis carcinogenesis in mice by reducing TNF-alpha and IL-6 levels and functions and by rebalancing intestinal microbiota', *BMC Microbiol*, vol. 19, no. 1, Nov 6, p. 246.

Wang, Y & LêCao, KA 2020, 'Managing batch effects in microbiome data', *Brief Bioinform*, vol. 21, no. 6, Dec 1, pp. 1954-1970.

Weersma, RK, Zhernakova, A & Fu, J 2020, 'Interaction between drugs and the gut microbiome', *Gut*, vol. 69, no. 8, pp. 1510-1519.

Wilkinson, JE, Franzosa, EA, Everett, C, Li, C, Hu, FB, Wirth, DF, Song, M, Chan, AT, Rimm, E, Garrett, WS & Huttenhower, C 2021, 'A framework for microbiome science in public health', *Nat Med*, vol. 27, no. 5, May, pp. 766-774.

Zhang, F, He, F, Li, L, Guo, L, Zhang, B, Yu, S & Zhao, W 2020, 'Bioavailability Based on the Gut Microbiota: a New Perspective', *Microbiology and molecular biology reviews : MMBR*, vol. 84, no. 2, pp. e00072-00019.

Zhang, YZ, Shi, M & Holmes, EC 2018, 'Using Metagenomics to Characterize an Expanding Virosphere', *Cell*, vol. 172, no. 6, Mar 8, pp. 1168-1172.

Zuo, T, Lu, X-J, Zhang, Y, Cheung, CP, Lam, S, Zhang, F, Tang, W, Ching, JYL, Zhao, R, Chan, PKS, Sung, JJY, Yu, J, Chan, FKL, Cao, Q, Sheng, J-Q & Ng, SC 2019, 'Gut mucosal virome alterations in ulcerative colitis', *Gut*, vol. 68, no. 7, p. 1169.

Appendix 1

Supplemental Material

for

Individuals with inflammatory bowel disease have an altered gut microbiome composition of fungi and protozoa

Gina L. Guzzo, Murthy N. Mittinty, Bastien Llamas, Jane M. Andrews, and Laura S. Weyrich

The following supplementary tables are contained in an electronic file at:

https://doi.org/10.25909/20523318

Table S1. Contaminants identified with Decontam with a prevalence score of 0.52 or greater.

Table S2. Results of EukDetect for the 1000IBD and 500FG studies before rarefying.

Table S3. Difference in EukDetect results for the 1000IBD and 500FG studies between pre- and post-rarefaction.

Table S4. Results of EukDetect for the faecal microbiota transplant study data.

Appendix 2

Supplemental Material

for

The utility of dry swabs over stool samples: a case study from a faecal microbiota transplant trial

Gina L. Guzzo, Samuel P. Costello, Jane M. Andrews, Chelsea Bickley, Murthy N. Mittinty, Bastien Llamas, and Laura S. Weyrich

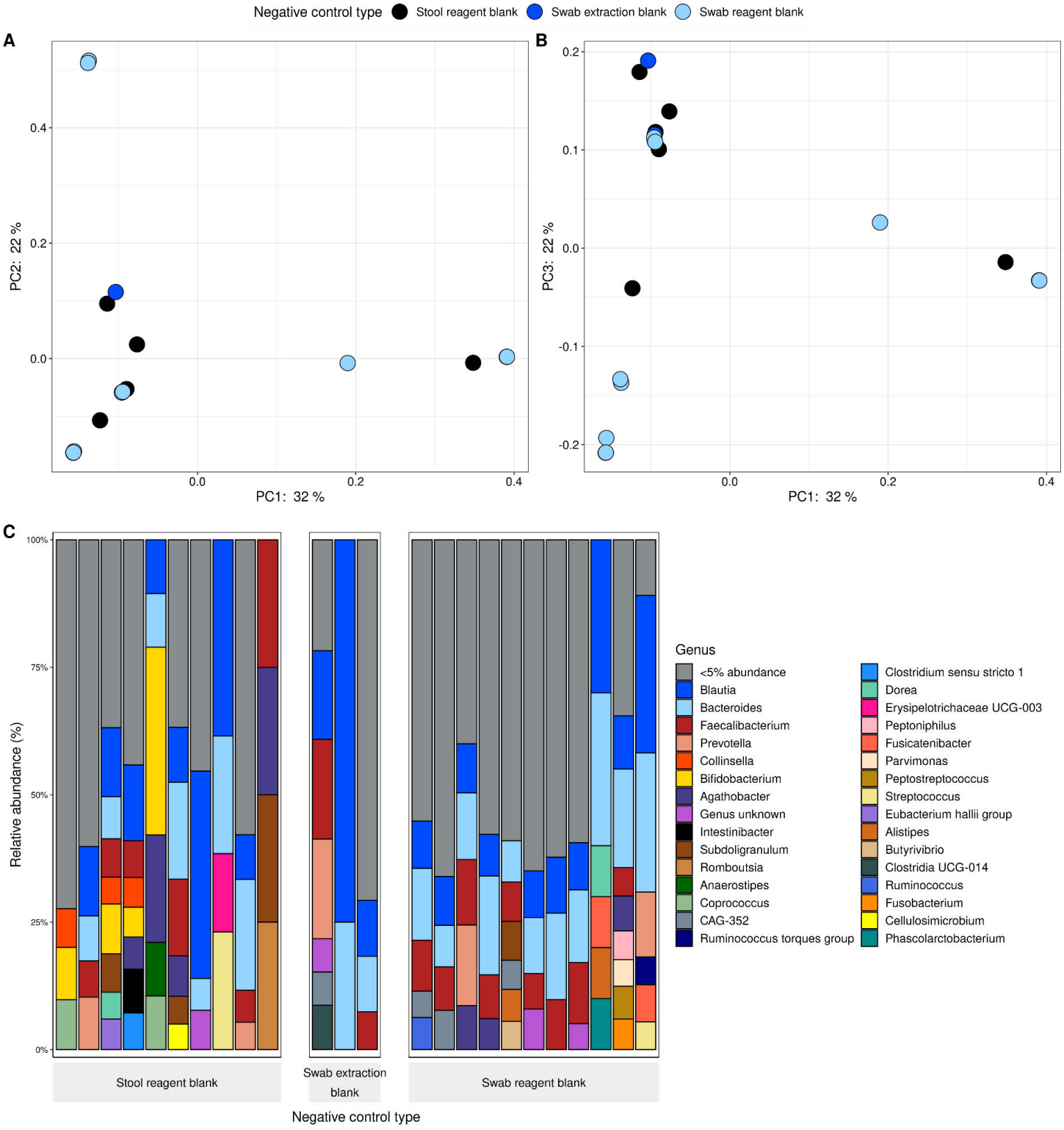

Figure S1. Negative control types did not show a distinct microbial signal. The composition of negative controls is shown with unweighted UniFrac distances displayed on PC1 and PC2 in (A) and PC1 and PC3 in (B), and with relative abundances of genera in (C).

Table S1. Contaminants identified with Decontam with a prevalence score of 0.52 or greater.

Located electronically at: https://doi.org/10.25909/20523030

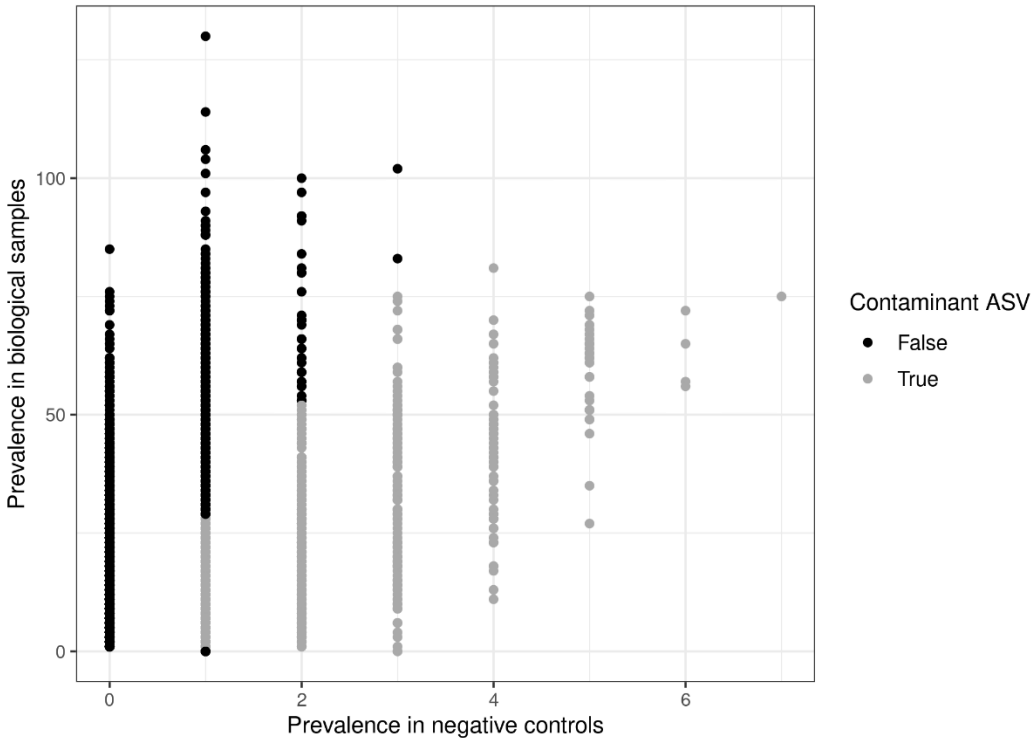

Figure S2. Prevalence of contaminant amplicon sequence variants (ASVs) with a Decontam score of 0.52 or greater in biological samples and negative control samples.

Table S2. Number of contaminants per bacterial phylum.

Phylum	**Count**	**Prop. (%)**
Firmicutes	1712	71.54
Bacteroidota	375	15.67
Actinobacteriota	128	5.35
Proteobacteria	93	3.89
Verrucomicrobiota	26	1.09
Desulfobacterota	22	0.92
Fusobacteriota	18	0.75
Euryarchaeota	13	0.54
Thermoplasmatota	4	0.17
Cyanobacteria	1	0.04
Synergistota	1	0.04
Total	**2,393**	**100**

Table S3. Results of decontamination by sample type.

	Total taxa	**Contaminant taxa**	**Proportion of contaminants (%)**	**No. samples before decontamination**	**No. samples after decontamination**
Swab	25,717	2,352	9.1	445	445
Stool	16,355	2,110	12.9	154	153

Table S4. Beta diversity significance testing of technical and participant characteristics.

Samples tested	Test group	No. per group	Degrees of freedom	Sum of squares	Mean squares	F model	R^2	Pr(>F)
Swab samples								
n = 431	**Year of FMT**	2014 (n=84) 2015 (n=277) 2016 (n=70)	2	3.84	1.92	9.42	0.05	**0.00**
	Recipient age	19-40 years (n=231) 41-62 years (n=200)	1	1.83	1.83	8.96	0.02	**0.00**
	Recipient sex	female (n=163) male (n=268)	1	1.50	1.50	7.35	0.02	**0.00**
	Residuals		369	75.29	0.20	NA	0.91	NA
	Total		373	82.46	NA	NA	1.00	NA
Stool samples								
n = 76	**Year of FMT**	2014 (n=14) 2015 (n=50) 2016 (n=12)	2	0.96	0.48	2.33	0.06	**0.00**
	Recipient age	19-40 years (n=45) 41-62 years (n=31)	1	0.27	0.27	1.32	0.02	0.08
	Recipient sex	female (n=31) male (n=45)	1	0.32	0.32	1.55	0.02	**0.03**
	Residuals		67	13.80	0.21	NA	0.90	NA
	Total		71	15.35	NA	NA	1.00	NA

Table S5. Alpha diversity significance testing of technical and participant characteristics

Covariate	Diversity metric	Test group		Kruskal-Wallis				
		Group 1	Group 2	H		p-value		q-value
Swab samples								
FMT year	Observed ASVs	2014 (n=74)	2015 (n=242)	41.78	44.70	**0.00**	**0.00**	**0.00**
			2016 (n=58)		9.55		**0.00**	**0.00**
		2015 (n=242)	2016 (n=58)		1.09		0.30	0.30
	Shannon diversity	2014 (n=74)	2015 (n=242)	38.42	41.15	**0.00**	**0.00**	**0.00**
			2016 (n=58)		6.43		**0.01**	**0.02**
		2015 (n=242)	2016 (n=58)		2.05		0.15	0.15

Recipient age	Observed ASVs	19-40 years (n=195)	41-62 years (n=179)	0.89		0.35		0.35
	Shannon diversity	19-40 years (n=195)	41-62 years (n=179)	0.10		0.76		0.76
Recipient sex	Observed ASVs	female (n=135)	male (n=239)	0.95		0.33		0.33
	Shannon diversity	female (n=135)	male (n=239)	3.39		0.07		0.07
Stool samples								
FMT year	Observed ASVs	2014 (n=14)	2015 (n=46)	14.08	14.25	**0.00**	**0.00**	**0.00**
			2016 (n=12)		5.85	**0.00**	**0.02**	**0.02**
		2015 (n=46)	2016 (n=12)		0.23		0.63	0.63
	Shannon diversity	2014 (n=14)	2015 (n=46)	13.95	13.86	**0.00**	**0.00**	**0.00**
			2016 (n=12)		5.84	**0.00**	**0.02**	**0.02**
		2015 (n=46)	2016 (n=12)		0.59		0.44	0.44
Recipient age	Observed ASVs	19-40 years (n=42)	41-62 years (n=30)	0.08		0.77		0.77
	Shannon diversity	19-40 years (n=42)	41-62 years (n=30)	0.01		0.92		0.92
Recipient sex	Observed ASVs	female (n=30)	male (n=42)	2.69		0.10		0.10
	Shannon diversity	female (n=30)	male (n=42)	3.13		0.08		0.08

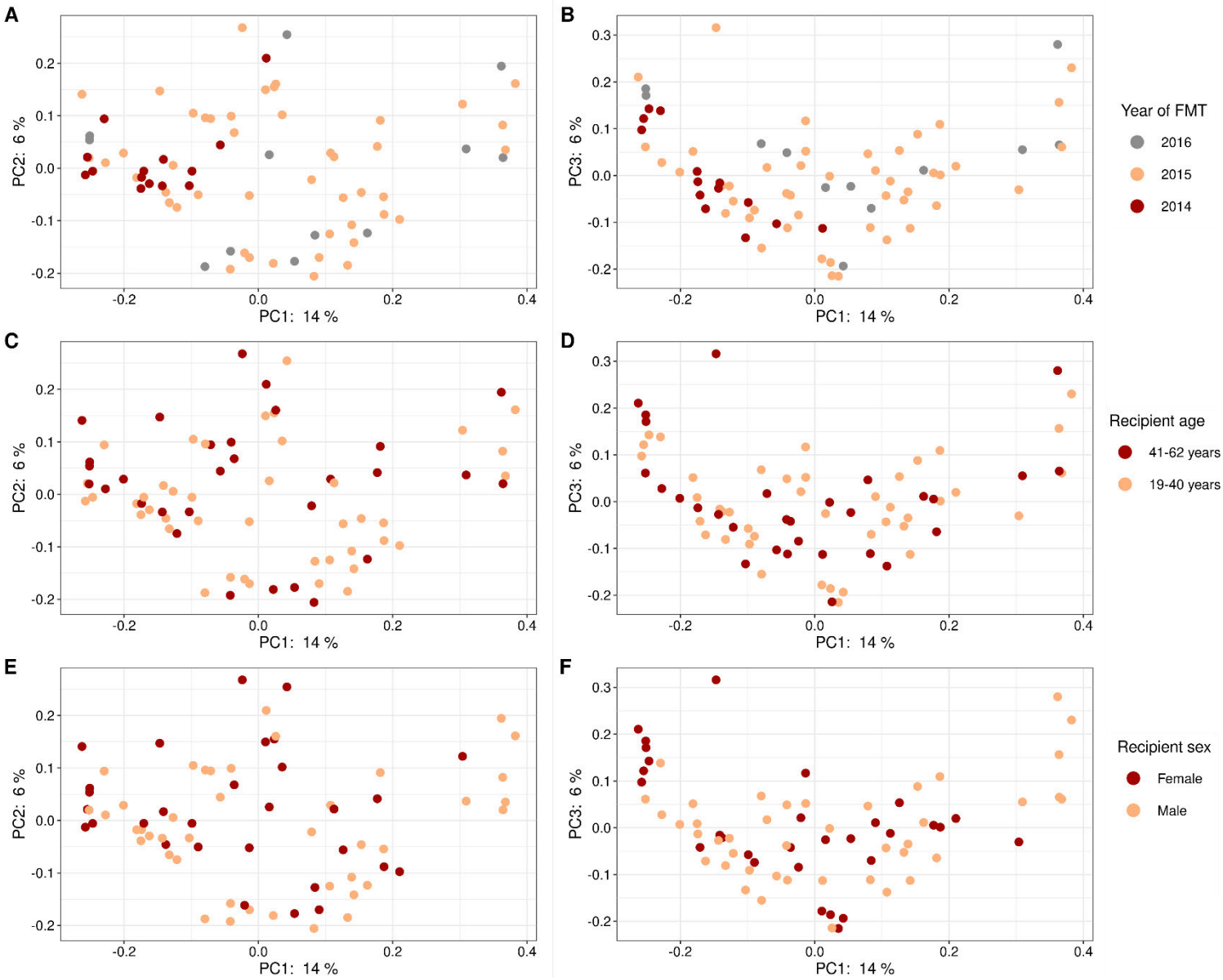

Figure S3. The year of FMT impacted the composition of stool samples, whereas the FMT recipient age and sex did not. Composition is shown with PCoA plots of unweighted UniFrac distances and samples are coloured by the year of FMT on (A) PC1 and PC2, and (B) PC1 and PC3. Samples are coloured by FMT recipient sex on (C) PC1 annd PC2, and (D) PC1 and PC3. PCoA plots were generated with unweighted UniFrac distances of samples rarefied to 1,019 16S rRNA sequences.

Table S6. Differentially abundant genera in swab samples across year of FMT.

Located electronically at: https://doi.org/10.25909/20523045

Table S7. Alpha diversity significance testing of sample type.

Test category	Test	Week 4				Week 8			
		Kruskal-Wallis			No. per group	Kruskal-Wallis			No. per group
		H	p-value	q-value		H	p-value	q-value	
Unpaired samples									
Sample type	Observed ASVs	1.55	0.21	0.21	Stool (n=22) Swab (n=20)	2.57	0.11	0.11	Stool (n=18) Swab (n=16)
	Shannon	0.01	0.94	0.94		0.58	0.45	0.45	
Paired samples									
Sample type	Observed ASVs	2.50	0.11	0.11	Stool (n=18) Swab (n=18)	0.85	0.36	0.36	Stool (n=11) Swab (n=11)
	Shannon	0.14	0.70	0.70		0.05	0.82	0.82	

Table S8. Adonis tests on unweighted UniFrac distances (rarefaction depth: 1,019 sequences).

Samples tested	Test group	No. per group	Degrees of freedom	Sum of squares	Mean squares	F model	R^2	Pr(>F)
Unpaired samples - week 4								
n=42	**Year of FMT**	2014 (n=8) 2015 (n=26) 2016 (n=8)	2	0.68	0.34	1.54	0.07	**0.01**
	Sample type (swab or stool)	stool (n=22) swab (n=20)	1	0.49	0.49	2.19	0.05	**0.00**
	Recipient age	19-40 years (n=23) 41-62 years (n=19)	1	0.28	0.28	1.28	0.03	0.09
	Recipient sex	female (n=17) male (n=25)	1	0.30	0.30	1.35	0.03	0.08
	Residuals		36	8.00	0.22	NA	0.82	NA
	Total		41	9.75	NA	NA	1.00	NA

Unpaired samples - week 8								
n=34	**Year of FMT**	2014 (n=7) 2015 (n=22) 2016 (n=5)	2	0.69	0.35	1.67	0.09	**0.00**
	Sample type (swab or stool)	stool (n=18) swab (n=16)	1	0.30	0.30	1.46	0.04	**0.04**
	Recipient age	19-40 years (n=18) 41-62 years (n=16)	1	0.28	0.28	1.34	0.04	0.08
	Recipient sex	female (n=13) male (n=21)	1	0.25	0.25	1.19	0.03	0.17
	Residuals		28	5.82	0.21	NA	0.79	NA
	Total		33	7.34	NA	NA	1.00	NA
Paired samples - week 4								
n=36	**Year of FMT**	2014 (n=6) 2015 (n=22) 2016 (n=8)	2	0.68	0.34	1.49	0.08	**0.01**
	Sample type (swab or stool)	stool (n=18) swab (n=18)	1	0.48	0.48	2.11	0.06	**0.00**
	Recipient age	19-40 years (n=20) 41-62 years (n=16)	1	0.31	0.31	1.35	0.04	0.08
	Recipient sex	female (n=14) male (n=22)	1	0.30	0.30	1.34	0.04	0.08
	Residuals		30	6.79	0.23	NA	0.79	NA
	Total		35	8.55	NA	NA	1.00	NA
Paired samples - week 8								
n=22	**Year of FMT**	2014 (n=4) 2015 (n=14) 2016 (n=4)	2	0.75	0.37	1.87	0.16	**0.00**
	Sample type (swab or stool)	stool (n=11) swab (n=11)	1	0.24	0.24	1.19	0.05	0.18
	Recipient age	19-40 years (n=12) 41-62 years (n=10)	1	0.28	0.28	1.39	0.06	0.07
	Recipient sex	female (n=8) male (n=14)	1	0.30	0.30	1.52	0.06	**0.03**
	Residuals		16	3.19	0.20	NA	0.67	NA
	Total		21	4.75	NA	NA	1.00	NA

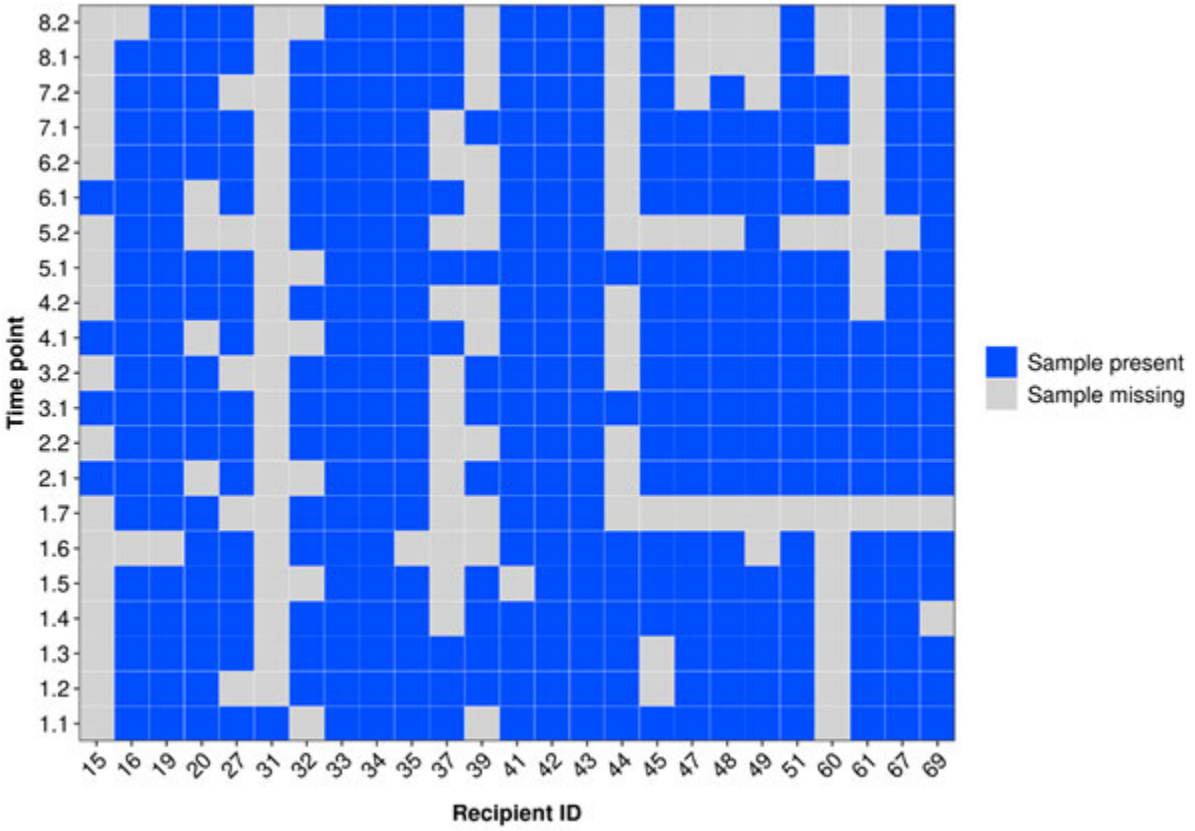

Figure S4. Swab sample collection across all time points and recipients in the study. Swabs were collected for each day of the first week after FMT administration and then two days in each week thereafter. The first number of each time point represents the week number and the decimal number indicates the day of each respective week. The two samples collected in week 2-8 were not collected on consecutive days.

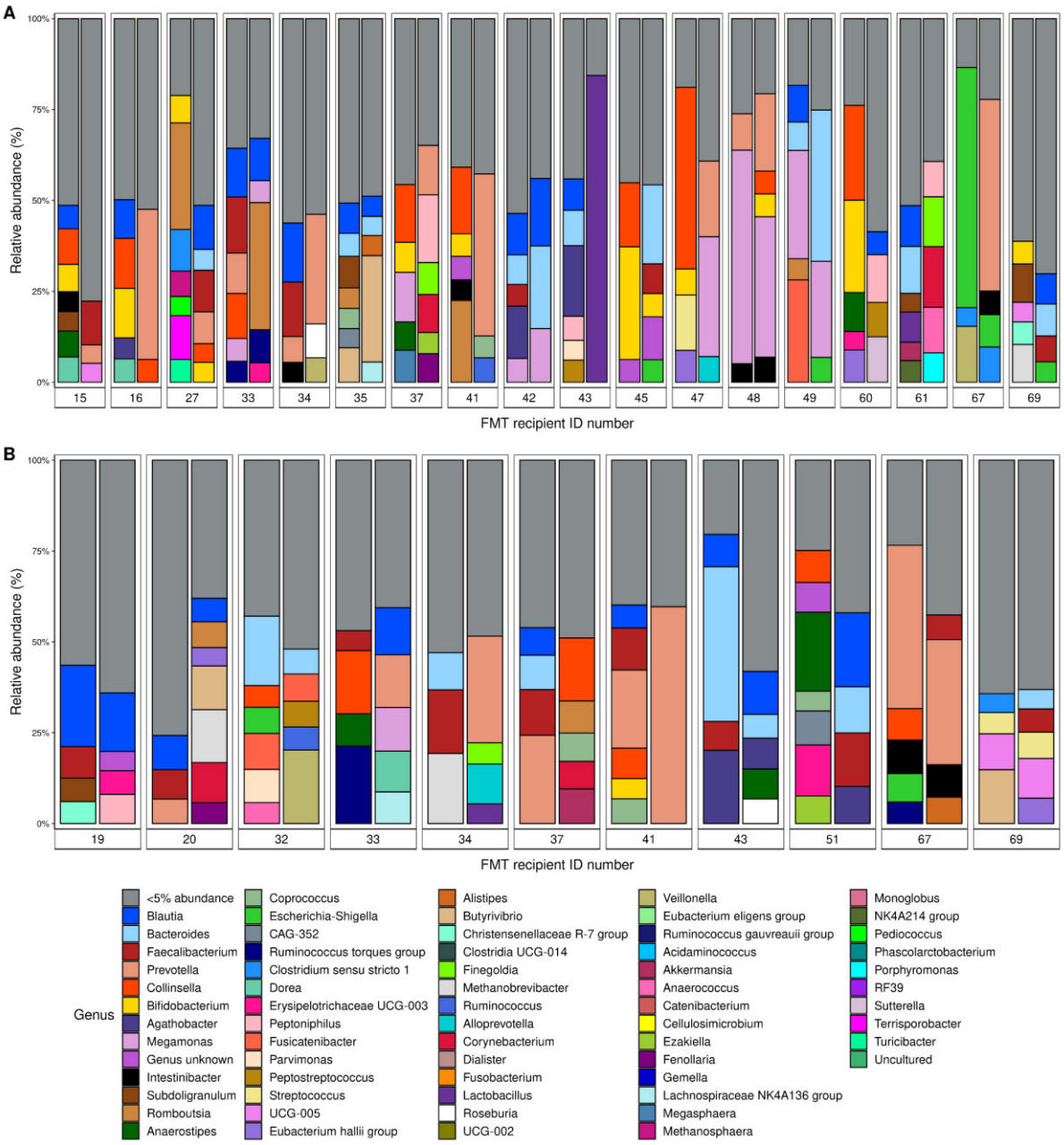

Figure S5. The relative abundance of bacterial genera in paired swab and stool samples collected from individuals with ulcerative colitis post-faecal microbiota transplant (FMT). Sample were collected at two time points: week 4 (A) and week 8 (B). Stool samples are shown on the left for each FMT recipient, and swab samples are on the right.

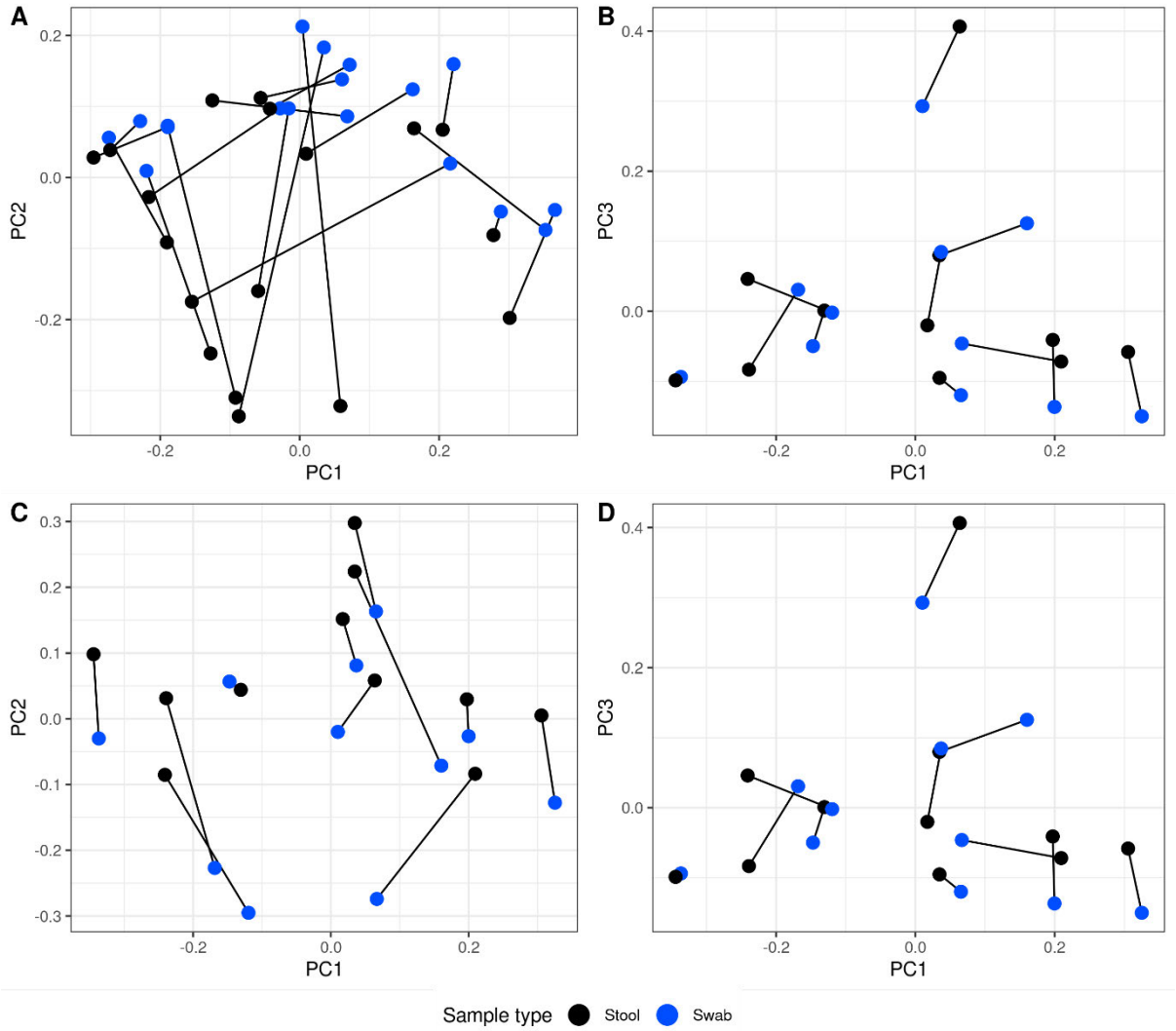

Figure S6. Pairwise microbiome compositions of swab and stool samples from individuals with ulcerative colitis after receiving faecal microbiota transplants. Microbial sample compositions are shown with unweighted UniFrac distances at two time points and three PC axes, (A) PC1 and PC2 at week 4, (B) PC1 and PC3 at week 4, (C) PC1 and PC2 at week 8, and (D) PC1 and PC3 at week 8. Samples from the same individuals are connected with lines. Swab samples are coloured in blue and stool samples are coloured in black. All samples were rarefied to 1,019 16S rRNA sequences.

Table S9. Multivariate correlation tests on unweighted UniFrac distances (rarefaction depth: 1,019 sequences) of paired swab and stool samples.

Test	Week	Total samples	Test statistic	Significance
Mantel	4	36	r = 0.25	**0.03**
	8	22	r = -0.00	0.48
Procrustes	4	36	Sum of squares = 0.15 Symmetric correlation = 0.92	**0.03**
	8	22	Sum of squares = 0.08 Symmetric correlation = 0.96	0.56

Table S10. Alpha significance tests of FMT recipient outcome.

Sample type	Test	Kruskal-Wallis		No. individuals per group	No. samples per group
		H	p-value		
Swab	Observed ASVs	38.38	**0.00**	Remission (n=6) No remission (n=13)	Remission (n=99) No remission (n=246)
	Shannon	59.39	**0.00**		

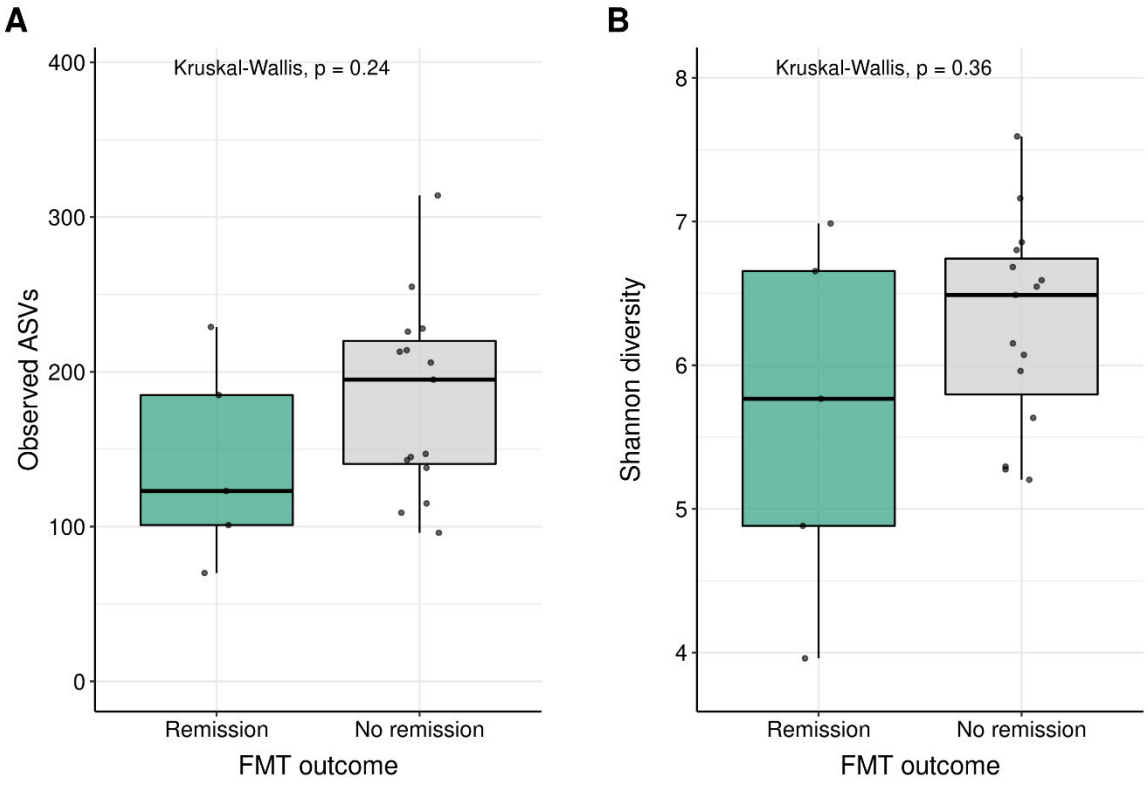

Figure S7. The alpha diversity of gut microbiomes in individuals with ulcerative colitis who responded to faecal microbiota transplantation (FMT) (n = 5) was lower even before receiving FMT, compared to those who did not respond (n = 15). Stool samples in FMT recipients were collected immediately before receiving FMT. Response to FMT was measured as clinical and endoscopic remission at week 8 post-FMT. Significance testing was performed with Kruskal-Wallis tests.

Table S11. Adonis tests of unweighted UniFrac distances (rarefaction depth: 1,019 sequences) of swab samples weeks 1-8 post-FMT from individuals with UC who received FMT.

Sample tested		No. per group	Degrees of freedom	Sum of squares	Mean squares	F model	R^2	Pr(>F)
n=345	**Year of FMT administration**	2014 (n=72) 2015 (n=224) 2016 (n=49)	2	3.77	1.89	9.65	0.05	**0.00**
	Recipient age	19-40 years (n=161) 41-62 years (n=184)	1	2.09	2.09	10.67	0.03	**0.00**
	Recipient sex	female (n=115) male (n=230)	1	1.58	1.58	8.08	0.02	**0.00**
	FMT type (active or placebo)	active (n=205) placebo (n=140)	1	2.52	2.52	12.89	0.03	**0.00**
	FMT outcome (remission or no)	remission (n=246) no remission (n=99)	1	1.04	1.04	5.32	0.01	**0.00**
	FMT type*outcome	NA	1	1.19	1.19	6.08	0.02	**0.00**
	Residuals		337	65.86	0.20	NA	0.84	NA
	Total		344	78.04	NA	NA	1.00	NA

Table S12. Adonis tests of unweighted UniFrac distances (rarefaction depth: 1,019 sequences) of stool samples weeks 4 and 8 post-FMT from individuals with UC who received FMT.

Sample tested		No. per group	Degrees of freedom	Sum of squares	Mean squares	F model	R^2	Pr(>F)
n=345	**Year of FMT administration**	2014 (n=7) 2015 (n=26) 2016 (n=7)	2	0.71	0.35	1.61	0.08	**0.00**
	Recipient age	19-40 years (n=23) 41-62 years (n=17)	1	0.26	0.26	1.21	0.03	0.15
	Recipient sex	female (n=18) male (n=22)	1	0.33	0.33	1.49	0.04	**0.04**
	FMT type (active or placebo)	active (n=26) placebo (n=14)	1	0.27	0.27	1.23	0.03	0.14
	FMT outcome (remission or no)	remission (n=18) no remission (n=12)	1	0.28	0.28	1.27	0.03	0.10
	FMT type*outcome	NA	1	0.27	0.27	1.24	0.03	0.13
	Residuals		32	7.02	0.22	NA	0.77	NA
	Total		39	9.14	NA	NA	1.00	NA