Master of Science by Research Thesis

Kingdom specific mucosal microbiota analysis

Ángela Del Castillo Izquierdo

June 28th, 2021

University of East Anglia

Quadram Institute Bioscience

Del Castillo Izquierdo, Ángela:
Kingdom specific mucosal microbiota analysis
Master of Science by Research Thesis
University of East Anglia, Quadram Institute Bioscience
Thesis period: October 1st, 2019 - June 28th, 2021

Abstract

Across the field of microbiome research, the ways in which bacteria influence the host in health and disease are well known. However, knowledge of the involvement of other kingdoms is still relatively poor and methods to isolate, identify and quantify them efficiently are needed. Nevertheless, the low abundance in which these microorganisms are estimated makes their study even harder.

Sample handling and DNA extraction methods are common factors that introduce variability to sample processing workflows. In this thesis these factors were examined using mucus isolated from mice colonic tissue. We tested three DNA extraction methods and compared their performance in terms of DNA quantity and quality, number of bacterial copies and microbial composition. We found significant differences between the kits tested which were attributed to different lysis and purification methods. Our results highlighted the importance of introducing controls to account for contaminants. We quantitatively compared two sample storage solutions: PBS and RNA*later*. Samples stored in RNA*later* yielded significantly more DNA and bacterial copies than those stored in PBS, making RNA*later* a good alternative to conserve samples.

These findings were corroborated during the study of bacterial and fungal communities in colorectal cancer. We analysed the differences in microbial communities between carcinoma and adjacent tissue using 16S rRNA gene and ITS1 region sequencing along with qPCR. We identified a large proportion of off-targets and contaminants, which hampered the detection and identification of fungi. Additionally, the observed bacterial and fungal communities were driven by the different DNA extraction methods used. We were unable to

detect differences between carcinoma and adjacent tissue. However, trends in our bacterial communities were observed, such as the increase of Fusobacteria in tumour samples, as previously reported.

The results of this thesis highlighted the importance of establishing standard methods for characterising microbial kingdoms in samples susceptible to host contamination.

Acknowledgements

A number of people have helped me along my masters, and I am grateful to each every one of them.

Firstly, I would like to thank my primary supervisor, Dr Falk Hildebrand, for giving me this opportunity in the first place, and many more that came up over this period. I have enjoyed and learned from each of them. Thanks for his trust, guidance, ideas and all the invaluable teachings, in tutorials and beyond. I am also deeply grateful to Dr Tamas Korcsmaros and Dr Isabelle Hautefort and their group, for their mentorship, their constant support and sharing their experience and knowledge.

I also wish to thank all the researchers who introduced me patiently to and provided me with feedback on the techniques and each kingdom discussed in this thesis: Dr Evelien Adriaenssens, Dr Gemma Kay, Dr Julia Kleinteich, Dr Mohammad Adnan Tariq, Dr Mohammad Bahram, Dr Naiara Beraza, Dr Raul Tito and Dr Steve James. I am also truly grateful to Dr Dimitra Lamprinaki and Dr Nathalie Juge for kindly providing the colorectal cancer samples and giving a second life to this thesis during COVID-19.

I also want to thank the whole Hildebrand group, its current and past members, for their unconditional support, help and shared laughter. From these two years I take with me great colleagues, but also great friends. Thanks to Dr Ezgi Özkurt for the numerous discussions, priceless feedback, movie nights and unforgettable moments in the lab. Joachim Fritscher for all the jokes, shared conversations and walks around Norwich. And Bernadette Breeze (Bernie) and Dr Rebecca Ansorge for their encouragement, help and sweet words in the past few months. I am also grateful to Dr Janis Bedarf, who

became my role model when it comes to working hard and being resilient in order to achieve my ambitions. And Dr Clémence Frioux who provided me with key guidance during the very beginning of this masters, her support and mentorship has been invaluable.

Besides the Hildebrand group, I would not have got through these two years as happily as I have, had it not been for Raven Reynolds and Dr Teagan Brown, and their care and help throughout this masters. I want to thank Raven for always making me laugh and having the patience to proofread my writing every time I needed it. She always finds the perfect words, inside and outside of research. And Teagan for being a great companion in the office, in the lab and during every dinner. Thanks for all the troubleshooting, and especially for looking after me.

A special thanks to my family and friends, for their love and always holding me in the good and bad. Specially to my sister, María, who helped me to produce the figures included in the first chapter. And Jorge, who always stood by my side, cheering me up and making me smile. Finally, I would like to express my infinite gratitude to Dr Daniela Barillà and Dr Tamara Maes, ex-supervisor and ex-boss, who gave me the best foundation I could have wish as a researcher and encouraged me to follow what I am passionate about. Gracias.

Contents

List of Figures

List of Tables

Abbreviations

AIEC	**A**dherent-**i**nvasive ***E**scherichia **c**oli*
ANOVA	**An**alysis **o**f **va**riance
ASV	**A**mplicon **s**equence **v**ariant
CAC	**C**olitis-**a**ssociated **c**olorectal cancer
CD	**C**rohn's **d**isease
CRC	**Colo**rectal **c**ancer
Ct	**C**ycle **t**hreshold
DNA	**D**eoxyribo**n**ucleic **a**cid
dsDNA	**d**ouble **s**tranded **DNA**
GC	**G**uanine-**c**ytosine
GIT	**G**astro**i**ntestinal **t**ract
IBD	**I**nflammatory **b**owel **d**isease
ITS	**I**nternal **t**ranscribed **s**pacer
ITS1	**I**nternal **t**ranscribed **s**pacer **1**
LCA	**L**east **c**ommon **a**ncestor
MUC2	**Muc**in **2**
OTU	**O**perational **t**axonomic **u**nit
PBS	**P**hosphate-**b**uffered **s**aline
PCoA	**P**rincipal **Co**ordinates **A**nalysis
PCR	**P**olymerase **c**hain **r**eaction
PERMANOVA	**Per**mutational **m**ultivariate **an**alysis **o**f **v**ariance
QMP	**Q**uantitative **mi**crobiome **p**rofiling
qPCR	**Q**uantitative **PCR**
RMP	**R**elative **mi**crobiome **p**rofiling
RNA	**R**ibo**n**ucleic **a**cid
rRNA	**R**ibosomal **RNA**
UC	**U**lcerative **c**olitis

Chapter 1

Introduction

1.1 Human gut microbiota

The human gastrointestinal tract (GIT) is a complex environment which is inhabited by a diverse and dynamic community of microorganisms; predominately composed of bacteria along with Archaea, Eukaryotes and viruses (Gouba et al., 2013; Lima et al., 2019; Parfrey et al., 2014). The collection of these microorganisms is referred as the gut microbiota, while the term microbiome describes the genomes from all these microbes. The relationship between the host and the microbiota is in general considered to be commensal, but single members can be pathobionts (Kamada et al., 2013). The host provides the infrastructure and living conditions for the gut microbiota, whereas microbial communities allow the host to access nutrients, provide vitamins, offer colonisation resistance and help to develop the host immune system among other interactions (Buffie et al., 2015; Hooper et al., 2012; Qin et al., 2010; Rowland et al., 2018).

The acquisition of microbes takes place at birth, mainly from maternal sources (Ferretti et al., 2018). This early microbiota rapidly evolves within the first three years of life, increasing in diversity until the microbial composi-

tion resembles an adult microbiome (Bäckhed et al., 2015; Ferretti et al., 2018; Koenig et al., 2011). The microbial colonization seems to be continuous along the GIT. Microbes from the mouth are transmitted constantly to the gut, making the oral cavity a potential source of microbial colonization for the intestinal environment, although the digestive machinery will limit transmission of microbes (Schmidt et al., 2019). Different physical, chemical and environmental conditions displayed along the GIT segregate the microbial composition and its function. In this way, different anatomical sites present different microbial composition (Hillman et al., 2017).

1.1.1 Colon and the mucus system

The large intestinal microbiota accounts for the majority of the microbial cells (around 10^{13} cells) in the human body (Schmidt et al., 2019; Sender et al., 2016). In this organ two main functions of the digestive system occur: the absorption of water and electrolytes through osmosis, and the breakdown of the remaining indigestible material by the microbiota through bacterial fermentation. These occur along the whole colon, which is divided in four sections: ascending, transverse, descending and sigmoid colon (**Figure 1.1A**). The intestinal wall is comprised by four different layers: mucosa, submucosa, muscularis and serosa (**Figure 1.1B**).

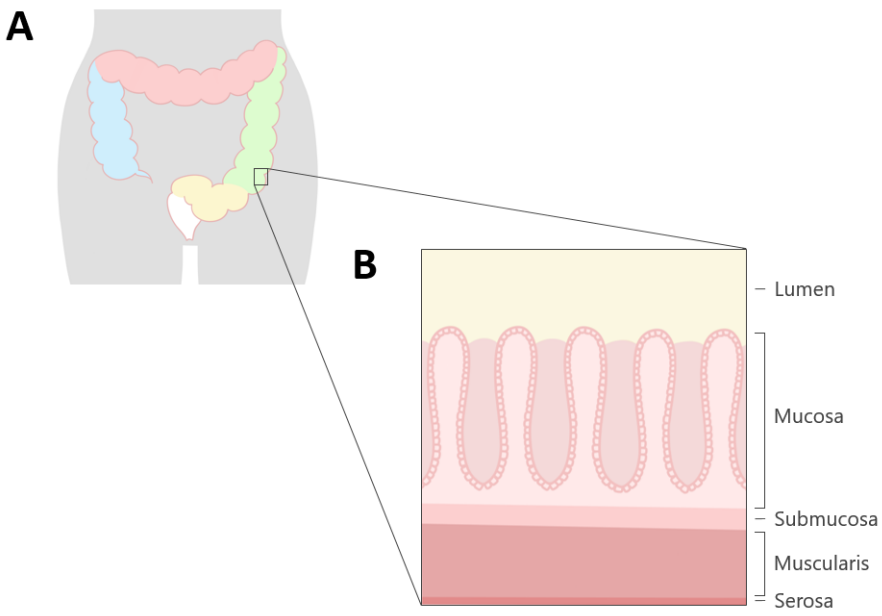

Figure 1.1. Schematic diagram of the different parts of the large intestine. (**A**) Simplistic representation of colon sections differentiated by colors: ascending (blue), transverse (red), descending (green) and sigmoid (yellow). (**B**) Layers of the colon. The wall of the colon is made up of the mucosa (innermost layer), submucosa, muscularis and serosa (outermost layer).

The mucosa is the outermost layer of the intestine and is therefore in direct contact with the luminal content (**Figure 1.1B**). For this reason, this region is key to ensure both protection from the gut microbiota and communication with the host epithelial cells. To ensure the maintenance of gut steady state, the host immune system has evolved to a state of tolerance towards certain commensal microorganisms (immunotolerance); while at the same time, the contact between the host cells and microbes is limited. The intestinal epithelium is covered with a thick mucus gel; a physical protective separation which provides a first defence layer against luminal microorganisms (Johansson, 2014). The colon has a two-layered mucus system. The inner layer is dense, stratified and depleted of microorganisms. It is later converted into the outer layer that is loose and colonised by commensal bacteria (M. E. Johansson et al., 2011) (**Figure 1.2**). In a healthy state, commensal microbes colonise

the outer mucus layer, contributing to its maintenance, while the inner layer is resistant to microbial penetration due to its high density.

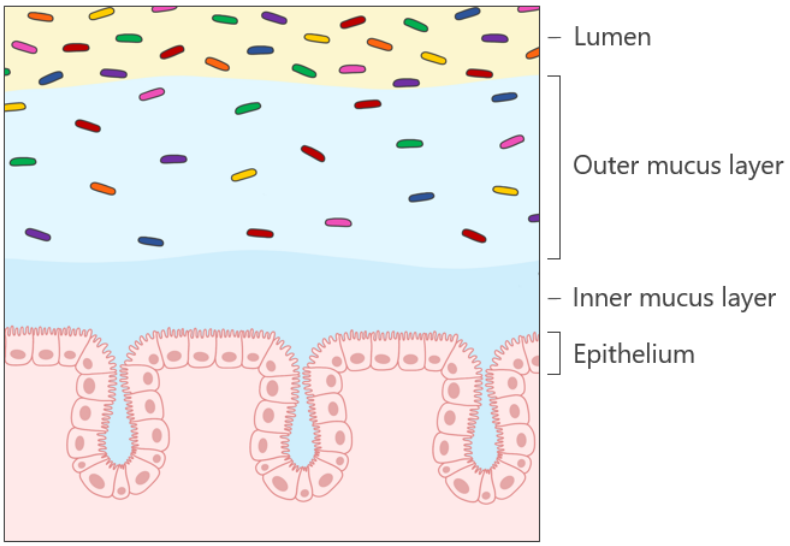

Figure 1.2. Structure of the two-layered mucus system in the large intestine. Schematic diagram of the mucus layers. The outer mucus layer is in contact with the luminal content and colonised by bacteria, while the inner layer is devoid of microbes. The intestinal epithelium is composed of a single layer of cells.

Both layers are mainly made up of mucin 2 (MUC2), but 21 different mucins have been identified (Hansson, 2020); these are the building blocks of mucus, as well as the attachment sites and substrates for intestinal microbes. A mucin is a heavily glycosylated protein. The variety of mucins provides different conditions for the residing microorganisms, in this way, the host is indirectly selecting them. Equally the microbiota has the potential to shape the mucus layer properties as part of the colonization process and to sustain the microbial growth (Engevik et al., 2019; Jakobsson et al., 2015; H. Wu et al., 2020). In short, the gut microbiota can be involved in the mucus layer synthesis, maturation, regulation, composition and degradation.

1.2 Microbial dysbiosis

The gastrointestinal microbial composition is shaped by host-specific and environmental factors; in brief, these include lifestyle, host genetics, the surrounded microbial community and microbiome-intrinsic factors (Schmidt et al., 2018).

Although the host and its microbiota maintain a mutualistic relationship, enabling intestinal homeostasis (Hooper & MacPherson, 2010), the microbiome composition can change due to both host and environmental factors which could lead to dysbiosis. Dysbiosis is the condition of having an imbalanced microbiota. This state has been associated with various diseases, including inflammatory bowel disease (IBD), colorectal carcinoma (CRC) and Parkinson's disease (Bedarf et al., 2017; Brennan & Garrett, 2016; Pittayanon et al., 2020). Although different diseases will show different dysbiotic signatures, dysbiosis is characterized by a reduction in bacterial diversity and the loss of beneficial bacteria such as butyrate-producing bacteria and *Bacteroides* strains. Finally, dysbiosis in diseases has been also associated with impaired epithelial barrier, bacterial translocation and local inflammation (Okumura & Takeda, 2017; Zeng et al., 2017).

1.2.1 Inflammatory bowel disease

IBD is the relapsing and chronic inflammation of the GIT, the etiology of which remains to be elucidated. The most accepted hypothesis suggests that it is triggered by environmental factors in a genetically prone host (Guan, 2019). This disease features mucus discontinuity and overall mucosal damage in inflamed areas. Bowel inflammation results from an altered interaction between the intestinal immune system and the microbiota, where the inflam-

mation is associated with a reduction in mucus thickness and an increased mucosal colonisation by bacteria (**Figure 1.3**) (Johansson, 2014; Van Der Post et al., 2019; Walker, Sanderson, et al., 2011). This leads to a distinct dysbiotic signature of the mucosal microbiome (Kleessen et al., 2002; Li et al., 2014).

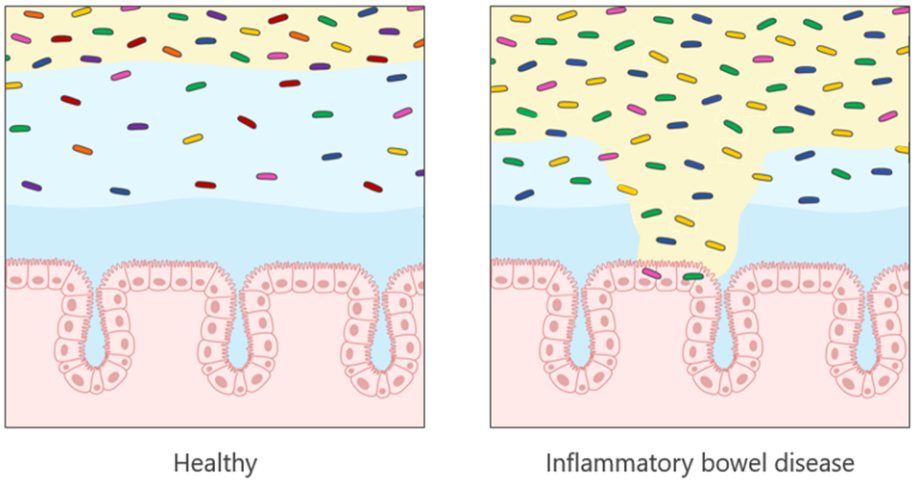

Figure 1.3. Differences in the inner and outer mucus layers lining the colon epithelium in healthy and IBD states. Schematic diagram of the colon mucus layers in healthy and IBD patients. The gut microbiota colonises the outer mucus layer in healthy state, while the inner layer is resistant to microbial penetration. The thickness of mucus layers is reduced in IBD patients, allowing luminal microbes to gain access to the intestinal epithelium and inducing a strong inflammatory response by the host immune system.

Crohn's disease (CD) and Ulcerative colitis (UC) are the main two subtypes of this condition. CD involves mucosal and muscle inflammation of any part along the GIT which appears as a patchy pattern. This differs from UC which displays a continuous ascending mucosal inflammation from the rectum and all the way along to the large intestine (Farmer et al., 1975) (**Figure 1.4**). Amongst their inflammation patterns, they also have pathological and clinical distinctive differences. However, in some patients it is not possible to distinguish between both disease subsets (Odze, 2015).

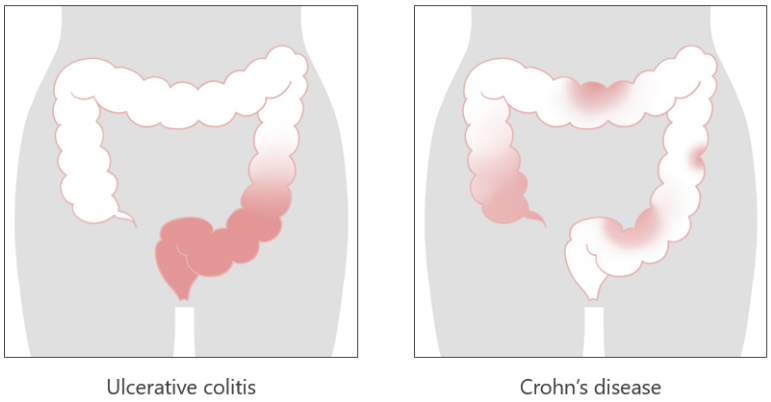

Figure 1.4. Overview of Ulcerative colitis and Crohn's disease inflammatory pattern along the colon. Schematic diagram of the affected areas in the colon during Ulcerative colitis and Crohn's disease. UC begins in the rectum and extends to the entire colon, while CD usually involves the beginning of the colon and can affect any other part in a patchy pattern.

IBD has been associated with an overall decrease in the gut bacterial diversity which corresponds to a decrease in the ecosystem stability (Pittayanon et al., 2020). It also displays a decrease of commensal bacteria such as butyrate-producing bacteria. This bacterial group is considered to have a 'protective' role and modulate the inflammatory response in the ecosystem (Aden et al., 2019; J. Chen & Vitetta, 2020; Ferrer-Picón et al., 2020; Magnusson et al., 2019). The IBD microbiota is further characterized by an increase in *Escherichia coli* abundance compared to healthy individuals, and often adherent-invasive *Escherichia coli* (AIEC) strains can be detected, known to induce inflammation (Palmela et al., 2018). Microbial composition is additionally different between the UC and CD subtypes compared to control patients (Pittayanon et al., 2020). Mucosa-associated microbiota also differs from inflamed and non-inflamed segments of the colon in patients with UC and CD (Ryan et al., 2020).

Ulcerative colitis

UC is characterised by the systemic inflammation of the colon and rectum, which increases in a proximal to distal gradient pattern from the anal verge. This inflammation damages the large intestine wall; however, it is limited to the mucosal layer. Other endoscopic features of UC include redness of the mucosa (erythema), loss of vascular pattern, superficial ulcers, erosion and bleeding around the inflamed area. While histological examination reveals changes in the mucosal architecture such as cryptitis, crypts distortion or pseudopolyp formation (DeRoche et al., 2014; Negreanu et al., 2019). These changes in the architecture of the mucosa are the result of chronic inflammation of the colon wall: cryptitis refers to inflammation of the colonic crypts that could lead to distortion of the crypts themselves. Finally, pseudopolyps are formed as a result of the healing process.

In terms of taxonomic composition, UC patients show decreased abundance of *Roseburia*, *Lactobacillus* and *Faecalibacterium* genera among other commensal bacteria with an anti-inflammatory impact. While other genera such as *Bifidobacterium*, *Fusobacterium* and *Campylobacter* are found in higher abundances (Pittayanon et al., 2020). The differences reported between controls and UC patients vary across studies, but most of them show a reduced bacterial diversity alongside an increase in potentially pro-inflammatory taxa.

Crohn's disease

In CD inflamed patches are scattered all along the GIT. In contrast to UC, CD inflammation damages all the layers of the mucosa (transmural inflammation). Deep inflamed patches create a cobblestone appearance that is presented lon-

gitudinally. Other endoscopic and histological evidences include ulcers, strictures, granulomas, crypt abscesses and fistulas.

During CD, pro-inflammatory AIEC correlate with disease location and activity (Palmela et al., 2018). Besides this, *Actinomyces* and *Veillonella* genera are increased in patients with CD, while *Faecalibacterium prausnitzii* is found decreased in several studies (Pittayanon et al., 2020). Similar to UC studies, the differences reported between CD patients and controls are inconsistent across studies (Pittayanon et al., 2020).

1.2.2 Colorectal cancer

Colorectal cancer (CRC) involves the development of malignant neoplasm in the colon, rectum and appendix. Most of the cases start with a polyp, that grows from adenoma to carcinoma over time. Its progression ends up originating metastatic tumors. Firstly, it is originated in the internal colorectal walls and progresses through the mucosa, submucosa and muscularis layers. The event of long-standing IBD can lead to the development of CRC, known as colitis-associated colorectal cancer (CAC) (Jess et al., 2012). As with IBD, CRC aetiology is unknown although several factors have been associated with its development such as age, genetics, diet and smoking (Harmon et al., 2017; Keum & Giovannucci, 2019; Law et al., 2019; Tsoi et al., 2009).

In the early stages, the appearance of neoplasms is often correlated to local inflammation, that contributes to tumour development and progression (Long et al., 2017). CRC displays aberrant mucus characteristics including altered and atypical mucin expression (Myerscough et al., 2001; Velcich et al., 2002; Xiao et al., 2013). In cancer, intestinal epithelial cells are surrounded by mucus, basolateral and apical secretion instead of only apical expression

(Verhulst et al., 2012). This atypical mucin expression reduces the efficacy of anti-cancer drugs and facilitates the attachment and colonization of luminal microbes (Coleman et al., 2018; Li et al., 2019).

Several bacteria have been identified and correlated with the microenvironments formed around tumours (Dejea et al., 2014). The microbiota profile evolves during cancer development, with significant distinct profiles between adenoma and carcinoma stage (Feng et al., 2015; Lu et al., 2016; Luan et al., 2015). Bacteria such as *Fusobacterium nucleatum*, *Bacteroides fragilis* and AIEC are found in higher abundance in CRC tissues. They can secrete pro-inflammatory toxins and alter host intestinal permeability (Denizot et al., 2012; Ellermann et al., 2015; Flanagan et al., 2014; Sears & Pardoll, 2011). Some of these bacteria have also been associated with other inflammatory conditions as IBD or appendicitis (Pittayanon et al., 2020; Strauss et al., 2011; Swidsinski et al., 2012). Overall, the gut microbiota is involved in CRC pathogenesis by modulating host's immune response, promoting tumour growth through inflammation (Coker et al., 2019; Li et al., 2019; Long et al., 2017; Wong & Yu, 2019). Different models have been proposed to understand how the microbiota could promote a chronic mucosal immune response and ultimately lead to CRC (**Figure 1.5**) (Van Raay & Allen-Vercoe, 2017).

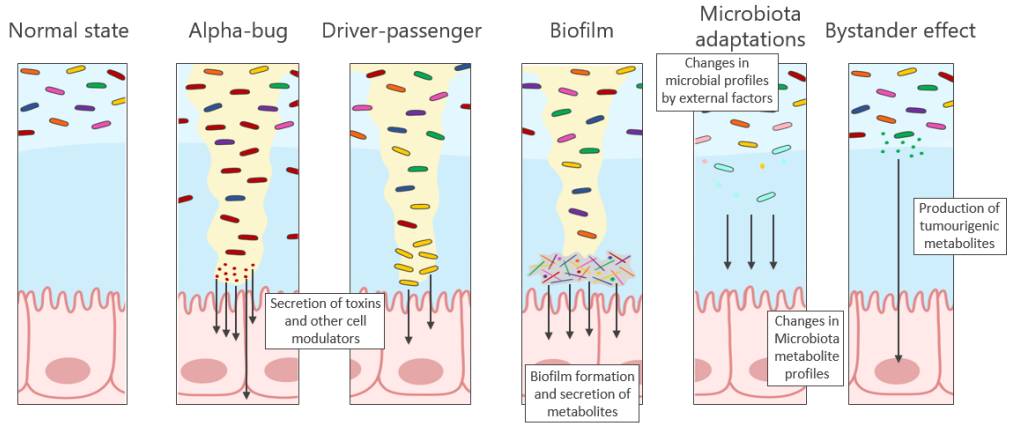

Figure 1.5. Proposed routes for microbial involvement in CRC. In **normal state**, the inner mucus layer is intact. In the **alpha-bug** hypothesis, an alpha microbe (red) alters the microbiota composition and disrupt the mucosal barrier integrity, through the secretion of molecules. In the **driver-passenger** hypothesis, a driver microbe (red) disrupts the mucus layer and create a metabolic environment favouring other bacterial species (passengers, yellow). These passengers compete with the alpha-bug microbe and drive a pro-tumour host response. In the **biofilm** hypothesis, certain bacteria form biofilms and alter host's metabolism. In the **microbiota adaptation** hypothesis, the microbiota composition changes due to external factors (e.g. diet and drugs) and produce pro-tumour metabolites. In the **bystander effect** hypothesis, bacteria produce certain metabolites which promote tumourigenesis. Design taken and adapted from (Van Raay & Allen-Vercoe, 2017).

For example, *Fusobacterium spp.* are proposed as passenger bacteria in CRC. They are found to be enriched in CRC patients. The metabolic signature in a tumour microenvironment benefits their growth, and they are positively associated with more advanced tumour stage (Amitay et al., 2017; Garza et al., 2020).

Finally, bacterial involvement in CRC includes the expression of tumour-associated factors, the recruitment of tumour-associated cells and the suppression of the local immune system (Dejea et al., 2018; Wong & Yu, 2019).

1.3 Prevalent microbial kingdoms in the human gut

The GIT is mainly colonised by bacteria, but it also harbours Archaea, Eukaryotes and viruses whose contribution to host's health is typically overlooked (Andersen et al., 2013; Horz, 2015; Kapitan et al., 2019). These build a complex microbial community with interdependencies that can influence the host's immune system and health. In recent years, the role of these different microbial 'kingdoms' to understand the human gut microbiota dynamics through different stages of life in health and disease.

1.3.1 Archaea

Archaea have gained much attention in the recent years due to their ubiquity, deep phylogenies and often extremophile lifestyle. However, the presence of this domain and its role in the GIT microbiota (archaeome) is often overlooked. The archaeome represents a small fraction of the human microbiome, and there are no known pathogenic archaeon that could motivate the investigation of this group (Gill & Brinkman, 2011). Due to this limited research attention, there are no standardise detection protocols and the already published methods show low reproducibility (Adams et al., 2015). Results vary across studies depending on the DNA isolation protocol, selection of primers and sequencing processing pipeline (Koskinen et al., 2017). Additionally, these microbes are usually difficult to cultivate, further complicating their study (Bang & Schmitz, 2018).

Archaeal diversity differs biogeographically along the human body and within the colon (Koskinen et al., 2017). They colonise the human GIT in the first months of life, although their presence and abundance are incon-

sistent and vary among individuals (Rao et al., 2021). Archaeal estimated abundance in the human gut microbiota varies from 0.1% to 20% (Eckburg et al., 2005; Kim et al., 2020). The phylum Euryarchaeota dominates the GIT, specially the *Methanobrevibacter* genus (Hoffmann et al., 2013). The presence of methanogens has been positively associated with high carbohydrates diets. Due to its unique methanogenic metabolism, *Methanobrevibacter smithii* is a keystone species in the gut microbiome, enabling fermentation of carbon dioxide and fermentative dihydrogen into methane (Camara et al., 2021). This metabolic process enhances the efficiency of bacterial short chain fatty acid fermentation. However, non-methanogenic halophilic and ammonia oxidizing archaea can also be detected in the gut, but this seems to depend on the host diet and geographical location (Kim et al., 2020; Nkamga et al., 2017).

In relation with other kingdoms, *Methanobrevibacter* genus is often positively associated with *Prevotella*, *Candida* and *Saccharomyces* genera (Hoffmann et al., 2013). *Nitrosophaera* -an ammonia-oxidizing archaeon- is negatively associated with *Candida* and *Saccharomyces* fungi. Both associations, from *Methanobrevibacter* and *Nitrosophaera*, seem mutually exclusive and hints on how diet is an important driving factor of the archeaome composition (Hoffmann et al., 2013).

1.3.2 Fungi

During the last two decades fungal research has moved from studying fungal pathogens in the GIT to their role as a part of the normal gut microbiota (termed mycobiota, or mycobiome). However, some challenges still have to be addressed when studying the gut mycobiome. For instance, fungal abundance is estimated to be around 0.1% of the gut microbiome, hindering its cultivation

and DNA isolation from a dense pool of host and bacterial DNA (Qin et al., 2010). Besides this, there is still a lot of missing taxonomic information in fungal databases and fungal detection varies strongly between DNA extraction methods and primer choice (Frau et al., 2019).

Fungi contributes to the early gut colonisation and host immune system development (Iliev & Leonardi, 2017; van Tilburg Bernardes et al., 2020). Although fungal dynamics are not predictable during the maturation of the microbiota, some fungi such as *Candida* spp. shape the assembly of bacterial communities during the first weeks of life by inhibiting the growth of some early bacterial colonizers such as *Klebsiella* and *Escherichia* genera (Rao et al., 2021). Among fungi, *Candida* and *Saccharomyces* genera are normally predominant in the gut, and their dominance is considered a signature of a healthy mycobiome (Hallen-Adams & Suhr, 2017). As the microbiome maturates, fungal diversity and absolute abundance decreases while those of bacteria increase. This plateau is kept over time but is reversed in elderly (Rao et al., 2021; Strati et al., 2016). This negative correlation suggests cross-kingdom competition between both kingdoms despite other positive associations between specific taxa. Finally, as happens with the most dominant genera from Archaea, *Candida* genus has been positively correlated with *Prevotella* and negatively with *Bacteroides* genera (Hoffmann et al., 2013).

In relation with microbiome-associated diseases, the fungal microbiota seems altered in IBD patients. Specifically, fungi are favoured at the expense of bacterial diversity during CD (Moyes & Naglik, 2012; Sokol et al., 2017). For CRC patients, the mycobiome composition differs at early and late stages of the disease. Fungal diversity is decreased in adenoma biopsies compared with control biopsy samples; however, diversity seems to increase in late-stage

CRC. Despite the diversity loss during CRC onset, the appearance of opportunistic pathogens, such as *Candida* and *Phoma* genera, has been reported in patients with adenoma, early and late CRC stages. Overall, It seems that disease stage is closely related to changes in the mycobiome (Gao et al., 2017; Luan et al., 2015).

1.3.3 Other Eukaryotes

Besides fungi, other micro-eukaryotes can be found in the human gut microbiota. These are commonly considered parasites that belong to Protozoa and Helminth groups. Most studies researching these are focused on the micro-eukaryotic potential to modulate the host immune response as part of their parasitic activity. Their participation in the gut communities as commensals remains controversial (Newbold et al., 2017; Rook, 2012).

A few studies have looked into their diversity and stability in the gut microbiota. They are found in low abundances and low diversity in newborns with high variability between subjects, as happens with Archaea (Wampach et al., 2017). Their colonisation does not seem stable during the microbiota maturation, but they are present and temporarily stable in adults GIT (Scanlan & Marchesi, 2008). Some protozoans are associated with increased bacterial diversity, and overall protozoans seem to have a bigger impact in the bacterial communities than helminths (Audebert et al., 2016). A recurrent protozoan in the human gut is *Blastocystis* spp., the successful colonization of which has been associated with both a healthy gut microbiome and gastrointestinal disorders (Audebert et al., 2016; Rossen et al., 2015; Scanlan et al., 2014; Toychiev et al., 2021). Microbial GIT communities dominated by *Ruminococcus* and *Prevotella* seem more prone to contain *Blastocystis*, while *Bacteroides*

dominant communities are negatively associated with the presence of this protozoan (Andersen et al., 2015). Finally, different *Blastocystis* subtypes has been positively and negatively associated with UC in several studies (Rossen et al., 2015; Toychiev et al., 2021; Z. Wu et al., 2014). For those subtypes found in high prevalence in UC patients, the combination of anti-parasitic and anti-inflammatory therapy achieved better mucosal healing than regular anti-inflammatory therapy (Tai et al., 2011; Toychiev et al., 2021). This highlights their involvement and negative impact as parasites in some diseases. For CRC, *Blastocystis* spp. have been proposed as opportunistic pathogens and associated with carcinogenesis (Chandramathi et al., 2010; Toychiev et al., 2018).

1.3.4 Viruses

Finally, the colon also harbours bacteriophages and small proportion of eukaryotic viruses (virobiota or virome) which can be associated with prokaryotes (prokaryotic virobiota), host cells and other eukaryotes (eukaryotic virobiota). Viral abundance is directly related to the environment and microbial colonization; thus, the main viral colonizers are DNA bacteriophages (Hoyles et al., 2014). Their association with commensal members and role in the host immune system is becoming steadily more recognised beyond their inherent infectious properties. However, due to small genome sizes (relative to bacteria and micro-eukaryotes), lack of universally conserved genomic region and their high genetic and morphologic variability makes it difficul to study this group (Reyes et al., 2012). In this way, large viruses and RNA viral genomes tend to be often overlooked in most experimental workflows (Conceição-Neto et al., 2015). Being such a highly heterogeneous group, their study requires the development of complex techniques to separate them from microbes and host

cells (Conceição-Neto et al., 2015). Despite the above-mentioned impediments, recent studies have shown the relation of the virome with the commensal bacterial microbiota and its association with host health.

The interactions between the commensal microbiota and virobiota can be beneficial and detrimental for the host health. For example, commensal bacteria can confer limited protection against viral infections by stimulating a host immune response, or can enable viral evasion of the host immune system (Kernbauer et al., 2014; Robinson et al., 2014; Wilks et al., 2015). Further, the virobiota indirectly serves to control the abundance of microbial populations, in particular bacteriophages that keep their target bacteria repressed.

Overall, the virobiota has the potential to enhance the intestinal immune response (Gogokhia et al., 2019). In microbiome-associated diseases such as IBD, the enteric virome composition is significantly different in diseased patients compared to healthy controls (Norman et al., 2015; Pérez-Brocal et al., 2015). Some bacteriophages have been also associated with the unbalance of the gut microbiota during IBD. The increased abundance of *Caudovirales* has been correlated with decreased bacterial diversity in CD patients (Norman et al., 2015). Viral families such as *Hepadnaviridae* and *Hepeviridae* are found enriched in UC and CD patients. The proteins associated to these viruses can indirectly impact host's transcription activity and alter the host immune response (Ungaro et al., 2019). Thus, the presence of specific eukaryotic viruses can trigger and promote local inflammation during IBD onset and progression. Eukaryotic viral infections are also proposed as a risk factor of CRC (Chen et al., 2019; Emlet et al., 2020). The virobiota can promote CRC pathogenesis through the integration and expression of oncogenic viral proteins among other mechanisms (Emlet et al., 2020; Khoury et al., 2013).

1.4 Sequencing-based community exploration

Just around 30% of the taxa recovered from human faecal samples, using metagenomic and 16S sequencing, corresponds to species cultivated in the laboratory (Qin et al., 2010; Walker, Ince, et al., 2011). Sequencing-based methods allow to identify the uncultivable fraction of the human microbiome, despite the recent efforts to improve cultivation strategies (Lewis et al., 2021). For this reason, sequencing-based approaches are a popular choice to study microbial compositions, for example in diseases associated to microbiome variation.

Current sequencing strategies include amplicon sequencing and shotgun metagenomics. Briefly, amplicon sequencing involves the PCR amplification and sequencing of marker genes, whilst shotgun metagenomics aims to capture all the genomic information contained on the sample. During shotgun metagenomics, instead of sequencing a specific marker gene, long DNA fragments from the sample are fragmented and sent for sequencing. Following this, sequenced fragments are assembled forming genes and ultimately genomes. Both approaches require the extraction of DNA from the sample.

Amplicon sequencing relies on the ribosomal RNA (rRNA) gene, encoded in DNA. This gene has a slow evolutionary rate and is ubiquitous across the three domains of life. Thus, the 16S rRNA (Bacteria and Archaea) and 18S rRNA (Eukaryotes) marker genes allows to distinguish between closely related taxa. Amplicon sequencing can also be used in other genomic regions to obtain significant taxonomic information about specific kingdoms, for example the internal transcribed spacer (ITS). This region corresponds to the spacer DNA located between different rRNA genes. Primers specific for ITS are often used

to identify and analyse fungal diversity.

The likelihood of an amplicon depicting an accurate representation of a microbial community is dependant on the amplified gene region and the primers choice when exploring microbial communities. Shotgun metagenomics overcomes PCR amplification bias, and allows a more precise delineation of the community taxonomic composition through the reconstruction of gene sequences. In summary, both approaches use taxonomic composition to quantify the relative contribution of different microbial taxa to the host health, and explore the association of microbial taxa with co-varying factors.

1.4.1 Amplicon sequencing pipelines

In amplicon sequencing, amplicons of the target genes are sequenced via 2nd or 3rd generation sequencers. To process the raw sequencing reads, bioinformatic pipelines have been developed that can semi-automatically quality control and cluster sequences, to estimate the phylogenetic diversity and abundance of microbes in a given sample. Amplicon sequencing pipelines are used to process the data from the sequenced 16S, 18S, ITS amplicons, although they can be used with other marker genes. The standard workflow of these amplicon processing pipelines includes demultiplexing, quality filtering, sequences clustering and taxonomic assignment (Estaki et al., 2020; Hildebrand et al., 2014; Schloss et al., 2009) These will be detailed in the following on the workflow of the LotuS pipeline.

First the sequences are quality filtered based on the sequence length, GC content, repeated nucleotides and quality of the base calls. Raw sequences are assigned to the sample they originated from, in a step known as demultiplexing. The resulted sequences can be clustered into operational taxonomic

units (OTUs). OTUs are clusters of reads that share 97% of similarity (Edgar, 2013). Instead of OTUs, reads can also be corrected for amplification and sequencing errors forming amplicon sequence variants (ASVs) (Callahan et al., 2016). Chimeric sequences are removed (Edgar et al., 2011). Following this, OTUs or ASVs are compared to a reference database for taxonomic assignment, using mappers such as lambda, Blast and vsearch (Altschul et al., 1990; Hauswedell et al., 2014; Rognes et al., 2016). Finally, a feature abundance matrix is created for different taxonomic levels.

Some 16S amplicon processing workflows include correcting bioinformatically for 16S copy number variations, as bacterial cells can contain more than one 16S copy (Louca et al., 2018). However, these copy variations are also found within the same species, complicating the data analysis

1.4.2 Quantitative microbiome profiling

After processing the sequenced data, microbiome datasets are often converted to relative abundance values or normalised prior to analysis. Total read counts cannot inform on the absolute abundance of microbes, as the total number of reads depends on the fixed capacity of the sequencing instrument. For this reason, ecological data needs to be treated as compositional data. Although we can detect microbiome variations using relative abundance values. Compositional data from taxonomic information can lead to false positive associations (Hawinkel et al., 2019; Vandeputte et al., 2017). For example, any shift in relative abundance is important for studying interaction networks based on co-occurrence, however, the co-occurrence patterns are highly susceptible to the microbial load (**Figure 1.6**). This applies to any differential abundance analysis when microbial abundances from sequencing data are reported in a

relative manner. To avoid such false positive associations, it is important to consider the total microbial load along with the relative microbiome profiling (RMP). This allows the absolute microbial quantification, known as quantitative microbiome profiling (QMP), which increases the specificity and resolution of sequencing-based analysis (Props et al., 2017; Satinsky et al., 2013; Stämmler et al., 2016).

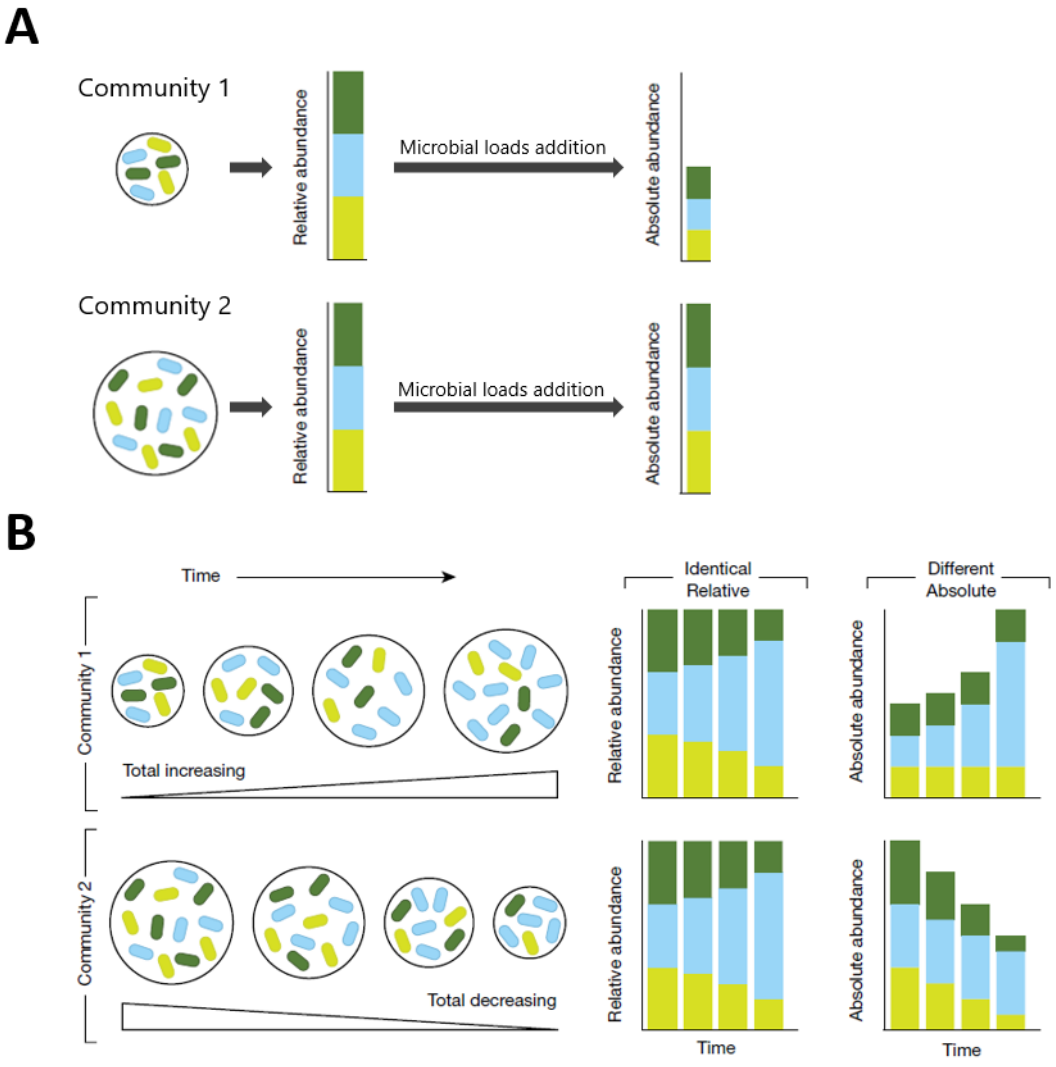

Figure 1.6. Relative versus quantitative microbiome profiling (**A**) Schematic illustrating how two communities with different microbial loads can have the same relative abundance profile but distinct absolute abundances. (**B**) Schematic illustrating how relative abundance data can mask community dynamics. Adapted from (Rao et al., 2021).

Techniques for absolute quantification: Real-time polymerase chain reaction

Several techniques are available to quantify the community members absolute abundances. These can be separated in two groups: cell-based and molecular-based techniques. The most popular techniques within these groups are flow cytometry and qPCR respectively.

Real-time polymerase chain reaction, or quantitative polymerase chain reaction (qPCR), is a variant of a PCR to amplify and quantify simultaneously the PCR reaction product. This technique allows the relative and absolute quantification of a gene of interest. The amplified DNA can be detected by adding non-specific fluorescent dyes that intercalates with double stranded DNA, or using specific DNA probes that contains fluorescent labels. Both approaches rely on the detection of the resulting fluorescent signal. This signal is later compared to the signal produced by a negative and positive control.

In microbial ecology, this technique is widely used to measure the number of copies of a gene of interest (absolute quantification). It requires a standard curve, which is constructed using dilutions of a sample of known concentration called standard. In this way, the amount of the sample target template is inferred by a known template quantity (Bonk et al., 2018).

Compared to flow cytometry, qPCR is cheaper, easier to perform and overall more accessible; while flow cytometry is laborious to prepare and requires specialized equipment and substantial technical expertise. However, qPCR results are susceptible to biases during DNA extraction and PCR amplification, steps that are not included in the flow cytometry workflow. Additionally, flow cytometry has a reduced less intra-sample variability and higher sensitivity than qPCR (Jian et al., 2021). Despite all these differences, both techniques generate comparable quantitative microbiome profiles (Jian et al., 2021; Rao et al., 2021; Vandeputte et al., 2017).

1.5 General objective

For this thesis, we aim to explore the microbial composition of the human gut using a broad and precise approach, enabling us to understand better the pathogenesis of gastrointestinal diseases associated with microbial mucosal colonic shifts. For this purpose, we plan to study the microbial intestinal mucus using a polykingdom approach which includes bacteria, fungi, archaea, protists, metazoan and viruses. Besides this, we plan to collect colonic tissue of different regions of the colon instead of faecal samples to obtain an accurate depiction of the colonic microbial communities. Lastly, we want to explore different variables from the experimental design and samples metadata that could potentially impact microbial distribution.

The microbial community members of the colon are identified using the amplicon sequencing technique instead of shotgun metagenomics to avoid the sequencing of host cells. In addition to amplicon sequencing, we use the qPCR technique to quantify the community members absolute abundances.

Finally, we expect to develop an experimental and analytical pipeline that allows a better understanding of the microbial dynamics occurring on the human colon. Therefore, we expect to obtain a distinct and more diverse microbial profile after accounting for different microbial kingdoms and confounders, and using a QMP approach in opposition to RMP.

Chapter 2

Methods for quantifying and sequencing polykingdom mucosal communities

2.1 Introduction

The gut mucosal-associated microbiome in IBD is severely understudied, with the majority of studies assessing the luminal microbiota composition using faecal samples. This is a shortcoming as it will poorly represent the mucosa-associated microorganisms in the GIT (Gevers et al., 2014; Pittayanon et al., 2020). In addition to this, most of the studies are focused on bacteria and tend to overlook the association of different kingdoms including fungi, archaea, protists, metazoan and viruses (polykingdom) with the gut microbiota, and ultimately within IBD pathogenesis.

In this project, we aim to study the microbial composition of the intestinal mucus in IBD using a polykingdom approach; specifically, we plan to col-

lect and analyse human mucosal biopsies along different regions of the colon, with samples collected from IBD patients and healthy control individuals. To address this, firstly, the intestinal mucus needs to be isolated from the gut samples to reduce host cell contamination. A washing procedure was adapted and used for this purpose on colonic tissue from mice. Laboratory optimisations will be based on murine samples and the optimised methods will be used to process the human samples. Once the mucus is isolated, amplicon sequencing and qPCR will be used in parallel to enable a QMP approach. The amplification of 16S, 18S and ITS1 marker genes will allow for the identification of prokaryotic, eukaryotic and fungal microbiome, respectively.

It is important to note that the above workflow could potentially bring about some issues; the DNA extraction method used to isolate the genetic material could affect the way in which the different microbial kingdoms are represented, bringing about biases in observation. Additionally, from the time the sample is collected to the point of DNA extraction, the microbial composition could change and differ from the samples original state. These issues highlight the importance of evaluating the ways in which different DNA extraction protocols and storage solutions can impact microbial representation, which are summarised in the following objectives:

Objective 1. Test and compare three different DNA extraction kits in terms of DNA quantity, DNA quality, number of 16S copies and biasing microbial composition.

Objective 2. Test phosphate-buffered saline (PBS) and RNA*later* storage solutions to preserve cells integrity prior to the washing procedure and their impact on downstream analysis steps.

2.2 Materials and Methods

2.2.1 Sample collection and preparation of mucus samples

Colon samples were obtained from 3 to 4 week old *C57/B6J* mice (Disease Model Unit at the University of East Anglia, UK). These were stored in sterile PBS or RNA*later* (Invitrogen) stabilization and storage solution. A piece of tissue (up to 10 mg) was collected from the proximal colon and stored in pre-weighted tubes. The number of samples collected from the colon was determined by the number of different DNA extraction protocols to test. Samples were kept on ice and processed after collection. The pieces of tissue were subjected to three washes to release the microbes from the mucus (washing procedure adapted from Li et al., 2003). These were performed by transferring the gut samples into a 15 mL conical tube containing 12 mL of saline solution with 0.1% Tween 80 (4 mL per piece of tissue). Then the sample was shaken for 1 minute and transferred into a new tube. Once finished, the pieces of tissue were removed and the three washes were pooled together and centrifuged at 11.200 g for 30 minutes at 4°C. The supernatant was discarded and the pellet containing the microbial cells was resuspended in 1.8 mL of sterile PBS (600 µL per each DNA extraction protocol) for further processing. Samples were prepared in duplicates.

2.2.2 Microbial DNA Extraction

Three different DNA extraction kits were used in duplicates to extract microbial DNA from the mucus samples.

FastDNA™ SPIN Kit for Soil (MP Biomedicals). The protocol was followed according the manufacturer's standard protocol with some modifica-

tions (*FastDNA™ SPIN Kit for Soil detailed protocol*) (**Figure 2.1**).

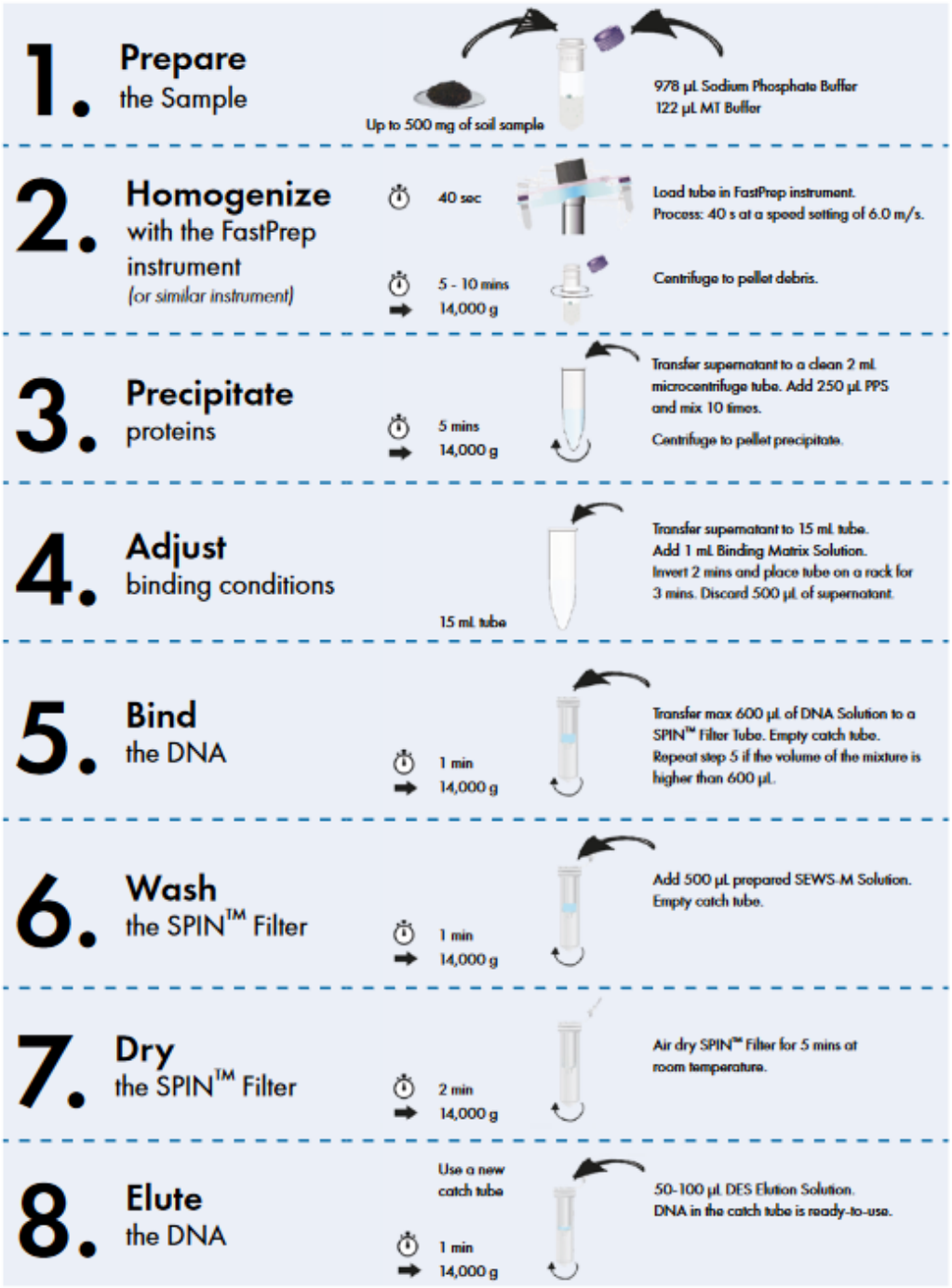

Figure 2.1. Diagram of FastDNA™ SPIN Kit for Soil detailed protocol. Workflow from FastDNA™ SPIN Kit for Soil manual (MP Biomedicals).

Once the MT buffer was added in step 1, the sample was incubated for 40 minutes at 4 °C. The homogenization in the Fastprep instrument (MP Biomedicals) during step 2 lasted 40 seconds and was repeated three times, with 1 minute of resting between each bead-beating repetition. After this step, the sample was centrifuged at 14,000 g for 10 minutes. In step 4, 800 µL of supernatant were removed and discard. Step 6 was repeated. Step 8 was also repeated using an elution volume of 50 µL.

QIAamp® PowerFecal® Pro DNA Kit (Qiagen). The protocol was followed according the manufacturer's standard protocol with modifications. The sample was homogenised using a Vortex Adapter at maximum speed for 10 minutes. In the step 16, 50 µL of Solution C6 was added.

Maxwell® RSC PureFood GMO and Authentication Kit (Promega). The protocol was followed according the manufacturer's standard protocol without modifications (*Purifying DNA on the Maxwell® Instruments*).

2.2.3 Evaluation of DNA quality and yield

DNA quality was determined spectrophotometrically using NanoDrop™ ND-2000 (NanoDrop Technologies). Purity and intactness were assessed with the 260/280 and 260/230 absorbance ratios. DNA yield was measured fluorometrically using Qubit® Fluorometer 3.0 with the High Sensitivity dsDNA kit (Invitrogen) according to the manufacturer's protocol. The DNA concentration obtained was normalised using the tissue weight of each sample.

2.2.4 Quantitative microbiome profiling

Quantitative PCR assessment of microbial loads

Samples were diluted to the lowest DNA concentration observed among the samples and used as a template for qPCR amplification of the V6 hypervariable region of the 16S rRNA (1048F: GTGSTGCAYGGYYGTCGTCA, 1194R: ACGTCRTCCMCNCCTTCCTC). Each reaction contained 3 µL of SYBR® Select Master Mix (Applied Biosystems), 1.5 µL of template DNA and two times 0.2 µL primer solution (10nM), and 1.6 µL PCR grade water, in a total reaction volume of 6.5 µL. The amplification was performed in a ViiA7 qPCR detection system (Applied Biosystems). The amplification program comprised of 2 stages; an initial denaturation step at 95 °C for 180 s, followed by 40 two-step cycles at 95 °C for 15 s and at 60 °C for 60 s. At the end of the run, a melting curve analysis was performed. Cycle threshold value was determined using the ViiA7 qPCR detection system software (v 1.2) (Thermo Fisher Scientific). A calibration curve was run with each sample set. This reference curve was generated by using serial dilutions from a stock of known 16S copies concentration. All reactions were run in duplicates.

Library preparation

For the first PCR, the 16S rRNA V4 amplicons were amplified using 0.4 µL of each primer (515F: GTGYCAGCMGCCGCGGTAA, 806R: GGACTACNVGGGTWTCTAAT) (Caporaso et al., 2011; Walters et al., 2016), 5 µL equimolar amounts of DNA as a template, 12.5 µL of GoTaq® G2 Green Master master mix (Promega) and 6.7 µL of PCR grade water. This makes a total reaction volume of 25 µL. The program used included 95 °C for 5 minutes, 30

cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s followed by a final extension at 72 °C for 5 minutes.

The following part was performed by Dave Baker (Quadram Institute Bioscience).

Following the first PCR and clean-up a second PCR master mix was made up using 4 ul kapa2G buffer, 0.4 µl dNTP's, 0.08 µl Polymerase, and 6.52 µl PCR grade water (contained in the Kap2G Robust PCR kit) per sample and added to each well to be used in a 96-well plate. 2 µl of each P7 and P5 of Nextera XT Index Kit v2 index primers (Illumina Catalogue No. FC-131-2001 to 2004) were added to each well. Finally, the 5 µl of the clean specific PCR mix was added and mixed. The PCR was run using 95 °C for 5 minutes, 10 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s followed by a final 72C for 5 minutes. Following the PCR reaction, the libraries were quantified using the Quant-iT dsDNA Assay Kit, high sensitivity kit (Catalogue No. 10164582) and run on a FLUOstar Optima plate reader. Libraries were pooled following quantification in equal quantities. The final pool was cleaned using 0.7X SPRI using KAPA Pure Beads. The final pool was quantified on a Qubit 3.0 instrument and run on a High Sensitivity D1000 ScreenTape (Agilent Catalogue No. 5067-5579) using the Agilent Tapestation 4200 to calculate the final library pool molarity.

The pool was run at a final concentration of 8pM on an Illumina MiSeq instrument using MiSeq® Reagent Kit v3 (600 cycle) (Illumina Catalogue FC-102-3003) following the Illumina recommended denaturation and loading recommendations which included a 20% PhiX spike in (PhiX Control v3 Illumina Catalogue FC-110-3001). The raw data was analysed locally on the MiSeq using MiSeq reporter.

Processing of sequencing reads

The LotuS pipeline (v 1.65) (Hildebrand et al., 2014) was used to process the V4 16S rRNA amplicon sequences. The steps include demultiplexing, quality-filtering, clustering of quality-filtered reads with uparse (Edgar, 2013) at 97% identity. Chimeric OTUs were filtered out with uchime2 (Edgar et al., 2011). For taxonomic annotation, OTU seed sequences were aligned to the reference database SILVA (v 123) and the LotuS LCA algorithm using lambda aligner (Hauswedell et al., 2014; Quast et al., 2013).

The command for LotuS utilised the following parameters:

-s sdm_miSeq.txt -xtalk 1 -keepUnclassified 1 -p miSeq -keepTmpFiles 0 -id 0.97 -simBasedTaxo lambda -refDB SLV -amplicon_type SSU -tax_group bacteria -derepMin 8:1,4:2,3:3 -saveDemultiplex 0 -CL unoise -thr 12 -exe 0

2.2.5 Statistical analysis

Statistical analyses were performed using R (v 3.6.1). After processing the sequencing reads with LotuS, we obtained a 16S OTU abundance table, containing the OTU frequencies for each sample, and taxa abundance tables for different taxonomic levels. Samples from the 16S OTU abundance tables were rarefied 200,000 reads per sample with R-package rtk (v 0.2.6.1) (Saary et al., 2017). The smallest number of sequences per sample observed was used to choose this rarefaction depth, rarefactions were repeated 10 times. Rarefactions were used to calculate the mean Shannon's diversity index and taxa richness estimate (chao1) of each sample from 16S OTU frequencies. To obtain QMP profiles, taxa abundance tables were normalised by the number of 16S copies for each sample, as estimated using qPCR. Shapiro-Wilk test of nor-

mality was used to assess whether the data are normally distributed. Multiple pairwise comparisons could not be performed between the DNA extraction methods due to the limited number of sequenced samples.

2.3 Results

2.3.1 Effect of DNA extraction method on DNA quantity and quality

We obtained colon samples from 2 different healthy mice both 3 – 4 weeks old and isolated the microbes from the mucus. These samples were pooled together and split across the different DNA extraction methods and storage treatments (n = 19). DNA from mucus samples was extracted using three different extraction methods (**Table 2.1**). Extraction analysis is presented in terms of quality and quantity.

Table 2.1. Summary of DNA extraction methods.

Extraction method	Abbreviation	Lysis method	Purification method	Elution volume (µL)
FastDNA™ SPIN Kit for Soil	MP	Mechanical - Matrix beating (homogenizer)	Spin silica column	100
Maxwell® RSC PureFood GMO and Authentication Kit	PR	Chemical - Lysis buffer	Magnetic beads	100
QIAamp® PowerFecal® Pro DNA Kit	QI	Mechanical - Bead beating (vortex) Chemical - Lysis buffer	Spin silica column	50

DNA Quantity

DNA quantity was characterised by fluorometry using Qubit high sensitivity dsDNA. This assay is highly selective for double-stranded DNA over RNA unlike the absorbance 260/280 ratio.

Overall, there was no significant difference between the three DNA extraction methods on the concentration of DNA obtained (**Figure 2.2A**). Although *PR* seems to yield the highest amount of DNA per tissue weight, and *MP* had the extracts with the lowest concentrations of DNA.

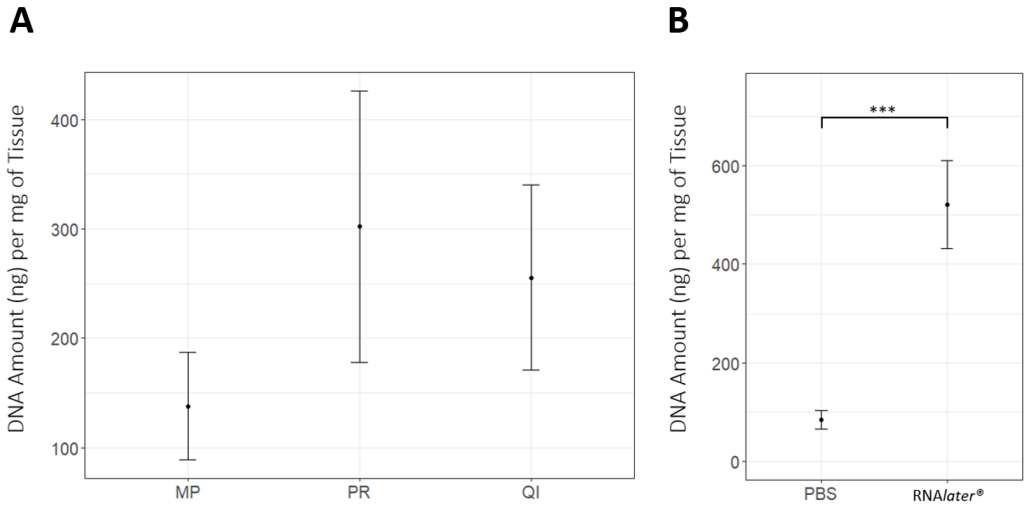

Figure 2.2. (**A**) **Differences in DNA amount per mg of tissue between DNA extraction methods (mean ± SEM).** Error bars represent standard error of the mean. Not significant differences in DNA amount per mg of tissue between DNA extraction methods. One-way ANOVA, $P > 0.05$. (**B**) **Differences in DNA amount per mg of tissue between PBS and RNA*later* buffers (mean ± SEM).** Error bars represent standard error of the mean. Significant differences in DNA amount per mg of tissue between storage solutions. Kruskal-Wallis test, *** $P < 0.001$.

The samples conserved on the RNA*later* buffer yielded a significantly higher amount of DNA per mg of tissue than samples in PBS (Kruskal-Wallis: X2(1) = 11.38, P = 7E-04) (**Figure 2.2B**). RNA*later* buffer-treated samples yielded approximately six times more DNA than PBS. All extracts among the three DNA extraction kits yield higher amount of DNA when RNA*later* was used (**Figure 2.3**). *PR* extraction method obtained the highest amount of DNA from extracts treated with RNA*later*, followed by *QI* and *MP* methods. At the same time, *PR* also had the highest standard deviation among the three kits when using RNA*later*. While *QI* method yielded higher concentration of DNA than the other kits when PBS was used. However, *QI* also had the highest standard deviation among the three kits when using PBS.

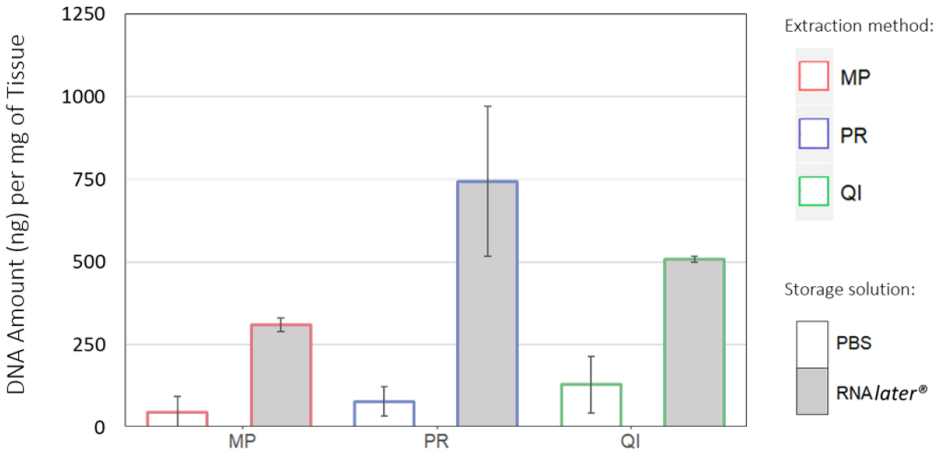

Figure 2.3. Amount of DNA per tissue weight among the three DNA extraction methods and storage solutions (mean ± SD). Error bars represent the standard deviation. Different DNA extraction methods are differentiated by the border colour: *MP*, red, *PR*, blue, *QI*, green. Storage buffers are differentiated by the fill colour: phosphate-buffered saline, white and RNA*later*, grey.

Finally, *PR* was the DNA extraction method with the highest DNA concentration, followed by *QI*. The latter kit yielded the highest DNA concentration with samples stored in PBS; however, this tendency changes when using RNA*later*.

DNA Quality

DNA quality is measured by DNA purity and DNA intactness, which indicate the level of contaminants and suitability for PCR amplification and sequencing. DNA purity and integrity was assessed using the absorbance 260/280 ratio across the three methods (**Figure 2.4**).

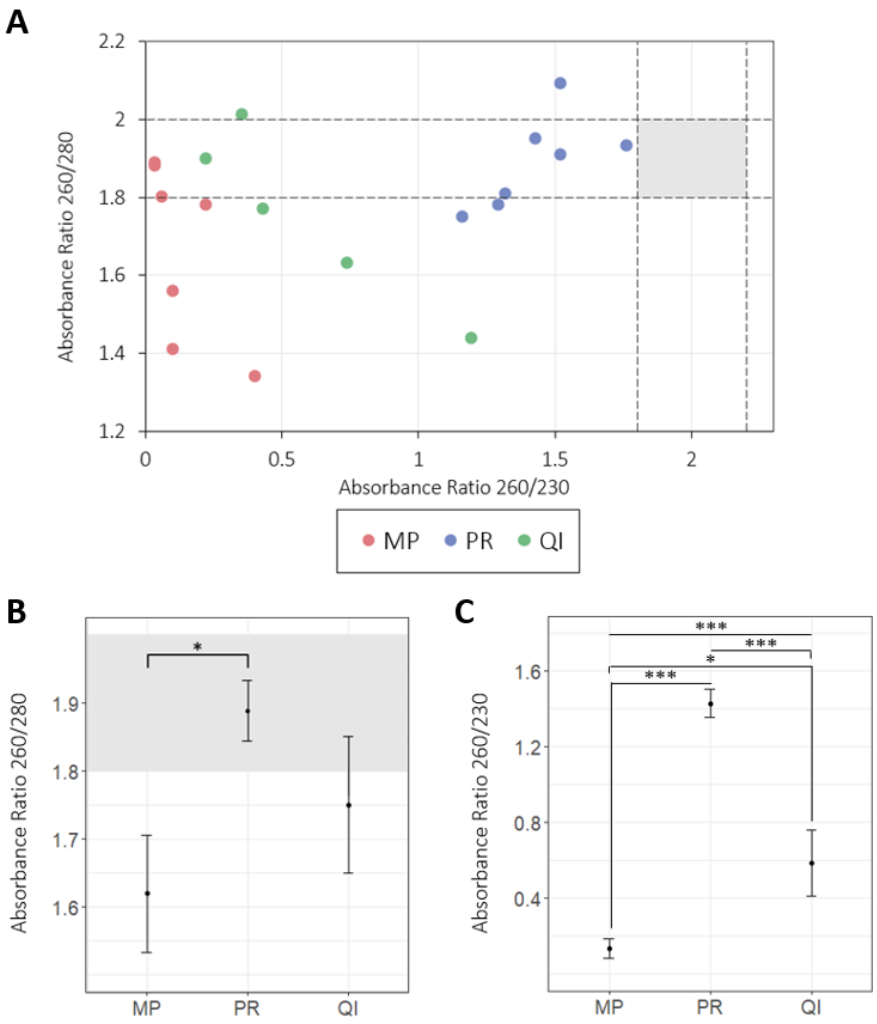

Figure 2.4. (**A**) **Scatter plot DNA extract quality (A_{260}/A_{280} nm vs. A_{260}/A_{230} nm).** The used DNA extraction methods are indicated by colour: FastDNA™ SPIN Kit for Soil (n = 7 extracts), red, Maxwell® RSC PureFood GMO and Authentication Kit (n = 7 extracts), blue, QIAamp® PowerFecal® Pro DNA Kit (n = 5 extracts), green. Dotted lines indicate the upper and lower limits of accepted absorbance ratios. Grey area delimits the accepted range for both ratios. (**B**) **Differences in A_{260}/A_{280} ratios between DNA extraction methods (mean ± SEM)**. Error bars represent standard error. Grey area delimits the accepted range (1.8 - 2.0). Significant differences in absorbance 260/280 ratios between *PR* and *MP* methods. Tukey's HSD test, $^{*}P < 0.05$. (**C**) **Differences in A_{260}/A_{230} ratios between DNA extraction methods (mean ± SEM)**. Error bars represent standard error. Samples out of the accepted range (accepted range corresponds to 1.8 – 2.2). Significance of differences in absorbance 260/230 ratios among different DNA extraction methods was tested using a One-way ANOVA, $^{***}P < 0.001$. Significant differences in absorbance 260/230 ratios between *PR* and *MP* methods. Tukey's HSD test, $^{***}P < 0.001$. Significant differences in absorbance 260/230 ratios between *MP* and *QI* methods. Tukey's HSD test, $^{*}P < 0.05$. Significant differences in absorbance 260/230 ratios between *PR* and *QI* methods. Tukey's HSD test, $^{***}P < 0.001$.

Overall, *PR* and *QI* extraction methods mean (± SEM) A_{260}/A_{280} ratios are within or near the accepted range of 1.8 to 2.0 (**Figure 2.4B**), where *PR* method (± SEM) had the highest proportion of extracts with accepted ratios (4/7) (**Figure 2.4A**). Although *QI* mean ratio was near the accepted range, only 1 out of 5 extracts has an accepted ratio. *MP* extraction method had the lowest mean (± SEM) A_{260}/A_{280} ratio among the three methods (1.62 ± 0.23), out of the accepted range; most of *MP* extracts had A_{260}/A_{280} ratios below 1.8 (4/7).

There was no significant difference between extraction methods in the absorbance 260/280 ratios obtained (ANOVA: F = 3.387; d.f. = 2, 16; P = 0.0594). However, *PR* method A_{260}/A_{280} ratios were significantly greater than *MP* method ratios (Tukey: P = 0.0479) (**Figure 2.4B**). Maxwell® RSC PureFood GMO and Authentication Kit seems to yield the highest purity and integrity DNA among the three kits.

UV absorbance 260/230 ratio was used as secondary measure to evaluate DNA extracts purity, this ratio indicates the presence of any residual carryover (non-proteic contamination). All extracts among the three different extraction methods had A_{260}/A_{230} ratios below the accepted range of 1.8 to 2.2 (**Figure 2.4A**). *PR* extraction method produced the highest A_{260}/A_{230} ratios (1.43 ± 0.20), significantly higher than *QI* (0.59 ± 0.39) (Tukey: P = 5.88E-05) and *MP* methods (0.13 ± 0.13) (Tukey: P = 1E-07) (**Figure 2.4C**). While *MP* extractions produced the lowest A_{260}/A_{230} ratios, significantly lower than *QI* extraction ratios (Tukey: P = 0.015). Finally, there was a significant difference in absorbance 260/230 ratios between different DNA extraction methods (ANOVA: F = 51.11; d.f. = 2, 16; P = 1.13E-07).

The use of RNA*later* or PBS did not affect the quality of DNA extracts

(Kruskal-Wallis: $P > 0.05$). The extracts were also loaded into a gel to assess their DNA integrity; however, they could not be visualised due to their low DNA concentration as the samples had a low microbial biomass.

2.3.2 Effect of DNA extraction method and storage buffer on prokaryotic absolute abundance

Mucus samples yielded an average of 500 16S gene copies per miligram of tissue, and the maximum number of copies was an average of 1271 16S gene copies, when using the *PR* extraction method combined with RNA*later* sample storage (**Figure 2.5**). While *QI* and *MP* yielded an average of 540 and 520 16S gene copies. When mucus samples were treated with PBS, the *QI* extraction method had the highest number of 16S copies (131 copies), followed by *MP* (98 copies) and *PR* (79 copies).

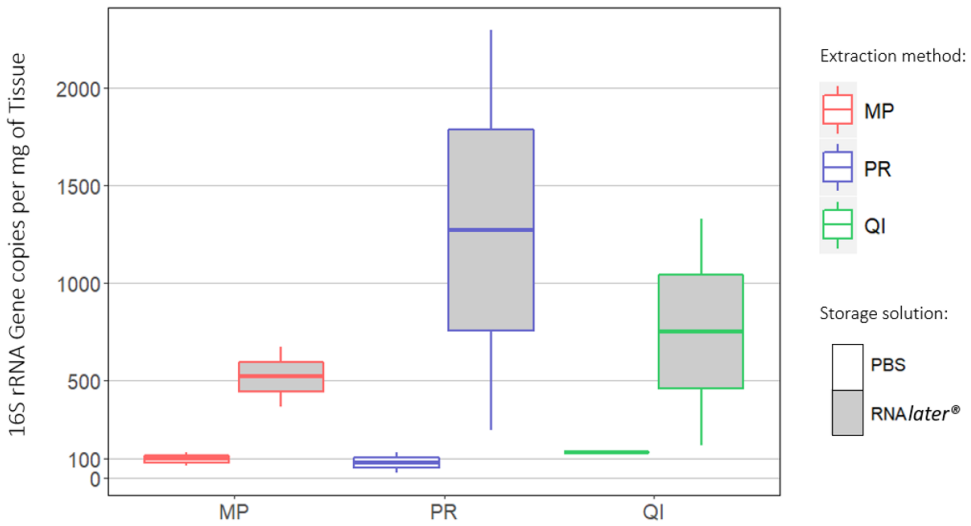

Figure 2.5. Box plot representation of 16S rRNA gene copies per mg of tissue distribution across the three DNA extraction methods and storage solutions. The box ranges from the first to the third quartile of the distribution. The line across the box indicates the median. Different DNA extraction methods are differentiated by the border colour: *MP*, red, *PR*, blue, *QI*, green. Storage buffers are differentiated by the fill colour: phosphate-buffered saline, white and RNA*later*, grey.

There was no significant difference in the number of 16S copies among the three DNA extraction methods (Kruskal-Wallis: $P > 0.05$) (**Figure 2.6A**), and it seems that *PR* extraction method yielded higher counts of 16S on RNA*later*-treated samples than the other extraction kits. However, this cannot be tested due to the limited number of samples.

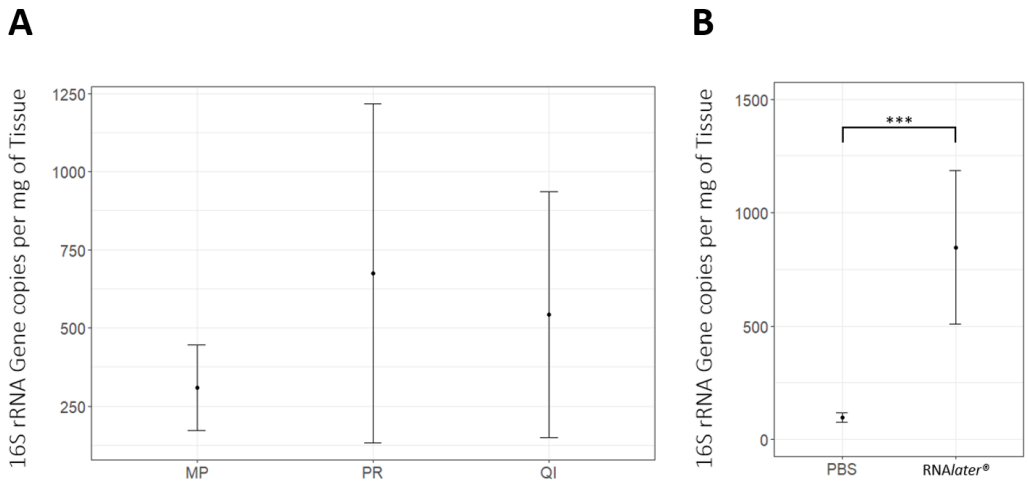

Figure 2.6. (**A**) **Differences in 16S copies per mg of tissue between DNA extraction methods (mean ± SEM).** Error bars represent standard error of the mean. Not significant differences in DNA amount per mg of tissue between DNA extraction methods. Kruskal-Wallis test, $P > 0.05$. (**B**) **Differences in DNA per mg of tissue between PBS and RNA*later* buffers (mean ± SEM).** Error bars represent standard error of the mean. Significant differences in DNA amount per mg of tissue between storage solutions. Kruskal-Wallis test, *** $P < 0.001$.

The mucus samples stored with RNA*later* solution had a significantly higher number of 16S gene copies than the samples under PBS treatment (Kruskal-Wallis: X2(1) = 7.5, P= 6E-03) (**Figure 2.6B**).

Overall, samples treated with RNA*later* buffer yielded higher concentration of DNA and higher number of 16S copies than samples stored with phosphate-buffered saline solution.

2.3.3 Effect of DNA extraction method on microbial composition

We sent a mucus extract from each DNA extraction method for amplicon sequencing. These extracts were derived from the same sample pool to ensure an equal comparison. Mucus microbiota diversity was higher in the *MP* extract (3.76), followed by the *QI* (3.59) and *PR* DNA extractions (3.08). Similarly, the *MP* extract showed higher richness estimate (159.5) than the extracts from *QI* (155.3) and *PR* (120.0).

The murine intestinal tract is dominated by two main bacterial phyla: Bacteroidetes and Firmicutes. Both phyla were abundant in the samples (**Figure 2.7**); their combined relative proportion was 68.9% in the extract from *MP*, 93.0% in the extract from *PR* and 90.2% in the extract from *QI*. The extract obtained from *QI* had the highest Bacteroidetes relative abundance (21.2%), followed by *PR* (19.5%) and *MP* (10.9%). While the extract from *PR* had the highest Firmicutes relative abundance (73.5%), followed by *QI* (68.9%) and *MP* (58.0%). QI also had the highest abundance of Actinobacteriota (5.0%).

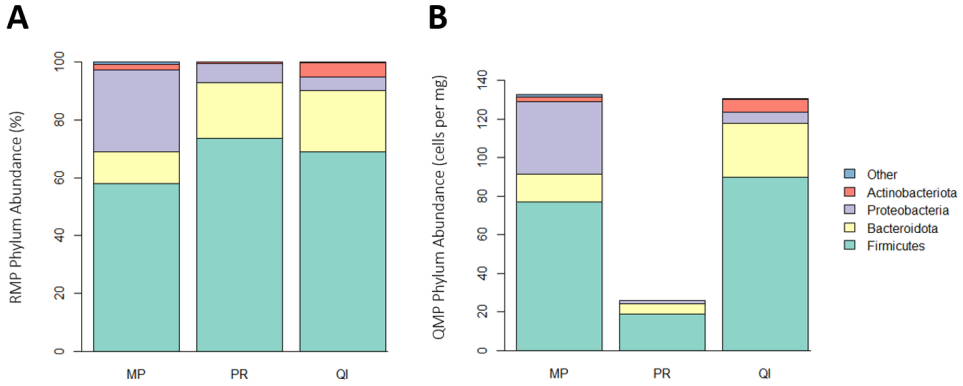

Figure 2.7. Relative versus quantitative microbiome profiling at phylum level. Composition plots of the top 4 phyla among the three samples, with all other phyla pooled into 'Other'. (**A**) **Phylum-level mucus relative microbiome composition among different DNA extraction methods.** Relative microbiota profiles obtained using standard microbiome sequencing methods. (**B**) **Phylum-level mucus quantitative microbiome composition among different DNA extraction methods.** Absolute microbiota profiles obtained using the relative microbial abundance along with prokaryotic cell counts.

The Firmicutes phylum was mostly represented by bacteria of the family Lachnospiraceae, contributing from 31.9% to 44.5% of the total taxonomic composition. This taxon is compromised by anaerobic bacteria that can ferment complex polysaccharides to a wide range of products including butyrate, acetate and ethanol. Thus, this family is commonly found in high abundance in the gut microbiota of mice and humans (Wang et al., 2019).

As expected from previous work on mice gut microbiota, prevalent genera such as *Lactobacillus*, *Bacteroides*, *Blautia*, *Bifidobacterium*, *Enterorhabdus*, *Roseburia* and *Marvinbryantia* were present on the mucus DNA extracts (Wang et al., 2019), where *Lactobacillus*, *Bacteroides*, *Blautia*, *Bifidobacterium* and *Enterorhabdus* were some of the most abundant genera among the samples (**Figure 2.8**).

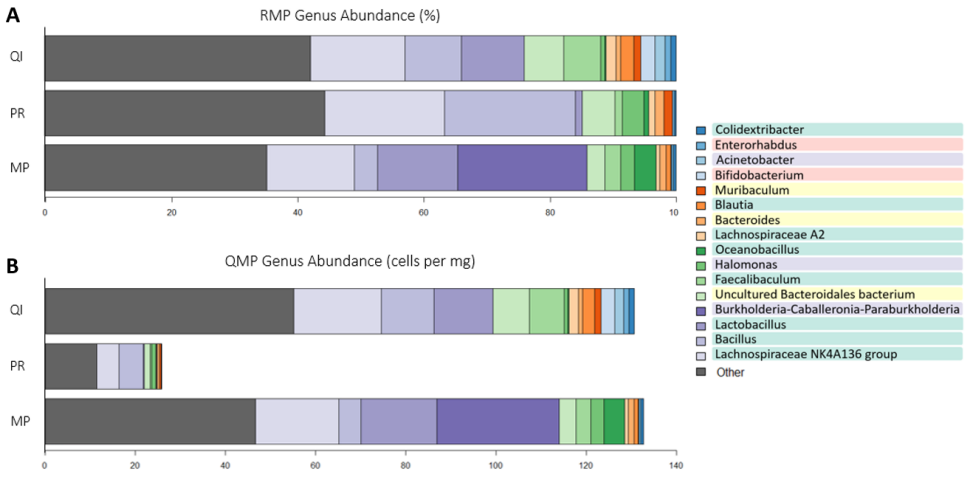

Figure 2.8. Relative versus quantitative microbiome profiling at genus level. Composition plots of the top 16 known genera among the three samples, with all other genera pooled into 'Other'. Taxonomies of the genera were coloured according to phylum: Firmicutes, green, Bacteroidota, yellow, Proteobacteria, lavender, Actinobacteriota, red. (**A**) **Genus-level mucus relative microbiome composition among different DNA extraction methods.** Relative microbiota profiles obtained using standard microbiome sequencing methods. (**B**) **Genus-level mucus quantitative microbiome composition among different DNA extraction methods.** Absolute microbiota profiles obtained using the relative microbial abundance along with prokaryotic cell counts.

Burkholderia/Caballeronia/Paraburkholderia genera stand out among the three DNA extraction methods. These genera are only present in the *MP* DNA extract in high abundance (**Figure 2.8B**), and constitutes most of the Proteobacteria phyla found in the *MP* extract (72.5%). The dominance of this group in the Proteobacteria phyla explains the high difference in the abundance of Proteobacteria between different DNA extraction kits (**Figure 2.7**).

2.4 Discussion

In order to assess the optimal approach for bacterial DNA extraction from mucus samples, we evaluated three DNA extraction methods in terms of the DNA quantity and quality, bacterial absolute abundance (estimated via number of 16S copies), and microbial composition. To identify and quantify the bacteria recovered in each method, we used primers for the 16S rRNA gene for the qPCR and amplicons sequencing. Additionally, we tested two different storage solutions during this process to decide on the best treatment to keep cells integrity.

We compared the DNA concentrations yielded under control (PBS) and RNA*later* treatments among the different DNA extraction methods. *QI* was the DNA extraction method with the highest DNA concentration in control conditions, followed closely by *PR*. RNA*later* treatment increased the overall amount of DNA extracted and altered the kit's performance. After RNA*later* treatment, *PR* was the method that achieved the highest DNA concentration, but it also had the highest standard deviation. When a treatment results in an increase in the mean, an increase in the standard deviation (Treatment effect*N(m,s) = N(m*Treatment effect, s*Treatment effect)) is also expected. However, the increase observed in the standard deviation of *PR* did not occur with the other extraction methods. Therefore, we could hypothesise that the presence of RNA*later* residues from the samples pre-treatment could have a negative impact on the performance of *PR*. This negative effect would explain *PR* inconsistent results.

RNA*later* has a high concentration of salt. Salt is used in DNA extraction methods for DNA neutralisation and further DNA precipitation. The use of salt also helps to remove proteic carryovers. However, it is necessary to remove any excess salt to prevent the salt from co-precipitating with the DNA or even causing DNA denaturation, after the DNA extraction. Any salt excess from the RNA*later* treatment could be a challenge for the purification step of the extraction methods tested.

Overall, the RNA*later* pre-treatment increases the amount of DNA yielded by all the kits. In theory, this might be due to RNA*later* functioning to preserve the integrity of the cells, preventing them from bursting during sample handling. However, how this conserving buffer interferes with the performance of each extraction method could not be tested due to the limited number of samples. The impact of RNA*later* residues on each kit could explain why *PR* performance was inconsistent compared to the other kits after RNA*later* pre-treatment. Finally, despite the higher standard deviation from *PR*, it seems that the kit from Promega yields the highest amount of DNA after RNA*later* pre-treatment, considering the elution volumes used in each kit.

Results in 260/230 absorbance ratios are similar to 260/280 absorbance ratios, where *PR* obtained the best results in terms of DNA quality. This kit offers an automated extraction, employing magnetic particles to isolate and move the sample around different solutions during the DNA extraction process. It seems that the automation of this process could have a positive impact on the sample's quality. In contrast, *MP* has the lowest ratios, whilst also being the method that requires the most user involvement. During *MP* workflow, the sample is transferred multiple times to different tubes. Additionally, the user has the task of mixing the sample with the binding matrix, this binding

matrix serves to isolate the DNA. These differences made *MP* a more manual extraction method compared to the two other kits evaluated, where magnetic beads or a silica-column are used instead, to facilitate the sample handling.

Despite the above mentioned, none of the samples had absorbance 260/280 and 260/230 ratios within the acceptable absorbance range (grey box in **Figure 2.4A**). These samples had a low microbial biomass which could potentially explain why their absorbance ratios are so low, as contaminant or residue from the extraction process or the samples themselves could affect the quality of the DNA extracts. Lastly, it would be useful to compare the DNA quality from those samples pre-treated with RNA*later* and PBS. However, the limited number of samples did not allow for this analysis.

When looking at the number of 16S copies, this value might be an indicator of the number of prokaryotic cells. *PR* obtained the largest number of prokaryotic cells , followed by *QI* and *MP*. However, some bacteria and archaea can have more than one copy in their genomes, which biases the estimation of cell counts. This heterogeneity can also bias the results when comparing different methods to extract DNA as each protocol would benefit specific prokaryotic groups. We find the same challenge when comparing the microbial compositions using the 16S counts. Multiple tools are available to predict the number of 16S gene copies based on the sample's microbial composition and correct for this bias. But no consensus exists on the procedure to be followed.

Again, RNA*later* pre-treatment appears to have a positive effect on the samples; the number of 16S copies obtained increases after RNA*later* pre-treatment in all of the kits. However, *QI* did not experience the same increase in 16S copies as the other kits after the treatment. We propose that the *QI* purification method was insufficient to cope with the residues left from the

RNA*later* pre-treatment. The results obtained with and without RNA*later* pre-treatment correlate with the DNA concentrations as expected. Finally, the obtained DNA quantities and number of 16S copies among the kits could be a result of the difference in DNA quality observed between the different extraction kits. Good quality DNA will be easier to detect and use in downstream applications.

Samples extracted with each kit were sent for sequencing to test their suitability for amplicon sequencing, despite being considered low microbial biomass samples. In addition, these results help to assess the differences between DNA extraction methods. The phyla and genus relative abundances were adjusted to obtain their absolute abundances using the microbial loads from qPCR results. *MP* had higher OTU diversity and richness, followed by *QI* and *PR*. Unlike the other DNA extraction kits (*PR* and *QI*), *MP* includes a lysis matrix which is made of particles of different shapes, sizes and materials . This lysis matrix could help to lyse a wider range of cell walls. The low microbial diversity found in *PR* DNA extraction could be explained by its low microbial load (around 5 times lower than the other DNA extracts). This sample was randomly selected for sequencing, which coincidentally happened to be the sample with the lowest number of 16S copies per mg of tissue from all the samples extracted with *PR*. Overall, the richness estimates reported among the DNA extracts correlate with other richness estimates from mouse microbiota studies using faecal samples (Hildebrand et al., 2013; Wang et al., 2019).

All the extracts were dominated by Firmicutes and Bacteroidota phyla as was expected from literature (Wang et al., 2019). However, *MP* showed a very high proportion of Proteobacteria, even higher than the Bacteroidota fraction and this is unexpected in a normal mice GIT. These Proteobacte-

ria were mostly of genera *Burkholderia/Caballeronia/Paraburkholderia* in *MP* kit. Species from all these three genera have been reported in humans (Gerrits et al., 2005; Vandamme et al., 2013). *Burkholderia* genus is known to include several animal pathogenic species, whereas species from *Caballeronia* and *Paraburkholderia* are not associated with human infections. *Burkholderia* has been also identified as an inherent contaminant in the FastDNA™ SPIN Kit for Soil (*MP*) extraction kit (Salter et al., 2014). Therefore, the presence of these high abundance genera only in the *MP* extract appears to be contamination from the DNA extraction kit itself.

Other studies have reported that within the Firmicutes phylum the greatest variability can be found when different DNA extraction methods are used (Abusleme et al., 2014; Momozawa et al., 2011). This is because Firmicutes and Actinobacteria species are considered difficult to lyse bacteria (Abusleme et al., 2014). Although *PR* has a higher relative abundance of Firmicutes among the three kits, *QI* could also recover other hard-to-lyse taxa as Actinobacteria in higher abundance than the other kits. This DNA extraction kit combines enzymatic lysis along with bead beating in its lysis step. Therefore, it seems that the double lysis step from *QI* is more efficient than the other kits against these bacteria. However, the notable difference in microbial loads between samples makes it difficult to conclusively answer this hypothesis.

In summary, *MP* yielded the lowest amount of DNA, with the lowest quality and the lowest number of prokaryotic cells. It also appears that the kit used, contained contaminants, in sufficient abundance to compete with the true taxa in the sample. Taking these observations into consideration *MP* appears to be a poor choice when extracting DNA from samples with low microbial biomass, as seen with our mucus samples. This was not the case for *PR* and *QI*, *PR*

extracted the highest quantity and quality of DNA, and also recovered the highest number of prokaryotic cells. Its performance was comparable to *QI*. Both DNA extraction kits showed similar taxonomic profiles and good bacterial recovery rates, making them good choices for extracting DNA from mucus samples.

2.5 Conclusion

We wanted to study different microbial kingdoms in human biopsies from IBD and healthy control samples. However, this aimed required the development of a protocol that would allow the efficient isolation of microbes from the host tissue. Furthermore, this method had to be reliable in isolating different microbial kingdoms, despite their specific characteristics.

We developed and evaluated different approaches in a murine model, using mice colonic samples. Once we were able to successfully isolate the microbes from the mucosa, we had to decide on a DNA extraction method. The experiments and analyses carried out to decide which method best suits our type of sample and objectives, were described in this thesis.

We tested three different DNA extraction kits: FastDNA™ SPIN Kit for Soil (MP Biomedicals), QIAamp® PowerFecal® Pro DNA Kit (Qiagen) and Maxwell® RSC PureFood GMO and Authentication Kit (Promega). To evaluate their performance, we compared them in terms of DNA quantity and quality, bacterial copy number and microbial composition. Besides this, we wanted to work with fresh samples, to avoid any microbial loss during the freeze and thaw cycle. For this reason, we also compared the DNA quantity and bacterial copy number yielded by two storage solutions. These storage solutions were PBS and RNAlater.

Our results showed that RNAlater pre-treatment increased the bacterial yielded compared with PBS in every scenario tested. We also observed that Maxwell® RSC PureFood GMO and Authentication Kit was the best kit in terms of DNA quality and quantity and number of copies. However, the sequencing results showed that QIAamp® PowerFecal® Pro DNA Kit (Qiagen)

extracted DNA from a wider range of microorganisms, these included difficult-to-lyse bacteria. We explained this outcome due to the double lysis step, mechanical and enzymatic, that Qiagen's kit includes and it is absent in the other extraction methods.

Although FastDNA™ SPIN Kit for Soil includes a mechanical lysis, its microbial composition was outcompeted by the presence of contaminants. The presence of contaminants can be easily reduced by reducing sample processing. Similarly, there will be less microbial loss the less the sample is processed. This was observed in our analysis; the kit with the best performance used a robot to process the sample in the fewest steps, while the kit that gave the worst results included the highest number of steps with the greatest user involvement among the kits tested. It is important to remember that these results are specific to the type of sample used. We worked with low biomass samples, but other type of samples could benefit from all the steps included in FastDNA™ SPIN Kit for Soil.

Finally, our results indicate that Maxwell® RSC PureFood GMO and Authentication Kit and QIAamp® PowerFecal® Pro DNA Kit are good options for studying mucus isolate bacteria. These DNA extraction kits together with an RNAlater pre-treatment, guaranteed the best results when isolating and identifying bacteria in samples dominated by host DNA. However, this analysis will have to be repeated for each microbial kingdom to reach a conclusion about which approach will best represent a polykingdom community.

Chapter 3

Bacterial and fungal microbiota analysis in colorectal cancer

3.1 Introduction

The pathogenesis of CRC has been associated with the gut microbiota (Lu et al., 2016; Wong & Yu, 2019). The microbiota composition of CRC patients changes during disease progression, with different bacterial profiles between adenoma and cancer stage (Feng et al., 2015). Different pathways have been proposed to explain bacterial involvement in disease development (Van Raay & Allen-Vercoe, 2017). These models have been explored and are supported by findings evolving around bacteria. Colon tissue from CRC patients is characterised by a reduced bacterial diversity compared to healthy tissue. In addition, bacterial species such as *Fusobacterium nucleatum*, *Bacteroides fragilis* and *Escherichia coli* are found in high abundance in the microbial profiles of these patients. However, other microbial kingdoms are interrelated and constitute the human gut microbiota, albeit to a lesser extent. One of these overlooked

kingdoms is constituted by fungi, where little data is currently available in its involvement in CRC onset and progression.

The 80% of sporadic CRC cases originate from the appearance of adenomas (Fischer et al., 2019). Initial studies of the mycobiome in patients with adenoma indicated that as with bacterial diversity, fungal diversity is reduced in adenomas when compared to adjacent tissue (Coker et al., 2019; Luan et al., 2015). Changes in the mycobiome are closely related to adenoma size and progression. However, fungal diversity appears to increase at later tumour stages (Gao et al., 2017). Similarly, patients with CRC show a distinct mycobiome profile at early and late stages of the disease. All these findings suggest that in tumour tissue the fungal community is different than in the adjacent non-tumour tissue, as shown in adenomas. This fungal composition would also be expected to be related to the bacterial composition given the reported interplay between both groups in health and dysbiosis (Sam et al., 2017; van Tilburg Bernardes et al., 2020). Based on these hypotheses, we defined the following objectives:

Objective 1. To characterize the bacterial quantitative microbiome profile of clinical cases of CRC using colonic resected tissue.

Objective 2. To characterize the fungal quantitative microbiome profile of clinical cases of CRC using the same samples as in **Objective 1**.

Objective 3. To compare the bacterial and fungal communities between tumour tissue (on-tumour) and tissue adjacent to the tumour (off-tumour).

To carry out the objectives described above, qPCR and amplicon sequencing were performed on resected tissue samples from CRC patients. These samples were extracted from different regions of the colon, always in pairs in order to compare on-tumour and off-tumour sites. The same primers were

used for both qPCR and amplicon sequencing. These were primers for the V4 region of the 16S rRNA gene, used to detect and identify bacteria, and primers for ITS1, which is specific for fungi. The results obtained from qPCR and amplicon sequencing were integrated together to obtain the quantitative microbiome profiles specific to each primer pair.

3.2 Materials and Methods

3.2.1 Microbial DNA extraction

Two different DNA extraction kits were employed to extract DNA in triplicates from colonic resected tissue from colorectal cancer patients. The samples from the first ten donors were processed with QIAamp® DNA Mini (Qiagen), while the Monarch® Genomic DNA Purification (New England Biosciences) kit was used with the samples from the three other donors. All the DNA extracts from each sample were pooled together. This was conducted in the lab of Dr Natalie Juge by Dr Dimitra Lamprinaki.

3.2.2 Evaluation of DNA yield

DNA concentration was measured fluorometrically with the High Sensitivity dsDNA kit (Invitrogen) on Qubit® Fluorometer 3.0 according to the manufacturer's guidance.

3.2.3 Quantitative microbiome profiling

Quantitative PCR assessment of microbial loads

The number of 16S rRNA gene copies in each sample was determined via qPCR using primers for V4 hypervariable region of the 16S rRNA gene (515F: GTGYCAGCMGCCGCGGTAA, 806R: GGACTACNVGGGTWTCTAAT) (Caporaso et al., 2011; Walters et al., 2016). For each reaction was used 3 µL of SYBR® Select Master Mix (Applied Biosystems), 2 µL of template DNA and two times 0.2 µL primer solution (10nM), and 1.1 µL PCR grade water, in a total reaction volume of 6.5 µL. The amplification program comprised of 2

stages; an initial denaturation step at 95 °C for 180 s, followed by 40 three-step cycles at 95 °C for 30 s, 55 °C for 30 s and at 72 °C for 30 s. A calibration curve was run with each sample set which consists of serial dilutions from a stock of known 16S copies concentration.

Primer for the ITS1 region were used to assess the proportion of fungal genetic material among the samples (ITS1F: CTTGGTCATTTAGAGGAAG-TAA, ITS2: GCTGCGTTCTTCATCGATGC) (Gardes & Bruns, 1993; White et al., 1990). Reactions contained 3 µL of SYBR® Select Master Mix (Applied Biosystems), 1.5 µL of template DNA and two times 0.15 µL primer solution (10nM), and 1.7 µL PCR grade water, in a total reaction volume of 6.5 µL. qPCR conditions were as follows: an initial denaturation step at 95°C for 180 s, followed by 40 two-step cycles at 95°C for 15 s and at 60°C for 60 s.

All reactions for 16S V4 and ITS1 regions were run in duplicates. The qPCR reactions were performed in a ViiA7 Real-time PCR detection system (Applied Biosystems). Cycle threshold value was determined using the ViiA7 Real-time PCR detection system software (v 1.2) (Thermo Fisher Scientific). The DNA concentration from each sample was used to normalize their 16S rRNA gene copies.

Library preparation

The library preparation was executed as described in *2.2.4 Library preparation* including primers for ITS1. Different amounts of ITS1 and 16S amplicons (2:1) were pooled to compensate the expected differences in starting material between fungal and bacterial DNA. Two reactions with sterile water were included per primer pair as negative controls for the sequencing process. This process was repeated with the same conditions for those samples with the

lowest number of reads for both 16S and ITS1.

Processing of sequenced data

The LotuS pipeline (v 2.01) (Hildebrand et al., 2014) was used to process the V4 16S rRNA and ITS1 region amplicon sequences. The steps include demultiplexing, quality-filtering, clustering of the reads and taxonomic classification. 16S amplicons were clustered using Dada2 (Callahan et al., 2016), while uparse at 97% identity was used for ITS1 amplicons (Edgar, 2013). Host contamination was detected and removed by aligning the reads to the human reference genome GRCh38.p13 using minimap2 (H. Li, 2018). Chimeric OTUs were filtered out with uchime2 (Edgar et al., 2011). Reads were aligned to their respective databases and the LotuS LCA algorithm using lambda aligner against reference database SILVA (v 138) for 16S rRNA and UNITE (v 8.0) for ITS1.

The commands for LotuS utilised the following parameters:

For 16S: *-s sdm_miSeq.txt -keepUnclassified 1 -p miSeq-keepTmpFiles 1 -id 0.97 -doBlast usearch -redoTaxOnly 0 -simBasedTaxo lambda -refDB SLV -amplicon_type SSU -tax_group bacteria -saveDemultiplex 2 -offtargetDB hg38.masked.fa -CL dada2 -buildPhylo 1 -thr 12 -exe 0*

For ITS1: *-s sdm_miSeq.txt -keepUnclassified 1 -p miSeq-keepTmpFiles 1 -id 0.97 -doBlast usearch -redoTaxOnly 0 -refDB UNITE -amplicon_type ITS -tax_group fungi -saveDemultiplex 2 -offtargetDB hg38.masked.fa -buildPhylo 1 -CL 3 -thr 12 -exe 0*

3.2.4 Contamination removal

The R package decontam was used to identify contaminants (Davis et al., 2018). Contaminants were detected based on their increased prevalence in negative controls using the methods isContaminant (prev) and isNotcontaminant, with default arguments. All contaminants found with the different methods were removed and new abundance matrices were produced for further analysis.

3.2.5 Statistical analysis

Statistical analysis was conducted using R (v 3.6.1). Sample depth from 16S samples was rarefied 10 times to 3,505 reads per sample with R-package rtk (v 0.2.6.1) (Saary et al., 2017), using the smallest number of sequences per sample observed. While sample depth from ITS1 samples was rarefied 10 times to 202 reads per sample, leaving 4 samples with lower total read numbers out. Rarefactions were used to calculate the mean Shannon's diversity index and taxa richness estimate (chao1) of each sample from 16S ASV and ITS1 OTU frequencies with the R-package vegan (Dixon, 2003). Counts for 16S taxa were normalised by the number of 16S copies per ng of DNA for each sample. Significance between paired groups of samples was performed with a Wilcoxon signed rank test. Significance between non-paired groups of samples was tested with a Kruskal–Wallis test. To calculate compositional differences, the permutational multivariate analysis of variance (PERMANOVA) was performed using the *adonis2* function included in the R-package vegan. Bray-Curtis distances and 999 permutations were used for this test. Multiple testing correction was carried out using the Benjamini–Hochberg false discovery rate (Benjamini & Hochberg, 1995).

3.3 Results

To study the differences in microbial composition between on-tumour and off-tumour sites in CRC clinical samples, 13 CRC cases were studied, making a total of 26 samples. These samples were provided by the Norfolk and Norwich University Hospitals (Norwich, UK), and came from middle-aged subjects (40-65) of different sex, ages and at different stages of colorectal cancer. The bacterial load among samples was determined by qPCR and described as 16S copies per ng of DNA. These results were used for the 16S rRNA gene sequencing analysis to obtain absolute taxonomic abundances relative to total input DNA. Additionally, fungal loads were also assessed via qPCR to include them in the ITS1 region sequencing analysis. Overall, there was no significant difference in bacterial loads between on-tumour and off-tumour sites (Kruskal-Wallis: X2(1) = 0.23, $P = 0.62$) (**Figure 3.1A**), though the median was slightly increased in on-tumour sites (median: 13,370 copies on-tumour, 12,165 copies off-tumour). There were not enough samples to compare the bacterial loads at different locations along the colon or with tumour staging. Fungal DNA was reliably detected in only few samples (n = 4); however, fungal presence was inconsistent across samples showing high standard deviation between technical replicates (**Figure 3.1C**). The low consistency across ITS1 Ct values is noticeable when compared to 16S technical replicates (**Figure 3.1B**). For this reason, fungal qPCR results were not further included in the analysis.

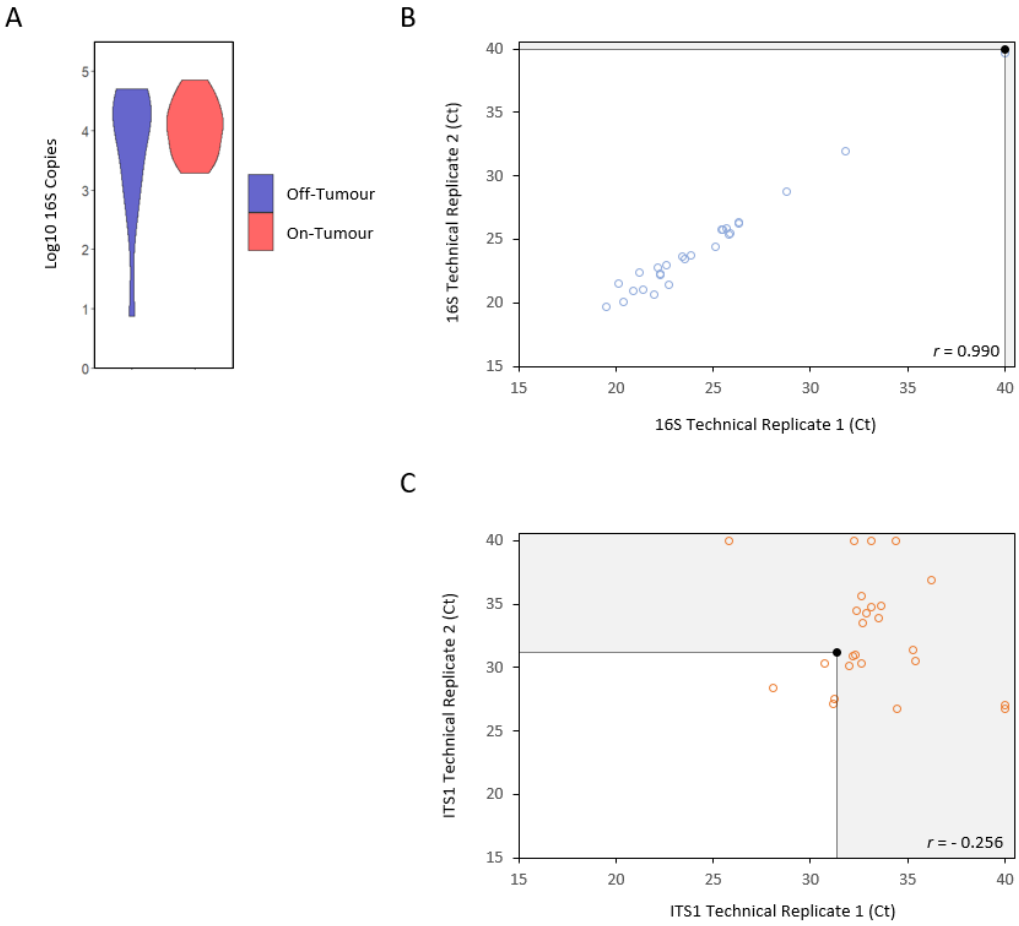

Figure 3.1. Bacterial and fungal loads across CRC samples. (**A**) **Violin plot showing no significant differences in prokaryotic loads between off-tumour and on-tumour sites.** Bacterial loads depicted as log10 scaled 16S rRNA copies per nanogram of DNA in each sample. (**B**) **Scatter plot showing the correlation between technical replicates from the quantitative amplification of the 16S V4 rRNA region in the samples.** We detected a high Pearson's correlation coefficient ($r = 0.99$) between the technical replicates (blue outlined dots). Negative control included for reference (black dot) which demarcates the area out of range in grey. (**C**) **Scatter plot showing the correlation between technical replicates from the quantitative amplification of the ITS1 region in the samples.** Low Pearson's correlation coefficient (r = -0.256) was detected between each sample technical replicates (orange outlined dots). Negative control included for reference (black dot) which demarcates the area out of range in grey

After sequencing the samples, those with the lowest number of reads were sent for re-sequencing, in order to have the necessary reads to rarefy the samples and further analysis. Sequencing of 16S rRNA V4 region returned a total of 1,181,582 reads (mean: 11,889 reads in the first run, 200,443 reads in the

second run), while sequencing of ITS1 region resulted in 164,533 reads (mean: 1,614 reads in the first run, 22,159 reads in the second run). Sequencing contaminant reads were detected by comparing the prevalence of ASVs/OTUs across samples and negative controls. Number of contaminant reads varied between samples (**Figure 3.2**).

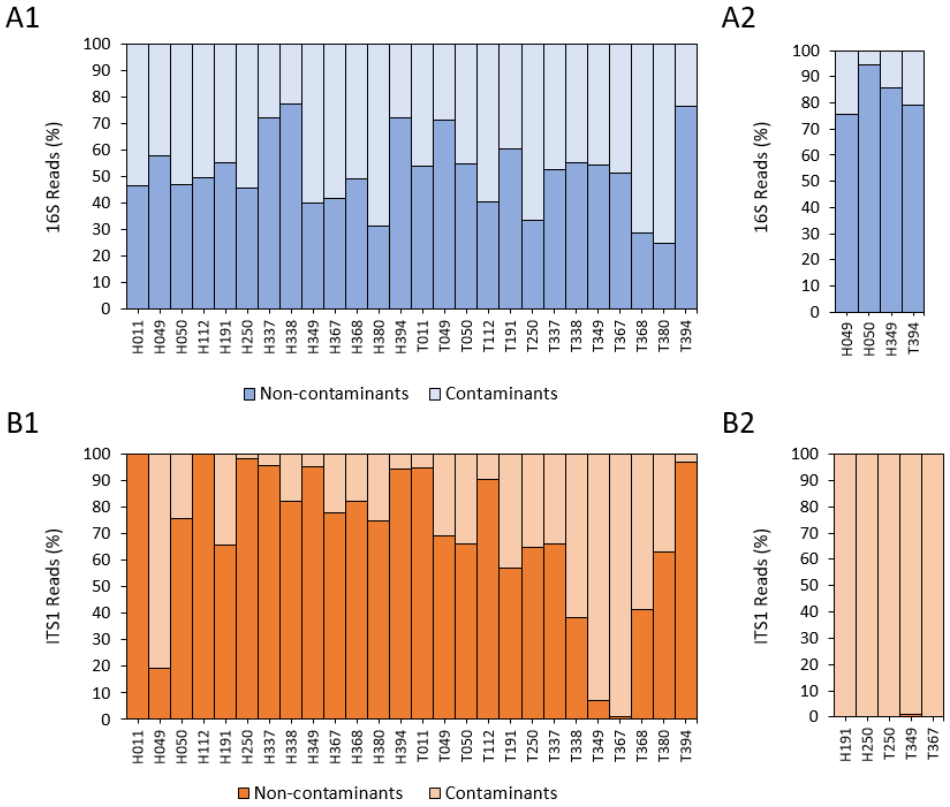

Figure 3.2. Compositional plot showing the proportion of sequencing contaminant versus non-contaminant reads across samples. (**A**) Proportion of 16S contaminants in the first (**A1**) and second (**A2**) sequencing run. (**B**) Proportion of ITS1 contaminants in the first (**B1**) and second (**B2**) sequencing run.

After the identification and removal of contaminants, a total of 836,090 16S rRNA and 32,166 ITS1 region sequence reads were assigned to 1,907 ASVs and 17 OTUs, respectively. To estimate the alpha diversity between off-tumour and on-tumour sites, samples were rarefied to 3,505 reads for 16S amplicons and 202 reads for ITS1. The Shannon diversity index for bacteria did not differ

significantly either between paired samples (Wilcoxon: V = 28, $P = 0.24$) or samples site (Kruskal-Wallis: X2(1) = 0.55 , $P = 0.45$) (**Figure 3.3A**). Likewise, no significant differences were found in bacterial richness between pairs (Wilcoxon: V = 48, $P = 0.89$) or sites (Kruskal-Wallis: X2(1) = 6.58 x 10^{-4}, $P = 0.97$) (**Figure 3.3B**).

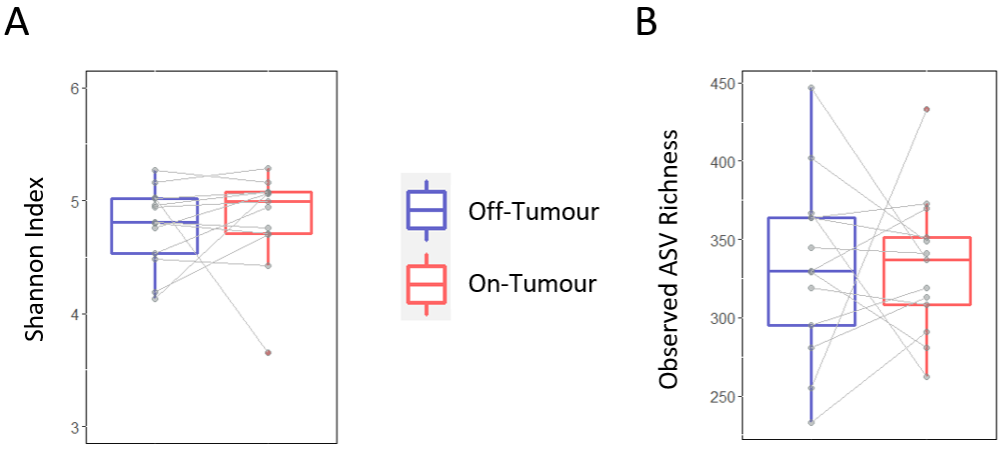

Figure 3.3. Paired box plot of alpha bacterial diversity in off-tumour and on-tumour sites. Paired samples collected from different disease states at adjacent sites in the colon: off-tumour and on-tumour. Different locations are differentiated by the border colour: blue for off-tumour and red for on-tumour site. Paired samples are linked (grey line). (**A**) **Shannon's index from 16S rRNA amplicons.** (**B**) **Observed ASV richness from 16S rRNA reads.** The box ranges from the first to the third quartile of the distribution. The line across the box indicates the median.

In contrast, fungal Shannon diversity index differed significantly between off-tumour and on-tumour regions (Kruskal-Wallis: X2(1) = 6.61, $P = 0.01$) (**Figure 3.4A**), although it did not differ significantly between pairs (Wilcoxon: V = 37, P = 0.09). For fungal richness, there was not significant differences between sites (Kruskal-Wallis: X2(1) = 2.40, $P = 0.12$) or pairs (Wilcoxon: V = 33, $P = 0.21$) (**Figure 3.4B**). Overall, fungal diversity and richness seem to decrease in on-tumour site when compared with off-tumour.

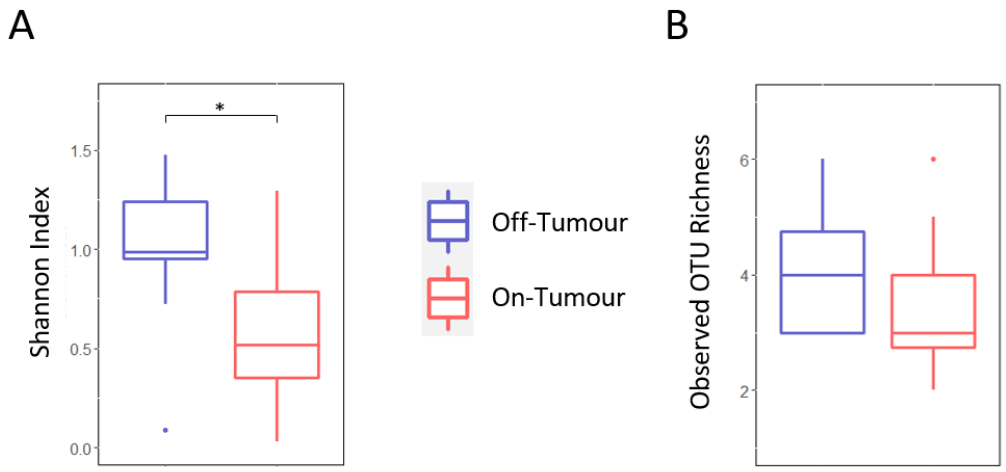

Figure 3.4. Paired box plot of alpha fungal diversity in off-tumour and on-tumour sites. Paired samples collected from different disease states at adjacent sites in the colon: off-tumour and on-tumour. Different locations are differentiated by the border colour: blue for off-tumour and red for on-tumour site. (**A**) **Shannon's index from ITS1 region amplicons.** Significant difference in diversity between off-tumour and on-tumour sites. Kruskal-Wallis test, $*P < 0.05$. (**B**) **Observed ASV richness from 16S rRNA reads.** Boxplots center lines represent the median, and the edges represent first and third quartiles.

In terms of bacterial composition, Firmicutes and Bacteroidetes were the dominant bacterial phyla among majority of the samples in both groups off-tumour and on-tumour sites (**Figure 3.5**). They accounted for 87.16% of the total number of sequences, whereas the phyla with the next highest relative abundance - Proteobacteria, Fusobacteriota, Campylobacterota and Actinobacteriota – harboured 12.5% of sequences. Overall, the human gut microbiome profiles are in agreement with other studies where Firmicutes and Bacteroidetes represent the 90% of the human gut microbiota (Arumugam et al., 2011).

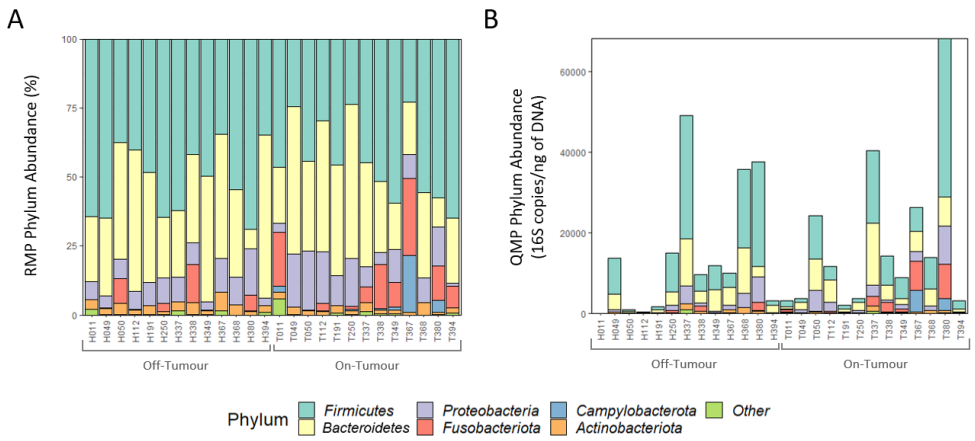

Figure 3.5. Relative versus quantitative bacterial microbiome profiling at phylum level. Composition plots of the top 6 most abundant phyla among the samples, with all other phyla pooled into 'Other'. (**A**) **Phylum-level relative microbiome composition.** (**B**) **Phylum-level quantitative microbiome composition.** Absolute microbiota profiles were obtained by multiplying relative microbial abundance with the 16S-qPCR based bacterial loads.

No phyla differed significantly between off-tumour and on-tumour sites when using the RMP approach (Kruskal-Wallis: $P > 0.05$, q > 0.1) (**Figure 3.6A**). However, the outcomes of this comparative analysis changed using the QMP approach (**Figure 3.6B**). Although no significant differences between sites could be detected, some trends are more pronounced including bacterial loads (Kruskal-Wallis$P > 0.05$, q > 0.1). For example, Campylobacterota and Fusobacteriota phyla were enriched in the tumour site, while in other phyla such as Actinobacteriota, Firmicutes or Bacteriodota, the difference observed between on-tumour and off-tumour sites was decreased. Similarly, in the QMP approach we could also not detect significant differences using lower taxonomic levels between sites (Kruskal-Wallis$P > 0.05$, q > 0.1).

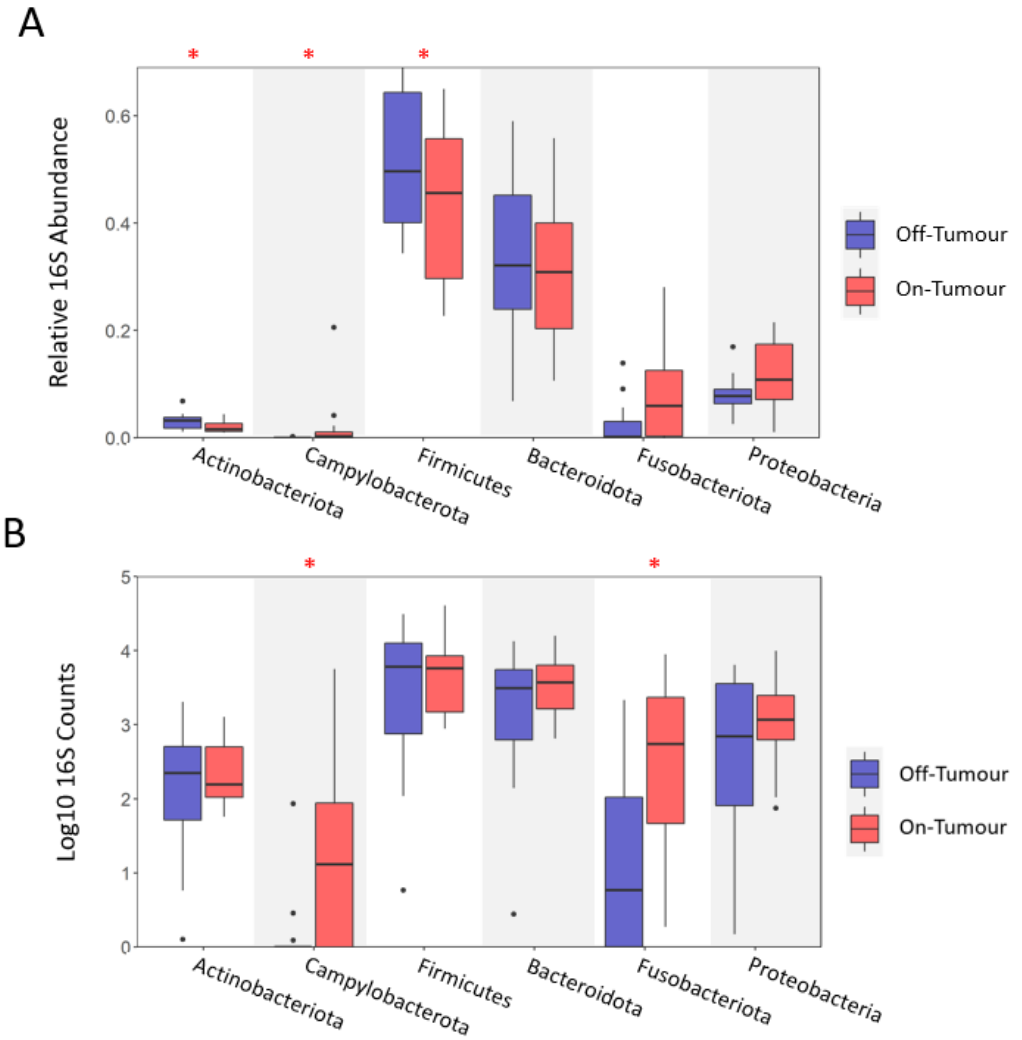

Figure 3.6. Boxplot showing the differences in bacterial abundances between off-tumour and on-tumour sites of the top 6 most abundant phyla among samples. Off-tumour and on-tumour sites are differentiated by colour: blue for off-tumour and red for on-tumour. (**A**) **Differences in bacterial relative abundances at phylum-level.** Red star (*) indicates no significant differences after multiple testing correction (Kruskal-Wallis test, * $P < 0.05$, q > 0.1). (**B**) **Differences in bacterial absolute abundances at phylum-level.** Abundances were normalized by the number of 16S rRNA copies of each sample and log10-transformed (log10(x+1)). Red star (*) indicates no significant differences after multiple testing correction (Kruskal-Wallis test, * $P < 0.05$, q > 0.1). The box ranges from the first to the third quartile of the distribution. The line across the box indicates the median.

The Firmicutes/Bacteroidetes ratio is considered an important marker for gut health versus intestinal dysbiosis. When looking at the Firmicutes/Bacteriodota ratio of CRC cases, it did not significantly differ between

off-tumour and on-tumour sites, although it seems to decrease in on-tumour site (Kruskal-Wallis: X2(1) = 0.016, $P = 0.89$) (**Figure 3.7**).

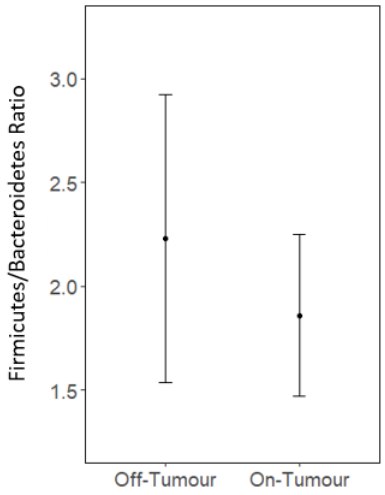

Figure 3.7. Firmicutes/Bacteroidota ratio boxplot showing no significant difference between off-tumour and on-tumour sites.

The bacterial composition of the samples was ordinated using a principal coordinates analysis (PCoA) to compare Bray-Curtis dissimilarities among samples (**Figure 3.8**). The PCoA did not show any clear group neither using the relative composition (**Figure 3.8A**) nor the absolute composition (**Figure 3.8B**). The four samples observed to the left of the graph in the relative composition PCoA corresponded to the samples that were sequenced a second time (**Figure 3.8A**). There was no differentiation between off-tumour or on-tumour sites. However, some variables from the available metadata explained some of the samples distances variation. These were the DNA extraction method and the samples tumour stage. The DNA extraction method explained 5.8% of the variance of the samples relative composition (PERMANOVA: $P = 0.030$, R2 = 0.058) and 6.8% of the absolute composition (PERMANOVA:

$P = 0.003$, R2 = 0.068). While, the samples tumour stage explained 13% of the variance of the absolute composition (PERMANOVA: $P = 0.043$, R2 = 0.135).

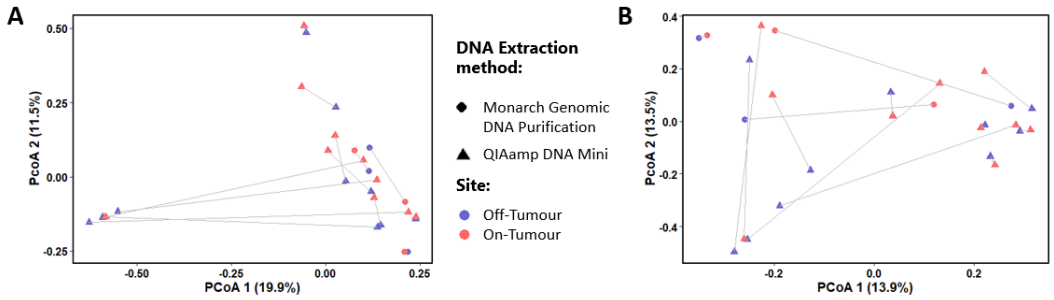

Figure 3.8. OTUs-level bacterial microbiome community variation, represented by principal coordinates analysis (Bray-Curtis dissimilarity PCoA) Points correspond to samples from the 13 CRC cases (n = 26). Samples were coloured according whether they are from off-tumour (blue) or on-tumour (red) sites. Shapes of the points indicate the DNA extraction method used: circle for Monarch Genomic DNA Purification from New England Biosciences, and triangle for QIAamp DNA Mini from Qiagen. Paired samples are linked. (**A**) **PCoA using relative microbiome profiles.** The first PCoA component explained 19.9% and second component explained 11.5% of the total variance among the samples. (**B**) **PCoA using quantitative microbiome profiles.** The first PCoA component explained 13.9% of the total variance among the samples, while second component explained 13.5%.

Moving on to fungal composition, the relative microbial profile varied greatly between samples at all taxonomic levels, using matrices with contaminant taxa already removed (**Figure 3.9**). There was no dominant fungal species identified in the samples, although the *Malasseziaceae* family was present in most of them.

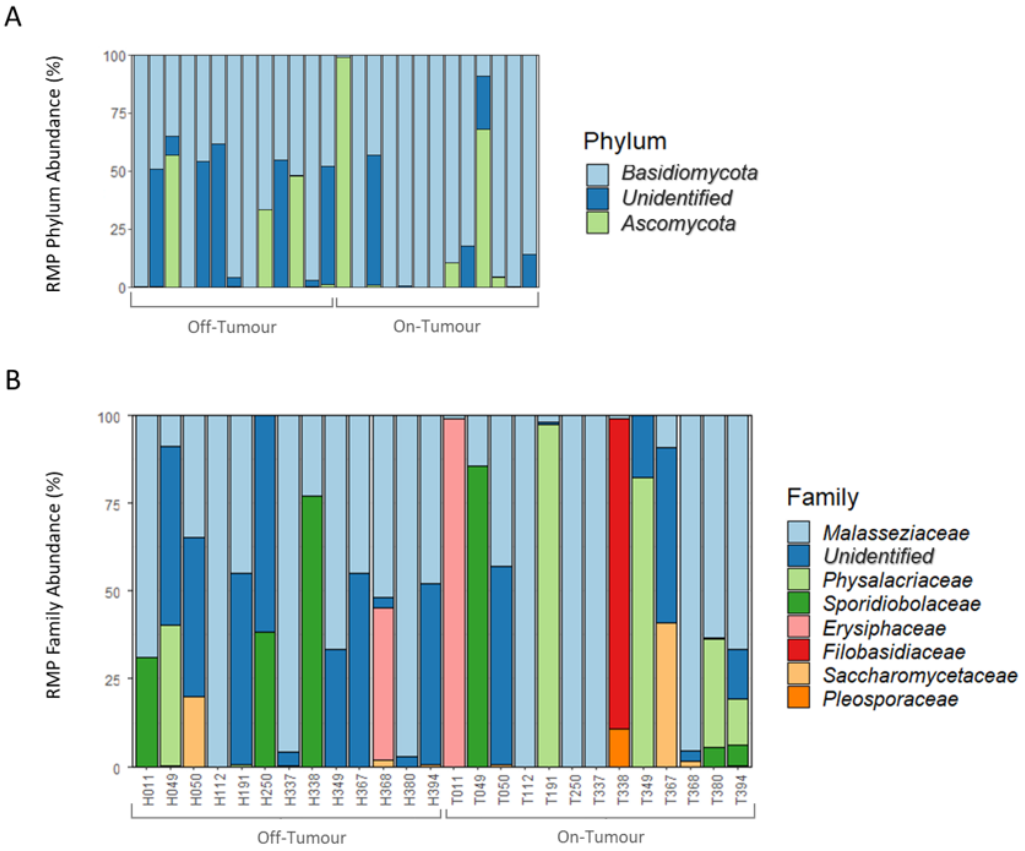

Figure 3.9. Composition plots of fungal relative abundances at phylum (A) and family (B) levels. Relative microbiota profiles obtained using standard microbiome sequencing methods.

When looking at the differences between sites at family-level, there was no significant differences between off-tumour and on-tumour sites (Kruskal-Wallis: $P > 0.05$, q > 0.1) (**Figure 3.10**). Overall, the relative abundance and presence of the families between sites was too low to reach any conclusive results. As with bacteria, there were no significant differences in abundance in lower taxonomic levels between sites (Kruskal-Wallis: $P > 0.05$, q > 0.1).

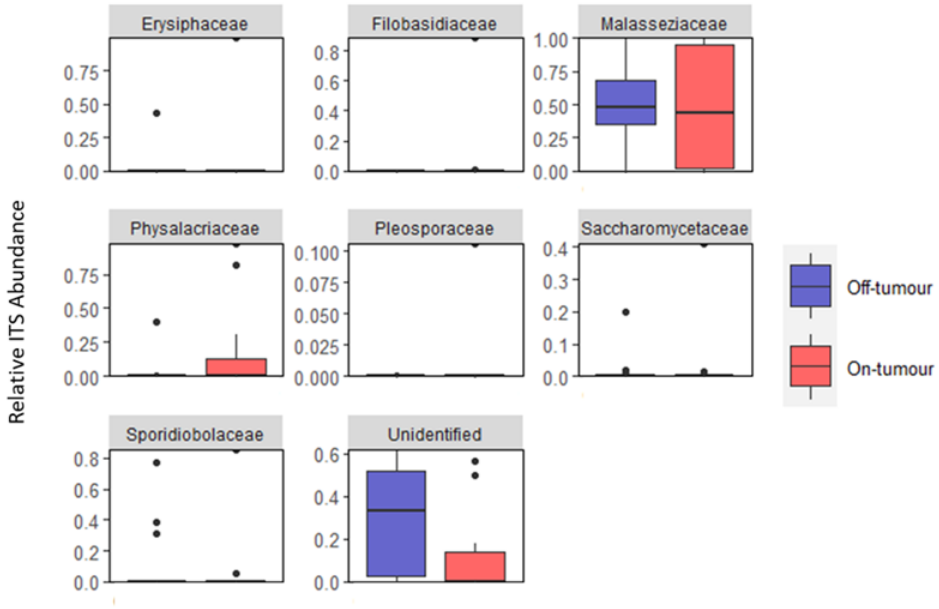

Figure 3.10. Boxplot showing no differences in fungal abundances between off-tumour and on-tumour sites at family-level among samples. Off-tumour and on-tumour sites are differentiated by colour: blue for off-tumour and red for on-tumour. The box ranges from the first to the third quartile of the distribution. The line across the box indicates the median.

Finally, an ordination was performed on the relative sample composition using a PCoA. This showed no clear cluster nor systematic compositional differences between on-tumour and off-tumour sites (**Figure 3.11**). Whether the sample belonged to off-tumour or on-tumour sites did not contribute significantly to the distribution of samples; however, the DNA extraction kit biased sample composition significantly (PERMANOVA: $P = 0.048$, R2 = 0.082). This could be traced to single taxa, e.g. samples extracted with the QIAamp DNA Mini kit from Qiagen had significantly higher abundance of Ascomycota phylum compared to those DNA samples extracted with the Monarch Genomic DNA Purification from New England Biosciences (Kruskal-Wallis: ** $P < 0.01$, q = 0.021).

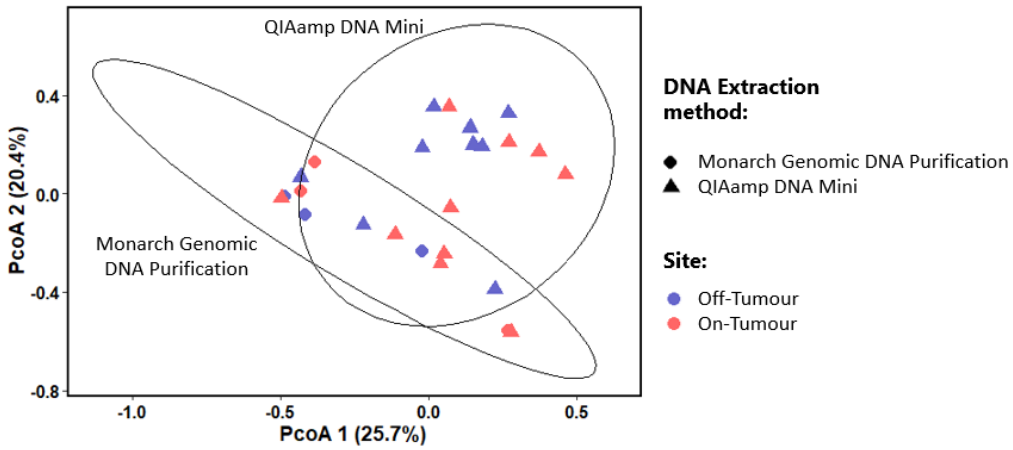

Figure 3.11. OTUs-level fungal microbiome community variation, represented by principal coordinates analysis (Bray-Curtis dissimilarity PCoA) Points correspond to samples from the 13 CRC cases (n = 26). Samples were coloured according whether they are from off-tumour (blue) or on-tumour (red) sites. Shapes of the points vary according the DNA extraction method used: circle for Monarch Genomic DNA Purification from New England Biosciences, and triangle for QIAamp DNA Mini from Qiagen. Data ellipses around the two clusters in black corresponding to samples extracted with different DNA extraction methods.

3.4 Discussion

We used 13 CRC clinical cases to study the microbial composition between off-tumour and on-tumour sites. These cases included subjects of different sex, age and at different stages of cancer. The focus of this project was to explore the differences between the two sites in a paired study. While at the same time, to consider other microbial kingdoms, specifically fungi. For this purpose, samples were sequenced using the well-known primer pairs for the V4 region of the 16S rRNA gene and primers for the fungal-specific internal transcribed spacer 1 (ITS1). In addition to this, the 16S and ITS1 copy numbers (microbial load) were estimated via qPCR, using the same primers used for sequencing. These results were integrated into the microbiome analysis which allowed quantitative microbiome profiling of the CRC cases.

Starting with fungi, we first calculated the fungal load of the samples. Most of the Ct values obtained from the CRC samples when performing qPCR were outside the limits set by the negative control. Thus, we concluded that either there are no fungi in those samples which Ct value exceeds the negative control or not enough fungi to be detected with the primers used in qPCR. Furthermore, the difference between technical replicates was higher than 0.5 Ct in most cases, indicating that the results are not reliable due to lack in consistent replication. This scenario can occur when in fact the fungal concentration is very low or non-existent. But it could also be due to other technical difficulties or related to the nature of the samples.

The primer pair used for ITS1 has previously been used successfully to detect and identify fungi in culture, environmental and faecal samples (Bokulich 2013, Heather 2015, Hoarau 2016, Walters 2016). In all these samples a high

proportion of bacterial or fungal genomic material is expected. However, the samples used in this study were human tissue resected from different locations of the colon during colectomies. Thus, we would expect that the DNA extracted from these samples would contain a high percentage of mainly human DNA, a small fraction of bacterial and an even smaller fraction of fungal DNA, if at all. The proportion of fungal DNA is estimated to correspond to 0.1% of the total human gut microbiome (Qin et al., 2010). This situation could lead to a competitive amplification between the highly available human or even bacterial DNA, compared to the low proportion of fungal DNA. In such a scenario, non-specific amplification of non-fungal DNA can happen (off-targets), despite the use of fungi-specific primers, or no amplification at all (Bedarf et al., 2021). Finally, one study might suggest otherwise, in this study they used the primers ITS1F and ITS2 to successfully identify fungi in human biopsies with the Illumina HiSeq 2000 sequencing platform (Luan et al., 2015). To improve the fungus phylum coverage, they included four ITS-forward primers containing the same length as ITS1F and starting from the ITS1F 5' end adjacent nucleotide during the library preparation. Perhaps the addition of these primers during qPCR would allow us to quantify a wider range of fungi.

The results obtained from qPCR indicate the possibility of obtaining a large proportion of host contamination. In this situation, we tried to compensate for the low abundance of fungi twice. The first time, during sequencing; using a higher amount of DNA during library preparation after ITS1 amplification. The second time was during amplicon data processing by removing off-target amplicons matching to the human genome with minimap2. During raw read processing, 226 OTUs were removed as these were identified as host DNA. Finally, a total of 164 ,533 reads assigned to 55 OTUs were obtained after initial

quality filtering of DNA reads, but only 32,166 reads (17 OTUs) remained after removing contaminants. The large proportion of contaminants shows how susceptible the samples are to contamination, possibly due to the low fungal biomass present in the samples (Salter et al., 2014); indeed, the qPCR results indicate that fungal DNA was too low abundant in our DNA extracts to be reliably detected (**Figure 3.1C**).

Looking at the relative fungal composition at phylum and family levels (**Figure 3.9**), the sample profiles suggest that there is no minimum threshold of reads necessary to characterize the fungal composition among samples. Additionally, the low relative abundance of different fungi and their inconsistent occurrence among our samples suggests that fungi are probably absent or extremely rare in these biopsies. Basidiomycota was the dominant phylum in the majority of samples (**Figure 3.9A**). This phylum has found enriched in CRC patients (Anandakumar et al., 2019). However, the fungal profiles from other studies also include a fraction of Ascomycota. This phylum was only present in a few of our samples, possibly due to the low amount of fungal cells. From the observed fungi at family level, *Malasseziaceae*, *Sporidiobolaceae*, *Erysiphaceae*, *Filobasidiaceae*, *Saccharomycetaceae* and *Pleosporaceae* were previously identified in colonic biopsies and faecal samples (Coker et al., 2019; Hoarau et al., 2016; Liguori et al., 2016; Luan et al., 2015; Sokol et al., 2017). Specifically, *Malasseziaceae* and *Erysiphaceae* families were found to be enriched in CRC patients (Coker et al., 2019). Only the *Physalacriaceae* family has not been reported as part of the human mycobiome in other studies.

In addition to these results, the primers used seem to overestimate fungal diversity and richness. We observed this trend by looking at the taxonomic profiles of well-defined mock communities (Unpublished data). This is an

important consideration given the low richness observed, although the decrease in fungal diversity observed at on-tumour sites has already been reported in similar studies in adenomas and at early tumour stage (Luan et al., 2015). Other studies indicate that diversity may increase in late tumour stages (Gao et al., 2017). The limited number of cases did not allow us to assess whether there was any correlation between diversity or richness with the samples tumour stage. Our dataset includes samples at different tumour stages, therefore it would also have been expected to observe no difference in diversity between sites.

On top of all these factors, the DNA extraction kits had an additional impact on fungi detected in the biopsies (**Figure 3.11**). The phylum Ascomycota was enriched in those samples extracted with the QIAamp DNA Mini kit from Qiagen, and this could be due to extraction performance, or contamination in these kits. Considering the large variability between samples, this should be validated on a larger set of samples, as only 3 out of 26 samples from this dataset were extracted with the kit from New England Biosciences. Altogether, it seems that our CRC biopsies did not contain a high enough concentration of fungal DNA to be efficiently detected via our ITS1 primers.

For the bacterial community, the bacterial load was estimated and calculated as the number of 16S copies per nanogram of DNA. Unlike fungi, there was very little variability between technical replicates, indicating that the measurements are reliable and not the result of non-specific amplification (**Figure 3.1B**). All samples had a higher bacterial load than the negative control. However, the difference in bacterial loads between samples was notable. This might be due to the resolution offered by the standard curve, which is determined and limited by its number of dilutions. Greater number of dilu-

tions allow better delimitation of the standard curve, and finally the number of copies. The lower the number of copies, the more inaccurate the standard curve will be as the Ct values increases exponentially between each cycle. The high difference in bacterial loads between samples could also be the result of non-specific amplification of host DNA (Bedarf et al., 2021). The exact mass of tissue used for DNA extraction was not known, otherwise a more accurate parameter could have been obtained. Finally, no difference in bacterial load was observed between off-tumour and on-tumour sites (**Figure 3.1A**).

After sequencing the samples, 434 ASVs were identified as off-targets (host DNA) and around 70% of the sequences were detected as sequencing contaminants. More than 50% of all reads were contaminants in 12 of 26 samples (**Figure 3.2A**). This shows how samples with low microbial biomass are more affected by contaminant microbes and non-specific amplification, and therefore the necessity to include appropriate negative controls during sample processing (Bedarf et al., 2021; Salter et al., 2014). The number of off-targets identified also highlights the possibility of getting non-specific amplification during 16S copies quantification.

Regarding the bacterial diversity and richness of the CRC samples, no significant difference was observed between on-tumour and off-tumour sites (**Figure 3.3**). While several studies have reported a significant decrease in diversity in CRC patients (Ai et al., 2019; Chen et al., 2012; Saffarian et al., 2019). Furthermore, no difference was found in the composition of the samples between the two sites at different taxonomic levels. Although some trends can be observed between different phyla (**Figure 3.6**). We first used relative microbiome profiles to explore differences between sites. And then we obtained the quantitative microbiome profiles of the samples to validate these

first results. As expected, the integration of the number of 16S copies affected the outcomes of the differential abundance analysis. Some differences observed in RMP were even more pronounced using the QMP approach. This was the case with the Fusobacteria phylum, that seemed increased in abundance at on-tumour sites (**Figure 3.6B**). *Fusobacterium* spp., specifically *Fusobacterium nucleatum*, have been extensively studied in the context of CRC. They have been found to be enriched in faeces and biopsies derived from CRC patients and are considered opportunistic pathogens (Garza et al., 2020). *Fusobacterium* has been recognised by its strong adhesive and invasive abilities conferred by the membrane protein FadA, these adhesive and invasive abilities facilitate its adherence to epithelial cells after mucosal barrier disruption (Rubinstein et al., 2013). Once it is internalised by the host cells, it can suppress host immune system and ultimately, promote tumour development (Brennan & Garrett, 2019).

The Campylobacterota phylum also seemed highly increased at on-tumour sites after accounting for the 16S number of copies. Some *Campylobacterota* ssp. have been found enriched in CRC tissue samples and proposed to promote intestinal inflammation (Mangifesta et al., 2018; Warren et al., 2013). Experiments with mice have showed the capacity of *Campylobacter* to promote CRC pathogenesis. Those mice colonised with human isolates of *Campylobacter jejuni* developed more and larger tumours in comparison with uninfected mice (He et al., 2019). The presence of Campylobacterota together with high abundances of Fusobacteria is not accidental. Co-occurrence networks analysis of CRC tissue showed the co-occurrence within individual tumours of *Fusobacterium* and *Campylobacter* species and their ability to form aggregates (Warren et al., 2013). This ability to form aggregates could facilitate the invasion of

the host cells as has been proven with other bacterial species (Edwards et al., 2006).

Additionally, the ratio of Firmicutes/Bacteroidetes was used to compare off-tumour and on-tumour sites. This ratio is considered to represent the state of health and dysbiosis of the gut microbiota. Although this ratio is often used as a marker in patients with obesity, a decrease in this ratio has also been related to the presence of adenomas and neoplastic lesions (Lu et al., 2016; Mori et al., 2018). In the clinical samples of this study there was no significant difference in the Firmicutes/Bacteroidetes ratio between off-tumour and on-tumour sites, or observable trend when compared the samples ratio with the available metadata (**Figure 3.7**).

Finally, as happened with fungi, bacterial composition appears to be affected by the DNA extraction method used. As showed in *Chapter 2*, the DNA extraction method is an important source of bias when studying microbial communities. Thus, using different DNA extraction methods for the same set of samples can introduce great variability across the samples microbial profiles and mask the effect of other important variables (Brooks et al., 2015). The inclusion of new reads from the second sequencing run could also be another source of bias that masks the distribution of the samples. This was indicated by the relative composition PCoA where those samples re-sequenced look like outliers (**Figure 3.8**).

In addition to the DNA extraction method, the tumour stage also drove the bacterial distribution among the samples. Although we were unable to study the taxonomic differences between tumour stages, due to the limited number of samples, other studies indicate that there are indeed different microbial profiles at different stages of the disease (Feng et al., 2015). Besides the tumour

stage, other variables such as the sex, age or diet of the patients could act as confounders. For example, it has been shown that there is a sexual dimorphism in the human gut microbiota from puberty to menopause in women, where the microbiota composition of menopausal women is closer to men, as opposed to menstruating women (Zhang et al., 2021). Again, we were unable to measure the variability in the microbial composition and abundance introduced by these confounding factors, due to the limited number of samples.

Finally, we cannot discern if there is a difference between on-tumor and off-tumor sites based on our results. Results from differential abundance analysis suggest that certain bacteria may be enriched in tumour tissue. However, we could not reach any statistical significance due to the limited number of clinical cases and the high heterogeneity of our dataset. Regarding the relationship between bacterial and fungal communities, we could not explore any pattern.

3.5 Conclusion

In this chapter, we wanted to analyse the fungal and bacterial microbiota in carcinomas (on-tumour) and compare it with adjacent tissue (off-tumour). For this purpose, we obtained paired DNA samples from colonic tissue of colorectal cancer patients.

Although fungi have been previously detected in human colonic biopsies, we could not detect fungi in our samples, at least, quantitatively. Instead, we were able to identify fungi using amplicon sequencing. However, the high number of off-targets and contaminants detected in the sample casts doubt on the actual existence of fungi in our colorectal samples. The presence of off-targets and contaminants is also common in samples with low microbial mass, which could also be the case with our samples. Previous studies and the fungi identified in our samples, indicate that indeed there is fungi present in our samples, but in too low of an abundance to be easily detected and amplified. Finally, these results did not allow us to explore the relationship between bacteria and fungi communities.

In contrast, we found no problems detecting and identifying bacteria in the samples, despite the large proportion of contaminants and off-targets found. Although there was no difference in the distribution of the samples or the number of bacterial copies between carcinoma and adjacent tissue, we could observe some trends in our data. Fusobacteriota and Campylobacterota appear to be enriched in carcinoma samples compared to adjacent tissue. While no difference was observed in the other phyla. These results are in agreement with other studies, where *Fusobacterium* is reported as an opportunistic pathogen which usually co-occurs with *Campylobacter*.

In conclusion, it was not possible to characterise the fungal mycobiota from these samples. Both bacterial amplicons and fungi contained a large percentage of contaminants and off-targets. Although this was not a problem when it came to exploring the bacterial microbiota, it was for fungi. Despite this, we could observe how the DNA extraction method affected the sample distribution in both kingdoms. This reinforce the need to keep consistency when processing samples.

Final discussion and conclusion

This thesis aimed to develop an experimental and analytical pipeline that allows a better understanding of the microbial dynamics occurring on the human colon in health and disease.

Most of the literature uses fecal samples when studying the human gastrointestinal microbiota due to their easy access. However, they can include microbes that do not end up colonizing the intestine introduced by the diet (Schmidt et al., 2019). For this reason, we used colonic tissue, a more representative sample of the microbes residing in the mucus. Additionally, tissue samples allow us to observe differences in microbial composition between different parts of the colon (Hillman et al., 2017).

To preserve tissue samples and stop microbial growth during sample collection, we study the effect of different storage solutions. PBS is commonly used for its isotonic and non-toxic features when working with tissue. However, it does not prevent cell bursting when freezing the samples, which can alter the microbial composition of the samples (Bahl et al., 2012). For this reason, we study the effects of using RNA*later* in DNA quality, quantity, and absolute bacterial abundance. RNA*later* minimizes the need to process tissue samples immediately or freeze them. This storage solution is used to work with RNA and preserves the integrity of the cells, preventing them from bursting during sample handling (Choi et al., 2012). In our experiments, samples preserved in RNA*later* had higher DNA concentrations and bacterial absolute abundance. Also, the storage solution did not affect DNA quality. However, we observed some variability in DNA concentration and bacterial abundance among samples depending on the DNA extraction method. For this reason, it is necessary

to watch out for the interaction between RNA*later* and the DNA extraction method.

Besides tissue preservation, the low abundance of microbes was another issue we faced when working with tissue. The density of bacterial cells is much higher in fecal samples than in intestinal tissue (Schmidt et al., 2019; Vandeputte et al., 2017). Meanwhile, there is a higher proportion of human cells in the tissue. The high ratio of human cells compared to bacteria in the tissue makes it difficult to identify microbes in low abundance, thus compromising the use of high-throughput sequencing technologies such as shotgun metagenomics. Shotgun metagenomics would achieve precise sequencing resolution (down to strain level) and insights into community functionality; however, it would also sequence the large proportion of human genetic material, 'masking' those microbes in lesser abundance. Considering the aim of this thesis, we chose to use amplicon sequencing. Although amplicon sequencing does not allow as much sequencing resolution and functional information as shotgun metagenomics; it allows sequencing to be targeted to a specific taxonomic group defined by the marker gene used. In this way, we could study the human microbiome from a polykingdom approach by using marker genes and reference databases specific to each kingdom. The possibility of studying different microbial kingdoms was relevant for the main aim of the thesis since most studies only focus on the role of bacteria in gastrointestinal diseases. For detecting bacteria and fungi, we used 16S and ITS1 marker genes along with the databases SILVA and UNITE. In parallel, we used qPCR to quantify the abundance of each taxon and enable a QMP approach.

In practice, we could not obtain consistent results with ITS1 primers from neither sequencing nor qPCR, even though we tried different reaction condi-

tions with ITS1 primers for qPCR and compensated for low fungal abundance during sequencing. Recent studies confirm the presence of fungi in the intestinal mucus and provide different options to identify and quantify fungi (Coker et al., 2019; Luan et al., 2015; Rao et al., 2021). These options include using four different ITS-forward primers to improve fungus phylum coverage and switching to a sequencing platform that generates a high number of reads, such as HiSeq 2000.

Initially, we planned to detect other eukaryotes, but this could not be optimized and implemented in the thesis due to time constraints. However, if we were aiming to identify other eukaryotes beyond fungi, we would have used 18S as a eukaryotic marker gene which also amplifies human DNA. To prevent the amplification of host genetic material, 18S primers can be used along with blocking primers (Vestheim & Jarman, 2008). Blocking primers for 18S have been used in other animals but never in humans. Given the growing attention to the human microbiota and its role in diseases, it would have been very innovative to include the detection of other eukaryotes using blocking primers in the thesis.

Despite using amplicon sequencing, there was a high proportion of off-targets from host contamination in the CRC tissue samples. We removed the off-target reads during the analysis. However, most studies do not account for this type of contamination inherent to the sample (Bedarf et al., 2021). These off-targets lead to an overestimation of microbial diversity. Besides host contamination, we also identified contamination from the sequencing process and removed it. However, we still expect other contaminant reads from sample handling to be present in the samples. These contaminant reads could be detected by including negative controls since samples reception, but most

experimental designs do not have these controls (Glassing et al., 2016; Salter et al., 2014).

Another important source of variability during sample processing is the DNA extraction method (Hallmaier-Wacker et al., 2018). We tested three different commercial kits for extracting DNA. Each kit offers a different combination of cell lysis and DNA isolation. As we expected, each kit yielded a different DNA concentration, DNA quality and 16S copies using the same tissue sample. We also observed differences among kits in microbial composition. Finally, we found significant variability introduced by the DNA extraction method in the microbial composition of tissue samples from CRC clinical cases. Overall, these results highlight the importance of selecting an appropriate DNA extraction method based on the working material and expected outcome. And more importantly, these results show how the DNA extraction method should not change throughout the study.

The last chapter of this thesis was an opportunity to understand the underlying microbial dynamics in CRC and explore the effects of different confounders. For this chapter, we wanted to compare the differences in microbial composition between carcinoma and adjacent tissue using paired samples and including fungi. Although we did not find differences between carcinoma and non-carcinogenic tissue, we observed previously reported trends such as the increase in the abundance of Fusobacteria in on-tumour samples (Luan et al., 2015), even after correcting by bacterial absolute abundance. Furthermore, we explored the impact of confounding factors such as the DNA extraction method, sample tumour stage, location, external contamination, subjects sex, or age. The potential impact of these variables has become more relevant in recent years (Feng et al., 2015; Schmidt et al., 2018; Zhang et al., 2021). We

could not conclude to what extent our results were affected by tumour stage, sex, age or sample location as in other studies. But we managed to explore the microbial community associated to colorectal cancer in a novel way by considering these often overlooked confounding factors.

Bibliography

Abusleme, L., Hong, B. Y., Dupuy, A. K., Strausbaugh, L. D., & Diaz, P. I. (2014). Influence of DNA extraction on oral microbial profiles obtained via 16S rRNA gene sequencing. *Journal of Oral Microbiology*, *6*(1). https://doi.org/10.3402/jom.v6.23990

Adams, R. I., Bateman, A. C., Bik, H. M., & Meadow, J. F. (2015). Microbiota of the indoor environment: a meta-analysis. *Microbiome*, *3*(1), 49. https://doi.org/10.1186/s40168-015-0108-3

Aden, K., Rehman, A., Waschina, S., Pan, W. H., Walker, A., Lucio, M., Nunez, A. M., Bharti, R., Zimmerman, J., Bethge, J., Schulte, B., Schulte, D., Franke, A., Nikolaus, S., Schroeder, J. O., Vandeputte, D., Raes, J., Szymczak, S., Waetzig, G. H., ... Rosenstiel, P. (2019). Metabolic Functions of Gut Microbes Associate With Efficacy of Tumor Necrosis Factor Antagonists in Patients With Inflammatory Bowel Diseases. *Gastroenterology*, *157*(5), 1279–1292. https://doi.org/10.1053/j.gastro.2019.07.025

Ai, D., Pan, H., Li, X., Gao, Y., Liu, G., & Xia, L. C. (2019). Identifying gut microbiota associated with colorectal cancer using a zero-inflated lognormal model. *Frontiers in Microbiology*, *10*(APR), 826. https://doi.org/10.3389/fmicb.2019.00826

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amitay, E. L., Werner, S., Vital, M., Pieper, D. H., Höfler, D., Gierse, I. J., Butt, J., Balavarca, Y., Cuk, K., & Brenner, H. (2017). Fusobacterium and colorectal cancer: Causal factor or passenger? Results from a large colorectal cancer screening study. *Carcinogenesis*, *38*(8), 781–788. https://doi.org/10.1093/carcin/bgx053

Anandakumar, A., Pellino, G., Tekkis, P., & Kontovounisios, C. (2019). Fungal microbiome in colorectal cancer: a systematic review. https://doi.org/10.1007/s13304-019-00683-8

Andersen, L. O., Bonde, I., Nielsen, H. B. B., & Stensvold, C. R. (2015). A retrospective metagenomics approach to studying Blastocystis. *FEMS Microbiology Ecology*, *91*(7). https://doi.org/10.1093/femsec/fiv072

Andersen, L. O., Vedel Nielsen, H., & Stensvold, C. R. (2013). Waiting for the human intestinal Eukaryotome. https://doi.org/10.1038/ismej.2013.21

Arumugam, M., Raes, J., Pelletier, E., Paslier, D. L., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., ... Zeller, G. (2011). Enterotypes of the human gut microbiome. *Nature*, *473*(7346), 174–180. https://doi.org/10.1038/nature09944

Audebert, C., Even, G., Cian, A., Blastocystis Investigation Group, Loywick, A., Merlin, S., Viscogliosi, E., Chabé, M., El Safadi, D., Certad, G., Delhaes, L., Pereira, B., Nourrisson, C., Poirier, P., Wawrzyniak, I., Delbac, F., Morelle, C., Bastien, P., Lachaud, L., ... Rabodonirina, M. (2016). Colonization with the enteric protozoa Blastocystis is associated with increased diversity of human gut bacterial microbiota. *Scientific Reports*, *6*. https://doi.org/10.1038/srep25255

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M. T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y. S., Kotowska, D., Colding, C., ... Jun, W. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host and Microbe*, *17*(5), 690–703. https://doi.org/10.1016/j.chom.2015.04.004

Bahl, M. I., Bergström, A., & Licht, T. R. (2012). Freezing fecal samples prior to dna extraction affects the firmicutes to bacteroidetes ratio determined by downstream quantitative pcr analysis. *FEMS Microbiology Letters*, *329*, 193–197. https://doi.org/10.1111/j.1574-6968.2012.02523.x

Bang, C., & Schmitz, R. A. (2018). Archaea: Forgotten players in the microbiome. https://doi.org/10.1042/ETLS20180035

Bedarf, J. R., Hildebrand, F., Coelho, L. P., Sunagawa, S., Bahram, M., Goeser, F., Bork, P., & Wüllner, U. (2017). Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. *Genome Medicine*, *9*(1). https://doi.org/10.1186/s13073-017-0428-y

Bedarf, J. R., Beraza, N., Khazneh, H., Özkurt, E., Baker, D., Borger, V., Wüllner, U., & Hildebrand, F. (2021). Much ado about nothing? Off-target amplification can lead to false-positive bacterial brain microbiome detection in healthy and Parkinson's disease individuals. *Microbiome*, *9*(1), 1–15. https://doi.org/10.1186/s40168-021-01012-1

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bonk, F., Popp, D., Harms, H., & Centler, F. (2018). PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. https://doi.org/10.1016/j.mimet.2018.09.015

Brennan, C. A., & Garrett, W. S. (2016). Gut Microbiota, Inflammation, and Colorectal Cancer. *Annual Review of Microbiology*, *70*(1), 395–411. https://doi.org/10.1146/annurev-micro-102215-095513

Brennan, C. A., & Garrett, W. S. (2019). Fusobacterium nucleatum — symbiont, opportunist and oncobacterium. *Nature reviews. Microbiology*, *17*, 156. https://doi.org/10.1038/S41579-018-0129-6

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., Reris, R. A., Sheth, N. U., Huang, B., Girerd, P., Strauss, J. F., Jefferson, K. K., & Buck, G. A. (2015). The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies Ecological and evolutionary microbiology. *BMC Microbiology*, *15*(1), 66. https://doi.org/10.1186/s12866-015-0351-6

Buffie, C. G., Bucci, V., Stein, R. R., McKenney, P. T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., Littmann, E., Van Den Brink, M. R., Jenq, R. R., Taur, Y., Sander, C., Cross, J. R., Toussaint, N. C., Xavier, J. B., & Pamer, E. G. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature*, *517*(7533), 205–208. https://doi.org/10.1038/nature13828

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

Camara, A., Konate, S., Tidjani Alou, M., Kodio, A., Togo, A. H., Cortaredona, S., Henrissat, B., Thera, M. A., Doumbo, O. K., Raoult, D., & Million, M. (2021). Clinical evidence of the role of Methanobrevibacter smithii in severe acute malnutrition. *Scientific reports*, *11*(1), 5426. https://doi.org/10.1038/s41598-021-84641-8

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(SUPPL. 1), 4516–4522. https://doi.org/10.1073/pnas.1000080107

Chandramathi, S., Suresh, K., & Kuppusamy, U. R. (2010). Solubilized antigen of Blastocystis hominis facilitates the growth of human colorectal cancer cells, HCT116. *Parasitology Research*, *106*(4), 941–945. https://doi.org/10.1007/s00436-010-1764-7

Chen, J., & Vitetta, L. (2020). Butyrate in Inflammatory Bowel Disease Therapy. *Expert Opin Drug Metab Toxicol*, *1*, 1026. https://doi.org/10.1053/j.gastro.2019.08.064

Chen, Kost, J., Sulovari, A., Wong, N., Liang, W. S., Cao, J., & Li, D. (2019). A virome-wide clonal integration analysis platform for discovering cancer viral etiology. *Genome Research*, *29*(5), 819–830. https://doi.org/10.1101/gr.242529.118

Chen, Liu, F., Ling, Z., Tong, X., & Xiang, C. (2012). Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE*, *7*(6). https://doi.org/10.1371/journal.pone.0039743

Choi, Y.-T., Son, E.-H., Seo, B.-K., Lee, E.-H., Ryu, E.-K., Ha, J.-K., Kim, G.-W., Kwon, J.-S., Nam, M.-R., Kim, J.-H., Lee, Y.-J., & Kim, K.-R. (2012). Effects of storage buffer and temperature on the integrity of human dna. *Korean Journal of Clinical Laboratory Science*, *44*, 24–30.

Coker, O. O., Nakatsu, G., Dai, R. Z., Wu, W. K. K., Wong, S. H., Ng, S. C., Chan, F. K. L., Sung, J. J. Y., & Yu, J. (2019). Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. *Gut*, *68*(4), 654–662. https://doi.org/10.1136/gutjnl-2018-317178

Coleman, O. I., Lobner, E. M., Bierwirth, S., Sorbie, A., Waldschmitt, N., Rath, E., Berger, E., Lagkouvardos, I., Clavel, T., McCoy, K. D., Weber, A., Heikenwalder, M., Janssen, K. P., & Haller, D. (2018). Activated ATF6 Induces Intestinal Dysbiosis and Innate Immune Response to Promote Colorectal Tumorigenesis. *Gastroenterology*, *155*(5), 1539–1552. https://doi.org/10.1053/j.gastro.2018.07.028

Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., Yinda, C. K., Lavigne, R., Maes, P., Ranst, M. V., Heylen, E., & Matthijnssens, J. (2015). Modular approach to customise sample preparation procedures for viral metagenomics: A reproducible protocol for virome analysis. *Scientific Reports*, *5*(1), 1–14. https://doi.org/10.1038/srep16532

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, *6*(1). https://doi.org/10.1186/s40168-018-0605-2

Dejea, C. M., Fathi, P., Craig, J. M., Boleij, A., Taddese, R., Geis, A. L., Wu, X., DeStefano Shields, C. E., Hechenbleikner, E. M., Huso, D. L., Anders, R. A., Giardiello, F. M., Wick, E. C., Wang, H., Wu, S., Pardoll, D. M., Housseau, F., & Sears, C. L. (2018). Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science*, *359*(6375), 592–597. https://doi.org/10.1126/science.aah3648

Dejea, C. M., Wick, E. C., Hechenbleikner, E. M., White, J. R., Mark Welch, J. L., Rossettid, B. J., Peterson, S. N., Snesrud, E. C., Borisy, G. G., Lazarev, M., Stein, E., Vadivelu, J., Roslani, A. C., Malik, A. A., Wanyiri, J. W., Goh, K. L., Thevambiga, I., Fu, K., Wan, F., ... Sears, C. L. (2014). Microbiota organization is a distinct feature of proxi-

mal colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(51), 18321–18326. https://doi.org/10.1073/pnas.1406199111

Denizot, J., Sivignon, A., Barreau, F., Darcha, C., Chan, H. F., Stanners, C. P., Hofman, P., Darfeuille-Michaud, A., & Barnich, N. (2012). Adherent-invasive Escherichia coli induce claudin-2 expression and barrier defect in CEABAC10 mice and Crohn's disease patients. *Inflammatory Bowel Diseases*, *18*(2), 294–304. https://doi.org/10.1002/ibd.21787

DeRoche, T. C., Xiao, S. Y., & Liu, X. (2014). Histological evaluation in ulcerative colitis. https://doi.org/10.1093/gastro/gou031

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., & Relman, D. A. (2005). Microbiology: Diversity of the human intestinal microbial flora. *Science*, *308*(5728), 1635–1638. https://doi.org/10.1126/science.1110591

Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), 996–998. https://doi.org/10.1038/nmeth.2604

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Edwards, A. M., Grossman, T. J., & Rudney, J. D. (2006). Fusobacterium nucleatum transports noninvasive Streptococcus cristatus into human epithelial cells. *Infection and Immunity*, *74*(1), 654–662. https://doi.org/10.1128/IAI.74.1.654-662.2006

Ellermann, M., Huh, E. Y., Liu, B., Carroll, I. M., Tamayo, R., & Sartor, R. B. (2015). Adherent-invasive Escherichia coli production of cellulose influences iron-induced bacterial aggregation, phagocytosis, and induction of colitis. *Infection and Immunity*, *83*(10), 4068–4080. https://doi.org/10.1128/IAI.00904-15

Emlet, C., Ruffin, M., & Lamendella, R. (2020). Enteric Virome and Carcinogenesis in the Gut. https://doi.org/10.1007/s10620-020-06126-4

Engevik, M. A., Luk, B., Chang-Graham, A. L., Hall, A., Herrmann, B., Ruan, W., Endres, B. T., Shi, Z., Garey, K. W., Hyser, J. M., & Versalovic, J. (2019). Bifidobacterium dentium fortifies the intestinal mucus layer via autophagy and calcium signaling pathways. *mBio*, *10*(3). https://doi.org/10.1128/mBio.01087-19

Estaki, M., Jiang, L., Bokulich, N. A., McDonald, D., González, A., Kosciolek, T., Martino, C., Zhu, Q., Birmingham, A., Vázquez-Baeza, Y., Dillon, M. R., Bolyen, E., Caporaso, J. G., & Knight, R. (2020). QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome

Data and Comparative Studies with Publicly Available Data. *Current Protocols in Bioinformatics*, *70*(1). https://doi.org/10.1002/cpbi.100

Farmer, R. G., Hawk, W. A., & Turnbull, R. B. (1975). Clinical Patterns in Crohn's Disease: A Statistical Study of 615 Cases. *Gastroenterology*, *68*(4), 627–635. https://doi.org/10.1016/S0016-5085(75)80270-8

Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., Zhang, D., Xia, H., Xu, X., Jie, Z., Su, L., Li, X., Li, X., Li, J., Xiao, L., Huber-Schönauer, U., Niederseer, D., Xu, X., Al-Aama, J. Y., ... Wang, J. (2015). Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nature Communications*, *6*(1), 1–13. https://doi.org/10.1038/ncomms7528

Ferrer-Picón, E., Dotti, I., Corraliza, A. M., Mayorgas, A., Esteller, M., Perales, J. C., Ricart, E., Masamunt, M. C., Carrasco, A., Tristán, E., Esteve, M., & Salas, A. (2020). Intestinal Inflammation Modulates the Epithelial Response to Butyrate in Patients With Inflammatory Bowel Disease. *Inflammatory Bowel Diseases*, *26*(1), 43. https://doi.org/10.1093/IBD/IZZ119

Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D. T., Manara, S., Zolfo, M., Beghini, F., Bertorelli, R., De Sanctis, V., Bariletti, I., Canto, R., Clementi, R., Cologna, M., Crifò, T., Cusumano, G., ... Segata, N. (2018). Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host and Microbe*, *24*(1), 133–145. https://doi.org/10.1016/j.chom.2018.06.005

Fischer, J., Walker, L. C., Robinson, B. A., Frizelle, F. A., Church, J. M., & Eglinton, T. W. (2019). Clinical implications of the genetics of sporadic colorectal cancer. https://doi.org/10.1111/ans.15074

Flanagan, L., Schmid, J., Ebert, M., Soucek, P., Kunicka, T., Liska, V., Bruha, J., Neary, P., Dezeeuw, N., Tommasino, M., Jenab, M., Prehn, J. H., & Hughes, D. J. (2014). Fusobacterium nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. *European Journal of Clinical Microbiology and Infectious Diseases*, *33*(8), 1381–1390. https://doi.org/10.1007/s10096-014-2081-3

Frau, A., Kenny, J. G., Lenzi, L., Campbell, B. J., Ijaz, U. Z., Duckworth, C. A., Burkitt, M. D., Hall, N., Anson, J., Darby, A. C., & Probert, C. S. (2019). DNA extraction and amplicon production strategies deeply influence the outcome of gut mycobiome studies. *Scientific Reports*, *9*(1), 1–17. https://doi.org/10.1038/s41598-019-44974-x

Gao, R., Kong, C., Li, H., Huang, L., Qu, X., Qin, N., & Qin, H. (2017). Dysbiosis signature of mycobiota in colon polyp and colorectal cancer. *European Journal of Clinical Microbiology and Infectious Diseases*, *36*(12), 2457–2468. https://doi.org/10.1007/s10096-017-3085-6

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology*, *2*(2), 113–118. https://doi.org/10.1111/j.1365-294X.1993.tb00005.x

Garza, D. R., Taddese, R., Wirbel, J., Zeller, G., Boleij, A., Huynen, M. A., & Dutilh, B. E. (2020). Metabolic models predict bacterial passengers in colorectal cancer. *Cancer & Metabolism*, *8*(1), 1–13. https://doi.org/10.1186/s40170-020-0208-9

Gerrits, G. P., Klaassen, C., Coenye, T., Vandamme, P., & Meis, J. F. (2005). Burkholderia fungorum septicemia. *Emerging Infectious Diseases*, *11*(7), 1115–1117. https://doi.org/10.3201/eid1107.041290

Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S. J., Yassour, M., Morgan, X. C., Kostic, A. D., Luo, C., González, A., McDonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., … Xavier, R. J. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host and Microbe*, *15*(3), 382–392. https://doi.org/10.1016/j.chom.2014.02.005

Gill, E. E., & Brinkman, F. S. (2011). The proportional lack of archaeal pathogens: Do viruses/phages hold the key? *BioEssays*, *33*(4), 248–254. https://doi.org/10.1002/bies.201000091

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., & Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathogens*, *8*(1), 24. https://doi.org/10.1186/s13099-016-0103-7

Gogokhia, L., Buhrke, K., Bell, R., Hoffman, B., Brown, D. G., Hanke-Gogokhia, C., Ajami, N. J., Wong, M. C., Ghazaryan, A., Valentine, J. F., Porter, N., Martens, E., O'Connell, R., Jacob, V., Scherl, E., Crawford, C., Stephens, W. Z., Casjens, S. R., Longman, R. S., & Round, J. L. (2019). Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation and Colitis. *Cell Host and Microbe*, *25*(2), 285–299. https://doi.org/10.1016/j.chom.2019.01.008

Gouba, N., Raoult, D., & Drancourt, M. (2013). Plant and Fungal Diversity in Gut Microbiota as Revealed by Molecular and Culture Investigations. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059474

Guan, Q. (2019). A Comprehensive Review and Update on the Pathogenesis of Inflammatory Bowel Disease. *Journal of Immunology Research*, *2019*. https://doi.org/10.1155/2019/7247238

Hallen-Adams, H. E., & Suhr, M. J. (2017). Fungi in the healthy human gastrointestinal tract. https://doi.org/10.1080/21505594.2016.1247140

Hallmaier-Wacker, L. K., Lueert, S., Roos, C., & Knauf, S. (2018). The impact of storage buffer, DNA extraction method, and polymerase on microbial

analysis. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-24573-y

Hansson, G. C. (2020). Mucins and the Microbiome. *Annual Review of Biochemistry*, *89*(1), 769–793. https://doi.org/10.1146/annurev-biochem-011520-105053

Harmon, B. E., Wirth, M. D., Boushey, C. J., Wilkens, L. R., Draluck, E., Shivappa, N., Steck, S. E., Hofseth, L., Haiman, C. A., Le Marchand, L., & Hébert, J. R. (2017). The Dietary Inflammatory Index Is Associated with Colorectal Cancer Risk in the Multiethnic Cohort. *The Journal of nutrition*, *147*(3), 430–438. https://doi.org/10.3945/jn.116.242529

Hauswedell, H., Singer, J., & Reinert, K. (2014). Lambda: the local aligner for massive biological data. *30*, 349–355. https://doi.org/10.1093/bioinformatics/btu439

Hawinkel, S., Mattiello, F., Bijnens, L., & Thas, O. (2019). A broken promise: microbiome differential abundance methods do not control the false discovery rate. *Briefings in bioinformatics*, *20*(1), 210–221. https://doi.org/10.1093/bib/bbx104

He, Z., Gharaibeh, R. Z., Newsome, R. C., Pope, J. L., Dougherty, M. W., Tomkovich, S., Pons, B., Mirey, G., Vignard, J., Hendrixson, D. R., & Jobin, C. (2019). Campylobacter jejuni promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut*, *68*(2), 289–300. https://doi.org/10.1136/gutjnl-2018-317200

Hildebrand, F., Nguyen, T. L. A., Brinkman, B., Yunta, R. G., Cauwe, B., Vandenabeele, P., Liston, A., & Raes, J. (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biology*, *14*(1), R4. https://doi.org/10.1186/gb-2013-14-1-r4

Hildebrand, F., Tadeo, R., Voigt, A. Y., Bork, P., & Raes, J. (2014). LotuS: An efficient and user-friendly OTU processing pipeline. *Microbiome*, *2*(1), 30. https://doi.org/10.1186/2049-2618-2-30

Hillman, E. T., Lu, H., Yao, T., & Nakatsu, C. H. (2017). Microbial ecology along the gastrointestinal tract. https://doi.org/10.1264/jsme2.ME17017

Hoarau, G., Mukherjee, P. K., Gower-Rousseau, C., Hager, C., Chandra, J., Retuerto, M. A., Neut, C., Vermeire, S., Clemente, J., Colombel, J. F., Fujioka, H., Poulain, D., Sendid, B., & Ghannoum, M. A. (2016). Bacteriome and mycobiome interactions underscore microbial dysbiosis in familial Crohn's disease. *mBio*, *7*(5). https://doi.org/10.1128/mBio.01250-16

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G. D., Lewis, J. D., & Bushman, F. D. (2013). Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0066019

Hooper, L. V., Littman, D. R., & Macpherson, A. J. (2012). Interactions between the microbiota and the immune system. https://doi.org/10.1126/science.1223490

Hooper, L. V., & MacPherson, A. J. (2010). Immune adaptations that maintain homeostasis with the intestinal microbiota. https://doi.org/10.1038/nri2710

Horz, H. P. (2015). Archaeal lineages within the human microbiome: Absent, rare or elusive? https://doi.org/10.3390/life5021333

Hoyles, L., McCartney, A. L., Neve, H., Gibson, G. R., Sanderson, J. D., Heller, K. J., & van Sinderen, D. (2014). Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Research in Microbiology*, *165*(10), 803–812. https://doi.org/10.1016/j.resmic.2014.10.006

Iliev, I. D., & Leonardi, I. (2017). Fungal dysbiosis: Immunity and interactions at mucosal barriers. https://doi.org/10.1038/nri.2017.55

Jakobsson, H. E., Rodríguez-Piñeiro, A. M., Schütte, A., Ermund, A., Boysen, P., Bemark, M., Sommer, F., Bäckhed, F., Hansson, G. C., & Johansson, M. E. (2015). The composition of the gut microbiota shapes the colon mucus barrier. *EMBO reports*, *16*(2), 164–177. https://doi.org/10.15252/embr.201439263

Jess, T., Rungoe, C., & Peyrin-Biroulet, L. (2012). Risk of Colorectal Cancer in Patients With Ulcerative Colitis: A Meta-analysis of Population-Based Cohort Studies. *Clinical Gastroenterology and Hepatology*, *10*(6), 639–645. https://doi.org/10.1016/j.cgh.2012.01.010

Jian, C., Salonen, A., & Korpela, K. (2021). Commentary: How to Count Our Microbes? The Effect of Different Quantitative Microbiome Profiling Approaches. https://doi.org/10.3389/fcimb.2021.627910

Johansson. (2014). Mucus layers in inflammatory bowel disease. *Inflammatory Bowel Diseases*, *20*(11), 2124–2131. https://doi.org/10.1097/MIB.0000000000000117

Johansson, M. E., Holmén Larsson, J. M., & Hansson, G. C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(SUPPL. 1), 4659–4665. https://doi.org/10.1073/pnas.1006451107

Kamada, N., Chen, G. Y., Inohara, N., & Núñez, G. (2013). Control of pathogens and pathobionts by the gut microbiota. https://doi.org/10.1038/ni.2608

Kapitan, M., Niemiec, M. J., Steimle, A., Frick, J. S., & Jacobsen, I. D. (2019). Fungi as part of the microbiota and interactions with intestinal Bacteria. *Current topics in microbiology and immunology* (pp. 265–301). Springer Verlag. https://doi.org/10.1007/82{_}2018{_}117

Kernbauer, E., Ding, Y., & Cadwell, K. (2014). An enteric virus can replace the beneficial function of commensal bacteria. *Nature*, *516*, 94–98. https://doi.org/10.1038/NATURE13960

Keum, N. N., & Giovannucci, E. (2019). Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. https://doi.org/10.1038/s41575-019-0189-8

Khoury, J. D., Tannir, N. M., Williams, M. D., Chen, Y., Yao, H., Zhang, J., Thompson, E. J., Meric-Bernstam, F., Medeiros, L. J., Weinstein, J. N., & Su, X. (2013). Landscape of DNA Virus Associations across Human Malignant Cancers: Analysis of 3,775 Cases Using RNA-Seq. *Journal of Virology*, *87*(16), 8916–8926. https://doi.org/10.1128/jvi.00340-13

Kim, J. Y., Whon, T. W., Lim, M. Y., Kim, Y. B., Kim, N., Kwon, M. S., Kim, J., Lee, S. H., Choi, H. J., Nam, I. H., Chung, W. H., Kim, J. H., Bae, J. W., Roh, S. W., & Nam, Y. D. (2020). The human gut archaeome: Identification of diverse haloarchaea in Korean subjects. *Microbiome*, *8*(1), 114. https://doi.org/10.1186/s40168-020-00894-x

Kleessen, B., Kroesen, A. J., Buhr, H. J., & Blaut, M. (2002). Mucosal and Invading Bacteria in Patients with Inflammatory Bowel Disease Compared with Controls. *Scandinavian Journal of Gastroenterology*, *37*(9), 1034–1041. https://doi.org/10.1080/003655202320378220

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T., & Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(SUPPL. 1), 4578–4585. https://doi.org/10.1073/pnas.1000081107

Koskinen, K., Pausan, M. R., Perras, A. K., Beck, M., Bang, C., Mora, M., Schilhabel, A., Schmitz, R., & Moissl-Eichinger, C. (2017). First insights into the diverse human archaeome: Specific detection of Archaea in the gastrointestinal tract, lung, and nose and on skin. *mBio*, *8*(6). https://doi.org/10.1128/mBio.00824-17

Law, P. J., Timofeeva, M., Fernandez-Rozadilla, C., Broderick, P., Studd, J., Fernandez-Tajes, J., Farrington, S., Svinti, V., Palles, C., Orlando, G., Sud, A., Holroyd, A., Penegar, S., Theodoratou, E., Vaughan-Shaw, P., Campbell, H., Zgaga, L., Hayward, C., Campbell, A., ... Dunlop, M. G. (2019). Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nature Communications*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-09775-w

Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z., & Ettema, T. J. (2021). Innovations to culturing the uncultured microbial majority. https://doi.org/10.1038/s41579-020-00458-8

Li, Gong, J., Cottrill, M., Yu, H., De Lange, C., Burton, J., & Topp, E. (2003). Evaluation of QIAamp® DNA Stool Mini Kit for ecological studies

of gut microbiota. *Journal of Microbiological Methods*, *54*(1), 13–20. https://doi.org/10.1016/S0167-7012(02)00260-9

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, Peppelenbosch, M. P., & Smits, R. (2019). Bacterial biofilms as a potential contributor to mucinous colorectal cancer formation. https://doi.org/10.1016/j.bbcan.2019.05.009

Li, Wang, C., Tang, C., He, Q., Li, N., & Li, J. (2014). Dysbiosis of gut fungal microbiota is associated with mucosal inflammation in crohn's disease. *Journal of Clinical Gastroenterology*, *48*(6), 513–523. https://doi.org/10.1097/MCG.0000000000000035

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., Campieri, M., & Sokol, H. (2016). Fungal dysbiosis in mucosa-associated microbiota of Crohn's disease patients. *Journal of Crohn's and Colitis*, *10*(3), 296–305. https://doi.org/10.1093/ecco-jcc/jjv209

Lima, M. T., Andrade, A. C. d. S. P., Oliveira, G. P., Nicoli, J. R., Martins, F. d. S., Kroon, E. G., & Abrahão, J. S. (2019). Virus and microbiota relationships in humans and other mammals: An evolutionary view. https://doi.org/10.1016/j.humic.2018.11.001

Long, A. G., Lundsmith, E. T., & Hamilton, K. E. (2017). Inflammation and Colorectal Cancer. https://doi.org/10.1007/s11888-017-0373-6

Louca, S., Doebeli, M., & Parfrey, L. W. (2018). Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome*, *6*(1), 41. https://doi.org/10.1186/s40168-018-0420-9

Lu, Y., Chen, J., Zheng, J., Hu, G., Wang, J., Huang, C., Lou, L., Wang, X., & Zeng, Y. (2016). Mucosal adherent bacterial dysbiosis in patients with colorectal adenomas. *Scientific Reports*, *6*(1), 1–10. https://doi.org/10.1038/srep26337

Luan, C., Xie, L., Yang, X., Miao, H., Lv, N., Zhang, R., Xiao, X., Hu, Y., Liu, Y., Wu, N., Zhu, Y., & Zhu, B. (2015). Dysbiosis of fungal microbiota in the intestinal mucosa of patients with colorectal adenomas. *Scientific Reports*, *5*. https://doi.org/10.1038/srep07980

Magnusson, M. K., Isaksson, S., & Öhman, L. (2019). The Anti-inflammatory Immune Regulation Induced by Butyrate Is Impaired in Inflamed Intestinal Mucosa from Patients with Ulcerative Colitis. *Inflammation*, *43*(2), 507–517. https://doi.org/10.1007/s10753-019-01133-8

Mangifesta, M., Mancabelli, L., Milani, C., Gaiani, F., de'Angelis, N., de'Angelis, G. L., van Sinderen, D., Ventura, M., & Turroni, F. (2018). Mucosal microbiota of intestinal polyps reveals putative biomarkers of colorectal cancer. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-32413-2

Momozawa, Y., Deffontaine, V., Louis, E., & Medrano, J. F. (2011). Characterization of Bacteria in Biopsies of Colon and Stools by High Throughput Sequencing of the V2 Region of Bacterial 16S rRNA Gene in Human (F. Falciani, Ed.). *PLoS ONE*, *6*(2), e16952. https://doi.org/10.1371/journal.pone.0016952

Mori, G., Rampelli, S., Orena, B. S., Rengucci, C., De Maio, G., Barbieri, G., Passardi, A., Casadei Gardini, A., Frassineti, G. L., Gaiarsa, S., Albertini, A. M., Ranzani, G. N., Calistri, D., & Pasca, M. R. (2018). Shifts of Faecal Microbiota during Sporadic Colorectal Carcinogenesis. *Scientific Reports*, *8*(1), 10329. https://doi.org/10.1038/s41598-018-28671-9

Moyes, D. L., & Naglik, J. R. (2012). The mycobiome: Influencing IBD severity. https://doi.org/10.1016/j.chom.2012.05.009

Myerscough, N., Sylvester, P. A., Warren, B. F., Biddolph, S., Durdey, P., Thomas, M. G., Carlstedt, I., & Corfield, A. P. (2001). Abnormal subcellular distribution of mature MUC2 and de novo MUC5AC mucins in adenomas of the rectum: Immunohistochemical detection using non-VNTR antibodies to MUC2 and MUC5AC peptide. *Glycoconjugate Journal*, *18*(11-12), 907–914. https://doi.org/10.1023/A:1022204626604

Negreanu, L., Voiosu, T., State, M., Voiosu, A., Bengus, A., & Mateescu, B. R. (2019). Endoscopy in inflammatory bowel disease: from guidelines to real life. https://doi.org/10.1177/1756284819865153

Newbold, L. K., Burthe, S. J., Oliver, A. E., Gweon, H. S., Barnes, C. J., Daunt, F., & Van Der Gast, C. J. (2017). Helminth burden and ecological factors associated with alterations in wild host gastrointestinal microbiota. *ISME Journal*, *11*(3), 663–675. https://doi.org/10.1038/ismej.2016.153

Nkamga, V. D., Henrissat, B., & Drancourt, M. (2017). Archaea: Essential inhabitants of the human digestive microbiota. https://doi.org/10.1016/j.humic.2016.11.005

Norman, J. M., Handley, S. A., Baldridge, M. T., Droit, L., Liu, C. Y., Keller, B. C., Kambal, A., Monaco, C. L., Zhao, G., Fleshner, P., Stappenbeck, T. S., McGovern, D. P., Keshavarzian, A., Mutlu, E. A., Sauk, J., Gevers, D., Xavier, R. J., Wang, D., Parkes, M., & Virgin, H. W. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell*, *160*(3), 447–460. https://doi.org/10.1016/j.cell.2015.01.002

Odze, R. D. (2015). A contemporary and critical appraisal of 'indeterminate colitis'. *Modern Pathology*, *28*(1), S30–S46. https://doi.org/10.1038/modpathol.2014.131

Okumura, R., & Takeda, K. (2017). Roles of intestinal epithelial cells in the maintenance of gut homeostasis. https://doi.org/10.1038/emm.2017.20

Palmela, C., Chevarin, C., Xu, Z., Torres, J., Sevrin, G., Hirten, R., Barnich, N., Ng, S. C., & Colombel, J. F. (2018). Adherent-invasive Escherichia coli in inflammatory bowel disease. https://doi.org/10.1136/gutjnl-2017-314903

Parfrey, L. W., Walters, W. A., Lauber, C. L., Clemente, J. C., Berg-Lyons, D., Teiling, C., Kodira, C., Mohiuddin, M., Brunelle, J., Driscoll, M., Fierer, N., Gilbert, J. A., & Knight, R. (2014). Communities of microbial eukaryotes in the mammalian gut within the context of environmental eukaryotic diversity. *Frontiers in Microbiology*, *5*(JUN). https://doi.org/10.3389/fmicb.2014.00298

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. *Inflammatory Bowel Diseases*, *21*(11), 2515–2532. https://doi.org/10.1097/MIB.0000000000000549

Pittayanon, R., Lau, J. T., Leontiadis, G. I., 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*, *158*(4), 930–946. https://doi.org/10.1053/j.gastro.2019.11.294

Props, R., Kerckhof, F. M., Rubbens, P., Vrieze, J. D., Sanabria, E. H., Waegeman, W., Monsieurs, P., Hammes, F., & Boon, N. (2017). Absolute quantification of microbial taxon abundances. *ISME Journal*, *11*(2), 584–587. https://doi.org/10.1038/ismej.2016.117

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., ... Zoetendal, E. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/nature08821

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Rao, C., Coyte, K. Z., Bainter, W., Geha, R. S., Martin, C. R., & Rakoff-Nahoum, S. (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature*, *591*(7851), 633–638. https://doi.org/10.1038/s41586-021-03241-8

Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F., & Gordon, J. I. (2012). Going viral: Next-generation sequencing applied to phage populations in the human gut. https://doi.org/10.1038/nrmicro2853

Robinson, C. M., Jesudhasan, P. R., & Pfeiffer, J. K. (2014). Bacterial lipopolysaccharide binding enhances virion stability and promotes environmental fitness of an enteric virus. *Cell host microbe*, *15*, 36–46. https://doi.org/10.1016/J.CHOM.2013.12.004

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *2016*(10). https://doi.org/10.7717/peerj.2584

Rook, G. A. (2012). Hygiene hypothesis and autoimmune diseases. *Clinical Reviews in Allergy and Immunology*, *42*(1), 5–15. https://doi.org/10.1007/s12016-011-8285-8

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. (2015). Low prevalence of Blastocystis sp. in active ulcerative colitis patients. *European Journal of Clinical Microbiology and Infectious Diseases*, *34*(5), 1039–1044. https://doi.org/10.1007/s10096-015-2312-2

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., & Tuohy, K. (2018). Gut microbiota functions: metabolism of nutrients and other food components. https://doi.org/10.1007/s00394-017-1445-8

Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G., & Han, Y. W. (2013). Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating e-cadherin/-catenin signaling via its fada adhesin. *Cell host microbe*, *14*, 195–206. https://doi.org/10.1016/J.CHOM.2013.07.012

Ryan, F. J., Ahern, A. M., Fitzgerald, R. S., Laserna-Mendieta, E. J., Power, E. M., Clooney, A. G., O'Donoghue, K. W., McMurdie, P. J., Iwai, S., Crits-Christoph, A., Sheehan, D., Moran, C., Flemer, B., Zomer, A. L., Fanning, A., O'Callaghan, J., Walton, J., Temko, A., Stack, W., ... Claesson, M. J. (2020). Colonic microbiota is associated with inflammation and host epigenomic alterations in inflammatory bowel disease. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-15342-5

Saary, P., Forslund, K., Bork, P., & Hildebrand, F. (2017). RTK: efficient rarefaction analysis of large datasets. *Bioinformatics (Oxford, England)*, *33*(16), 2594–2595. https://doi.org/10.1093/bioinformatics/btx206

Saffarian, A., Mulet, C., Regnault, B., Amiot, A., Tran-Van-Nhieu, J., Ravel, J., Sobhani, I., Sansonetti, P. J., & Pédron, T. (2019). Crypt- and mucosa-associated core microbiotas in humans and their alteration in colon cancer patients. *mBio*, *10*(4). https://doi.org/10.1128/mBio.01315-19

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*(1), 1–12. https://doi.org/10.1186/s12915-014-0087-z

Sam, Q. H., Chang, M. W., & Chai, L. Y. A. (2017). The fungal mycobiome and its interaction with gut bacteria in the host. https://doi.org/10.3390/ijms18020330

Satinsky, B. M., Gifford, S. M., Crump, B. C., & Moran, M. A. (2013). Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods in enzymology* (pp. 237–250). Academic Press Inc. https://doi.org/10.1016/B978-0-12-407863-5.00012-5

Scanlan, P. D., & Marchesi, J. R. (2008). Micro-eukaryotic diversity of the human distal gut microbiota: Qualitative assessment using culture-dependent and -independent analysis of faeces. *ISME Journal*, *2*(12), 1183–1193. https://doi.org/10.1038/ismej.2008.76

Scanlan, P. D., Stensvold, C. R., Rajilić-Stojanović, M., Heilig, H. G., De Vos, W. M., O'Toole, P. W., & Cotter, P. D. (2014). The microbial eukaryote Blastocystis is a prevalent and diverse member of the healthy human gut microbiota. *FEMS Microbiology Ecology*, *90*(1), 326–330. https://doi.org/10.1111/1574-6941.12396

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schmidt, T. S., Hayward, M. R., Coelho, L. P., Li, S. S., Costea, P. I., Voigt, A. Y., Wirbel, J., Maistrenko, O. M., Alves, R. J., Bergsten, E., de Beaufort, C., Sobhani, I., Heintz-Buschart, A., Sunagawa, S., Zeller, G., Wilmes, P., & Bork, P. (2019). Extensive transmission of microbes along the gastrointestinal tract. *eLife*, *8*. https://doi.org/10.7554/eLife.42693

Schmidt, T. S., Raes, J., & Bork, P. (2018). The Human Gut Microbiome: From Association to Modulation. https://doi.org/10.1016/j.cell.2018.02.044

Sears, C. L., & Pardoll, D. M. (2011). Perspective: Alpha-bugs, their microbial partners, and the link to colon cancer. https://doi.org/10.1093/jinfdis/jiq061

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8). https://doi.org/10.1371/journal.pbio.1002533

Sokol, H., Leducq, V., Aschard, H., Pham, H. P., Jegou, S., Landman, C., Cohen, D., Liguori, G., Bourrier, A., Nion-Larmurier, I., Cosnes, J., Seksik, P., Langella, P., Skurnik, D., Richard, M. L., & Beaugerie, L. (2017). Fungal microbiota dysbiosis in IBD. *Gut*, *66*(6), 1039–1048. https://doi.org/10.1136/gutjnl-2015-310746

Stämmler, F., Gläsner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P. J., Gessner, A., & Spang, R. (2016). Adjusting microbiome profiles for

differences in microbial load by spike-in bacteria. *Microbiome*, *4*. https://doi.org/10.1186/s40168-016-0175-0

Strati, F., Di Paola, M., Stefanini, I., Albanese, D., Rizzetto, L., Lionetti, P., Calabrò, A., Jousson, O., Donati, C., Cavalieri, D., & De Filippo, C. (2016). Age and gender affect the composition of fungal population of the human gastrointestinal tract. *Frontiers in Microbiology*, *7*(AUG). https://doi.org/10.3389/fmicb.2016.01227

Strauss, J., Kaplan, G. G., Beck, P. L., Rioux, K., Panaccione, R., Devinney, R., Lynch, T., & Allen-Vercoe, E. (2011). Invasive potential of gut mucosa-derived fusobacterium nucleatum positively correlates with IBD status of the host. *Inflammatory Bowel Diseases*, *17*(9), 1971–1978. https://doi.org/10.1002/ibd.21606

Swidsinski, A., Dörffel, Y., Loening-Baucke, V., Tertychnyy, A., Biche-Ool, S., Stonogin, S., Guo, Y., & Sun, N. D. (2012). Mucosal invasion by fusobacteria is a common feature of acute appendicitis in Germany, Russia, and China. *Saudi Journal of Gastroenterology*, *18*(1), 55–58. https://doi.org/10.4103/1319-3767.91734

Tai, W. P., Hu, P. J., Wu, J., & Lin, X. C. (2011). Six ulcerative colitis patients with refractory symptoms co-infective with Blastocystis hominis in China. *Parasitology Research*, *108*(5), 1207–1210. https://doi.org/10.1007/s00436-010-2164-8

Toychiev, A., Abdujapparov, S., Imamov, A., Navruzov, B., Davis, N., Badalova, N., & Osipova, S. (2018). Intestinal helminths and protozoan infections in patients with colorectal cancer: prevalence and possible association with cancer pathogenesis. *Parasitology Research*, *117*(12), 3715–3723. https://doi.org/10.1007/s00436-018-6070-9

Toychiev, A., Navruzov, B., Pazylova, D., Davis, N., Badalova, N., & Osipova, S. (2021). Intestinal protozoa and helminths in ulcerative colitis and the influence of anti-parasitic therapy on the course of the disease. *Acta Tropica*, *213*, 105755. https://doi.org/10.1016/j.actatropica.2020.105755

Tsoi, K. K., Pau, C. Y., Wu, W. K., Chan, F. K., Griffiths, S., & Sung, J. J. (2009). Cigarette Smoking and the Risk of Colorectal Cancer: A Meta-analysis of Prospective Cohort Studies. *Clinical Gastroenterology and Hepatology*, *7*(6), 682–688. https://doi.org/10.1016/j.cgh.2009.02.016

Ungaro, F., Massimino, L., Furfaro, F., Rimoldi, V., Peyrin-Biroulet, L., D'Alessio, S., & Danese, S. (2019). Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease. *Gut Microbes*, *10*(2), 149–158. https://doi.org/10.1080/19490976.2018.1511664

Van Der Post, S., Jabbar, K. S., Birchenough, G., Arike, L., Akhtar, N., Sjovall, H., Johansson, M. E., & Hansson, G. C. (2019). Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis patho-

genesis. *Gut*, *68*(12), 2142–2151. https://doi.org/10.1136/gutjnl-2018-317571

Van Raay, T., & Allen-Vercoe, E. (2017). Microbial Interactions and Interventions in Colorectal Cancer. *Microbiology Spectrum*, *5*(3). https://doi.org/10.1128/microbiolspec.bad-0004-2016

Vandamme, P., De Brandt, E., Houf, K., Salles, J. F., van Elsas, J. D., Spilker, T., & LiPuma, J. J. (2013). Burkholderia humi sp. nov., Burkholderia choica sp. nov., Burkholderia telluris sp. nov., Burkholderia terrestris sp. nov. and Burkholderia udeis sp. nov.: Burkholderia glathei-like bacteria from soil and rhizosphere soil. *International Journal of Systematic and Evolutionary Microbiology*, *63*(PART 12), 4707–4718. https://doi.org/10.1099/ijs.0.048900-0

Vandeputte, D., Kathagen, G., D'Hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R. Y., De Commer, L., Darzi, Y., Vermeire, S., Falony, G., & Raes, J. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature*, *551*(7681), 507–511. https://doi.org/10.1038/nature24460

van Tilburg Bernardes, E., Pettersen, V. K., Gutierrez, M. W., Laforest-Lapointe, I., Jendzjowsky, N. G., Cavin, J. B., Vicentini, F. A., Keenan, C. M., Ramay, H. R., Samara, J., MacNaughton, W. K., Wilson, R. J., Kelly, M. M., McCoy, K. D., Sharkey, K. A., & Arrieta, M. C. (2020). Intestinal fungi are causally implicated in microbiome assembly and immune development in mice. *Nature Communications*, *11*(1), 1–16. https://doi.org/10.1038/s41467-020-16431-1

Velcich, A., Yang, W. C., Heyer, J., Fragale, A., Nicholas, C., Viani, S., Kucherlapati, R., Lipkin, M., Yang, K., & Augenlicht, L. (2002). Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science*, *295*(5560), 1726–1729. https://doi.org/10.1126/science.1069094

Verhulst, J., Ferdinande, L., Demetter, P., & Ceelen, W. (2012). Mucinous subtype as prognostic factor in colorectal cancer: A systematic review and meta-analysis. https://doi.org/10.1136/jclinpath-2011-200340

Vestheim, H., & Jarman, S. N. (2008). Blocking primers to enhance pcr amplification of rare sequences in mixed samples - a case study on prey dna in antarctic krill stomachs. *Frontiers in Zoology*, *5*, 1–11. https://doi.org/10.1186/1742-9994-5-12/FIGURES/3

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., Brown, D., Stares, M. D., Scott, P., Bergerat, A., Louis, P., McIntosh, F., Johnstone, A. M., Lobley, G. E., Parkhill, J., & Flint, H. J. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME Journal*, *5*(2), 220–230. https://doi.org/10.1038/ismej.2010.118

Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N., Rayment, N., Brostoff, J., Parkhill, J., Dougan, G., & Petrovska, L.

(2011). High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiology*, *11*(1), 7. https://doi.org/10.1186/1471-2180-11-7

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. A., Jansson, J. K., Caporaso, J. G., Fuhrman, J. A., Apprill, A., & Knight, R. (2016). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*, *1*(1). https://doi.org/10.1128/msystems.00009-15

Wampach, L., Heintz-Buschart, A., Hogan, A., Muller, E. E., Narayanasamy, S., Laczny, C. C., Hugerth, L. W., Bindl, L., Bottu, J., Andersson, A. F., de Beaufort, C., & Wilmes, P. (2017). Colonization and succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life. *Frontiers in Microbiology*, *8*(MAY). https://doi.org/10.3389/fmicb.2017.00738

Wang, J., Lang, T., Shen, J., Dai, J., Tian, L., & Wang, X. (2019). Core Gut Bacteria Analysis of Healthy Mice. *Frontiers in Microbiology*, *10*(APR), 887. https://doi.org/10.3389/fmicb.2019.00887

Warren, R. L., Freeman, D. J., Pleasance, S., Watson, P., Moore, R. A., Cochrane, K., Allen-Vercoe, E., & Holt, R. A. (2013). Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome*, *1*(1), 16. https://doi.org/10.1186/2049-2618-1-16

White, T., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. *In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR protocols: a guide to methods and applications. United States: Academic Press. pp. 315–322.*

Wilks, J., Lien, E., Jacobson, A. N., Fischbach, M. A., Qureshi, N., Chervonsky, A. V., & Golovkina, T. V. (2015). Mammalian lipopolysaccharide receptors incorporated into the retroviral envelope augment virus transmission. *Cell host microbe*, *18*, 456–462. https://doi.org/10.1016/J.CHOM.2015.09.005

Wong, S. H., & Yu, J. (2019). Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. https://doi.org/10.1038/s41575-019-0209-8

Wu, H., Rebello, O., Crost, E. H., Owen, C. D., Walpole, S., Bennati-Granier, C., Ndeh, D., Monaco, S., Hicks, T., Colvile, A., Urbanowicz, P. A., Walsh, M. A., Angulo, J., Spencer, D. I., & Juge, N. (2020). Fucosidases from the human gut symbiont Ruminococcus gnavus. *Cellular and Molecular Life Sciences*, *78*(2), 675. https://doi.org/10.1007/s00018-020-03514-x

Wu, Z., Mirza, H., & Tan, K. S. W. (2014). Intra-Subtype Variation in Enteroadhesion Accounts for Differences in Epithelial Barrier Disruption and Is Associated with Metronidazole Resistance in Blastocystis Subtype-7. *PLoS Neglected Tropical Diseases*, *8*(5). https://doi.org/10.1371/journal.pntd.0002885

Xiao, X., Wang, L., Wei, P., Chi, Y., Li, D., Wang, Q., Ni, S., Tan, C., Sheng, W., Sun, M., Zhou, X., & Du, X. (2013). Role of MUC20 overexpression as a predictor of recurrence and poor outcome in colorectal cancer. *Journal of Translational Medicine*, *11*(1). https://doi.org/10.1186/1479-5876-11-151

Zeng, M. Y., Inohara, N., & Nuñez, G. (2017). Mechanisms of inflammation-driven bacterial dysbiosis in the gut. https://doi.org/10.1038/mi.2016.75

Zhang, X., Zhong, H., Li, Y., Shi, Z., Ren, H., Zhang, Z., Zhou, X., Tang, S., Han, X., Lin, Y., Yang, F., Wang, D., Fang, C., Fu, Z., Wang, L., Zhu, S., Hou, Y., Xu, X., Yang, H., ... Ji, L. (2021). Sex- and age-related trajectories of the adult human gut microbiota shared across populations of different ethnicities. *Nature Aging*, *1*, 87–100. https://doi.org/10.1038/s43587-020-00014-2