Deciphering the Identity of Oral and Gut bacteria in Colorectal Cancer

Lisa Alves Azevedo

Thesis to obtain the Master of Science Degree in

Biological Engineering

Supervisors: Dr Leonilde de Fátima Morais Moreira
Dr Paul O'Toole

Examination Committee

Chairperson: Dr Gabriel António Amaro Monteiro
Supervisor: Dr Leonilde de Fátima Morais Moreira
Member of the Committee: Dr Nuno Filipe Santos Bernardes

November 2018

‘’Porque eu sou do tamanho do que vejo

E não do tamanho da minha altura”

‘”Because I am the size of what I see

And not the size of my height”

Alberto Caeiro

Abstract

Colorectal cancer (CRC) is consistently the third deadliest cancer worldwide. Recent research has established that an altered community of bacteria (dysbiotic microbiota) is a feature of CRC and has questioned if these bacteria can be derived partly from the oral cavity.

In this work, CRC patients were recruited, and a broad range approach was developed that includes histology and immunology techniques; microbiota metagenomic profiling; and a complete culture, isolation and identification methodology exploring multiple conditions that targeted faecal, oral and tumour bacteria of CRC patients. Oral colonization in CRC tumours was investigated by culturing and isolating bacteria from colonic tumours and oral swabs of 2 CRC patients; identifying bacteria by sequencing an amplicon from the full length 16S rRNA gene by Sanger sequencing; and performing random amplified polymorphic DNA (RAPD) genetic fingerprinting of strains to compare the genomic structure of oral and colonic isolates of the same species.

Oral colonization in CRC tumours was demonstrated by successful culture from both locations of a considerable range of isolated species - 15 species belonging to 6 different families and 2 phyla - whose ability to elicit inflammation, cytokine production and cell proliferation can now be investigated. This project allowed the creation of a defined phylogenetic collection with 500 pure archived cultures, belonging to 95 different species (containing 10 strains considered until now uncultured and unsequenced by the Human Microbiome Project), 28 families and 5 phyla. Their future study might bring new insights into CRC disease mechanisms.

Keywords: Colorectal cancer, Microbiota, Dysbiosis, Oral colonization, Broad culture-based approach

Resumo

O cancro do colo-retal (CRC) é consistentemente o terceiro cancro mais mortífero em todo o mundo. Estudos recentes estabeleceram que uma comunidade alterada de bactérias (microbiota disbiótico) é uma característica do CRC e questionaram se parte destas bactérias serão derivadas da cavidade oral.

Neste trabalho, recrutaram-se pacientes com CRC, e desenvolveu-se uma abordagem alargada, incluindo técnicas de histologia e imunologia; análise metagenómica do microbiota; e uma metodologia completa de cultura e identificação de microrganismos, testando múltiplas condições, que permitiu isolar bactérias fecais, orais e de tumor destes pacientes. A colonização de tumores por bactérias orais foi investigada através do cultivo e isolamento de bactérias provenientes desses tumores e esfregaços orais de 2 pacientes com CRC. Seguiu-se a identificação das espécies por sequenciação automática de Sanger (amplificação da sequência inteira do gene 16S rRNA) e por obtenção dos perfis de RAPD (*random amplified polymorphic DNA*) de forma a comparar a estrutura genómica dos isolados orais e tumorais da mesma espécie.

A colonização dos tumores por bactérias orais foi observada em 15 espécies pertencentes a 6 famílias diferentes e a 2 filos. Esta identificação é o primeiro passo, seguindo-se futuramente estudos sobre a sua capacidade para provocar inflamação, produção de citoquinas e proliferação celular. Este projeto permitiu criar uma coleção filogenética definida contendo 500 culturas puras pertencentes a 95 espécies diferentes (contendo 10 estirpes consideradas não cultiváveis até ao momento), 28 famílias e 5 filos, e que poderá contribuir para melhor compreender os mecanismos desta doença.

Palavras-chave: Cancro do colo-retal, Microbiota, Disbiose, Colonização oral, Abordagem diversificada de cultura

Acknowledgements

Several people contributed to this master thesis project and I would like to thank them. Please forgive me if I miss someone.

First, I must begin with Ana Almeida who made this journey to Ireland possible. I met her last year, as my supervisor, during a summer internship in Paul O'Toole's lab at APC Microbiome Ireland. This internship in Cork was supported by Calouste Gulbenkian Foundation-PARSUK Xperience 2017 grant. It was here that I had the first contact with colorectal cancer research. The opportunity of doing my thesis project in the same spot came easily. Dr Ana is a passionate (and workaholic) researcher that I admire professionally and personally. She is a woman with brain and guts. A fighter with a huge strength of will and with lots of positive energy, enough to give and to sell (as the Portuguese use to say). I want to thank her for her support and her friendship. This thesis is dedicated to her and to Rafael, her son.

I'm grateful for the opportunity of developing my project in Dr Paul O'Toole's lab. I want to thank him especially for his trust (in me and in my work). He gave me the chance to build my own path and to perform my work independently, which meant a lot to me. My internship didn't go in the usual way, because I didn't have a supervisor in the lab to guide me. Besides the hard work and the many sleepless nights, this unforeseen change of plans allowed me to be involved in all the steps of this project; to talk with researchers, doctors and nurses; to analyse, plan ambitiously and decide what I wanted to explore/research. And I absolutely loved it. I think I made the most out of this experience and because of that I know today what it feels like to be a researcher. Dr Paul also helped me with all the bureaucratic issues, despite being a busy man. I was a student and a worker at the same time, and for that, I need to thank the financial support given by APC (and authorized by Dr Paul) which allowed me to work and live in Cork without spending my parents' money (I think they are also grateful for that).

I would like to acknowledge all members of O'Toole's lab for their support in the lab and outside. I've learned everything with them. I'm deeply grateful to Alexandra Ntemiri and Marta Perez for the training in the anaerobic cabinet and for explaining me the steps and protocols within the culture, isolation and identification of pure cultures context. They were my true mentors through the experimental work and, therefore, the ones that most contributed to this thesis. The majority of the protocols presented in Methods section (2nd part) were adapted from Dr Alexandra and Dr Marta's protocols. I thank Werner for his help in the Illumina sequencing part and for his patience in answering my silly questions. I thank Maurice Barrett that plotted the microbiota profiles of patients from whose samples I've extracted gDNA last summer. Klára Vlčková that showed me the oral swab and faecal extraction protocols. Marta Neto for her guidance and kindness (I feel that she is the mother of everyone in the lab). Céline Ribière (the bioinformatics superwoman) that did an amazing job with the metagenomic analysis, providing me with the beautiful graphs presented in this context. I need to thank her also for investigating if some of my isolates were present in the 'Most Wanted' list of the Human Microbiome Project. Aida for her friendship and genuineness. Maxence Bourin that gave me his beads. Jean François for the encouraging words. Tam and Janaki for the funny talks and cinema sessions.

Besides work, the APC Microbiome Ireland also showed me the good environment that a research institute can have. People are so nice and always available to help. Interesting talks, workshops, courses occur every week and are available for every member (and sometimes with coffee, snacks and pizza at the end!). Being a student of University College Cork gave me some benefits that I won't forget, especially free entry in the biggest sports complex in Ireland (Mardyke Arena).

Outside the lab, there are some people that I must mention. Dr Clodagh Murphy – responsible for the recruitment of the two CRC patients and respective collection of the oral swab, saliva and tumour samples (used in the culture protocol). I would like to say thanks to a wonderful lady - the pathologist Juliet Barry from Cork Cancer Research Centre - who taught me everything that I needed to know about the staining methodologies. It was a pleasure being in her working environment with so many equipments and techniques that were unknown to me.

I would like to thank Sophie, Adrian, Alissa and Ambre for the good moments together after work and on the weekends. And also to my Irish family – Niamh, her three lovely kids and her friendly parents – for their hospitality and kindness.

I want to acknowledge my supervisor at Técnico, Dr Leonilde Moreira, for her support, trust and friendship. I know Dr Leonilde from my Genetic Engineering classes (the coolest that I had) and from a summer internship which I've applied to in her lab, regarding cystic fibrosis research. I thank her for inviting me to participate in her paper as a co-author. I enjoyed so much the work that we were developing at her lab, that I forgot to go to the party at the end of Alameda (in front of my university) to cheer the Portuguese football team, winner of the 2016 European Championship. She encouraged me to go outside and I'm thankful for that.

I also want to thank people that have inspired me to follow, from a young age, research of diseased states influenced or promoted by microorganisms. Especially, Dr Miguel Prudêncio and Dr Maria Mota from Instituto de Medicina Molecular (IMM), which allowed me to do volunteer work in their lab (and Insectarium with the mosquitoes) within the scope of malaria research.

Among my university friends, a special thanks goes to my buddy João Santana for all the support during these years. We formed a super team together. Among my closest friends, a special hug goes to my three companions António, Catarina and Filipa for their support and old friendship.

This is the end of 5 years in Instituto Superior Técnico, as a student of Biological Engineering. And what a journey. I couldn't be better prepared and eager to change the world!

Finally, some special words for my family. I want to thank my amazing parents for their support and love. They taught me to be never satisfied. To never judge what I don't know. And to be humble. To my intelligent and brave sister who taught me to be open-minded and brave too. To my stubborn grandmother which was a second mother to me and who taught me kindness. To my love and best friend Telmo, for the most important and beautiful journey of my life. Indeed, it's true that we are the size of what we see, and not the size of our height.

Contents

List of Figures

List of Tables

List of Abbreviations

ASC	Ascending
BHI	Brain Heart Infusion
CDT	Cytolethal distending toxin
CRC	Colorectal cancer
DAPI	4',6-diamidino-2-phenylindole
DESC	Descending
FISH	Fluorescent In-situ Hybridization
FITC	Fluorescein isothiocyanate
FOBT	Faecal occult blood test
H&E	Haematoxylin & Eosin
IF	Immunofluorescence
NCBI	National Center for Biotechnology Information
NOS	Nitrogen species
OTU	Operational taxonomic unit
PRRs	Pattern recognition receptors
RAPD	Random amplified polymorphic DNA
ROS	Reactive oxygen species
SCFA	Short chain fatty acids
TLR	Toll-Like receptors
TRANS	Transverse
TS	Tryptic Soy
UD	Un-diseased
YCFA	Yeast extract-casitone-fatty acid medium
CAG	Bacterial coabundance groups

Chapter 1

Introduction

1.1 Colorectal Cancer: Origin, Impact, Treatment

Cancer is characterized by the continuous and uncontrolled growth of abnormal cells beyond their usual boundaries that can invade and spread into surrounding tissues. Essentially, old or damaged cells are able to survive when they should die, and new cells form when they are not needed, which gives rise to growths called tumours (*What Is Cancer? - National Cancer Institute*, 2018). Colorectal cancer (CRC) is a type of cancer that starts in the colon (the longest part of the large intestine) or the rectum (the last several inches of the large intestine before the anus) (Figure 1.1). Significant risk factors for CRC include diets rich in fat and red meat, obesity and chronic inflammation of the gastrointestinal tract (Tilg *et al.*, 2018) - factors closely associated with changes in the composition and function of the gut bacterial community (the gut microbiota).

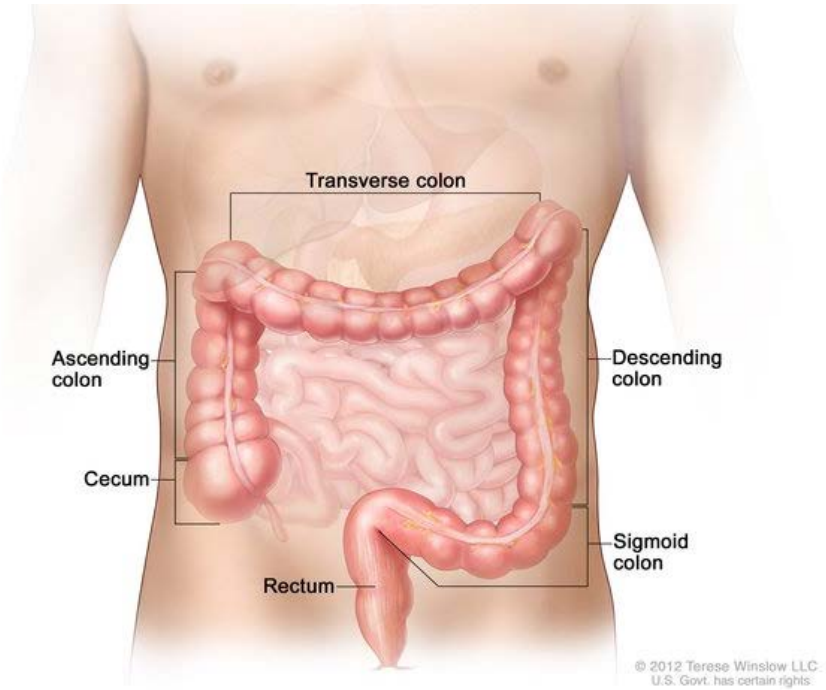


Figure 1.1 – Different parts of the large intestine (*Colon Anatomy - National Cancer Institute*, 2018).

1.1.1 Facts and numbers

Cancer is the second leading cause of death worldwide and is responsible for an estimated 9.6 million deaths in 2018. CRC is one of the most common cancers, estimated to affect 1.80 million people and kill 862 000 globally, in 2018 (*World Health Organization – Cancer Stats*, 2018). In industrialized nations, approximately 5% of people will be diagnosed with CRC at some point during their lifetime (*Colorectal Cancer - Cancer Stat Facts*, 2018). In addition, the lifetime risk of developing an adenoma (a noncancerous colon tumour that can develop into CRC) is approximately 20% (Sears and Garrett, 2014).

In 2018, 140 250 new cases of CRC and 50 630 deaths are estimated for this disease in the United States (*Colorectal Cancer - Cancer Stat Facts*, 2018). In Portugal, the incidence rate of colon and rectum malignant tumours is higher than the standard rate for the rest of Europe. CRC is the third deadliest cancer in Portugal with 4 000 people dying from the disease each year (Miranda and Portugal, 2016). By 2030, the global burden of CRC is expected to increase by 60% to more than 2.2 million new cases and 1.1 million deaths (Arnold *et al.*, 2017).

CRC is most frequently diagnosed among people aged 65-74 (*Colorectal Cancer - Cancer Stat Facts*, 2018). When the disease is local or confined, cure rates range from 70%–90% (Sears and Garrett, 2014). Indeed, 64.5% is the percentage of patients that survive 5 years or more after being diagnosed with CRC. However, advanced CRC has a high mortality rate, being the second leading cause of cancer death in the United States and consistently the third around the globe (*Colorectal Cancer - Cancer Stat Facts*, 2018).

1.1.2 Genetic alterations and progression of CRC

Tumours evolve from benign to malignant lesions by acquiring a series of mutations over time. CRC is essentially characterised by genomic instability either in the form of microsatellite instability or chromosomal instability. Microsatellite instability is the result of mutation of mismatch repair genes or their silencing through promoter methylation (Amaro *et al.*, 2016). Chromosomal instability is defined as a persistently high rate of loss and gain of whole chromosomes and its molecular causes are less well characterized (Thompson *et al.*, 2010).

A four-step progression of gene alterations in the colonic epithelium is well established (Figure 1.2). Mutations in several genes lead to colonic epithelial cells proliferation that slowly progresses over 10 to 40 years, resulting in adenoma or adenomatous polyp (a benign tumour), proceeding to *in situ* adenocarcinoma (a malignant tumour), and ultimately to the invasive and metastatic tumour. Although all adenomas have the potential to become cancerous, only ∼10% are estimated to progress to invasive cancer (Armaghany *et al.*, 2012).

It remains unknown which events precipitate either the initiation or progression of CRC. However, it's suggested that changes in the composition and function of one or several elements of the microbiota may play a crucial role in CRC progression (Sears and Garrett, 2014).

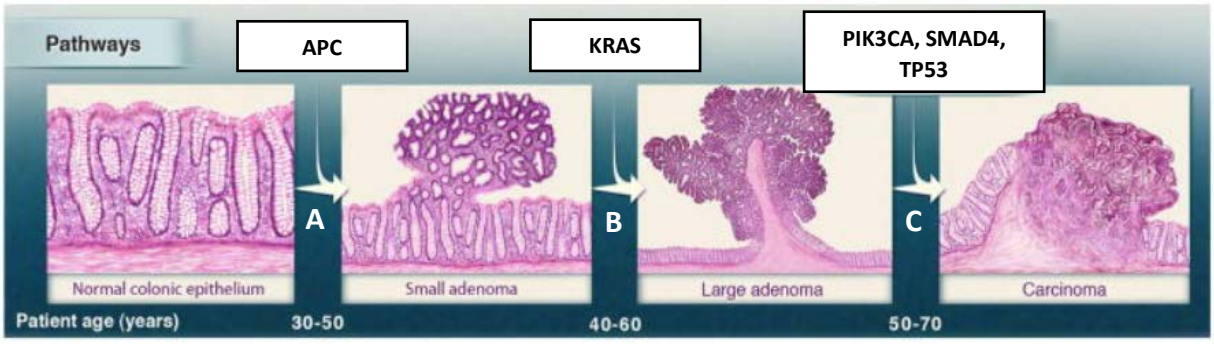


Figure 1.2 – Genetic alterations and progression of CRC. The signalling pathways that drive tumorigenesis are shown between each tumour stage (A, B and C transitions). Patient age indicates the time intervals during which the driver genes are usually mutated. A - The first mutation in the APC gene provides a selective growth advantage to a normal epithelial cell, allowing it to outgrow the cells that surround it. The result is the formation of a small adenoma that grows slowly. B - A second mutation in another gene, such as KRAS, unleashes a second round of clonal growth that allows an expansion of cell number, resulting in the development of a large adenoma. C - The previous process of mutation followed by clonal expansion continues, with mutations in genes such as PIK3CA, SMAD4 and TP53, which leads to the generation of a malignant tumour. This carcinoma is able to invade the underlying basement membrane and to metastasize to lymph nodes and distant organs such as the liver, in a later phase. Scheme adapted from (Vogelstein *et al.*, 2013).

1.1.3 Diagnosis and treatment

Tests used to detect and diagnose CRC examine the colon and the rectum. Some of these procedures include faecal occult blood test (FOBT) (test to check for occult blood in the stool); colonoscopy (a thin tube is inserted through the rectum into the colon to look for polyps, abnormal areas and cancer) (Figure 1.3.A); and biopsy (removal of cells or tissues so they can be viewed under a microscope by a pathologist to check for signs of cancer) (*Colon Cancer Treatment - National Cancer Institute*, 2018).

The most current screening approaches for CRC are FOBT and colonoscopy. As referenced before, most cases of CRC begin with the development of benign intestinal polyps which are relatively common in a more ageing population (> 50 years). In some cases, these finger-like growths that protrude into the lumen or the rectum may become cancerous and metastasize. They can be fragile and bleed intermittently, for instance, when food waste brushes against them. Usually, the blood released is not visible in the stool, but it can be detected with FOBT (non-invasive screening). There are two types of FOBTs - Guaiac FOBT (a small stool sample is placed in a special card that changes colour if there is blood present) and immunochemical FOBT (uses antibodies to detect the blood) (*Colon Cancer Treatment - National Cancer Institute*, 2018). It is recommended to perform these tests on at least three stool samples collected on different days. Although FOBT is a rapid screening method to detect advanced adenomas and carcinomas, it presents very low sensitivity for detecting early lesions which leads to the necessity of studying more reliable biomarkers (Flemer, Lynch, *et al.*, 2017).

There are different types of CRC treatment which include surgery such as resection of the colon (removing the cancer and a small amount of healthy tissue around it) (Figure 1.3.B); radiofrequency ablation (use of a special probe with tiny electrodes that kill cancer cells); chemotherapy (use of drugs to stop the growth of cancer cells); radiation therapy (use of high-energy X-rays or other types of radiation to stop the growth of cancer cells); and targeted therapy such as monoclonal antibodies and angiogenesis inhibitors (use of drugs or other substances to identify and attack specific cancer cells without harming normal cells) (*Colon Cancer Treatment - National Cancer Institute*, 2018).

A combination of immunotherapy drugs has been approved (by the Food and Drug Administration) for metastatic CRC, which includes ipilimumab and nivolumab – immune checkpoint inhibitors. These drugs are effective in patients whose tumours present one of two genetic features that prevent cells from repairing damaged DNA – high microsatellite instability or mismatch repair deficiency. Tumours with these features have high numbers of DNA mutations, some of which lead to the production of many abnormal antigens that are targeted by immune cells. Checkpoint proteins are found on T-cells and cancer cells, and they can keep T-cells from killing cancer cells. Checkpoint inhibitors block these proteins that stop the immune system from attacking the cancer cells, releasing, this way, the brakes on the immune system and allowing T-cell activation (*Immunotherapy - CRC*, 2018).

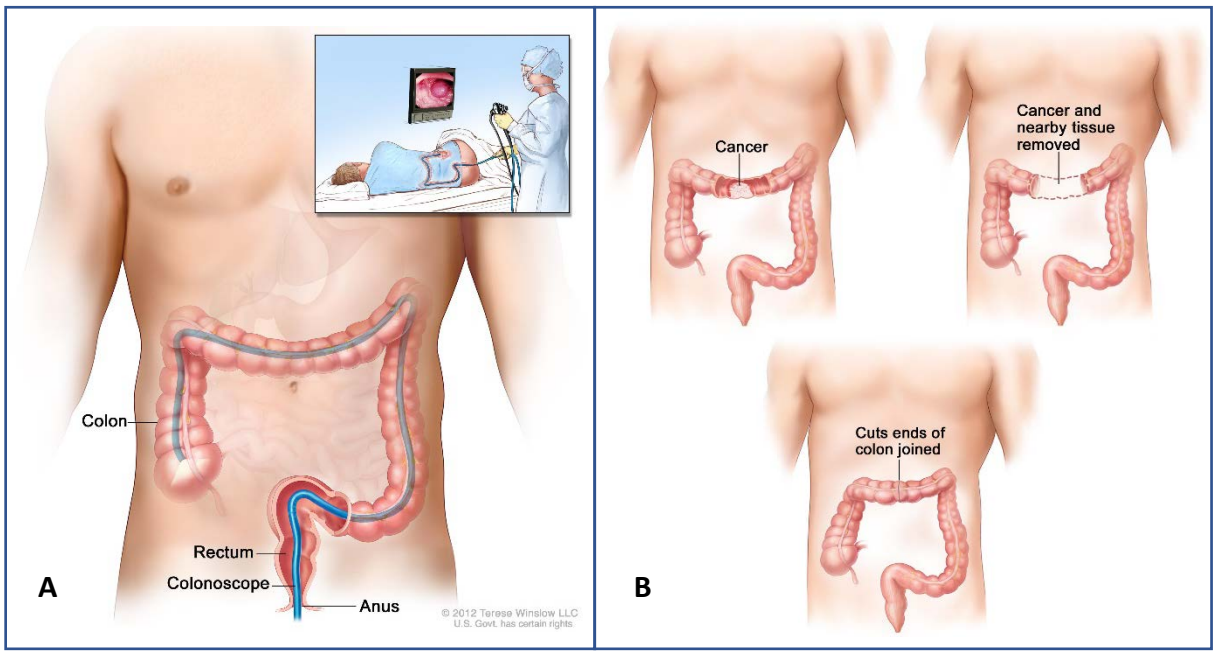


Figure 1.3 – Examples of screening and treatment approaches in CRC. A – Colonoscopy. B - Resection of the colon with anastomosis (part of the colon containing the cancer and nearby healthy tissue is removed, and then the cut ends of the colon are joined). Adapted from (*Colon Cancer Treatment - National Cancer Institute*, 2018).

1.2 Gut microbiota and Colorectal Cancer

The human gastrointestinal tract harbours trillions of bacteria that have co-evolved to play important roles in health and disease, conferring immunological, metabolic and neurological benefits to the host (Rooks and Garrett, 2016). The colon contains approximately 10^{14} bacteria which represent 70% of all the microorganisms that inhabit our body (Gagnière *et al.*, 2016). In addition, the colon is a common site for adenocarcinomas and also one of the most densely populated microbial ecosystems on our planet (Sears and Garrett, 2014). Our microbiota often referred to as the "forgotten organ", contains a metagenome that exceeds our own genome by 100-fold (Gill *et al.*, 2006). All these numbers suggest a potential role for the gut microbiota in colorectal carcinogenesis.

Intestinal colonization begins at birth during the first stages of life through the contact with the commensal flora from mother's vagina, skin and milk. From here, the gut microbiota matures during the first two years and its development is the result of interactions between the physiological process in the host and new microorganisms, which are introduced from the environment. The early acquisition of a balanced and diverse bacterial community seems to be critical for the development and maturation of a healthy immune system in the host since germ-free mice raised in bacteria-free conditions acquire immune abnormalities (Gagnière *et al.*, 2016). After this initial stage, the gut microbiota stabilizes and maintains a consistent composition dominated by three primary phyla – Firmicutes (30% to 50%), Bacteroidetes (20 to 40%) and Actinobacteria (1 to 10%). Firmicutes and Bacteroidetes represent more than 90% of the bacterial phyla in the colon (Gagnière *et al.*, 2016), whereas Actinobacteria, Proteobacteria and Verrucomicrobia comprise up to around 2%, 1% and 0.1%, respectively.

The human intestinal microbiota harbours a diverse bacterial community required for our health, nutrition and wellbeing, containing 100 to 1000 different bacterial species. Compositional diversity between individuals makes the gut microbiota unique as a fingerprint for each one of us (Browne *et al.*, 2016). A typical intestinal microbiota is dominated by strict anaerobes belonging to the genera *Bacteroides, Eubacterium, Bifidobacterium, Fusobacterium, Peptostreptococcus* and *Atopobium*, and presents a minority of facultative anaerobes such as Lactobacilli, Enterococci, Streptococci and Enterobacteriaceae (Gagnière *et al.*, 2016).

1.2.1 Gut microbiota imbalance and CRC

Microbes have been implicated in the pathogenesis of several cancers (approximately 20%), most strikingly in the case of *Helicobacter pylori* and gastric carcinoma, where *H. pylori* is considered carcinogenic and a preclinical risk factor for this type of cancer. Contrary to gastric carcinogenesis, which results from a single pathogen, tumour-promoting effects of the microbiota in CRC appear to be caused by altered host-microbiota interactions and/or by dysbiosis. Dysbiosis is a state of microbial composition that is characterized by differences in the relative abundance of bacterial subgroups in the gut microbiota compared with the regular proportion in a healthy state (Flemer, Lynch, *et al.*, 2017).

There are two main hypotheses that are in the lead to explain the contribution of microbiota in CRC. One of them states that a dysbiotic bacterial community presenting pro-carcinogenic characteristics is able to remodel the microbiota as a whole in order to initiate pro-inflammatory responses and epithelial cell transformation, leading ultimately to CRC. The other hypothesis is more recent and is called "driver passenger" theory. This theory divides the intestinal bacteria into two groups – "bacterial drivers", which initiate CRC by inducing epithelial DNA damage and tumorigenesis, and that in turn promote the proliferation of another group of bacteria termed "passenger bacteria" which have a growth advantage in the tumoral microenvironment (Gagnière *et al.*, 2016).

16S rRNA sequencing of bacteria from the digestive tissues and from faeces allowed to establish a positive correlation between colonic dysbiosis and CRC patients, in numerous clinical studies (Gagnière *et al.*, 2016). However, no unifying CRC-associated dysbiotic signature in the microbiota structure across these multiple studies has been determined (Flemer, Lynch, *et al.*, 2017). On the other hand, there are several well-studied bacterial species that have been identified in the human colon cancer tissue and are suspected to play a role in CRC. These include *H. pylori* (a carcinogen that promotes directly gastric cancer)*, Fusobacterium* spp. (includes *Fusobacterium nucleatum* – the most consistent anaerobic gram-negative pathogenic bacterium associated with CRC), *Escherichia coli* (genotoxic *E. coli* belonging to phylogroups B2 and tightly adherent *E. coli*), *Streptococcus gallolyticus* (the first bacterium associated with CRC), *Bacteroides fragilis* (anaerobic bacterium that represents less than 1% of the gut microbiota) and *Enterococcus faecalis* (facultative anaerobic commensal bacterium of the oral cavity and gastrointestinal tract) (Sears and Garrett, 2014). The suspected mechanisms by which these single microbes modulate CRC (summarized in Table 1.1) will be discussed later.

Table 1.1 – Suspected factors and mechanisms used by single bacteria to promote CRC. This table was designed using as references: (Schwabe and Jobin, 2013), (Gagnière *et al.*, 2016) and (Sears and Garrett, 2014).

Individual bacteria	Factor/Mechanism	Consequences
Helicobacter pylori	CagA and VacA virulence factors; direct and indirect production of ROS and NOS	Activation of inflammation pathways and cellular proliferation
Fusobacterium nucleatum	FadA adhesin	Invasion, stimulation of cell proliferation and tumour growth
Escherichia coli	Afa and eae adhesins; CDT and colibactin toxins	Invasion of the intestinal epithelium; DNA damage, inflammation and cell proliferation
Streptococcus gallolyticus	Pilus protein with a collagen-binding domain	Translocation, inflammation and pro-carcinogenic pathways
Bacteroides fragilis	*B. fragilis* toxin (BFT) also known as fragilysin	Colonic carcinoma cell proliferation, inflammation and induction of DNA damage
Enterococcus faecalis	Extracellular superoxide anions, ROS	Induction of mucosal macrophages to produce chromosomal-breaking factors that mediate DNA damage

1.2.2 Host-microbiota interactions and CRC development

Microbiota and host have co-evolved to form a complex super-organism. Both parts benefit from each other through a symbiotic relationship able to maintain a homeostatic state. The gut microbiota plays a large role in several immune, structural, protective and metabolic functions in the intestine (Gagnière *et al.*, 2016). However, abnormalities in the host regulatory pathways involved in homeostasis and bacterial sensing, or imbalances in the microbiota caused by environmental changes (diet, lifestyle or infection), can disturb this symbiotic relationship and promote a disease state (Schwabe and Jobin, 2013).

Pathogens activate robust innate and adaptive immune responses when entering the host ecosystem, while microbial communities that benefit or do not harm the host are allowed to establish a permanent niche and reside in a state of immune tolerance with the host. This state of homeostasis is maintained by an anatomical separation with multi-level barriers between microbes and host compartments, by a eubiotic microbiota (a healthy state of microbial composition in which population abundances are found in normal proportions) and by a state of low inflammation. The barriers in the gut are constituted by an intact epithelial lining, by sensing systems able to detect and combat invading bacteria, and a mucus layer. In addition, Paneth and goblet cells monitor the number and location of bacteria, and secrete antibacterial peptides in order to regulate them. Apart from host mechanisms, the microbiota by itself is a functional luminal barrier by maintaining epithelial cell turnover, by producing mucin and by competing for resources which suppresses the growth of undesirable pathobionts (Schwabe and Jobin, 2013).

Barrier defects (physically at the level of tight junctions or mucus layer, or at the level of antibacterial defence systems), immune failure and subsequent inflammation, and the loss of eubiosis (qualitative and sometimes quantitative changes in the microbiota) are associated with microbially driven carcinogenesis (Figure 1.4) (Schwabe and Jobin, 2013).

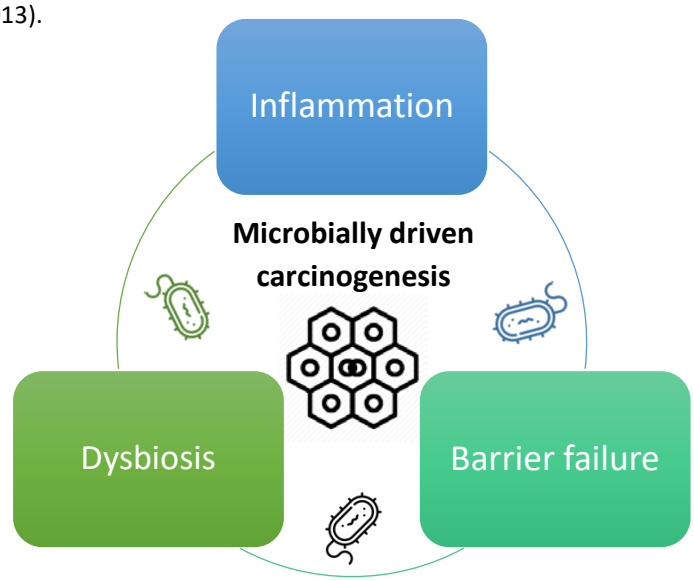


Figure 1.4 – Microbially driven carcinogenesis. Perturbation of the balanced mechanisms that control host-microbiota interactions (eubiotic microbiota, multi-level barrier and a state of low inflammation) lead to a cancer-promoting state with dysbiosis, barrier failure and inflammation.

Which factors are responsible for the failure of these control mechanisms? In the case of barrier failure, this can result from primary defects in genes that encode essential proteins involved in the maintenance of a functional barrier, or from secondary defects owing to infection and inflammation (Schwabe and Jobin, 2013).

Multiple pattern recognition receptors (PRRs), proteins expressed mainly by host cells of the innate immune system, sense bacteria and monitor microbial status and barrier integrity, initiating regulatory responses. They control the microbiota through antibacterial mediators and thereby suppress cancer. By contrast, the bacterial microbiota rarely triggers the degree of innate immune activation required for anti-tumour immune response. In this case, PRRs may promote resistance to cell death (one of the hallmarks of cancers) and may induce cancer-promoting low-grade chronic inflammation (Schwabe and Jobin, 2013). Innate immunity receptors like Toll-Like Receptors (TLRs) and Nod-Like Receptors (NLRs) recognize pathogen patterns/molecular motifs and their activation prompts signalling pathways. These pathways induce the expression of antimicrobial peptides and pro-inflammatory cytokines that promote cell proliferation and survival, all of which are involved in the development of an inflammatory response (Gagnière *et al.*, 2016).

Dysbiosis may occur through changes in diet, innate immune responses and inflammation, or infections - all perturbations able to affect microbial composition, richness and the metagenome (Schwabe and Jobin, 2013).

1.2.3 Microbiota mechanisms involved in carcinogenesis

Several mechanisms by which the bacterial microbiota modulates carcinogenesis include modulation of host defences and inflammation, bacterial-derived genotoxins and virulence factors, oxidative stress induction and microbial-derived metabolism (Figure 1.5).

1.2.3.1 Bacterial-derived genotoxins and virulence factors

The ability of some bacteria to induce chronic inflammation contributes to their tumorigenic potential. Additionally, bacteria also have the capacity to directly modulate carcinogenesis through specific toxins that induce DNA damages responses or virulence factors that confer pathogenicity (Schwabe and Jobin, 2013). This pathogenicity can be expressed by the ability that certain bacteria have to penetrate the gut mucosal barrier, and to adhere and invade intestinal epithelial cells through flagellum, pili or adhesins (Gagnière *et al.*, 2016).

Examples of specific bacterial pathogens are *H. pylori* expressing the virulence factor cytotoxin-associated gene A (CagA) or the vacuolating cytotoxin A (VacA) that may induce the activation of inflammation pathways and cellular proliferation; *F. nucleatum* using the virulence factor FadA; and some CRC-associated *E.coli* strains using *afa* and *eae* adhesins (Table 1.1).

Highlighting *F. nucleatum,* this is an adhesive bacterium that coaggregates with several microbial species, being among the most prevalent species in CRC tissues and metastatic lesions (Bullman *et al.*, 2017). A mechanism explaining colonization, dissemination and induction of host responses of this pathogen relies on FadA adhesin

that binds to cell-junction molecules, the cadherins on endothelial and epithelial cells, causing increased cell permeability and therefore allowing bacteria to penetrate through (Han, 2015). The binding of FadA adhesin to host cadherin also stimulates potentially oncogenic beta-catenin signalling which activates inflammatory responses and oncogene expression (Han, 2014).

Another pathogenic species that colonizes colonic tumours is *Streptococcus gallolyticus* which possesses a pilus protein with a collagen-binding domain (Table 1.1). Its virulence mechanisms allow *S. gallolyticus* to translocate through the intercellular space between colonic epithelial cells, to exhibit competitive growth advantage under metabolic conditions linked to CRC and to induce inflammation and pro-carcinogenic pathways (Gagnière *et al.*, 2016).

As discussed previously, barrier failure can allow luminal bacteria to access the epithelium and therefore enables direct contact with host cells and the delivery of toxins. Cyclomodulins such as cytotoxic necrotizing factor 1, cytolethal distending toxin (CDT) and colibactin produced by *E.coli*; and *Bacteroides fragilis* toxin are examples of bacterial toxins (Table 1.1). Colibactin and CDT exert direct double-strand DNA damage responses and genomic instability (crucial cellular responses implicated in tumorigenesis), being the only toxins considered truly genotoxic (Gagnière *et al.*, 2016).

Enterotoxigenic *B. fragilis* expresses *B. fragilis* toxin (BFT), also known as fragilysin. This zinc-dependent metalloprotease binds to epithelial receptors, altering colonic epithelial cells structure and function by cleavage of E-cadherin (a tumour suppressor protein) which results in the activation of nuclear Wnt/β-catenin and nuclear factor-kappa B (NF-κB) signalling pathways. While the former leads to increased colonic carcinoma cell proliferation and expression of pro-oncogenes; the latter causes the release of pro-inflammatory mediators such as cytokines by colonic epithelial cells (which contribute to the initiation of mucosal inflammation) and induces DNA damage (Sears and Garrett, 2014).

Most Gram-negative bacteria involved in CRC as *E. coli* produce CDT (Table 1.1). Upon infection, the CdtA and CdtC subunits form an anchor between the bacterium and the host cell; subsequently, the subunit CdtB is delivered into the cytoplasm and travels to the nucleus where it acts as a DNase and damages DNA (Nešić *et al.*, 2004).

Colibactin is encoded in the 54 kb polyketide synthase (*pks*) genotoxicity island (Table 1.1). The *pks* island is found primarily in the Enterobacteriaceae family, with phylogenetic group B2 *E. coli* representing the main carrier (Schwabe and Jobin, 2013). This group is involved in inflammation pathways and cell proliferation fuelling. *Klebsiella pneumoniae*, another bacterium associated with CRC, also carries the *pks* island. This toxin induces double-strand DNA breaks, ROS formation, associated DNA damage responses, cell cycle arrest and genomic instability (Gagnière *et al.*, 2016).

Although toxin-producing bacteria constitute a minority of the colonic microbiota, it is interesting to highlight their relevance as analysis performed on human CRC tissue samples at the metatranscriptome level revealed high expression of these toxins in the colon (Dutilh *et al.*, 2013).

1.2.3.2 Oxidative stress

Oxidative stress is characterized by an imbalance between the levels of pro-oxidative molecules, such as reactive oxygen species (ROS) and nitrogen species (NOS), and the efficacy of anti-oxidative host defences that help to regulate their production and repair defective mechanisms. Oxidative stress results in cell damage, including protein aggregation or fragmentation, DNA breaks, and cellular membrane dysfunction. Important to highlight that, in cases of bacterial infection this balance between pro- and anti-oxidative compounds is lost and the DNA repair mechanisms are altered in CRC (Gagnière *et al.*, 2016).

ROS can be generated either by cells during infection and inflammation as a defence mechanism to eliminate bacteria or directly by gut bacteria. Examples of single microbes that produce ROS are *Enterococcus faecalis* and *H. pylori* (Table 1.1). *E. faecalis* strains producing extracellular superoxide anions (hydroxyl radicals that are powerful mutagens) trigger carcinogenesis through induction of mucosal macrophages to produce diffusible clastogens (chromosomal-breaking factors) that mediate DNA damage (Sears and Garrett, 2014). Furthermore, using as an example what happens in gastric cancer scenario, *H. pylori* can both produce and induce the production of ROS by immune cells, affecting many signal transduction pathways in gastric cells and thereby promoting cancer (Handa *et al.*, 2010).

The activation of macrophages in the inflammation response (triggered by the gut microbiota) leads to host production of nitric oxide and its secondary NOS. Besides, Lactobacilli and Bifidobacteria groups are able to directly generate high levels of NOS and therefore induce genomic instability in CRC (Gagnière *et al.*, 2016).

1.2.3.3 Microbial-derived metabolism

Human metabolism is a combination of human and microbial enzyme activities. Note that the bacterial metagenome is enriched for genes that are relevant for nutrient, xenobiotic and bile acid metabolism, and for the biosynthesis of vitamins. Therefore, microbial-derived metabolism impacts CRC carcinogenesis *via* several processes such as metabolic activation or inactivation of xenobiotics; hormone metabolism; modification of inflammation pathways; and generation of CRC-promoting secondary bile acids (Schwabe and Jobin, 2013).

Compounds from chemical carcinogens to pharmaceuticals, generally referred to as xenobiotics, are metabolized in the liver and undergo further metabolism by the gut bacteria. This way, intestinal microbiota influences the activity and the side effects of drugs used for antitumour therapies. In the liver, the metabolism is conducted via oxidation and conjugation reactions, while in the colon the gut microbiota tends to favour reduction and hydrolytic processes. Some carcinogens are inactivated in the liver by glucuronic acid-mediated conjugation which makes them more water-soluble and facilitates their excretion in the urine or faeces (Sears and Garrett, 2014). These inactivated carcinogens are initially excreted *via* bile in the digestive tract. However, they can be reactivated in the colon by bacterial β-glucuronidase activity which hydrolyses these conjugates. The reactivation of these toxic components is involved in the initiation and progression of CRC, and can be prevented with bacterial β-glucuronidase inhibitors or antibiotics (Gagnière *et al.*, 2016).

High-fat diets increase bile acid production and subsequently the risk of CRC. There is significant interplay between host and gut microbiota in bile acid metabolism. Primary bile acids (conjugated) excreted into the gut are converted into secondary bile acids such as lithocholic and deoxycholic acids (the two dominant faecal bile acids) (Sears and Garrett, 2014). This process of deconjugation is performed by gut bacteria through hydrolase activities responsible for the removal of polar groups such as taurine and glycine. These deconjugated secondary bile acids are used by microorganisms as an energy source but are also known to be involved in many CRC carcinogenesis-linked processes, such as cell proliferation, apoptosis, DNA damage and tumour promotion. In addition, lithocholic and deoxycholic acids are pro-inflammatory and can elicit ROS and NOS production. Besides, a strong antimicrobial bile acid activity leads to significant changes in the gut microbiota composition with an increase in some Bacteroidetes and Gammaproteobacteria, previously associated with CRC (Gagnière *et al.*, 2016).

Microbial carbohydrate fermentation benefits the host through the generation of short chain fatty acids (SCFA), such as butyrate, whereas microbial protein fermentation generates potentially toxic and pro-carcinogenic metabolites involved in CRC, such as ammonia, amines, phenols, nitrosamines and sulphides (Schwabe and Jobin, 2013). In the colon, species belonging to Bacteroidetes and Firmicutes phyla produce nitrogenous compounds *via* degradation of amino acids. These compounds present potential carcinogenic effects through DNA alkylation which leads to mutations reported in Western diet-linked CRC (Gagnière *et al.*, 2016). Furthermore, gut bacteria, mostly Fusobacteria, can reduce dietary sulphate to hydrogen sulphides – a genotoxic gas capable to induce ROS formation and DNA damage (Schwabe and Jobin, 2013). Summarizing, high-protein and low-carbohydrate diets can lead to increased levels of toxic metabolites and to decreased levels of cancer-protective metabolites which increases the risk of CRC.

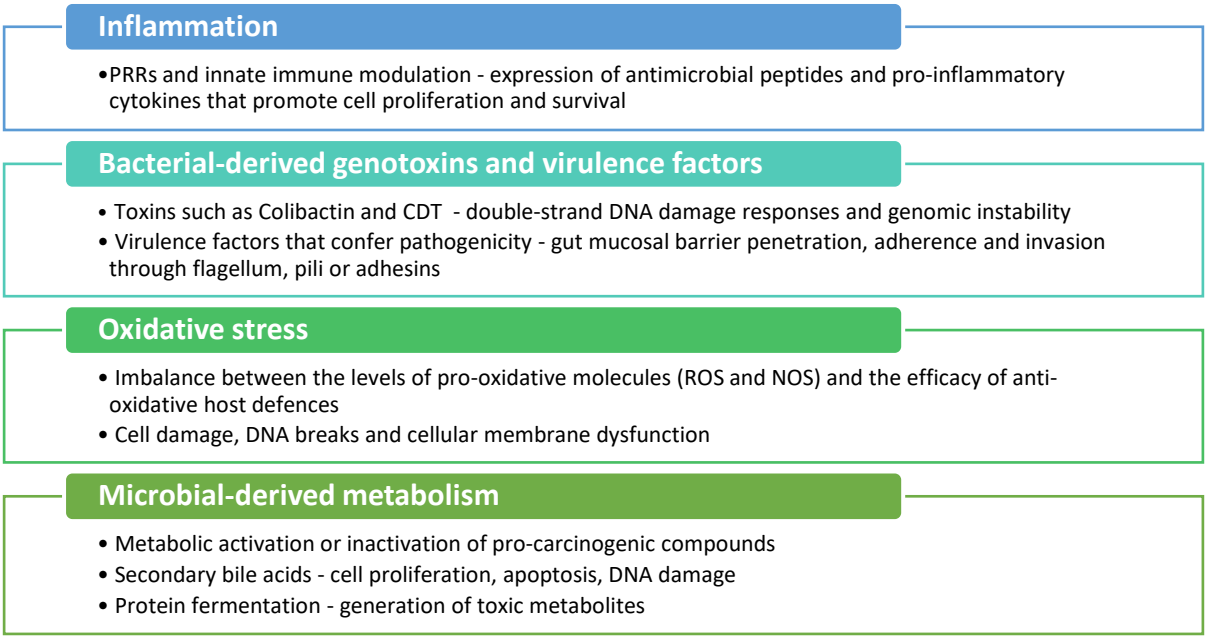


Figure 1.5 – Mechanisms by which the microbiota directly modulates CRC.

1.3 Oral microbiota and Colorectal Cancer

Compared with the gut microbiota, the oral microbiota has been relatively overlooked. However, it should be highlighted that the latter comprises more than 2 000 bacterial taxa, including an interesting large number of opportunistic pathogens associated with systemic infections and inflammation, including periodontal, cardiovascular and respiratory diseases, adverse pregnancy outcomes, rheumatoid arthritis, inflammatory bowel disease and CRC (Han and Wang, 2013). It is also considered the second most diverse community following gut microbiota in terms of species-richness (Wade, 2013). 687 oral species belonging to 185 genera and 12 phyla (Firmicutes, Fusobacteria, Proteobacteria, Actinobacteria, Bacteroidetes, Chlamydiae, Chloroflexi, Spirochaetes, SR1, Synergistetes, Saccharibacteria and Gracilibacteria) have been identified. For each species, a reference 16S rRNA sequence is available in the Human Oral Microbiome Database (Chen *et al.*, 2010).

1.3.1 Oral bacteria in disease

The homeostatic balance of oral bacterial flora can be lost, under certain circumstances. This results in a dysbiosis state characterised by increased abundance of pathogenic bacteria and expression of virulence properties which leads to dental caries and periodontal infections. A consortium of microbes rather than one species is usually involved in causing disease (Darveau, 2010). In caries, the ecological imbalance favours the growth of acidogenic lactic acid species such as Bifidobacteria, mutans Streptococci and Lactobacilli. In the case of periodontal diseases, members of the so-called red complex, such as *Porphyromonas gingivalis*, *Treponema denticola* and *Tannerella forsythia*, as well as some new taxa such as oral Synergistetes and Saccharibacteria, are the main bacteria at periodontal destruction sites. These are proteolytic bacteria capable of destabilizing the host inflammatory response (Perera *et al.*, 2016).

Oral bacteria may contribute to oral carcinogenesis via several mechanisms that include activation of cell proliferation, induction of chronic inflammation, promotion of cellular invasion and production of carcinogens (Perera *et al.*, 2016). Recent studies suggest that oral pathogens, such as *Fusobacterium nucleatum* and *Porphyromonas gingivalis,* which present a causative link with periodontal disease, may play an important role in the pathogenesis of not only oral cancer, but also cancer in the digestive system (Kato *et al.*, 2016). Therefore, it is hypothesized that the oral cavity may serve as a reservoir for potential intestinal pathobionts.

For instance, Nakajima *et al.* showed that single oral administration of *Porphyromonas gingivalis,* periodontopathic bacterium, induced significant alteration of phylogeny and diversity of gut microbiota which in turn allowed the dissemination of endotoxic Enterobacteriaceae to the liver and decreased mRNA expression of tight junction proteins (which enhances barrier failure) (Nakajima *et al.*, 2015). This is an important paper because it proves that oral influx of *P. gingivalis* can cause systemic inflammation through the promotion of changes in the gut microbiota.

Several mechanisms have been proposed in order to explain the influence of oral bacteria in extra-oral infections and inflammation, including transient bacteremia, systemic inflammation and systemic injury induced by soluble antigens and by free toxins (Han and Wang, 2013). Given that the oral cavity is at the beginning of the digestive tract, it is not surprising that microbial species find their way down the path. Indeed, note that in average a person generates and ingests around 1.5 litres of saliva per day which contain an enormous number of diversified oral-resident bacteria (Edgar *et al.*, 2004).

1.3.2 Oral microbiota is distinctive and predictive in CRC

A study performed by Kato *et al.* was the first to identify significant differences in the bacteria present in oral samples from CRC patients when compared with healthy controls. The study highlighted that CRC was associated with increased presence of genus *Lactobacillus* and *Rothia* (Kato *et al.*, 2016).

In 2017, a study from Paul O'Toole's group revealed that oral microbiota is distinctive and predictive in CRC (Flemer, Warren, *et al.*, 2017). The authors profiled the microbiota in oral swabs, colonic mucosae and stool from individuals with CRC (99), colorectal polyps (32) and controls (103). They identified *Streptococcus*, *Haemophilus*, *Neisseria*, *Prevotella*, *Fusobacterium*, *Veillonella*, *Leptotrichia*, *Rothia*, *Actinomyces* and *Porphyromonas* as the ten most abundant bacterial genera across all oral swab samples, in descending order of abundance, respectively. The overall oral profile of bacterial OTUs was significantly different between CRC patients and healthy controls. For instance, eight oral microbiota OTUs were differentially abundant in these two conditions (*Haemophilus*, *Parvimonas*, *Prevotella*, *Alloprevotella*, *Lachnoanaerobaculum*, *Neisseria* and *Streptococcus).* It's interesting to see that seven of these eight oral OTUs were less abundant in the oral cavity of individuals with CRC than in healthy controls. Comparing the oral microbiota of individuals with polyps and healthy controls, the authors identified four bacterial OTUs differentially abundant between the two groups (Flemer, Warren, *et al.*, 2017). These results are indicative that oral microbiota is significantly different in CRC (Figure 1.6.A).

Additionally, by combining the oral microbiota and stool microbiota datasets, they created a powerful tool capable of distinguishing those with cancer and those with colonic polyps from the controls, with a relatively high specificity and sensitivity (Flemer, Warren, *et al.*, 2017). Indeed, they identified 16 oral microbiota operational taxonomic units (OTUs) capable of distinguishing individuals with CRC from healthy controls and 12 oral OTUs that allow the detection of colorectal polyps. The oral swab microbiota classification model presented lonely a sensitivity of 53% and a specificity of 96%. The previous model developed by the same group based only in faecal microbiota presented a much lower sensitivity in detecting individuals with CRC (22%). The new combined oral and stool microbiota model increased the sensitivity up to 76%. Due to the prognostic and therapeutic importance of early discovery of colonic disease, it looks promising the high sensitivity for the detection of adenomas (88%) achieved by this CRC classifier (Flemer, Warren, *et al.*, 2017). Therefore, these results suggest that analysis of the oral microbiota could potentially be used as a screening method for CRC and polyp detection (Figure 1.6.B). Note that mouth is a good spot for convenient and widespread screening, since

sampling oral microbiota through mouthwash, oral swab or saliva is more amendable for epidemiologic research than collecting faecal samples and less invasive than obtaining mucosal biopsy through endoscopy.

The authors were also able to prove that putatively oral bacteria are in over-abundance on CRC biopsies and polyps (Figure 1.6.C), which is in accordance with other previous studies ((Flemer, Lynch, *et al.*, 2017), (Nakatsu *et al.*, 2015) and (Warren *et al.*, 2013)). Two tumour-associated bacterial coabundance groups (CAGs) were identified. Bacteria of these CAGs were meaningfully more abundant on tumour samples than in healthy tissue. The first CAG comprised oral pathogens like *F. nucleatum*, *Parvimonas micra*, *Peptostreptococcus stomatis*, *Dialister pneumosintes* – bacteria that are linked with late colonization of oral biofilms and with CRC. The second CAG comprised dominant bacteria in early dental biofilm formation and associated with healthy tooth pockets, including *Actinomyces*, *Haemophilus*, *Rothia*, *Streptococcus* and *Veilonella* spp.. The authors also identified similar bacterial networks in colonic microbiota (CRC biopsies, undiseased tissue samples from CRC patients and colonic polyps) and oral microbiota, including putative oral biofilm forming bacteria (Figure 1.6.D). These results indicate that these networks are already established before CRC starts to develop and can play a role in the initiation of CRC (Flemer, Warren, *et al.*, 2017).

Another interesting result presented in the same paper of Paul O'Toole's group links high abundance of Lachnospiraceae, such as *Anaerostipes*, *Blautia* and *Roseburia*, with low colonization of colonic tissue by oral bacterial networks (Figure 1.6.E). This suggests a protective role played by this family against CRC, hypothetically by conferring colonization resistance to CRC-associated oral pathogens (Flemer, Warren, *et al.*, 2017).

It is also important to refer that, contrary to expectations, bacteria usually present in more abundance on CRC tumours and found in the mouth and colon simultaneously, like *Fusobacterium*, *Parvimonas* and *Porphyromonas,* were less abundant in the oral cavity of CRC patients compared with healthy individuals. Besides, the number of oral pathogens comprised in the first CAG was also lower in the oral mucosa of CRC patients. In these cases, oral microbiota doesn't reflect gut microbiota in terms of quantitative distribution (Flemer, Warren, *et al.*, 2017).

Overall, the authors were able to establish a positive association between colonic colonization by a putative pathogenic oral community and CRC. This community forms bacterial coabundance networks like those found in the oral cavity, being tightly associated with the colon mucosa probably by forming oral-like biofilms. All in all, the results suggest that the inclusion of oral microbiota information can enhance the performance of current diagnostic tests.

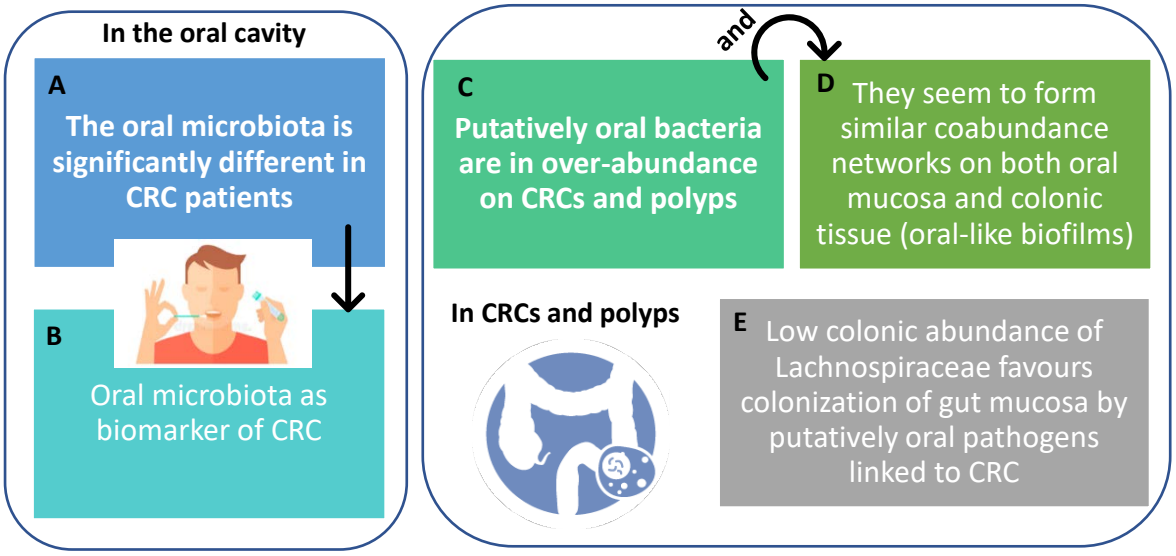


Figure 1.6 – Resume of the important conclusions achieved by (Flemer, Warren, *et al.,* 2017) about oral bacteria and CRC.

1.3.3 Oral bacteria and colonization of CRC

The clear message to be taken from the previous sections is that oral microbiota alterations are linked with CRC and that increased levels of putative oral bacteria have been reported on colonic tumour biopsies in previous metagenomic studies (Figure 1.6.A and 1.6.C). However, a causal role of oral bacteria ectopically colonizing the intestine remains unclear. To achieve this goal, it's mandatory to culture and isolate bacteria from oral swabs/saliva and to do the same with CRC tumour samples from the same CRC patient. Only the comparison of the genomic profile of these pure cultures from the two types of samples can truly demonstrate if colonic mucosa-associated taxa are indeed orally derived.

Within the scope of gut colonization by oral bacteria, two articles seem to show substantiated evidence of this premise until now, with only one of them acquiring the culture-dependent methodology presented in the previous paragraph in CRC patients (Komiya *et al.*, 2018), while the other uses transplantation of saliva from patients with Crohn's disease in germ-free mice (Atarashi *et al.*, 2017).

Atarashi *et al.* highlighted that strains of *Klebsiella* spp. isolated from the salivary microbiota are able to colonize the gut of germ-free mice and potently induce chronic intestinal inflammation (Atarashi *et al.*, 2017). The isolation and the sequencing techniques used in this study, make it one of the few reliable papers that are able to prove the colonization of oral bacteria in the gut (Cao, 2017). The authors administered salivary samples from two patients with Crohn's disease (chronic inflammatory condition of the gastrointestinal tract) to germ-free mice. After six weeks, they compared the community composition of the oral microbiota before transplantation and the faecal microbiota of the colonized mice, by using 16S rRNA gene sequencing. Interestingly, the majority of bacteria species present in the mice stool samples were minor components of the salivary microbiota (only

approximately 20 kinds of putative oral bacteria overlapped in both samples), which indicates that a small fraction of oral bacteria manages to expand and colonize the gut, and a subset of these can induce the accumulation of T helper 1 immune cells (T_H1). These cells express interferon-gamma protein, which, in one hand, activates microphage to enhance bactericidal activity, but, in other, can represent an onset of autoimmune diseases when excessive (Atarashi *et al.*, 2017).

By performing anaerobic culture of the caecal contents of 3 mice (from group CD#2 – Figure 1.7), using several culture media, the authors picked 224 colonies with different colony appearances. Sequencing of the 16S rRNA genes revealed that these colonies contained eight different strains from diverse genera - including *Streptococcus, Gemella, Escherichia, Veillonella, Fusobacterium, Anaerococcus, Bifidobacterium and Klebsiella* (Figure 1.7). These strains were subsequently examined for T_H1 cell-inducing capability. *Klebsiella pneumoniae* alone was the only one able to strongly promote T_H1 cell induction (greater in the colon than in the small intestine), which further elicits inflammation. Remarkably, this bacterium is usually a normal resident in the mouth of healthy individuals, however, it can colonize the gut when antibiotics like ampicillin promote a dysbiotic microbial state of the former. Besides being proinflammatory, *K. pneumoniae* is also multi-drug resistant. Further studies in this area can help the development of narrow-spectrum antibiotics or lytic bacteriophages specifically targeting *K. pneumoniae.* Another interesting approach is to try to identify members of the normal gut microbiota that can provide colonization resistance against this kind of orally derived bacteria (Atarashi *et al.*, 2017).

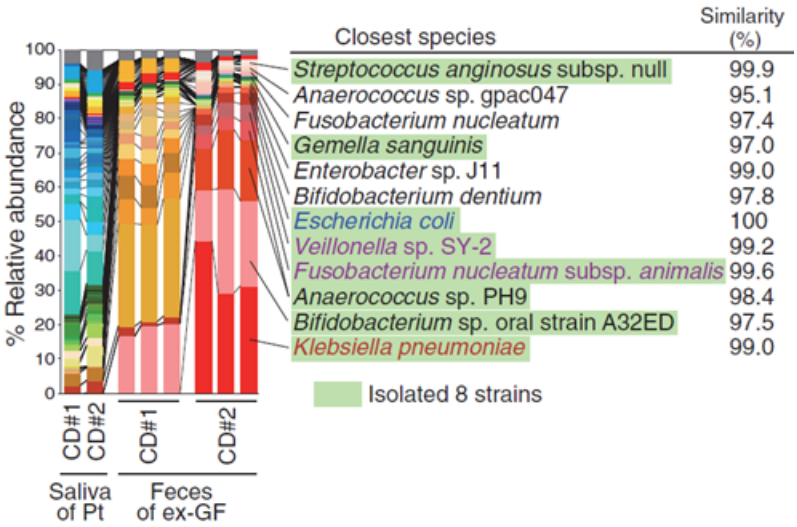


Figure 1.7 – Previous isolated oral species in the gut. Pyrosequencing results of the 16S rRNA genes from the saliva microbiota of 2 patients (Pt) with Crohn's disease (CD#1 and CD#2) and from the resulting faecal microbiota of the ex-germ-free mice (ex-GF) (3 mice per group) transplanted with the patients' saliva. In green is highlighted the different strains isolated from the group of mice transplanted with saliva samples from patient 2 (CD#2). The sequence similarity with their closest species in database was also presented by the authors (96% was the limit established). Adapted from (Atarashi *et al.*, 2017).

The other study was able to prove that patients with CRC present identical strains of one species (*Fusobacterium nucleatum)* in their colorectal cancer and oral cavity, by isolating this bacterium on *Fusobacterium*-selective agar.

They chose to target this single species since *F. nucleatum* is tightly associated with CRC development and pathogenicity, still lacking evidence of its oral origin. The authors collected CRC and saliva samples from 14 patients with CRC. These specimens were cultured within 24 h after collection in anaerobic conditions for 48-72 h at 37 ^{o}C. Approximately 50 colonies were randomly picked, purified and the resulting isolates were lysated and then analysed by polymerase chain reaction (PCR) using *Fusobacterium* spp.- and *F. nucleatum*-specific primers. This step allowed them to detect *F. nucleatum* in 8 of 14 patients from CRC samples and in all patients from saliva samples. Four *F. nucleatum* subspecies were identified by 16S rRNA gene sequencing - *animalis*, *nucleatum*, *polymorphum* and *vincentii* (Komiya *et al.*, 2018).

The next step was to identify equal *F. nucleatum* isolates between CRC and saliva samples at the strain level. For this purpose, the authors performed random amplified polymorphic DNA (RAPD) or arbitrarily primed PCR (AP-PCR) DNA fingerprinting, targeting the previously referred 8 patients (Komiya *et al.*, 2018). This technique uses as primers single oligonucleotides of arbitrarily chosen sequence with no known homology to the target genome and allows to obtain arrays of DNA fragments that are strain-specific and reproducible. Thus, AP-PCR (RAPD) is a sensitive and efficient method to distinguish strains of the same species (Berg *et al.*, 1994) – equal patterns on the gel identify equal strains. Komiya *et al.* detected identical *F. nucleatum* strains in both CRC and saliva in 6 of the 14 patients (42%), which suggests that *F. nucleatum* in CRC originates in the oral cavity.

1.3.4 Translation of oral diseases to CRC

Fusobacterium nucleatum is frequently identified by 16S rRNA sequencing in CRC tumours, often as a consortium with other putative oral microbes such as *Porphyromonas* spp.. Indeed, *F. nucleatum* and *Porphyromonas gingivalis* are two periodontal pathogens well known to promote oral cancer progression in a synergistic way (Gallimidi *et al.*, 2015). Both present several virulence mechanisms that include disruption of oncogene signalling and of cell-cell adhesion; promotion of inflammation, tumour proliferation and progression; inhibition of natural killer and cytotoxic T cells; and more recently discovered, ability to invade the gut submucosa and epithelium (Whitmore and Lamont, 2014). As previously mentioned, putative oral commensal and periodontopathic bacteria have been found enriched in the cancerous colon. Two opportunistic factors associated with the development of periodontal pathogenesis are the formation of biofilms by oral bacteria and their anaerobic asaccharolytic metabolism (Flynn *et al.*, 2016). These two factors have not yet been appreciated with respect to oral bacteria in CRC. In this way, the already known mechanisms of oral bacteria in periodontal disease, oral cancer and caries can help us to hypothesize and disclose the mechanisms by which putative oral bacteria may modulate colon tumorigenesis.

1.3.4.1 Periodontal biofilms and oral polymicrobial interactions

Periodontal biofilms form in three stages. First, early colonizers such as *Streptococcus* spp. and *Actinomyces* spp. bind to teeth and play the role of bridging organism where other early and intermediate colonizers, like *F. nucleatum*, can bind. In turn, these intermediate species express multiple adhesins, proteolytic enzymes, and

proinflammatory surface ligands that invade oral epithelial cells, enhancing the virulence of later-arriving species like *P. gingivalis* at which they aggregate*. P. gingivalis* is considered a "driver" pathogenic bacteria since is able to produce virulence factors that continue to modify the biofilm community structure, bacterial load and gene expression of the "passenger" species, ending up subverting the host immune response (Hajishengallis, 2015). A periodontal biofilm is formed. At this point, the periodontopathic community is able to support itself nutritionally and to increase proinflammatory tissue destruction (Flynn *et al.*, 2016).

Periodontal species are "specialists" in coaggregation using adhesins in the anaerobic subgingival space and, therefore, they seem well equipped for a life in the colon since this also presents an anaerobic environment. The presence of biofilms containing putative oral microbes in both healthy colonic mucosa and CRC tumours was first proved by Dejea *et al*. (2014). In the same study, by performing fluorescent in situ hybridization (FISH) and 16S rRNA gene sequencing, a higher abundance of invasive *F. nucleatum* was detected in CRC and adenoma polymicrobial biofilm samples, but not in healthy colon biofilms (Dejea *et al.*, 2014).

F. nucleatum is also considered a "driver" bacterium in CRC. Reminding the previously presented "driver-passenger" model, bacterial "drivers" of CRC are intestinal bacteria with pro-carcinogenic features that may initiate CRC development, while CRC bacterial "passengers" are relatively poor colonizers of a healthy intestinal tract, but have a competitive advantage in the tumour microenvironment, being able to outcompete with the driver ones (Tjalsma *et al.*, 2012).

A co-occurrence network of putative oral anaerobes such as *Leptotrichia, Campylobacter and F. nucleatum* in CRC and normal tissues was determined by Warren and co-workers (Warren *et al.*, 2013). In the case of CRC tissue, this network is additionally associated with *Porphyromonas.* The authors cultured and isolated *Campylobater* species from the previous samples and showed its ability to coaggregate with *F. nucleatum* which reinforces the idea that this bacterium facilitates colon tissue infection by acting as a vector for oral microbes with compatible adhesins (Warren *et al.*, 2013). This study highlights the carcinogenic potential of oral polymicrobial interactions.

There is evidence that *F. nucleatum* cannot stably colonize the gut on its own, as mouse studies using *F. nucleatum* alone require the administration of other bacteria daily for 8 weeks to establish tumours (Kostic *et al.*, 2013). Therefore, future approaches to assess oral microbe activity in the colon should use metatranscriptomic analysis on bacteria isolated from human CRC samples, probing for biofilm matrix proteins, quorum-sensing molecules, adhesins and other virulence proteins that may promote a stabilized oral community in the colon. Afterwards, a mouse model of infection can be used to directly test antibiofilm therapeutics or probiotic interventions (Flynn *et al.*, 2016).

1.3.4.2 Proteolytic and asaccharolytic metabolism

Periodontitis is characterized by a microbial community change from protective facultative anaerobes to strict anaerobic bacteria which are usually proteolytic and asaccharolytic species. This means that they do not metabolize carbohydrates, relying on other carbon and nitrogen sources for energy like peptides and amino

acids. For instance, subgingival bacteria metabolize nitrogenous compounds from gingival crevicular fluid (GCF). This mechanism creates a neutral pH in the subgingival space and enables the release of SCFA and ammonia. This environmental imbalance stimulates GCF efflux and keeps supporting the proteolytic activity of several species such as *Prevotella intermedia*. In turn, these species enhance this vicious cycle by continuing to promote neutral pH to the oral environment and, in this way, supporting the colonization of pathobionts like *P. gingivalis* (Flynn *et al.*, 2016).

This propitious environment allows the establishment of a much robust microbial consortium presenting bacteria from the genera *Eubacterium*, *Campylobacter*, *P. intermedia*, *F. nucleatum* and *P. gingivalis.* These bacteria use membrane-bound or secreted proteases to keep feeding of proteins in the GCF and epithelial cells. In turn, the metabolism of this cooperative microbial community promotes a nutrient-rich disease state that supports the growth of the community (Takahashi, 2005). Interestingly *P. intermedia* and *F. nucleatum* preferentially degrade amino acids, while *P. gingivalis* prefers dipeptides to smaller molecules. Indeed, the species from this microbial consortium use the same nutrient source, but their metabolic pathways differ, which can explain the reduced competition in these oral biofilms. Additionally, proteolytic species like *P. gingivalis* are able to provide amino acids to *F. nucleatum* which enhances the vicious circle presented previously, with both of them creating a neutral pH environment (Flynn *et al.*, 2016). Moreover, the continued breakdown of host proteins stimulates an immune response that then provides more nutrients for the maintenance and enhancement of the biofilm. (Hajishengallis, 2015).

The previous metabolic synergy is crucial for the development of periodontitis. Indeed, despite high interpatient variability in oral microbial composition and despite high variability in metabolic gene expression presented by individual species, transcriptomic studies of periodontal tissues showed that periodontal disease-associated communities display a highly conserved metabolic profile and virulence gene expression (Jorth *et al.*, 2014). In this case, the metabolic properties of the pathogenic community are more significant than the metabolic and virulence properties of only one species.

Along the same line, a similar synergistic effect may happen when putative oral microbes colonize the colon. Note that colon mucosa is anaerobic and more neutral in pH than the mouth or stomach. In another hand, gut mucosa and epithelium are always being sloughed off, releasing nutrients and providing binding sites for adhesive bacteria. After mucus layer being penetrated, *F. nucleatum* can cause epithelial junction disruption. This mucosal environment shift is able now to support the growth of other asaccharolytic anaerobes such as *Porphyromonas* and *Peptostreptococcus*. Further, an inflammatory response can be triggered by the occurrence of degradation of host proteins either in mouth and gut which in turn provides more peptides to be metabolized by gut and putative oral bacteria. It was showed that this chronic inflammatory state can promote the development of CRC (Louis *et al.*, 2014). In summary, it was presented here that oral bacteria can colonize the gut and be associated with CRC by using the same metabolic and virulence strategies that promote their survival and perseverance in the mouth.

1.4 Culture of the human microbiota

Culture was the first tool for the exploration of the digestive bacterial ecosystem (Moore and Holdeman, 1974). It was only when anaerobic culture techniques started to be used, that strict anaerobes and extremely oxygen-sensitive were identified. However, the comparison between the microscopic cell counts and the actual number of colony forming units from the same sample emphasised the large number of microorganisms resistant to culture (Sommer, 2015). This gap between the bacteria actually present in a given sample and those effectively cultivated is the so-called "great plate count anomaly" (Staley, 1985). The limited ability of researchers to cultivate microorganisms may be due to complex cross feeding relationships between the microbes, differential nutrient requirements or very slow growth rates (Sommer, 2015).

An entire new era of microbiology began with metagenomics (the study of metagenomes by using high-throughput sequencing directly from a complex ecosystem) strongminded on the enumeration and characterisation of the previously unknown microbial diversity. Indeed, the first metagenomic studies highlighted that 80% of the gut bacteria identified by pyrosequencing targeting the 16S rRNA gene had not yet been cultured (Eckburg, 2005). In this way, it was commonly accepted that this percentage of gut bacteria was unculturable (Turnbaugh *et al.*, 2007).

Culture-independent, genomic approaches revolutionized our understanding of the gut microbiota, by providing insights into its community composition, how it changes with environmental perturbation and its potential role in a myriad of diseases (Lau *et al.*, 2016). Despite the huge taxonomic diversity (100 to 1 000 different species), metagenomic sequencing has highlighted that health-associated intestinal microbiota codes for highly conserved gene families and pathways associated with basic bacterial physiology and growth. Many of these functions and phenotypes, such as basic microbiota functions related to homeostasis, immune system development, digestion, pathogen resistance and microbiota transmission, remain to be discovered, validated and deciphered, since the majority of intestinal bacteria are largely recalcitrant to culture and isolation (Browne *et al.*, 2016). Culture can associate phenotypes to specific strains or elucidate physiological interactions, and can define the roles of gut bacteria in causing or maintaining disease and healthy states (Sansonetti, 2018).

Since the gut microbiota is dominated by strict anaerobic bacteria which are extremely sensitive to oxygen, it is not clear how these bacteria survive to its environmental exposure outside the body when they are transmitted between individuals. Certain Firmicutes, such as the diarrhoeal pathogen *Clostridium difficile*, produce metabolically dormant and highly resistant spores, which facilitate both persistence within the host and environmental transmission, for example between family members (Browne *et al.*, 2016). Cultivation allows the germination of these spores (Rettedal *et al.*, 2014).

In addition, culture determines the viable population in a community, while most molecular approaches are not able to distinguish between DNA obtained from dead or live cells. Moreover, by using selective media, culture techniques enhance growth and detection of less abundant bacteria which are usually missed by insufficient

sequencing depth in metagenomics culture-independent studies (Lagkouvardos *et al.*, 2017). Indeed, one of the major biases of metagenomic studies is the depth bias. For instance, in a complex ecosystem like the gut microbiota, where we have approximately 10^{12} bacteria per gram of stool, current metagenomics is unable to detect bacteria at concentrations lower than 10^{5} bacteria per gram (Lagier *et al.*, 2016). In another study, Hugon *et al.* highlighted that molecular studies, such as pyrosequencing of 16S rRNA gene, neglect nearly 15% of Gram-negative populations in the human gut microbiota (Hugon *et al.*, 2013). Therefore, clinically relevant minority populations are missed, including potentially pathogenic bacteria (Lagier *et al.*, 2012).

This way, both genomic and culture-dependent approaches have their limitations, which enhances the need for polyphasic approaches. Accordingly, the freshness of gut microbiota cultivation approaches these days is substantially due to the integration of the later phylogenetic profiling methods with new cultivation conditions (Sommer, 2015). Advances in culture techniques include encapsulation of bacteria into microdroplets or gel particles, microfabricated cultivation chips, diffusion chambers simulating the natural environment of the samples, design of more effective culture media and use of alternative solidifying agents to agar (Rettedal *et al.*, 2014). These techniques enabled cultivation of new species, however, they require access to complex microfluidic and microfabrication technology, not easily implemented in most biology laboratories. Thus, it remains the need for simple cultivation procedures.

Within the scope of culture of "unculturable" gut bacteria, Browne *et al* presented a surprising article where they demonstrated that a substantial proportion of the bacteria within the faecal microbiota is culturable by using a single growth medium. The authors combined metagenomic sequencing and a simple bacterial culturing approach to define the resident bacterial community of six healthy humans, from their fresh faecal samples. They isolated 137 bacterial species in YCFA agar solid medium (complex and broad-range bacteriological medium) which were archived as pure cultures. Next, they profiled and compared the species present in the original faecal samples to the ones that they have isolated in culture. Based on the average relative abundance across the six donors, these isolates represented 96% of the bacterial abundance at the genus level and 90% of the bacterial abundance at the species level (Browne *et al.*, 2016).

There is the need to use culture-dependent approaches in CRC research for several reasons: 1. They allow us to study drug metabolism and its collateral effects on the gut microbiota. 2. They can expand reference genome catalogues with rare and new species and possibly species whose genomes have only been assembled from metagenomics datasets. 3. They are able to explore causal relationships through the administration of the studied strain into cell cultures, germ-free mice and colonized animals followed by specific biomarkers or monitorization of disease progression. Additionally, cultivation can study microbial interaction networks by co-cultivation of strains, allowing to clarify cross-feeding interactions, targeted antibiotic mechanisms and identification of strains capable of competing in the complex gut environment.

1.5 Targeting the microbiota for therapeutic modulation of CRC

As previously referred, gut microbiota modulates CRC through several mechanisms that include inflammation, genotoxicity and metabolism. The better understanding of these mechanisms provides possibilities to target the microbiota for cancer prevention strategies (Figure 1.8). For instance, prebiotic fibres, probiotics and microbiota transplants (such as faecal transplantation) can be suitable strategies to tackle CRC-associated dysbiosis by restoring eubiosis in chronic disease states, which consequently promotes the reduction of bacterial-associated genotoxicity and inflammation (Schwabe and Jobin, 2013).

Prebiotic fibre is a non-digestible part of foods like bananas, the skin of apples and beans. These fibres go through the small intestine undigested and are fermented when they reach the colon. The products of fermentation feed and increase the number of beneficial bacteria in our colon. On the other hand, probiotics are defined as live bacteria that when administered in adequate amounts, confer a health benefit on the host. These have already a long history of use and mainly include *Lactobacillus, Bifidobacterium, Saccharomyces* and *Escherichia coli*. Next-generation probiotics are microorganisms that have not been used as agents to promote health to date like *Akkermansia*, *Bacteroides* and *Faecalibacterium* (O'Toole *et al.*, 2017). Zhou and co-workers showed that *Actinobacteria* like *Streptomyces*, *Firmicutes* like *Brevibacterium* or *Bacillus*, and *Proteobacteria* like *Acinetobacter* may be suitable next-generation probiotic candidates since they exhibited a potent anti-carcinogenic activity on a broad range of solid cancers (Zhou *et al.*, 2017).

In addition, dietary approaches seem to be the most cost-effective and reasonable approach, since we know that diet modulates gut microbiota, being strongly involved in the evolution of CRC (O'Keefe, 2016). Tilg *et al* highlighted two dietary components that present promising results within this topic (Tilg *et al.*, 2018). One is resistant starch which presents anti-inflammatory and anti-cancer properties, being able of reducing tumour loads in rat models of inflammation-associated CRC, through modulation of the microbiota. The other constituent is a plant-derived component called berberine which was able to rescue tumour formation induced by *Fusobacterium nucleatum* (Tilg *et al.*, 2018).

Genotoxins and virulence factors are also targets of interest in the context of CRC treatment. An example is the case of boronic acid compounds, which inhibit serine, a crucial enzyme in the process of synthesis of the colibactin toxin. This way, the genotoxic activity of colibactin-producing *E.coli*, which presents a major role in CRC-associated genomic instability, is suppressed *in vitro* and *in vivo* (Gagnière *et al.*, 2016). Limited-spectrum antibiotics are also a possible approach to target genotoxic and pathogenic bacteria.

In combination with matched diets, the use of a genetically altered microbiota lacking or expressing specific enzymes is another approach with potential to decrease levels of tumour promoting substances/bacteria. For instance, the deletion in several strains of lipoteichoic acid, a TLR ligand, was able to normalize innate pathogenic immune responses and to decrease cancer-promoting inflammation in CRC (Gagnière *et al.*, 2016). These

inflammatory pathways that are activated by the bacterial microbiota can also be targeted directly by drugs in order to reduce inflammation (Schwabe and Jobin, 2013). In addition, *Bifidobacterium breve* and *Lactobacillus rhamnosus* appear to be good candidates in CRC prevention since they inhibit the production of pro-inflammatory cytokines and decrease histone acetylation and DNA methylation (events directly involved in CRC) (Gagnière *et al.*, 2016).

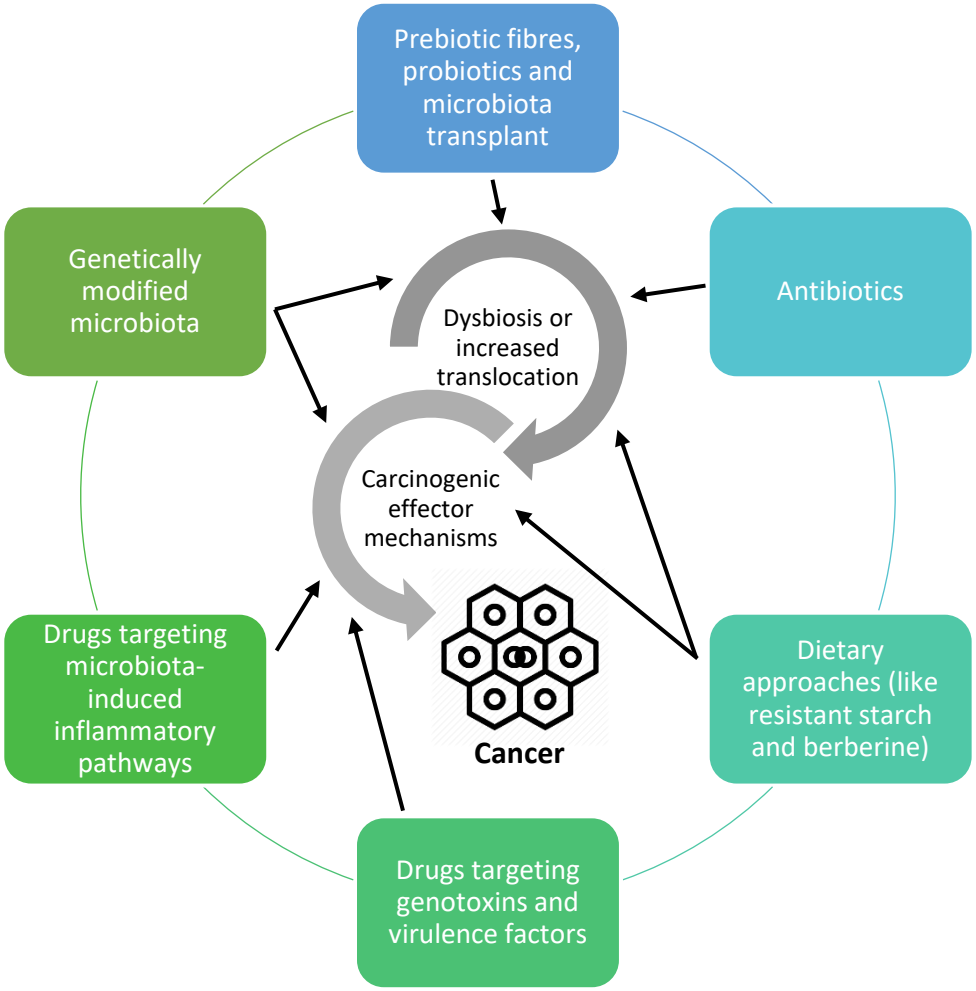


Figure 1.8 – Targeting the microbiota for therapeutic modulation of CRC.

1.6 Motivation

Colorectal cancer is one of the most common cancers, with 1.4 million new cases diagnosed in 2012 and 2.2 million expected by 2030 worldwide (Arnold *et al.*, 2017). Indeed, 1 in 20 people will be diagnosed with CRC and 1 in 5 will develop an adenoma (Sears and Garrett, 2014). Advanced CRC has a high mortality rate and is one of the top three cancer killers in men and women around the world (*Colorectal Cancer - Cancer Stat Facts*, 2018). The colon is also the part of our body with the largest number of microbes, which suggests a role for bacteria or its metabolites in colorectal carcinogenesis (Sears and Garrett, 2014).

Tumour-promoting effects of the microbiota in CRC may be caused by altered host-microbiota interactions and by dysbiosis (Flemer, Lynch, *et al.*, 2017). Certain microorganisms have been shown to be associated with an increased risk of CRC. And several mechanisms by which these single bacteria modulate CRC are well studied and

include inflammation, bacterial-derived genotoxins and virulence factors, oxidative stress and bacterial-derived metabolism. However, contrary to gastric carcinogenesis, CRC does not seem to result from a single pathogen but rather from a consortium which presents intricated relationships (Gagnière *et al.*, 2016).

Effective prevention of CRC requires identification of biomarkers of risk. However, currently available diagnostic strategies such as FOBT and colonoscopy are only able to detect lesions in an already developed stage (polyps and carcinomas). Due to the prognostic and therapeutic importance of early discovery of colonic disease, analysis of the gut and oral microbiota/specific pathogens are emerging as potential sensitive biomarkers/screening approaches predictive of the risk of CRC development (Flemer, Warren, *et al.*, 2017).

As previously referred, recent research has established that an altered community of bacteria is a feature of CRC (Flemer, Lynch, *et al.*, 2017). However, no unifying CRC-associated microbiota structure has been determined (Flemer, Warren, *et al.*, 2017). Within the scope of this topic, in the first part of this project, 16S rRNA Illumina sequencing was performed in order to catalogue biopsies (and nearby tissues) of 39 CRC patients and 5 healthy controls accordingly to their microbiota profiles. Biopsy samples of a selection of CRC patients were used to perform histological staining procedures in order to mainly validate or not the presence of bacteria and bacterial biofilms, and to understand how far they can penetrate into these tissues. This first part constitutes the basis for the second part of this project.

Interestingly, recent data suggests that part of these bacteria are derived partly from the oral microbiota since high abundance of putative oral bacteria have been found on CRC tumours and polyps by metagenomic profiling (Flemer, Warren, *et al.*, 2017). A working hypothesis is that sporadic colon cancer occurs in individuals who have accumulated a burden of mutations, the last of which may be triggered by environmental factors including specific bacteria. Whether or not this gut mucosal colonization by oral bacteria is allowed to occur probably depends on the ability of the gut microbiota to withstand such colonization, but this is still hypothetical. We also do not currently know how oral bacteria might accelerate tumorigenesis. To help advance this theory, there is the need to grow these bacteria in pure culture, to prove that they are truly derived from the oral cavity (and this can only be achieved with pure isolates) and to study their ability to interact with the innate immune system and with cultured cells in a later step. By understanding the mechanisms by which these single bacteria promote/protect CRC, therapeutic strategies can be engineered to target the microbiota for cancer prevention.

Are oral bacteria able to colonize CRC tumours? This is the main hypothesis investigated in this thesis. So far only one paper managed to prove this for one oral species - *Fusobacterium nucleatum* – by using a targeted culture approach (Komiya *et al.*, 2018). There is the interest in trying to discover more oral species with this capacity for all the reasons presented above. Several objectives were established in order to test our hypothesis - culture bacteria from colonic tumours and oral swabs of 2 CRC patients, by using a complete culture-based methodology with a selection of different incubation conditions and a diverse panel of media that include YCFA; identify bacteria by sequencing an amplicon from the full length 16S rRNA gene (Sanger sequencing); perform RAPD to compare the genomic structure of oral and colonic isolates of the same species; and identify candidates for further analysis by full genomic sequencing.

Chapter 2

Materials and Methods

1st Part

2.1 Sampling

Samples used for identification of microbiota profiles (V3/V4 variable region of 16S rRNA gene Illumina sequencing), Gram and H&E staining, FISH and IF

CRC patients: In total, 39 CRC individuals were scheduled for colonic resection at the Mercy University Hospital, in Cork. Individuals were not treated with antibiotics or probiotics in the month prior to surgery but were administered antibiotics intravenously within a few hours of the resection. Exclusion criteria included patients that had been subjected to radio or chemotherapy prior to surgery. After resection of the colon, 3 samples were collected in 4 mL of RNAlater (Qiagen): ON tumour (tumour mass), OFF tumour (around 3 cm away from the tumour mass), UDP or UDD (un-diseased proximal or distal; around 10 cm away from the tumour mass). Samples were stored at 4 ^{o}C for 12 h and then stored at -20 ^{o}C, for further analysis. Paired faecal samples from the same patients were collected before surgery and frozen at -80 ^{o}C. Biopsies of all patients were obtained in duplicates and collected in 12 mL of Methacarn (60% methanol absolute, 30% chloroform and 10% acetic acid glacial). Samples were then fixed for 48h at 4 ^{o}C and embedded in paraffin (Johansson and Hansson, 2012). 3-4 µm serial sections were obtained in several slides in order to perform staining protocols (Haematoxylin & Eosin staining, Gram staining, Fluorescent In-situ Hybridization and Immunofluorescence) in seven patients' samples - CRC031, CRC038, CRC042, CRC044, CRC045, CRC056 and CRC057 patients.

Healthy controls: In total, biopsy samples from 5 healthy individuals (controls) were collected in 4 mL of RNAlater. Three biopsies were obtained per individual via endoscopy from ascending (ASC), transverse (TRANS) and descending (DESC) colon (Figure 2.1).

* For better understanding of the "Results and Discussion" section, it must be highlighted that from the 39 CRC patients a division was done for distinct analysis: **1.** the microbiota profiles, posteriorly obtained, of 20 CRC patients were compared with the profiles of the 5 healthy patients; **2.** the microbiota profiles of the remaining 19 CRC patients were compared with the profiles of the 2 patients whose samples were used in the broad-range culture approach.

2.2 DNA and RNA extraction and 16S rRNA amplicon sequencing

Genomic DNA and total RNA were extracted from biopsy samples using the AllPrep DNA/RNA kit (Qiagen). Briefly, samples previously collected on RNAlater were thawed on ice. Appropriated number of bead-beating tubes containing 250 µL of 0.1 mm sterile glass beads and three 3-4 mm sterile glass beads were prepared in advance. Next, the lysis buffer was added to each one of them - 600 µL of RLT buffer containing 0.143 M β-Mercaptoethanol (Merck).

Each biopsy was removed from the RNAlater with sterile forceps and placed into one half of a Petri dish where it was cut with a sterile scalpel. One part of the tissue was cut into even smaller pieces in order to assist homogenization and added to a labelled bead-beating tube. The amount of tissue for each extraction was estimated (0.3-0.5 cm^3). Tubes were placed on ice. The remaining large piece of tissue was placed into a labelled 1.5 ml barcoded cryotube and stored at -80˚C, for further analysis.

Samples were homogenized twice for 15 s in a MagnaLyzer (Roche) at 6500 rpm. In between the homogenisation steps, the samples were kept on ice for 30 s. The following steps were carried out according to the AllPrep DNA/RNA extraction kit (Qiagen) instructions. The purified DNA was quantified (ng/µL) using a Nanodrop 2000 (Thermo Scientific).

16S rRNA gene amplicon sequencing was performed by Dr Werner Frei. 16S Metagenomic Sequencing Library Preparation protocol developed by Illumina was employed. 100 ng of DNA extracted from each biopsy (or 20 ng for faecal samples) was amplified employing primers targeting the V3/V4 variable region of the 16S rRNA gene. These primers are 16S amplicon PCR forward primer (V3 region): 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACG GGNGGCWGCAG; and 16S amplicon PCR reverse primer (V4 region): 5' GTCTCGTGGGCTCG GAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC. The products were purified and forward and reverse barcodes were attached (Nextera XT V.2 Index Kit sets A and D, Illumina). Amplicons were purified again and quantified using Qubit™ dsDNA HS Assay Kit (Invitrogen). Equimolar amounts of DNA per amplicon were then pooled and sequenced at National Irish Sequencing Centre (Teagasc Food Research Centre, Moorepark, Co. Cork).

Analysis of the 16S amplicon sequencing data was performed by Dr Céline Ribière. Paired-end reads were joined using the Fast Length Adjustment of SHort reads to improve genome assemblies (FLASH) programme (Magoc and Salzberg, 2011). Sequences were quality filtered using split_libraries_fastq.py script from the open source software package Quantitative Insights Into Microbial Ecology (QIIME v. 1.9.1) (Caporaso *et al.*, 2010). Forward and reverse primers were removed using cutadapt (Martin, 2011) and QIIME's script truncate_reverse_primer.py, respectively. The USEARCH sequence analysis tool (Edgar, 2010) was used for further quality filtering and OTU clustering. Sequences were filtered by length, sorted by size, single unique reads removed, and the remaining reads were clustered into OTUs (97% of identity). Chimeras were removed with UCHIME using the GOLD reference database. The original quality-filtered sequences were mapped onto the OTUs with 97% similarity. OTU representative sequences were taxonomically classified with an 80% confidence

threshold from phylum to genus level using classify.seqs command within the mothur suite of tools (v1.36.1) (Schloss *et al.*, 2009) with the RDP reference database (training set 14) and to species level using SPINGO with default parameters (Allard *et al.*, 2015). The R statistical software package (version 3.4.1) was used for data visualisation (ggplot2 2.2.1) into relative abundance bar charts (microbiota profile) at family and genus level.

2.3 Histological evaluation and tissue morphology: H&E staining

Haematoxylin & Eosin (H&E) staining was performed in paraffin-embedded human biopsy sections. The slides were incubated in histoclear (xylene substitute solution) twice for 5 min and then hydrated in solutions with decreasing concentration of ethanol (100, 95, and 70%) and in distilled water. The slides were incubated for 5 min in each bath. The samples were stained with haematoxylin Harris (Gill II) for 1 min, rinsed in running tap water for 5 min, differentiated with acid alcohol (0.37% HCl in 70% ethanol) for 3 sec, rinsed again in tap water for 5 min, stained with eosin 0.2% for 10 min and rinsed in tap water for 5 min. The final step included the sequential incubation in (90% and 100%) ethanol and in histoclear for 5 min twice. Slides were mounted.

2.4 Validation of bacteria presence an identification of bacterial distribution in the colonic tissue: Gram staining

The Gram stained protocol was performed in the biopsy sections by the following order. The slides were dewaxed and hydrated to water, stained with 1% aqueous Crystal Violet for 2 min, rinsed in distilled water, incubated with a drop of Gram Iodine for 2 min, rinsed briefly in distilled water and left to dry. After this first stage, the sections were decolourised in 30% acetone in alcohol for 1 min, rinsed briefly in distilled water, counterstained in 1% Neutral Red solution for 1 min, rinsed very briefly in water, airdried and mounted.

2.5 Identification of biofilms and local immune responses: Fluorescent In-situ Hybridization (FISH) and Immunofluorescence (IF) of CRC samples

2.5.1 FISH

The sections were de-waxed by an initial incubation of the slides at 60 ^{o}C for 10min. The slides were incubated in histoclear twice for 10min with the first solution prewarmed to 60 ^{o}C and subsequently incubated in 100% ethanol for 5 min. They were air-dried for 2 min. The incubation of the slides with the probes was done at a concentration of 2pmol/µl of each probe in pre-warmed (to 50 ^{o}C) hybridization buffer (900mM NaCl, 20mM Tris-HCl pH 7.4, 0.01% SDS). Next, the slides were incubated at 50 ^{o}C in a humid dark chamber for 16 hours.

After the hybridization step, the slides were rinsed in a pre-warmed washing buffer (900 mM NaCl, 20 mM Tris-HCl pH 7.4, 5 mM EDTA) for 15 min, at 48 ^{o}C. They were rinsed again with phosphate buffer saline (PBS) 3 times for 10 min to remove unbound probes. The slides were covered for 10 s with 4',6-diamidino-2-phenylindole

(DAPI) (0.125 µg/ml in PBS), posteriorly washed 3 times in PBS and allowed to dry completely. The coverslips were mounted using ProLongGold antifade reagent (Life Technologies) and allowed to set at room temperature. This protocol was adapted from (Dejea *et al.,* 2014) and (Pédron *et al.,* 2012).

The FISH probes used were Eub338 (5' GCTGCCTCCCGTAGGAGT) and Non338 (5' ACTCCTACGGGAGGCAGC) which were covalently linked with fluorescein isothiocyanate (FITC) and Cyanin 3 (Cy3) at their 5' ends, respectively. The former targets the conserved 16S ribosomal RNA bacterial domain (position 338-355), allowing the detection of all bacterial populations. The latter is a nonsense probe and it was used to control non-specific binding of Eub338, i.e., as a negative control.

2.5.2 CD45 anti-human antibody protocol (IF)

CD45 marker (lymphocyte common antigen) was chosen to identify the presence of immune cells. Briefly, the sections were de-waxed by an initial incubation of the slides at 60 ^{o}C for 10 min and subsequently incubated in histoclear twice for 5 min with the first solution prewarmed to 60 ^{o}C. The sample was rehydrated in solutions with decreasing concentrations of ethanol (100, 95, and 70%) twice for 3 min. The sections were washed in PBS and the tissue in the centre was marked around with a PAP pen. This is followed by a blocking step (of non-specific binding) where 1x blocking buffer (50 µl goat serum, 250 µl 10% BSA, 700 µl 1x PBS with 0.1% Tween (or 0.1% Triton X-100)) was added to the slides and these were incubated in darkness in a humid chamber for 2 hours at room temperature. Next, they were washed in PBS for 5 min.

The second part of this protocol concerns primary and secondary antibody staining**.** The primary antibody (CD45 marker) was diluted in 1x blocking buffer and added to the slide which was subsequently incubated in the dark in a humidified chamber for 4-24 hours at 4 ^{o}C. After this step, the slides were washed in PBS for 3x for 30 min. From this point forward, the slides were kept in the dark whenever possible. The secondary antibody was diluted in 0.5x blocking buffer (1mL 1x blocking buffer + 1mL 1x PBS with 0.1% Tween) and added to the sample. Incubation was performed in the dark in a humid chamber at room temperature for 30min-2h. The slides were washed in PBS 3x for 3 min and air-dried. Next, the slides were covered for 1-5min with DAPI (0.5 µg/ml in PBS), washed 2x in PBS for 3 min and allowed to dry completely.

Coverslips were mounted using ProLongGold antifade reagent by adding 1 drop of this mounting medium on top of the tissue. A coverslip was placed carefully on top of the drop, avoiding air bubbles. The slides dried 30-60 minutes in the dark and their edges were sealed.

In the 4 protocols (H&E and Gram staining, FISH and IF), slides were examined under a high-resolution Olympus BX3 microscope equipped with a charge-coupled-device (CCD) camera (Olympus DP71). Images were processed with the cell^F Imaging system and optimized with Image J software.

A scheme summarizing the methodology explored in the 1st part of this project is presented in Figure 2.1.

1

Sample collection

39 CRC individuals (colonic resection):

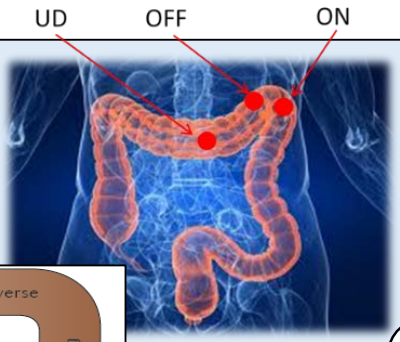
ON tumour (tumour mass), OFF tumour (around 3 cm away from the tumour mass), UDP and/or UDD (un-diseased proximal or distal; around 10 cm away from the tumour mass).

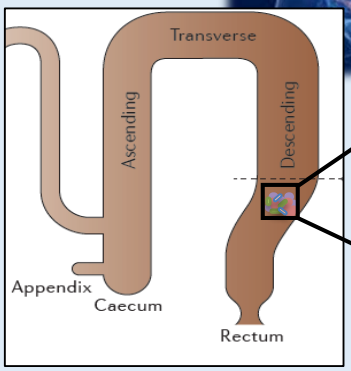
5 Healthy controls:

3 biopsies via endoscopy from ascending (ASC), transversal (TRANS) and descending (DESC) colon.

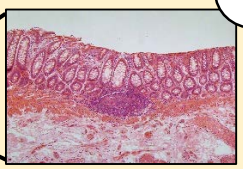
Collection medium – RNAlater (for gDNA extraction) and Methacarn (for serial sections preparation).

Samples in RNAlater were stored at 4°C for 12 h and then stored at -20 °C. Faecal samples from the same patients were frozen at -80 °C.





3



H&E staining – histological evaluation and tissue morphology

2

DNA/RNA extraction and V3/V4 16S rRNA Illumina sequencing

gDNA and total RNA extraction: adapted AllPrep DNA/RNA kit (Qiagen) protocol.

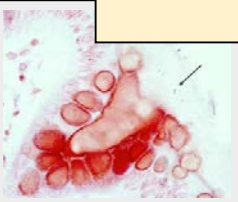
16S rRNA gene amplicon sequencing: 16S Metagenomic Sequencing Library Preparation protocol developed by Illumina.

Metagenomic analysis results:

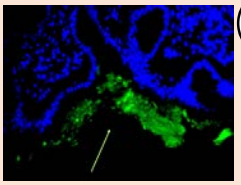
- Relative abundance bar charts (microbiota profile) plotted by family and genus level.

4

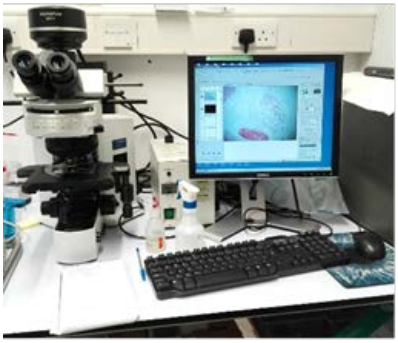
Gram staining - validation of bacteria presence, identification of bacteria distribution and density in the colonic tissue



5



FISH - validation of bacteria/biofilms presence



6

CD45 IF - identification of local immune responses

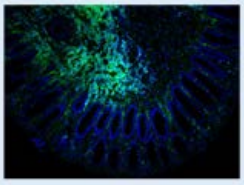


Figure 2.1 – Schematic of the experimental procedures (1st part)

2nd Part

2.6 Patients selected for the culture approach (patients 1 and 2)

CRC patients: 2 CRC individuals were scheduled for colonic resection at Mercy University Hospital, Cork. 3 types of samples (described below) were collected – oral swab (2x), saliva and colonic tumour sample (2x). Oral swabs and saliva sample were collected prior to the surgery in the morning. The colonic biopsy was collected in the afternoon.

Patient 1 is a male with an age of 62 presenting a rectal adenocarcinoma diagnosed in 2016. Ex-smoker, previous alcohol excess and hypertension are some comorbidities that can be relevant to present. Patient 1 wasn't treated with antibiotics, however, he had chemotherapy during the last 2 years.

Patient 2 presents male gender, 89 years old and a caecal adenocarcinoma which was diagnosed in May 2018. Only 1 month had passed between the collection and the diagnosis. This patient had prostate cancer in 2004 which was treated with radiotherapy. The patient has arthritis and myocardial infarction. Neither antibiotics treatment nor chemotherapy were performed in the case of this subject.

*Faecal samples of patients CRC038 and CRC057 were also used in the culture and isolation protocol as training. They allowed establishing the optimized culture and isolation conditions used in patient 1 and patient 2.

*Patient 1 and patient 2 are sometimes abbreviated to pat 1 and pat 2.

2.7 Collection of samples and transport

Oral swabs (2x): Oral samples were obtained by rubbing the inside of both cheeks with a swab. Sterile cotton swabs (Sterilin, Thermo Fisher) were used to collect the sample and the tip was placed into a cryotube with 1 mL of reduced PBS (Sigma) + 20% glycerol. This sample was used for culture purposes. The duplicated sample was placed in the original sterile cotton swab tube and stored at -80 °C. Genomic DNA extraction was performed in this sample by using the PowerSoil DNA Isolation Kit (MO BIO Laboratories).

Saliva sample: The patient had to spit on a 20 mL sterile container (Sterilin, Thermo Fisher). This sample was used in the culture procedure in case the oral swab was poor in bacterial load.

Colonic tumour samples (2x): The sample used in the culture protocol was collected into a cryotube with 1 mL of reduced PBS + 20% glycerol. Moreover, an extra piece of the same biopsy was placed in a 20 mL tube (Sarstedt) containing 4 mL of RNAlater and stored at -20 °C. This extra sample was used for DNA extraction by using the All Prep DNA/RNA Mini Kit (Qiagen).

The transport of samples was performed immediately from Mercy University Hospital to the lab. The previous cryotubes and containers with samples were transported inside an anaerobic box. To generate an anaerobic

atmosphere, the box contained an AnaeroGen W-Zip Compact bag system (Oxoid) that was ripped immediately after the collection of samples. Culture protocol was started 20-30 min after collection.

2.8 Culture, isolation and purification of bacteria from tumour biopsy and oral swab of the same CRC subject

2.8.1 Culture conditions

Culture, incubation and isolation work was carried out 1) under strictly anaerobic conditions (Whitley DG250 Workstation, Don Whitley Scientific - 10% H_2, 10% CO_2, 80% N_2); 2) under microaerophilic conditions (Incubator CO2 Forma Steri-cycle, Thermo Scientific - 5% CO_2); and 3) under aerobic conditions (Incubator MkII, LEEC) at 37 ^{o}C. Culture conditions selected include eleven different media, anaerobic pre-incubation in blood bottles for 7 and 17 days, heat shock (80 ^{o}C during 30 min-1 h), microaerophilic incubation in respective liquid medium for 7 days prior to plating and selective filtration with 0.2 µm pore filter. All conditions were tested simultaneously in colon tumour and oral samples.

The culture media selected rely on 3 types of medium approaches presented below:

1. Complex medium able to isolate a huge range of bacterial genera. YCFA medium (pH 7.5) is an anaerobic rich medium prepared under CO_2 flow and is used as the principal medium in this experiment. Some variations with kanamycin or/and vancomycin were used.

2. Media suitable for bacteria linked with caries; periodontal disease; and oral bacteria genera identified in more abundance on CRC tumour samples by (Flemer, Warren, *et al.,* 2017). These media are: BBL Brucella agar (BD Biosciences, pH 7, *Haemophilus* species); Bile Esculin agar (Merck, pH 6.6, *Streptococcus, Enterococcus, Klebsiella, Peptostreptococcus*); Tomato Juice agar (Fluka Analytical, pH 6.7, Aciduric bacteria); Todd Hewitt agar (Fluka Analytical, pH 7.8, *Streptococcus*), MRS agar (Oxoid, pH 6.2, *Lactobacillus*), M17 agar (Oxoid, pH 6.9, Lactic *Streptococcus*), Sabouraud 2% glucose (Merck, pH 5.6, Aciduric bacteria). Lower pH and M17 (with lactose) medium were used with the specific purpose of isolate bacteria related with caries. Neutral pH and media whose composition contains peptone were used to isolate bacteria linked with oral cancer and periodontal disease.

3. Media used to isolate general oral bacteria. 3 media were selected: Tryptic Soy agar (Merck) with or without 5% horse blood (pH 7.3), Brain Heart Infusion agar (Oxoid) with or without 5% horse blood (pH 7.4), and Fastidious Anaerobe agar (LAB M) with or without 5% horse blood (pH 7.2).

2.8.2 Preparation of YCFA medium

YCFA medium (1L) was prepared as follows. Casitone (18g), yeast extract (2.5 g), $NaHCO_3$ (4 g), glucose (2g), maltose (2 g), cellobiose (2 g), mineral solution 1 (150 mL), mineral solution 2 (150 mL), vitamin solution 1 (1 mL),

vitamin solution 2 (1 mL), haemin solution (10 mL) and distilled water (up to 995 mL) were added to an Erlenmeyer. Resazurin solution was added as an oxygen indicator (1mL). Once the medium is reduced resazurin turns colourless. The solution was brought to boil and allowed to cool down. 3 mL of VFA (volatile fatty acids) solution was added into the fume hood and the pH was adjusted to 7.5 ± 0.2. The solution was flushed with CO_2 and the reducing agent L-cysteine-HCL (1g) was added. Flushing was kept until medium lost the pink colour (oxygen was completed consumed by this stage). Serum bottles were flushed while the medium was being flushed. The medium was aliquoted into the serum bottles with a syringe and quickly sealed. To prepare YCFA agar, agar was added directly to the bottles at 2% before adding the reduced medium. Once the medium is autoclaved the serum bottles can only be opened again in the anaerobic cabinet. This ensures that the medium is anoxic. Plates were poured at least 24h in the anaerobic cabinet before the start of the culture protocol.

For detailed composition of YCFA's solutions, see supplementary Table A.1.

2.8.3 Starting a culture from a colonic tumour biopsy and oral swab sample in the anaerobic cabinet

The cryotube with 1 mL of reduced PBS + 20% glycerol containing the biopsy/oral swab was vortexed during 20s.

In the case of the oral swab tube, the cotton swab was removed with a clamp and spread into each solid culture medium plate. In the case of the biopsy, the tissue and the PBS solution were placed on a Petri dish. A piece of this tissue was cut with a sterile scalpel and with a sterile clamp spread into each solid culture medium plate. 2 mL of more reduced PBS was added to the Petri dish and the remaining tissue was chopped in very small pieces with a sterile scalpel and homogenised.

From the previous homogenised/mixed biopsy/oral swab solution, 1:10 serial dilutions were performed in YCFA liquid medium from 10^0 to 10^{-9}. The suspensions were vortexed for 5 s between dilutions. 100 µl of each dilution was plated. In the case of YCFA solid medium, all dilutions were plated. In the case of the other 10 remaining media, the dilutions plated were 10^{-1} and 10^{-2} for biopsy samples; 10^{-2} and 10^{-3} for oral swab samples (these dilutions were determined previously as the most suitable for isolation in a pilot trial performed with similar samples). The backup saliva sample was also plated but only in YCFA medium in the same dilutions used for the oral swab.

Additionally, in the case of patient 1, 100 µL of the initial biopsy/oral swab suspension was added to Falcons with 8 mL of each liquid medium and incubated for 4 days inside the anaerobic cabinet and microaerophilic incubator. After the 4 days, the Falcon tubes were centrifuged at 5 000 rpm for 10 minutes (ST 16R Centrifuge, with TX-400 rotor, Thermo Scientific), the supernatant removed, and the final pellets stored at -20 ^{o}C. V3/V4 16S rRNA amplicon sequencing was performed in these samples in order to show the selective power of each medium and the coverage of the isolates' library.

2.8.4 Isolation, purification and storage of the isolated strains

The plates were incubated in the conditions referred previously, for 7 days. After this period, colonies were compared in terms of morphology between biopsy and oral swab plates. Single colonies were picked from the above plates, re-streaked in new solid medium plates and grown for 2 days. Purified colonies were picked and grown in 8 mL of respective liquid medium for 2 days. Next, 20% glycerol stocks were prepared with each pure culture and stored at -80 ^{o}C. The isolates' cultures were centrifuged at 5 000 rpm for 10 minutes, the supernatant removed, and the final pellets stored at -20 ^{o}C.

2.9 Extraction of genomic DNA

2.9.1 Oral swab

Genomic DNA was extracted using the DNeasy PowerSoil Kit (Qiagen). The oral swab was placed into the PowerBead tube and vortexed briefly. The volume added of elution solution C6 was 20 µL. The rest of the DNA extraction was carried out according to the extraction kit procedure. The eluted DNA was quantified with Nanodrop (ng/µL) and stored at -20 ^{o}C.

2.9.2 Tumour biopsy

Genomic DNA was extracted using the AllPrep DNA/RNA kit (Qiagen) by following the adapted protocol presented before.

2.9.3 Isolates from oral swab and tumour biopsy

Genomic DNA was extracted using the reagents from Blood and Tissue kit (Qiagen). Briefly, the pellets of the isolates were thawed and 500 µL of Inhibitex buffer (Qiagen) was added to each Falcon tube containing the cultures. The pellet was resuspended and transferred into a tube containing 2x 250 µL of 0.1 mm and 1mm sterile glass beads (Thistle Scientific). The tubes were bead-beaten twice for 1 min using a Mini-BeadBeater (Biospec Products). In between the homogenisation steps, the samples were kept on ice for 1 min. Next, the tubes were centrifuged for 1 min at maximum speed (Centrifuge 5415D, Eppendorf). 15 µL of proteinase K, 200 µL of supernatant from previous sample and 200 µL of AL buffer were added by this order to sterile 1.5 mL Eppendorf tubes. An incubation step at 70 ^{o}C for 15 min was performed in the block heater (Stuart Scientific). After this step, 200 µL of pure ethanol was added and the mixture was vortexed. The entire resultant volume was pippeted into a DNeasy Mini spin column placed in a 2 mL tube. The columns were centrifuged 1 min at full speed and the flow through was discarded. 500 µL of AW1 buffer (binds DNA) was added and centrifuged for 1 min. The flow-through was discarded. 500 µL of AW2 buffer (washing buffer) was added and centrifuged for 3 min. The flow-through was discarded. Next, columns were placed in new 2 mL collection tubes and centrifuged for 3 min in order to remove completely the ethanol from the membrane. Columns were placed in new sterile Eppendorf

tubes and 50 µL of AE buffer was added to the centre of the membrane. These were incubated 5 min at room temperature and centrifuged for 1 min at full speed. The DNA was eluted, quantified with Nanodrop (ng/µL) and stored at -20 ^{o}C.

This procedure was also used to extract the DNA from the mixed bacterial pellets from patient 1 that grew in each liquid medium (selective power experiment).

2.10 PCR protocol (full-length 16S rRNA gene)

The universal bacterial primers employed for the PCR were 27F2 forward primer (5' AGAGTTTGATYMTGGCTC) and 1492R3 reverse primer (5' GGNTACCTTGTTAYGACTT) from Eurofins Genomics. These primers amplify 1500 bp of the 16S rRNA gene (the full sequence).

Briefly, MyTaq Mix (Bioline), primers and DNA samples were thawed on ice. 1:10 and 1:100 DNA dilutions were performed when DNA concentrations were higher than 20 and 200 ng/µL, respectively. The reaction mix (with 10% of volume excess) was prepared by the following way per reaction: Nuclease-free water (10 µL), 5 µM Forward and Reverse primers (5 µL each) and 2x MyTaq Mix (25 µL). 45 µL of reaction mix was added into the 96-well PCR plate. Next, 5 µL of DNA samples was added into the corresponding wells and mixed. The final volume of reaction was 50 µL. The plate was centrifuged for a few seconds and placed into the PCR Thermal Cycler (Veriti 96 Well Thermal cycler, AB Applied Biosystems). The PCR program adapted to this protocol was: Initialisation (95 ^{o}C, 1 min, no repeats), Denaturation (95 ^{o}C, 15 sec, x35 cycles), Annealing (59 ^{o}C, 15 s, x35 cycles), Extension (72 ^{o}C, 1 min, x35 cycles), Elongation (72 ^{o}C, 1 min, no repeats), and 4 ^{o}C ∞.

After the PCR, an electrophoresis was performed using a 2% agarose gel prepared in TAE buffer with 6% SYBR Safe DNA gel stain (Thermo Fisher Scientific). 5 µL of each PCR product and 1 µL of 5x DNA Loading Buffer Blue (Bioline) was mixed and loaded in each gel well. 5 µL of HyperLadder 1Kb (Bioline) was loaded as a DNA size marker. The negative control was also added. The gel was run for 45 min at 100 V and digitally photographed under UV light (ChemiDoc XRS Gel Documentation system, Bio-Rad).

2.11 Purification of PCR products

After checking the quality of the PCR products in the previous gel, the amplicons needed to be purified before being sent to Sanger sequencing. The High Pure PCR Product Purification Kit (Roche) was used for this purpose. Briefly, 55 µL of Nuclease-free water, 45 µL of PCR product and 500 µL of Binding buffer were added to 1.5 mL sterile Eppendorf tubes. The samples were vortexed and applied into a High Pure Filter Tube, centrifuged at full speed 1 min. The flow-through was discarded. 650 µL of Washing buffer was added to the upper reservoir and the tubes were centrifuged at full speed 1 min. The flow-through was discarded and a new collection tube was reconnected to the Filter tube. An additional 2 min centrifugation step was performed. The filter tube was inserted into a new sterile Eppendorf tube, 50 µL of Elution buffer was added and the tubes centrifuged at full speed 1 min. The purified DNA was quantified with Nanodrop (ng/µL) and stored at – 20 ^{o}C.

2.12 Preparation of samples to be sent for Sanger Sequencing

Purified PCR products were diluted in order to present a final concentration of 80 ng/µL. 5 µL of 5 µM forward primer and 5 µL of diluted purified PCR product were added to a 1.5 mL Eppendorf tube. The LIGHTRUN barcodes (Eurofins Genomics) were attached to the tubes and samples were sent for Sanger sequencing (GATC Biotech, Eurofins Genomics) in double bags (*LIGHTRUN Tube sample requirements*, 2018).

2.13 Analysis of Sanger sequencing data

After receiving the sequenced results from GATC Biotech, the ABI chromatogram of each isolate's sequence was opened in BioEdit Sequence Alignment Editor version 7.2.6 (Hall, 1999). The quality of the samples was checked and the ones that presented contaminations were discarded. Next, the obtained trimmed sequences in FASTA format were compared with the NCBI 16S rRNA sequence database using nucleotide BLAST (*Nucleotide BLAST: Search nucleotide databases using a nucleotide query*, 2018). Identification to the species level was defined as the top BLAST hit with maximum sequence similarity of at least ~99%.

The same was done in the Ribosomal Database Project (RDP) (*RDP Sequence Match*, 2018) to confirm the taxonomic designations to the genus level. From here, a table in Excel was plotted with the following entries: Phylum, Family, Species, apparent Strain, % ID, Isolate's code, Growth conditions, Original Isolation medium, Comments, and Box (where the isolates' stocks were kept) (Table 2.1). Bar charts were plotted to show percentage and numbers of phylum, family, species and apparent strains obtained for each patient.

Table 2.1 – Example of isolated species' resume table. This beginning of table identifies some isolates recovered from the Oral swab of patient 2. Respective isolates' taxonomy, % of identity with the closest relatives in database, codes by which they were archived in stocks, growth conditions and isolation medium are indicated.

Phylum	Family	Species ID (NCBI Blastn)	Strain ID (NCBI Blastn)	% ID (NCBI Blastn)	Isolate code	Growth conditions	Initial Isolation medium	Comments	Box
Actinobacteria	Actinomycetaceae	Actinomyces sp.	ICM39	99	O33	Microaerophilic (5% CO2), 37 °C	Brain Heart agar		Lisa_N
Actinobacteria	Actinomycetaceae	Actinomyces sp.	A11	99	O102	Anaerobic, 37 °C	YCFA		Lisa_P
Firmicutes	Bacillaceae	Bacillus sp.	PS	95	O55	Microaerophilic (5% CO2), 37 °C	Tryptic Soy agar	Contaminated	Lisa_N
Proteobacteria	Enterobacteriaceae	Cronobacter sakazaki	PM468	100	OE	Microaerophilic pre-incubation in each liquid medium for 4 days, Microaerophilic (5% CO2), 37 °C	Tryptic Soy agar + 5% sheep blood		Lisa_P
Actinobacteria	Propionibacteriaceae	Cutibacterium avidum	44067	99	ORAL 007	Anaerobic pre-incubation in blood for 17 days, Microaerophilic, 37 °C	Tryptic Soy agar + 5% horse blood		Lisa_Q
Proteobacteria	Neisseriaceae	Eikenella sp.	NML130454	99	ORAL 016	Anaerobic pre-incubation in blood for 17 days, Microaerophilic, 37 °C	Brain Heart agar + 5% horse blood		Lisa_Q
Firmicutes	Enterococcaceae	Enterococcus faecalis	HCD42-1	99	OB	Microaerophilic pre-incubation in each liquid medium for 4 days, Microaerophilic (5% CO2), 37 °C	Tryptic Soy agar + 5% sheep blood		Lisa_P
Proteobacteria	Enterobacteriaceae	Escherichia coli	RM9975	99	OD	Microaerophilic pre-incubation in each liquid medium for 4 days, Microaerophilic (5% CO2), 37 °C	Tryptic Soy agar + 5% sheep blood		Lisa_P

**Access to these tables will be granted upon request. Feel free to contact me.*

Phylogenetic trees were plotted by species and apparent strain level for each patient. Only one of the many isolates for each apparent strain was selected. The sequences of the isolates' closest relatives in database were aligned by MUSCLE (Multiple Sequence Comparison by Log-Expectation). The evolutionary history was inferred using the Neighbor-Joining method. Associated taxa were clustered together in the bootstrap test (1000 replicates). The evolutionary distances were computed using the Maximum Composite Likelihood method. All

positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. Evolutionary analyses were conducted in MEGA 6 and MEGA X (Kumar *et al.*, 2018). The phylogenetic trees were edited in iTOL (Letunic and Bork, 2011).

In addition, a local Blast was performed by Dr Céline Ribière in order to compare the sequence of the isolated species with the sequences present in the Human Microbiome Project's "Most Wanted" list of previously uncultured and unsequenced microbes (*NIH Human Microbiome Project - Most Wanted Genomes*, 2018). Figures 3.11, 3.12.B, 3.13 and 3.14 are from Dr Céline's authorship.

2.14 RAPD

RAPD was performed to compare the genomic structure of oral and colonic isolates of the same species. RAPD with the M13 primer (5' GAGGGTGGCGGTTCT) was employed in a first stage to see which strains between both samples presented equal patterns. A second RAPD with the 1254 primer (5' CCGCAGCCAA) was used to confirm that strains that matched between oral and colonic samples in the previous M13 RAPD had equal gel patterns again.

2.14.1 M13 RAPD

Briefly, MyTaq Red Mix (Bioline), M13 primer (Eurofins Genomics) and DNA samples were thawed on ice. DNA dilutions were performed to a final concentration of 25 ng/µL. The reaction mix (with 10% of volume excess) was prepared by the following way per reaction: Nuclease-free water (4 µL), 5 µM M13 primer (4 µL) and 2x MyTaq Red Mix (10 µL). 45 µL of reaction mix was added into the 96-well PCR plate. Next, 2 µL of 25ng/µL DNA samples were added into the corresponding wells and mixed. The final volume of reaction was 20 µL. The plate was centrifuged for a few seconds and placed into the PCR Thermal Cycler. The AP-PCR program adapted to this protocol was: Initialisation (94 ^{o}C, 15 min, no repeats), Denaturation (94 ^{o}C, 1 min, x40 cycles), Annealing (45 ^{o}C, 20 s, x40 cycles), Extension (72 ^{o}C, 2 min, x40 cycles), Elongation (72 ^{o}C, 5 min, no repeats), and 4 ^{o}C ∞ (Stenlid *et al.*, 1994).

After the PCR, an electrophoresis was performed using a 2% agarose gel (prepared in TAE buffer with 6% SYBR Safe DNA gel stain). 8 µL of each RAPD PCR product and 1 µL of loading buffer was mixed and loaded in each gel well. One well was loaded with 8 µL of 1Kb ladder. The negative control was also added. The gel was run for 90 min at 90 V.

2.14.2 1254 RAPD

This protocol was carried out according to the M13 AP-PCR protocol presented before but using the 1254 primer (Eurofins Genomics). RAPD PCR conditions were as follows: 4 cycles of 94 ^{o}C 5 min, 36 ^{o}C 5 min, 72 ^{o}C 5 min; 30 cycles of 94 ^{o}C 1 min, 36 ^{o}C 1 min, 72 ^{o}C 2 min; 72 ^{o}C 10 min and 4 ^{o}C ∞ with no repeats (Akopyanz *et al.*, 1992). A scheme summarizing the methodology explored in the 2nd part of this project is presented in Figure 2.2.

+ Faecal samples from patient CRC038 and CRC057

1 **Samples from patients 1 and 2**

- 2x CRC tumour (colonic resection)
- 2x Oral swab
- Saliva sample

Collection medium:

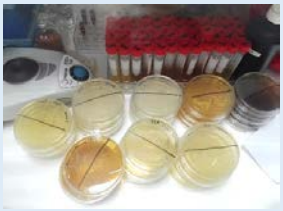
- Reduced PBS + 20% glycerol (culture purposes)
- RNAlater and storage at -20 °C (only tumour samples for subsequent gDNA extraction)

Transport done in anaerobic atmosphere.

2 **Culture (20-30 min after collection)**

Main conditions tested in both oral and tumour samples (37 ºC):

- Anaerobic (anaerobic cabinet -10% H_2, 10% CO_2, 80% N_2), Microaerophilic (5% CO_2) and Aerobic incubations
- Anoxic YCFA medium prepared under CO_2 flow + 10 more different media
- Anaerobic pre-incubation in blood bottles
- Heat shock



Serial dilutions, platting and incubation for 7 days.

Colony purification and liquid culture.

20% glycerol stocks and -80 ºC storage.

3 **Extraction of genomic DNA**

From oral and tumour samples (for microbiota profiling - V3/V4 16S rRNA Illumina sequencing):

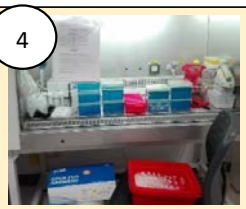
Oral swab – adapted DNeasy PowerSoil kit (Qiagen) protocol.

Tumour biopsy – adapted AllPrep DNA/RNA kit (Qiagen) protocol.

From cultured isolates:

Adapted Blood and Tissue kit procedure (Qiagen).

4 **PCR (full length 16S rRNA gene), Purification and Preparation of samples for Sanger sequencing**



Universal bacterial primers: 27F2 and 1492R3

Purification of PCR products: adapted High Pure PCR Product Purification kit (Roche) protocol.

5 µL of 5 µM forward primer and 5 µL of diluted purified PCR product (80 ng/µL) were added to a 1.5 mL Eppendorf tube with LIGHTRUN barcodes and sent for Sanger sequencing (GATC Biotech, Eurofins Genomics).

6 **RAPD**

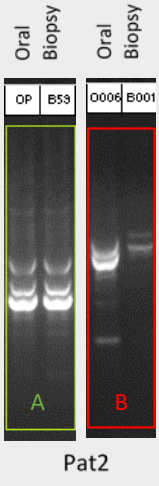
Comparison of the genomic structure of oral and colonic isolates of the same species (band pattern).

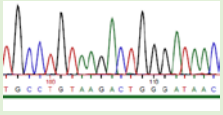
M13 RAPD in a 1st stage:

Same band pattern (A) – identical strains

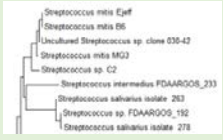
Different band pattern (B) – different strains

1254 RAPD in a 2nd stage to confirm equal strains.





BLAST ® » blastn suite



5 **Analysis and Identification of isolates**

ABI chromatogram of each isolate sequence was opened in **BioEdit** for quality analysis.

Sequences were compared with the 16S rRNA sequence database using **nucleotide BLAST** and RDP. Identification to the species level (top BLAST hit with maximum sequence similarity ~99%).

Phylogenetic trees: generated in MEGA 6 and MEGA X by using the Neighbor-Joining method with 1000 bootstrapping and edited in iTOL.

Figure 2.2 – Schematic of the experimental procedures (2nd part)

Chapter 3

Results and Discussion

3.1 Bacteria and biofilm presence in CRC

3.1.1 Tissue morphology (H&E staining) and identification of local immune responses (IF) in CRC

In the first part of this project, histological evaluation by Haematoxylin & Eosin (H&E) staining was performed in sections of CRC biopsies, in order to detect malignant areas with excessive big nucleus and necrotic zones with already dead tissue (an example is showed in Figure 3.1.A for patient CRC031). This is the usual procedure to be done by pathologists after a biopsy collection to confirm the presence or not of cancer and to evaluate its stage in positive cases. Haematoxylin is a basic dye which is positively charged and therefore interacts with nucleic acids. It gives a blue/purple colour to the slides. Eosin is an acidic dye negatively charged which interacts with amino groups in the cytoplasm, including proteins, cytoplasmic filaments, intracellular membranes. It is responsible for the pink staining in the slides.

Simultaneously, identification of local immune responses by CD45 IF (Figure 3.1.B) was performed in identical slides of the same section which allowed to detect agglomeration of lymphocytes in some parts of the tissue, indicative of inflammation. In order to perceive the morphology of the analysed CRC slides, a scheme of a colonic section with identification of structures and layers is showed in Figure 3.1.C, with special attention to mucosa. The mucus layers of colon have been ignored in histological analysis, since mucus is highly transparent and highly hydrated, which means that is very sensitive to regular fixation methods which dry the mucus and consequently promote its collapse (Johansson and Hansson, 2012). Here, fixation of CRC sample tissues was done in Methacarn (methanol-Carnoy) (instead of the typically used formaldehyde), which doesn't dehydrate the mucus, allowing us to obtain a thick mucus layer attached to the epithelia where we can search for the presence of bacteria (Figure 3.2.B and 3.2.C).

We saw previously that the normal intestinal microbiota inhabits the colon in a state of homeostasis without triggering inflammatory responses or inciting changes in the tissue biology, essentially because there is an anatomical separation between bacteria and the host epithelial layer – the mucus layer. Mucus is a highly regenerative and protective lubricant glycoprotein sheet (lubricates the transit of intestinal contents and is itself a microbial carbon source) which is secreted by the intestinal goblet cells (Figure 3.1.C). In fact, the mucus is organized in two layers - an inner layer firmly adherent to the epithelium and an outer non-attached layer

(Johansson *et al.*, 2011). While the outer layer contains intestinal bacteria, the inner layer is considered sterile. This layer is dense and doesn't allow bacteria to penetrate, keeping the epithelial cell surface isolated from them (Li *et al.*, 2015). Applying Gram staining and FISH in CRC slides (ON Tumour, OFF and UD samples), it was several times detected the presence of bacteria in this inner layer, the direct contact with epithelium and inclusively the invasion of this barrier.

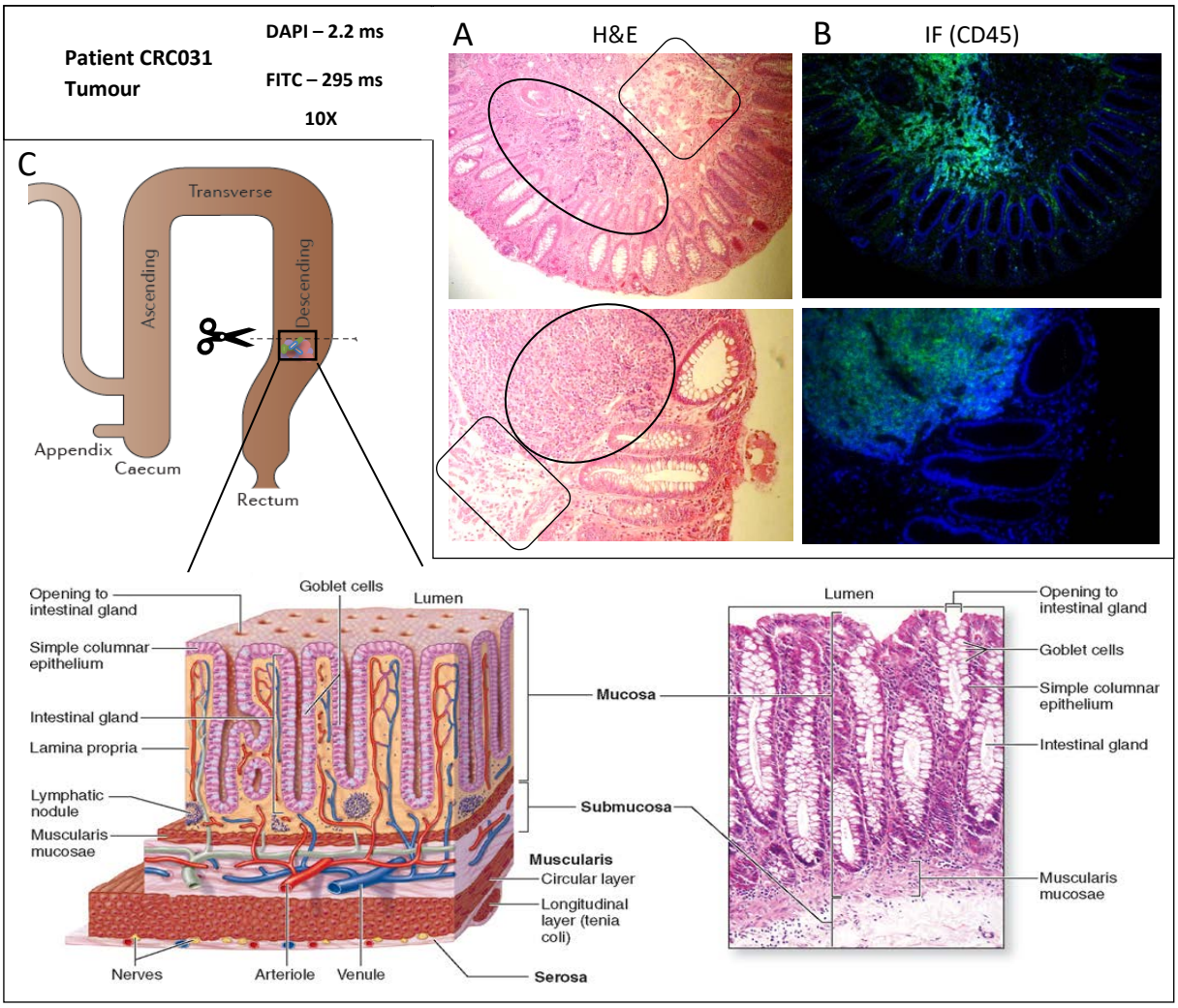


Figure 3.1 – Example of histological evaluation (H&E staining) and identification of local immune responses (IF) on a CRC sample section - CRC031 Tumour sample. A - H&E images (10X). Hematoxylin stains cell nuclei purple, while Eosin stains cytoplasm and connective tissue pink. The circles demarcate already malignant parts with bigger nucleus and the squares the disorganized and acellular areas without defined nucleus. B – IF images (10X). CD45 marker identifies the presence of lymphocytes (green areas). The section was counterstained with the nuclear stain DAPI (blue). Exposure times for DAPI and FITC were indicated. C – Structure of a colon section with closeup of mucosa and submucosa areas in an H&E-stained section (Adapted from (Mescher, 2013)). The mucus layer situated above the epithelial layer and in direct contact with the lumen was not represented in the figure. See supplementary Figure D.3 (CRC042 UDP sample) for colonic structure clarification.

3.1.2 Bacterial distribution in the colonic tissue (Gram staining and FISH) - adherence and invasion

Although Gram staining is an old and basic staining method for bacteria detection, and maybe that's why articles focusing microbiota-CRC interaction do not use it, preferring more sophisticated methods such as FISH and sequencing analysis, it allows a good detailed visualization and distinction of the several colonic layers and therefore the localization and/or invasion of bacteria in or within those structures. Thus, it is an easy method to access rapidly and to validate in a first instance the presence of bacteria in colonic tissue of CRC patients and their distribution. If Gram staining gives positive results for bacteria detection, FISH can be performed in a second stage in order to confirm this presence and to identify specific groups of bacteria (by using targeted phylum and family probes). This was the approach followed.

Indeed, Gram-stained CRC slides allowed to validate bacteria presence in all samples and to establish a distribution pattern more or less coherent presented by the seven CRC patients analysed (CRC042 bacterial distribution and density pattern in Figure 3.2 is used as model example). A selection of images from 3 of the 7 patients (CRC038, CRC042 and CRC057) is presented (Figure 3.3). In ON Tumour samples, a co-occurrence of bacteria is usually visualized inside the tumoral mass (Figure 3.3.A). In OFF samples, a dense community of bacteria which resembles an intricated bacterial biofilm is detected in the mucus layer, in close contact with epithelial tissue and sometimes even infiltrated in it (Figure 3.3.B). Finally, in UD samples, the results are more discordant. Some patients present a clear mucus layer punctuated by disperse bacteria well separated from mucosa, while others present a high abundance of agglomerated bacteria colonizing the mucus, resembling the OFF typical distribution (Figure 3.3.C). Supplementary Figures D.1 to D.4 show images of the analysed Gram-stained slides with different ampliations for more detailed perception of bacteria location within the tissue.

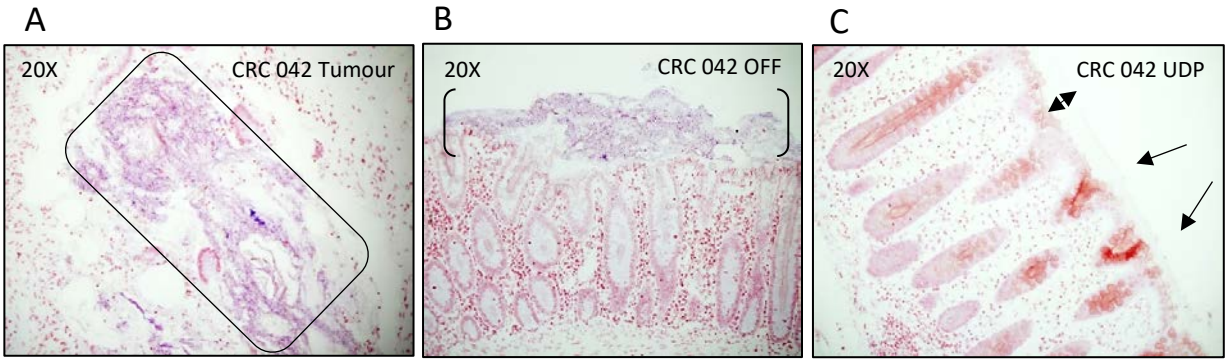


Figure 3.2 – Detection and usual distribution pattern of bacteria in Gram-stained sections (20X) of a CRC patient - CRC042 ON Tumour (A), OFF (B) and UDP (C). Purple-stained and red-stained Rods and Cocci represent Gram-positive and Gram-negative bacteria, respectively. A – In the middle of the necrotic tumour section, there is a massive invasion and accumulation of bacteria (rectangle) intercalated with dead tissue and malignant cells. B – Around 3 cm away from the tumour mass (OFF), the same patient presents a high density of bacteria infiltrated in the mucus layer (demarcated by the brackets), with the bacteria in the bottom directly contacting the colonic epithelial cells. C – Around 10 cm from the tumour, this UDP section presents a well-defined mucus layer (denoted by the double arrow) punctuated by Gram-positive bacteria (black arrows) presenting a disperse distribution without direct contact with epithelial cells. Single columnar epithelium, colonic crypts, Goblet cells, intestinal glands, lamina propria and muscularis mucosae are easy to visualize in sections B and C.

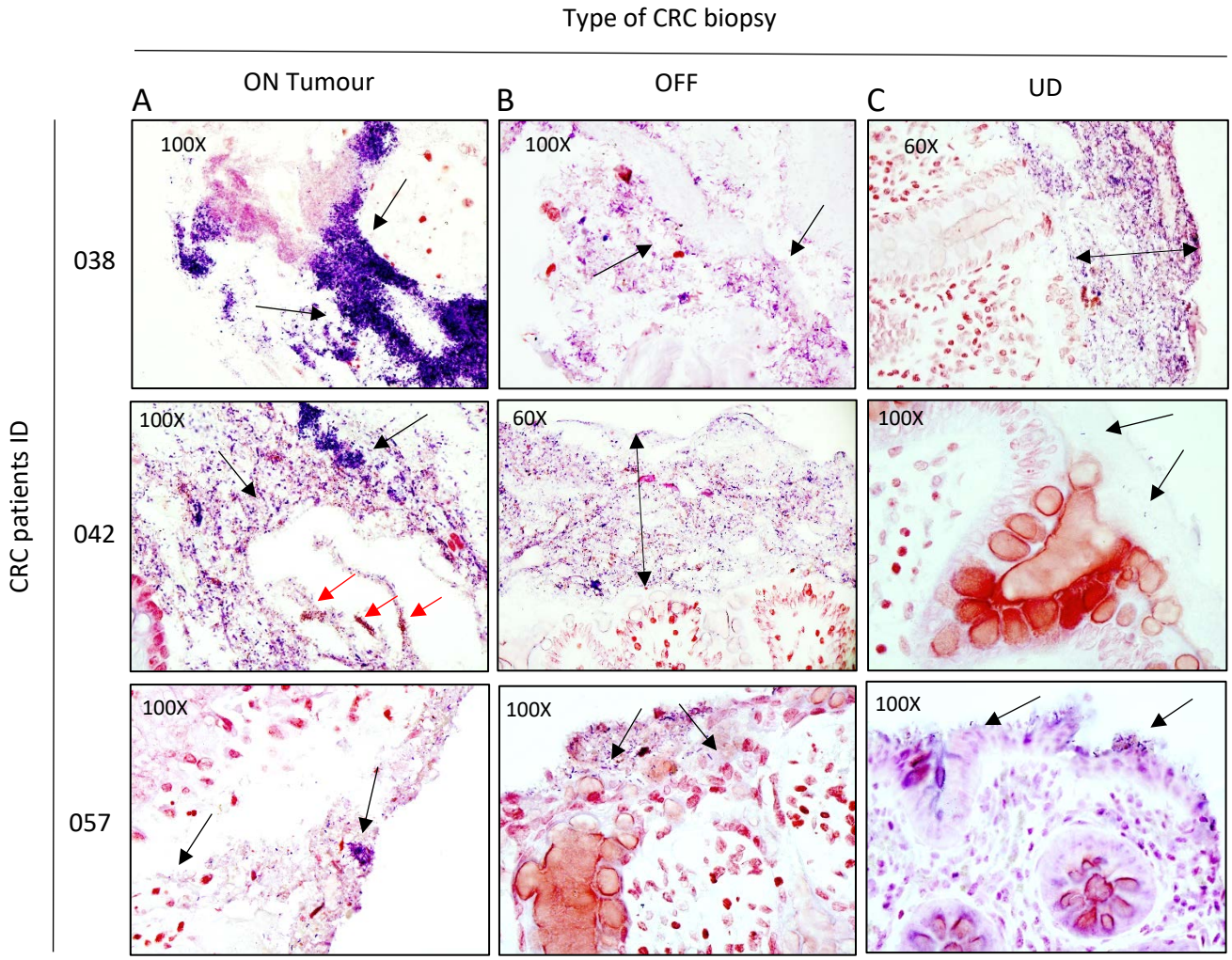


Figure 3.3 – Validation of bacteria presence and identification of bacteria distribution in colonic tissue (Tumour (A), OFF (B) and UD (C)) of CRC patients (CRC038, CRC042 and CRC 057). A - In tumour sections, the tissue presents a falling apart and disorganized structure, with big nucleus and with some acellular areas, however, some cells are still more or less intact and the glandular structure is still preserved. In all patients, a high accumulation of bacteria is found inside the tumour, with highlight to patient CRC038 which possesses the densest microbial community. B – The mucus layer is populated by a considerable number of bacteria (CRC038 and CRC042). In patient CRC057, direct contact with the defective epithelial layer and invasion into the lamina propria is clearly visualized. In patient CRC038, bacteria were localized in mucosa and submucosa. C – With exception of patient CRC038, which presents a high abundance of bacteria colonizing its colonic mucus layer, usually in UD samples, single bacteria distributed along this layer are visualized. The lens ampliation used is indicated on the top of each photo. Double and normal arrows demarcate the localization of single bacteria or groups of them. Although Gram-positive bacteria seem to be present in more abundance (purple-stained), Gram-negative bacteria are also in great number in these slides (see the agglomeration of red bacteria pointed by the red arrows in section CRC042 Tumour), being more difficult to detect due to similar tissue colour.

Previous studies have demonstrated that colonoscopy biopsies from healthy individuals generally present a mucus layer devoid of bacteria (less than 7% of biopsies), with no adherence or invasion into mucosa region (Dejea *et al.*, 2014), (Swidsinski *et al.*, 2007). Although the thickness of the mucus layer may vary, the 30 µm mucus zone close to mucosa is germ-free (Swidsinski *et al.*, 2007). From the seven patients analysed previously,

CRC038 and CRC057 patients, whose samples presented a high abundance of bacteria, were selected to perform FISH, using one universal probe (Eub338) for the dominium Bacteria and a nonsense probe (Non338) as negative control. Prepared bacterial cultures of MCC354 *Clostridum symbiosum*, MCC266 *Bifidobacterium longum*, MCC505 *E.coli* (Figure 3.4.D) and MCC110 *Bacteroides fragilis* were used for FISH validation. FISH results confirmed biofilm presence and invasion of the epithelial barrier in UD sections of these two patients (example of CRC057 is showed in Figure 3.4.A); and presence of these bacterial networks inside the tumoral formations (ON Tumour sections) (Figure 3.4.B and 3.4.C).

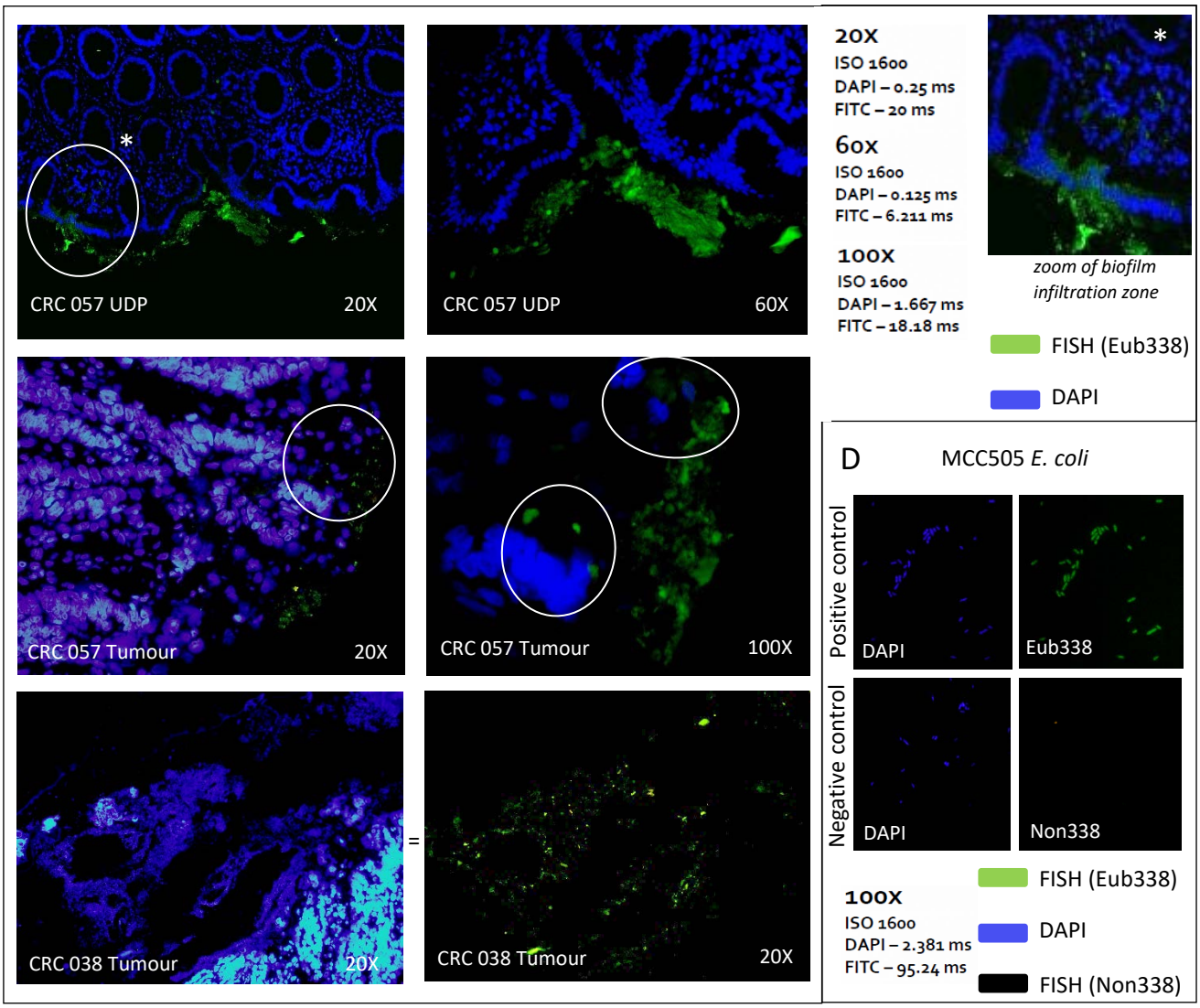


Figure 3.4 – Detection of bacterial biofilms on colon tumours and nearby tissues. FISH of all bacteria (green) on cancers (B and C) and UDP section (A). Sections were counterstained with DAPI (blue). A – A long bacterial biofilm can be visualized in the UDP sample of patient CRC057, which presents close contact with epithelial cells and even penetration and invasion into lamina propria in one zone (white circle). B – Close contact between bacteria and cancer cells was found in CRC057 Tumour section (white circles). C – In CRC038 Tumour section, high abundance of bacteria was confirmed inside the tumour tissue (both photos represent the exact same region of the CRC). D – Positive and negative control performed in a prepared bacterial culture of MCC505 *E.coli* to confirm feasibility of the FISH procedure. Lens ampliation, resolution (ISO) and exposure times for the captured images are indicated above, respectively. A and B images were merged using Image J software.

Biofilms are associated with several chronic mucosal diseases outside the colon such as cystic fibrosis pneumonia, infective endocarditis, rhinosinusitis, otitis, urethritis and vaginitis (Hall-Stoodley *et al.*, 2004), and inside such as inflammatory bowel disease (Dejea *et al.*, 2014). Biofilms able to invade the colonic mucus layer and to contact directly the epithelial cells are indicative of a pathologic state. Indeed, biofilm formation is associated with reduced epithelial E-cadherin which increases epithelial permeability and therefore enhances direct access of bacteria and their antigens/mutagens to a vulnerable and unprotected epithelial surface, which by their turn promote tissue inflammation and enhancement of cellular proliferation (a basic feature of oncogenic transformation) (Dejea *et al.*, 2014).

Although, there is a group of well-studied single bacteria known to promote chronic inflammation in CRC and although we know that a feature of CRC is a dysbiotic microbiota, bacterial biofilms (a driver of tissue inflammation) have not been reported/investigated in CRC. The biofilm presence in Gram stained and FISH CRC slides detected here is in accordance with a recent study that showed for the first time that these biofilms can be intrinsically associated with CRCs, by analysing colon tissues with FISH probes (Dejea *et al.*, 2014). In addition to this article, Gram staining allowed us to detect dense aggregations of bacteria communities not only in the periphery of CRC tissues (mucus layer, in near contact and invading the epithelial layers), but also in the middle of CRC ON Tumour samples.

3.2 Culture and identification of isolates from faecal, oral swab and tumour samples of CRC patients

3.2.1 Taxonomy of isolated species

After showing positive association of bacterial biofilms with CRC samples and the presence of bacteria inside CRC tumours, we can now move on to the second part of this project. In this part, bacteria from oral swab and colonic tumour samples from 2 CRC patients (patient 1 and patient 2) were isolated and identified in an attempt to find if the previous bacteria detected inside and around CRC tumours have in part an oral origin. In order to practice the anaerobic culture, isolation, purification and identification process and in order to define the best culture conditions (presented in Methods) to be tested in patient 1 and 2, faecal samples of patient CRC038 and patient CRC057 were used as pilot study culture specimens. Note that previous evaluation of their tissue sections allowed to detect a high abundance and invasion of bacteria in Tumour, OFF and UD samples (Figure 3.3 and 3.4). After obtaining the isolates from the four patients, extraction of gDNA, PCR protocol (amplifying full-length 16S rRNA gene) (supplementary Figure B.1) and purification, Sanger sequencing and sequence analysis with subsequent species identification were performed. To infer evolutionary relationships of taxa, the different species (and respective families and phyla) isolated from each sample were organized in phylogenetic trees (Figure 3.5 to 3.7).

Remembering the well-studied single microbes associated with CRC carcinogenesis referred in Introduction (Table 1.1), isolates of all of them were recovered in this project, excepting *H. pylori* and *Streptococcus gallolyticus.* The oral pathogens *Fusobacterium* sp. were obtained from CRC038 (*Fusobacteriaceae bacterium*), CRC057 (*Fusobacterium necrophorum* subsp. *funduliforme*), and patient 1 oral swab (*Fusobacterium nucleatum* subsp. *Animalis*). Although there is a strong association of this genus with CRC, no Fusobacteria was obtained in CRC patient 1 and 2 tumour samples. Species belonging to the Enterobacteriaceae family (which present *pks* island) were recovered from all samples, excluding patient 1 oral swab. Within this category *Escherichia coli* was isolated from CRC038 (6 isolates), CRC057 (26), patient 1 tumour (4), patient 2 oral cavity (1), and patient 2 tumour (16); and *Klebsiella* sp. (1) and *Klebsiella pneumoniae* (4) were recovered from patient 2 tumour. The latter is an important strain highlighted by *Atarashi et al.* (2017), since it is a multi-drug resistant oral bacterium capable of inducing potent chronic intestinal inflammation. *Bacteroides fragilis* was recovered only from patient 1 tumour (56). *Enterococcus faecalis* was obtained from patient 2 tumour (2), patient 2 oral (1), and patient 1 tumour (4). Pure cultures belonging to Lactobacillaceae were recovered from all samples, with exception of patient 1 tumour sample; and isolates belonging to Bifidobacteriaceae from patient 2 and patient 1 tumour samples. These two families are associated with direct NOS production and, consequently, with genomic instability. They are also formed by acidogenic lactic acid species associated with caries development. *Streptococcus mutans* which presents a central role in caries formation was obtained in all samples with exception of patient 1 tumour sample (see Figure 3.5 to 3.7 and supplementary Tables C.1 to C.6).

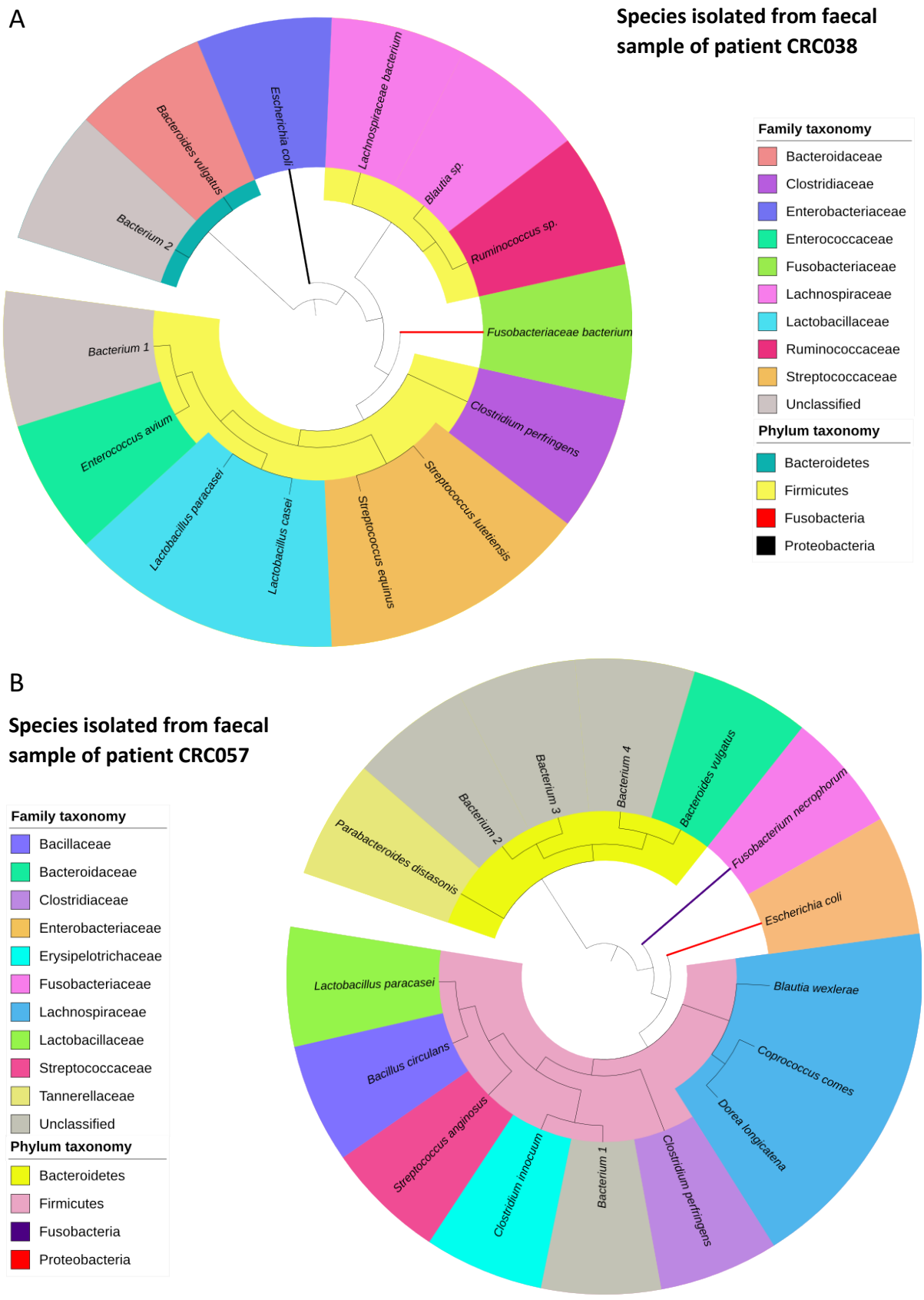


Figure 3.5 – Phylogeny of faecal isolates. Full-length 16S rRNA gene phylogeny illustrating the taxonomic relationship of species isolated and cultured from faecal samples of CRC038 (A) and CRC057 (B) patients. Coloured ranges and branches indicate distinct families and phyla.

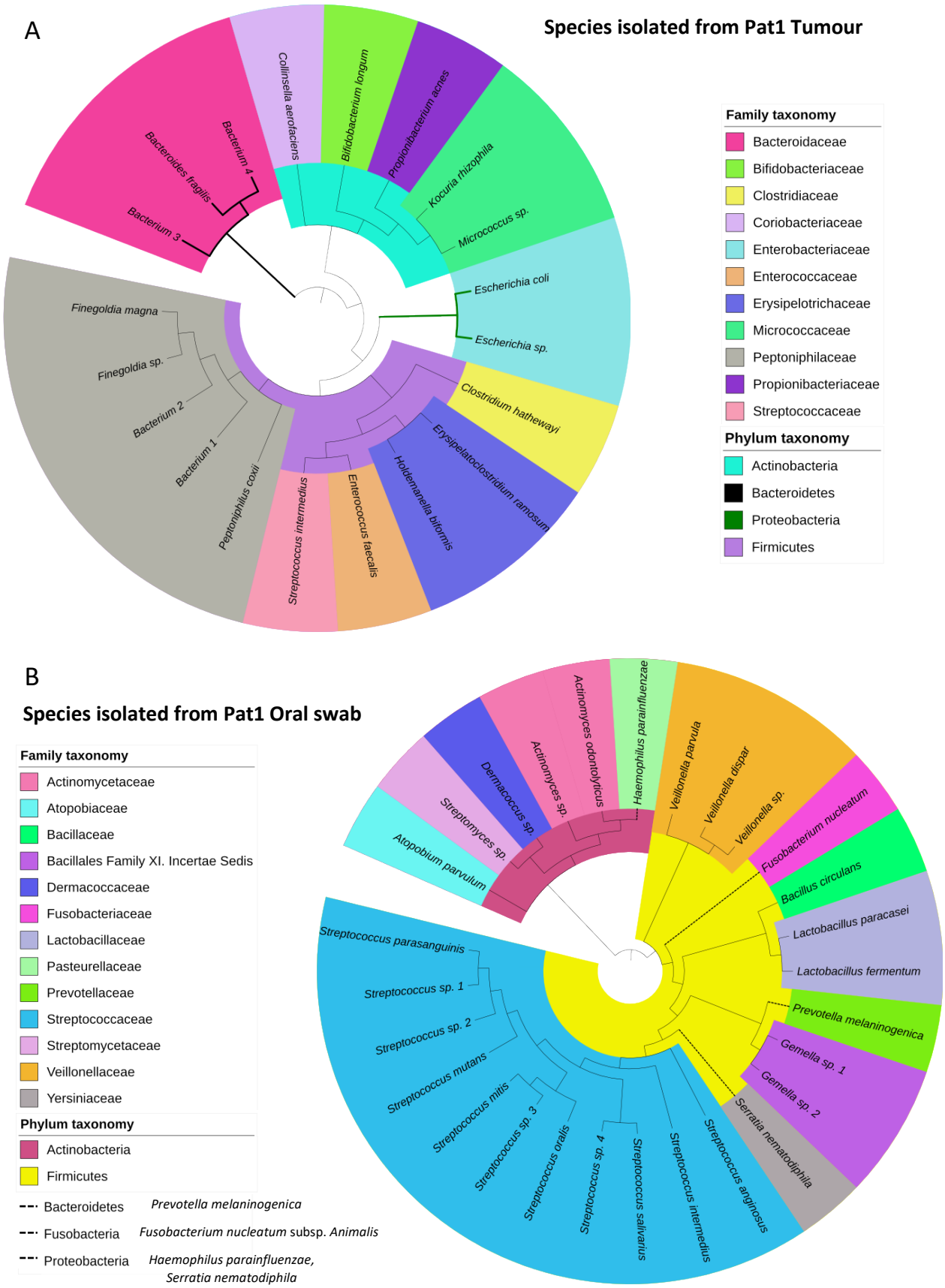


Figure 3.6 – Phylogeny of CRC patient 1 isolates. Full-length 16S rRNA gene phylogeny illustrating the taxonomic relationship of species isolated and cultured from paired CRC Tumour (A) and Oral swab (B) samples of patient 1. Coloured ranges and branches indicate distinct families and phyla.

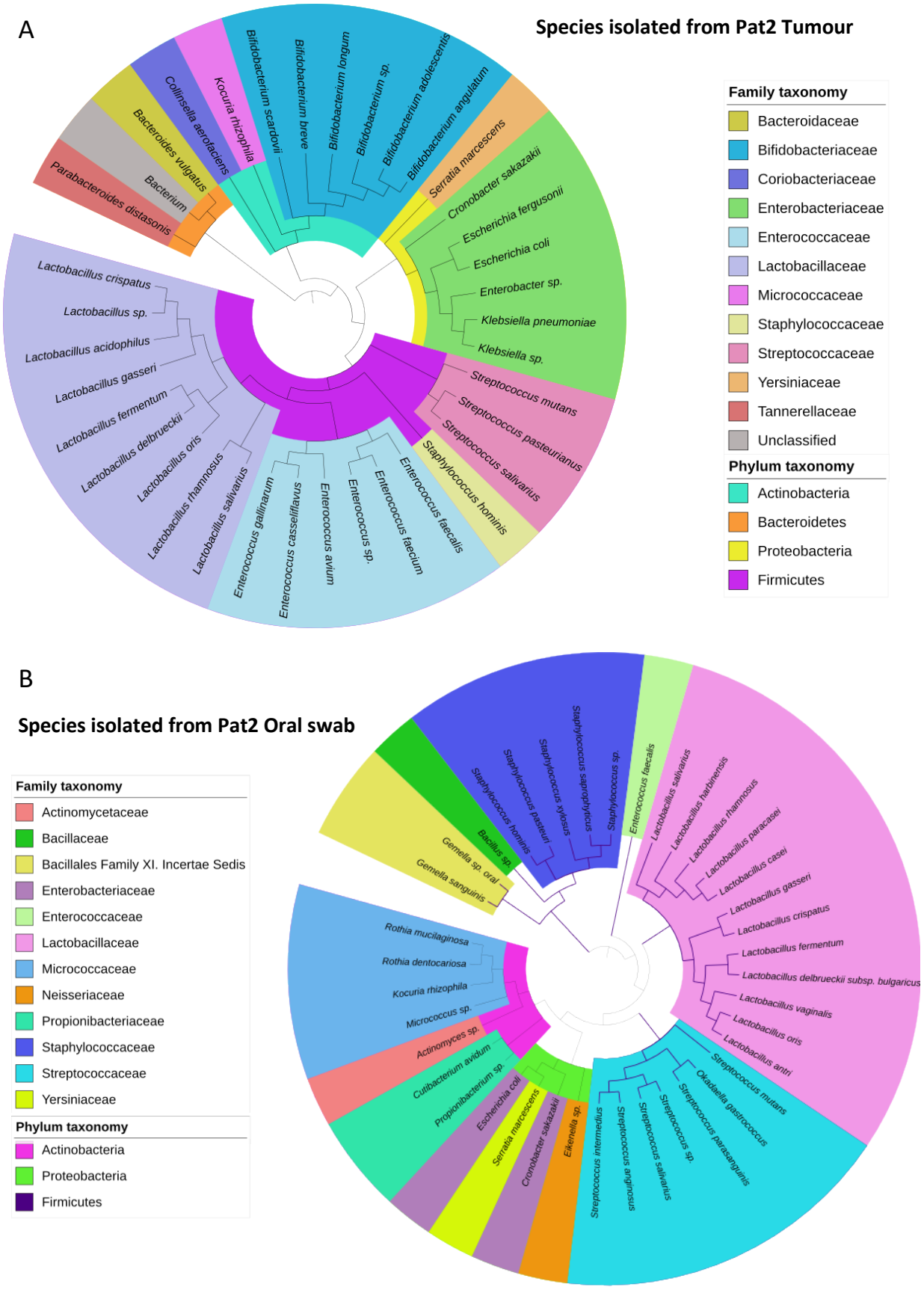


Figure 3.7 – Phylogeny of CRC patient 2 isolates. Full-length 16S rRNA gene phylogeny illustrating the taxonomic relationship of species isolated and cultured from paired CRC Tumour (A) and Oral swab (B) samples of patient 2. Coloured ranges and branches indicate distinct families and phyla.

Several unclassified bacteria were isolated (11 in total). The taxonomic evaluation (gey coloured ranges in Figures 3.5.A, 3.5.B, 3.6.A and 3.7.A) of these isolates allowed to discover at least the phylum category where they fit in.

Reminding that the ten most abundant bacterial genera across oral swab samples are *Streptococcus*, *Haemophilus*, *Neisseria*, *Prevotella*, *Fusobacterium*, *Veillonella*, *Leptotrichia*, *Rothia*, *Actinomyces* and *Porphyromonas* (Flemer, Warren, *et al.*, 2017), all genera were recovered from patient 1 and 2 oral cavities, with exception of *Leptotrichia* and *Porphyromonas*. Flemer, Warren, *et al.* showed that putatively oral bacteria were in over-abundance on CRC biopsies and polyps and identified two CAGs (groups of bacteria meaningfully more abundant on tumour samples than in healthy tissue). One is the oral pathogen CAG (oral pathogens associated with late colonization of oral biofilms and with diseases including CRC) and the other is the biofilm CAG (dominant bacteria in early dental biofilm formation). Bacteria from both CAGs were isolated in this project – *Fusobacterium* and *Gemella* in the former; and *Actinomyces*, *Haemophilus*, *Rothia*, *Streptococcus* and *Veilonella* spp. in the latter.

Isolates from the Lachnospiraceae family - which seems to present a protective role against CRC, by conferring colonization resistance to CRC-associated putative oral bacterial networks (Flemer, Warren, *et al.*, 2017) - were recovered from CRC038 and CRC057 faecal samples, including *Blautia* sp*., Blautia wexlerae, Lachnospiraceae bacterium, Coprococcus comes and Dorea longicatena.* Additionally, *Bifidobacterium breve* and *Lactobacillus rhamnosus,* previously described as good candidates in CRC prevention (Gagnière *et al.*, 2016), were recovered from patient 2 tumour and oral samples, respectively.

The group of isolated species represented in each phylogenetic tree (Figure 3.5 to 3.7), from the six studied samples, constitute a defined, phylogenetic diverse artificial faecal, oral and tumour community collection.

3.2.2 Qualitative and quantitative analysis of recovered isolates

For easier interpretation, a summary of the different phyla, families, species, apparent strains and number of isolates for each patient sample is presented in supplementary Tables C.1 to C.6, and supplementary Figure C.1. The most recovered species were *Enterococcus avium* (6 in 34 isolates) and *E. coli* (6 in 34) in CRC038 faecal sample; *E. coli* (26 in 41) in CRC057 faecal sample; *Streptococcus salivarius* (12 in 74) and *Veillonella parvula* (10 in 74) in patient 1 oral swab sample; *Bacteroides fragilis* (56 in 90) in patient 1 tumour sample; *Steptococcus salivarius* (31 in 135) from the oral cavity of patient 2; and *E.coli* (16 in 126), *Lactobacillus salivarius* (15 in 126) and *Streptococcus pasteurianus* (14 in 126) in tumour sample of patient 2 (supplementary Figure C.1). Important to note that a quantitative analysis of isolates recovered from the four patients can be performed, although in any step it reflects the real abundance of these bacteria in the original collection sites. Note that the culture conditions and the selection of colonies is a randomized process where the main goal was to re-streak and purify colonies, in first place, with similar morphology between oral swab and tumour plates, and, in second place, to isolate the most diverse possible panel of bacterial species. Original phylogenetic trees plotted by species and by apparent strains (only one of the many isolates for each apparent strain is presented) with node statistics and original branch lengths constitute supplementary Figure C.2 to C.6.

Heading to phylum and family distribution, classification, percentages and numbers for each sample were summarized in Figure 3.8. Firmicutes accounted for up to 71%, 77%, 85% and 60% of the total of isolates obtained from CRC038 faecal sample, patient 1 oral swab, patient 2 oral swab and patient 2 tumour sample, respectively, being the best represented phylum in quantitative terms (4 of 6 samples). In CRC057 faecal sample, 63% of the bacteria recovered belonged to phylum Proteobacteria. In the case of patient 1 tumour sample, the most abundant phylum corresponded to Bacteroidetes with 62% of the isolates belonging to this taxonomic category (Figure 3.8.A).

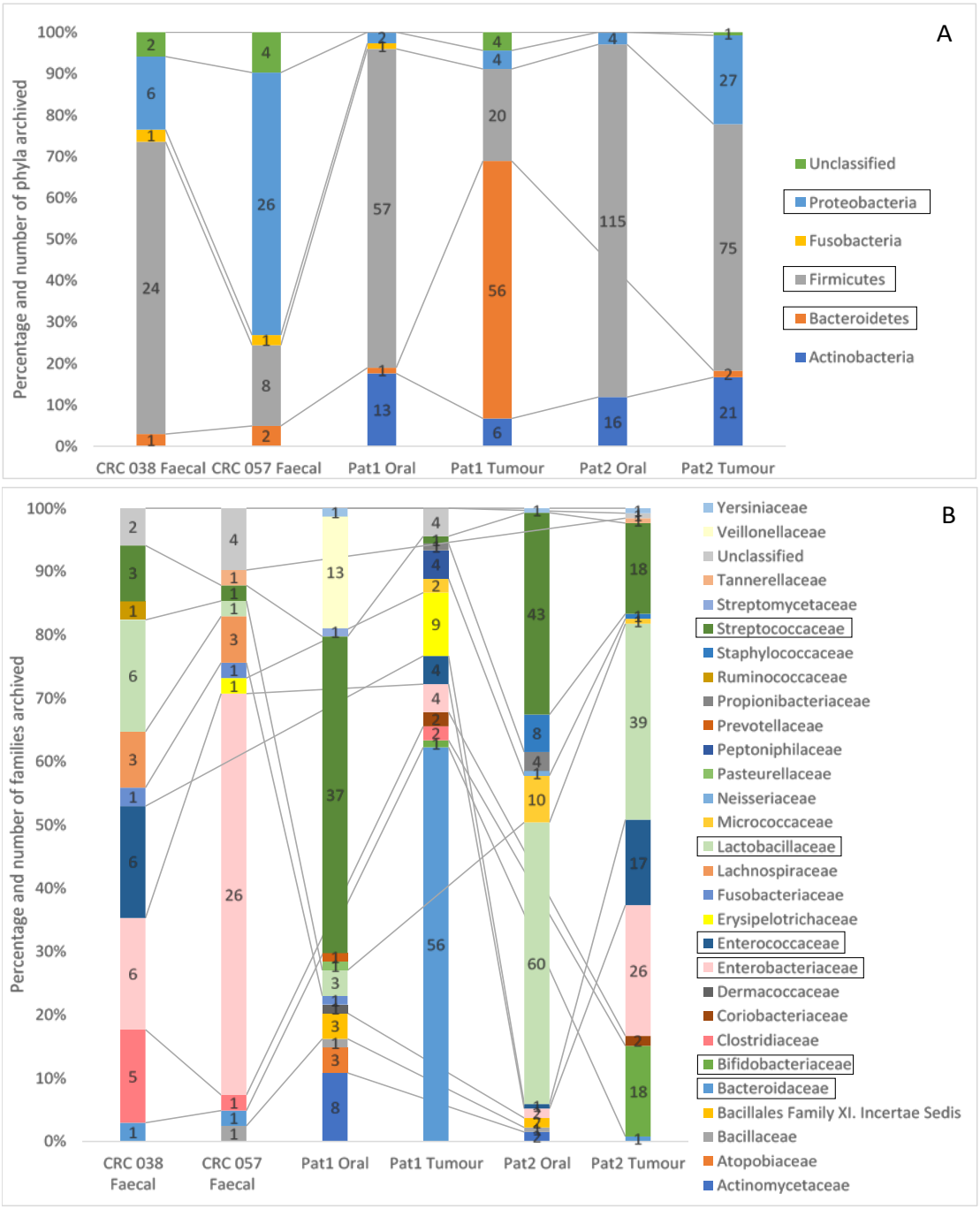


Figure 3.8 – Phylum and family quantitative distribution of the isolates recovered from the four CRC patients (CRC038, CRC057, patient 1 and patient 2). Stacked column charts with percentage and number of isolated phyla (A) and families (B) in each CRC sample (Stool, Oral swab and Tumour) are presented. The most abundant categories are highlighted by rectangles. The grey lines connect equal phylum/family categories between samples to facilitate comparison.

The most represented families by sample were: Enterococcaceae (18%), Enterobacteriaceae (18%) and Lactobacillaceae (18%) in CRC038; Enterobacteriaceae (63%) in CRC057; Streptococcaceae (50%) in patient 1 oral swab; Bacteroidaceae (62%) in patient 1 tumour; Lactobacillaceae in patient 2 oral swab (44%) and tumour (31%) samples (Figure 3.8.B).

3.2.3 Overview and detection of equal species in both oral swab and tumour samples of the same CRC patient

The isolation, purification and identification process done in this project allowed to create a wide collection of isolates from faecal, oral and tumour samples of 4 CRC patients. In total, 500 pure cultures were archived belonging to 95 different species, 28 families and 5 phyla (Figure 3.9.B). The number of isolates recovered from each sample varied, being higher in the case of patient 1 and 2 (Figure 3.9.A), since the objective here was not only to obtain interesting different species to form an artificial diverse collection for each body site (high number of different culture conditions used in all samples), but also to compare oral and CRC colonies for equal identity interrogation. This way, an effort was done to pick the highest number of different colonies from the culture medium plates (with special attention to similar colonies between paired oral and tumour specimens), re-streak them, purify them, archive them and processed them until the obtention of the Sanger sequencing results and their analysis. Important to highlight that several different species were isolated, but, within these, more apparent different strains were recovered (Figure 3.9.A). This is also proved, in the following section, by the different AP-PCR patterns obtained within the same species which indicate different strains identity.

After identification of species classification, equal species both present in oral and tumour samples of the same CRC patient were detected (Figure 3.9.C). In patient 1, only one paired species was identified (*Streptococcus intermedius*). The low amount of positive matches in this patient can be due to chemotherapy dysbiotic influence in the microbiota (Montassier *et al.*, 2015), since patient 1 performed this treatment during the last 2 years, or simply because the culture process failed to recover more potential equal isolates. Regarding patient 2, the results were more promising with 15 equal species detected in both oral and tumour specimens. *Enterococcus faecalis* and *E.coli* (single pathogens associated with CRC); *Streptococcus mutans, Streptococcus salivarius* and more seven species of *Lactobacillus* genus (bacteria linked with caries development (Becker *et al.*, 2002), (Aas *et al.*, 2008)); *Kocuria rhizophila*, *Cronobacter sakazaki, Serratia marcescens* and *Staphylococcus hominis* constitute this list (Figure 3.9.C). At this point, AP-PCR was performed to compare the genomic structure of oral and colonic isolates of the same species and, this way, to search for equal paired strains.

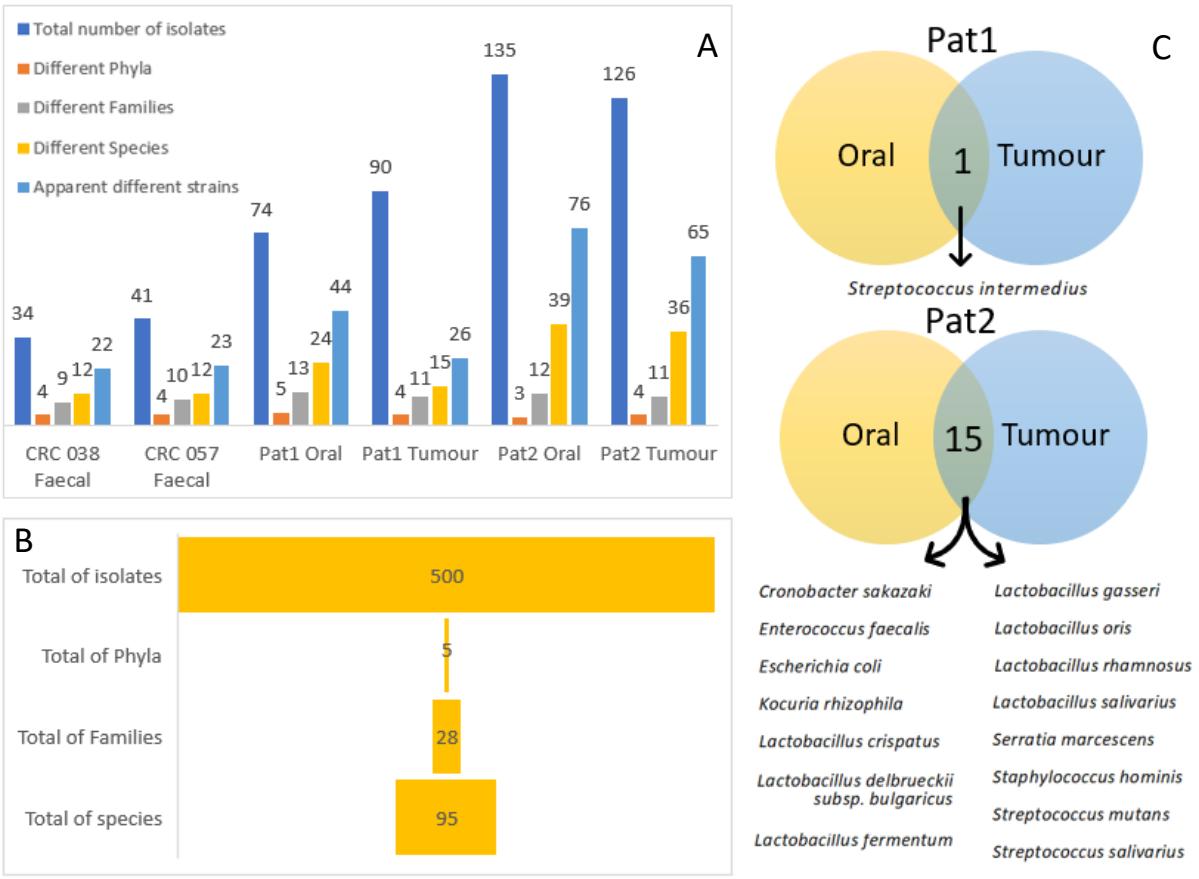


Figure 3.9 – Quantitative overview of the pure cultures isolated from the four CRC patients. A – Bar charts pointing out the total number of isolates, different phyla, families, species and apparent strains recovered from culture of patients CRC038 and CRC 057 Faecal samples, patient 1 and patient 2 Oral swab and Tumour samples. B – Total number of isolates, different phyla, families and species obtained in this project. C – Number of positive matches between Oral swab and Tumour samples from patient 1 and patient 2 with indication of respective paired isolated species. The term "apparent" different strains is applied, because nucleotide BLAST in NCBI, although indicates the closest relative strain in database for each isolated species, it is only feasible up to the subspecies level in terms of full-length 16S rRNA gene sequence analysis context. Identification at the strain level is only achieved by the AP-PCR strain typing method.

3.3 CRC patients have identical strains in their oral cavity and CRC tumour

3.3.1 AP-PCR patterns and paired strains identification

Evaluation of AP-PCR patterns - M13 AP-PCR (Figure 3.10.A) and 1254 AP-PCR (Figure 3.10.B) - allowed to identify equal isolates from oral and tumour samples of CRC patients at the strain level. All paired species identified before and tested here presented at least one identical strain, with exception of *Kocuria rhizophila* (patient 2). These results allow to conclude that oral bacteria are able to colonize CRC tumours – one positive result in patient 1 and 14 in patient 2. Comparing both AP-PCR gels, *Staphylococcus hominis* is the only strain susceptible of pattern similarity doubt. Although *Lactobacillus delbrueckii* and *Streptoccoccus intermedius* present patterns in

1254 gel that seem more diffuse in one sample compared to the other (Figure 3.10.B), they present clear equal patterns in M13 gel that prove their strains equality. Some differences in brightness of patterns can be due to small fluctuations in the volume dispensed in each well.

The oral strains discovered in the CRC tumours represent new species within the oral colonization context in CRC and their role in CRC tumorigenesis can now be studied, since pure cultures were archived. Indeed, *Streptococcus* strains were the only obtained genus referred in the list of putative oral tumour-associated CAGs established by Flemer, Warren, *et al.* (2017) in CRC. Additionally, it was proven before that an oral strain of *Escherichia coli* was able to colonize the intestine of germ-free mice (Atarashi *et al.*, 2017). Here, we prove that on tumours of CRC patients.

This is the first time, to our knowledge, that oral colonization in CRC was proven with a considerable range of isolated species - 15 species in total belonging to 6 different families and 2 different phyla. *E.coli* and *Cronobacter sakazaki* are members of the Enterobacteriaceae family and together with *Serratia marcescens* (Yersiniaceae family from Enterobacterales order) form the isolates of Proteobacteria phylum. *Staphylococcus hominis* (Staphylococcaceae), *Enterococcus faecalis* (Enterococcaceae), *Streptococcus mutans, Streptococcus salivarius* and *Streptococcus intermedius* (Streptococcaceae), and the seven *Lactobacillus* spp. (Lactobacillaceae) belong to the Firmicutes phylum.

The final list of isolated identical strains in paired oral and tumour samples of the 2 CRC patients is presented in Table 3.1. Growth conditions and initial isolation media which allowed to obtain these isolates are also indicated.

Looking at the panel of conditions tested in this project, some considerations can be done in order to facilitate future isolation work in CRC oral and gut context. Anoxic YCFA is recommended to isolate strict anaerobic bacteria and allows to obtain a wide range of species (if the goal is to create a representative artificial colonic and oral collection). The use of Brain Heart and Tryptic Soy agar is highly recommended, either to isolate oral or gut bacteria, and can substitute the majority of the other selective media. The addition of blood to culture medium formulation, pre-incubation in blood and in other liquid media prior to incubation, and usage of microaerophilic conditions, allowed targeting the majority of the strains both present in oral and tumour CRC samples.

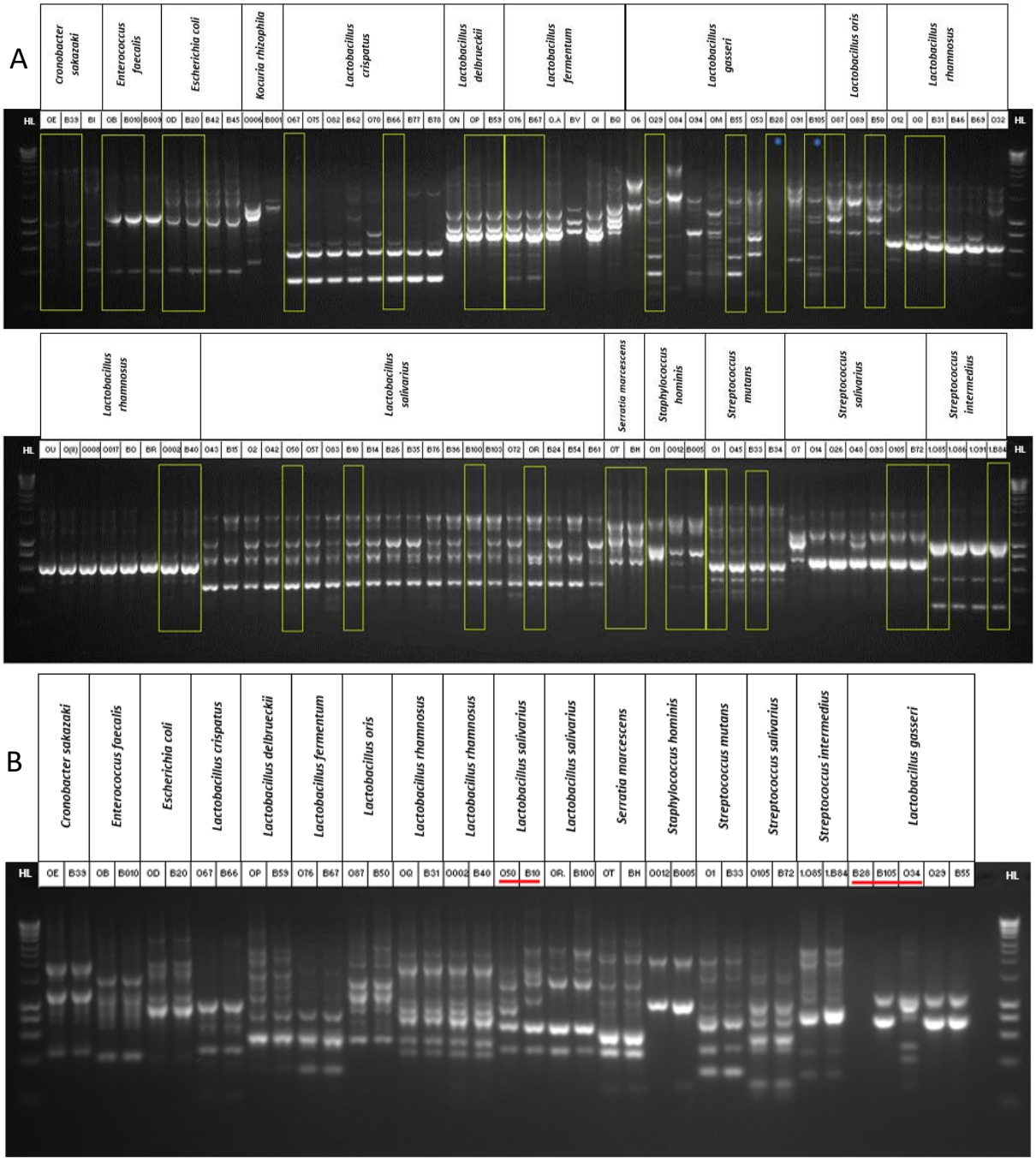


Figure 3.10 – Agarose gel electrophoresis showing the detection of equal strains in paired CRC tumour and oral swab samples. A - AP-PCR patterns detected with M13 primer to access if there are equal strains within the isolated species previously confirmed to be both present in oral swab and tumour samples of the same CRC patient (patient 1 and patient 2). Within these species, a selection of isolates was selected as target of the AP-PCR, giving primacy to the apparent identical strains outlined by NCBI during the sequencing data analysis. Each tested species is indicated in the white boxes above the isolates' codes, where O applies to Oral swab isolates and B to Biopsy Tumour isolates (see example in Figure 2.1 in Methods section). The identical pairs selected for subsequent confirmation of equal pattern profile are highlighted by the green rectangles (excepting the two rectangles with a blue dot which don't present equal AP-PCR patterns but were tested with an extra isolate in the following confirmation AP-PCR). *Streptococcus intermedius* is the only species tested from patient 1, all the others belong to patient 2. All species tested presented at least one identical strain, with exception of *Kocuria rhizophila* (patient 2). B - AP-PCR patterns detected with 1254 primer to confirm the previous identified positive identical strains from both specimens. Equal patterns were detected in all of them (15 species), with exception of the isolated strains highlighted by the red lines. HL, HyperLadder 1Kb (10037, 8000, 6000, 5000, 4000, 3000, 2500, 2000, 1500/1517, 1000, 800, 600, 400 and 200 bp).

Table 3.1 – Final list of species that present equal strains in tumour and oral swab samples of 2 CRC patients, identified by AP-PCR patterns comparison. Only *Streptococcus intermedius* strain belongs to patient 1. All the others were recovered from patient 2. Respective growth conditions and initial media used for isolation purposes are indicated. Single colonies were always re-streaked and purified in YCFA, Brain Heart agar or Tryptic Soy agar solid medium.

Identical strains found in CRC tumour and oral swab samples	Growth conditions and initial isolation media
Cronobacter sakazaki	• Microaerophilic pre-incubation in Tryptic Soy liquid medium for 4 days, Microaerophilic growth, 37 °C, Tryptic Soy agar + 5% sheep blood • Microaerophilic (5% CO2), 37 °C, Tryptic Soy agar
Enterococcus faecalis	• Microaerophilic pre-incubation in each liquid medium for 4 days, Microaerophilic growth, 37 °C, Tryptic Soy agar + 5% sheep blood • Anaerobic pre-incubation in blood for 17 days, Microaerophilic growth, 37 °C, Tryptic Soy agar + 5% horse blood
Escherichia coli	Microaerophilic, 37 °C, Brain Heart agar
Lactobacillus crispatus	Anaerobic, 37 °C, Todd Hewitt agar or Tryptic Soy agar
Lactobacillus delbrueckii* subsp. *bulgaricus	• Microaerophilic pre-incubation in Brain Heart liquid medium for 4 days, Microaerophilic growth, 37 °C, Brain Heart agar • Anaerobic, 37 °C, Sabouraud 2% glucose
Lactobacillus fermentum	Anaerobic, 37 °C, Tomato Juice agar
Lactobacillus gasseri	• Microaerophilic, 37 °C, Brain Heart agar • Anaerobic, 37 °C, Tryptic Soy agar
Lactobacillus oris	Anaerobic, 37 °C, Bile Esculin agar or Brain Heart agar
Lactobacillus rhamnosus	• Anaerobic pre-incubation in blood for 17 days, Microaerophilic, 37 °C, Tryptic Soy agar + 5% horse blood • Microaerophilic, 37 °C, Tryptic Soy agar
Lactobacillus salivarius	• Microaerophilic pre-incubation in M17 liquid medium for 4 days, Microaerophilic growth, 37 °C, M17 agar • Anaerobic, 37 °C, YCFA
Serratia marcescens	Microaerophilic pre-incubation in Todd Hewitt liquid medium for 4 days, Microaerophilic growth, 37 °C, Todd Hewitt agar
Staphylococcus hominis	Anaerobic pre-incubation in blood for 17 days, Microaerophilic growth, 37 °C, Brain Heart agar or Tryptic Soy agar + 5% horse blood
***Streptococcus intermedius* (Pat 1)**	• Anaerobic, 37 °C, Brain Heart agar • Microaerophilic, 37 °C, Brain Heart agar
Streptococcus mutans	• Microaerophilic, 37 °C, YCFA • Anaerobic pre-incubation in blood for 17 days, Microaerophilic, 37 °C, Tryptic Soy agar + 5% horse blood
Streptococcus salivarius	Anaerobic, 37 °C, YCFA or M17 agar

3.3.2 Characterization of the obtained identical paired species

As previously described, Enterobacteriaceae family members present the *pks* genotoxicity island where colibactin is encoded. The *pks* island carriers of these family are involved in inflammation pathways and cell proliferation in the CRC context (Schwabe and Jobin, 2013). *E. coli* belongs to this group of bacteria and its toxin-based mechanisms in CRC are well studied.

Regarding the isolate *Cronobacter sakazaki,* this pathogen causes infection mainly in premature infants (and also in hospitalized patients) and is associated with milk-based powdered formulas consumption (these formulas aren't sterilized unlike liquid formulas and a substantial percentage of neonates in neonatal intensive care units is fed with this milk-based formula). Bacteremia (presence of bacteria in the blood), sepsis (a severe infection that has spread via the bloodstream), meningitis and enterocolitis (inflammation of both the small intestine and the colon) have been reported in association with *Cronobacter sakazaki* (Smith and Garcia, 2011). This species can adhere to and invade brain endothelial cells and presents notorious antimicrobial resistance to ampicillin and cephalosporins. However, little is known about its virulence mechanisms (Donnenberg, 2014).

Along the same line, *Serratia marcescens* is an enteric pathogenic bacterium, also associated with cases of enterocolitis. This species produces several virulence factors such as fimbriae, and has been associated with a potent cytotoxin and an extracellular hemolysin that lyses blood cells (Ochieng *et al.*, 2014). Recently, it was demonstrated that *Serratia marcescens* elicits inflammation, cytotoxicity, adherence and invasion when it interacts with intestinal epithelial cells in culture (Ochieng *et al.*, 2014).

Enterococcus faecalis is a transient constituent of the oral microbiome and is the predominant human enterococcus (Komiyama *et al.*, 2016). *E. faecalis* CRC modulation mechanisms are also well studied regarding ROS production (Sears and Garrett, 2014), as explained in Introduction. Additionally, this bacterium presents resistance to several antibiotics, it has been related to oral diseases, such as caries and periodontitis, and presents the capacity to form biofilms (Komiyama *et al.*, 2016).

Staphylococcus hominis belongs to the coagulase-negative Staphylococci (CNS) group – a group recognised to be the major cause of infections in neonatal intensive care units such as sepsis (Chaves *et al.*, 2005). Although CNS group is mainly found in skin, *Staphylococcus* species are also present in the oral cavity (Smith *et al.*, 2001). This group produces a capsular polysaccharide adhesin which can be used in biofilm formation (Srinivasan and Evans, 2018).

Interestingly, 9 of the 15 species which proved to present identical strains, in tumour and oral cavity of CRC patients, are associated with dental caries and plaque. This can suggest a synergetic role on CRC tumorigenesis of these species, by similar pathogenic mechanisms already verified in the oral cavity. Several studies have shown that *Streptococcus mutans* is the main responsible species in caries pathogenesis, due to its ability to produce lactic acid (the acidic environment creates the risk for cavities). *Lactobacillus* species have also consistently been implicated as secondary pathogens in deep caries lesions, being also strong acid producers. The isolated

Lactobacillus fermentum, *Lactobacillus gasseri*, *Lactobacillus rhamnosus* and *Lactobacillus salivarius* are included in the dominant *Lactobacillus* species present in both adult and children caries. *Lactobacillus oris* is also dominant but in lesser prevalence, and *Lactobacillus crispatus* is less common (Caufield *et al.*, 2015). Regarding *Lactobacillus delbrueckii*, this was detected previously as being the predominant strictly homofermentative species on oral mucous surfaces (tongue surface and gums of children with tooth decay) (Badet and Thebaud, 2008). *Streptococcus salivarius* is also part of the community found in caries (Becker *et al.*, 2002). Adherence to surfaces, production of acidic metabolites, the build-up of glycogen reserves and the synthesis of extracellular polysaccharides are some of the capacities that enable these oral bacteria to colonize the dental surface and cause damage in tooth structure (Forssten *et al.*, 2010).

Streptococcus intermedius was the only species in paired samples of patient 1. This aerotolerant anaerobe belongs to the normal flora of the oral cavity and gastrointestinal tract and is associated with subgingival plaque. It comprises the *Streptococcus anginosus* group, being the most pathogenic within its 3 constituents (*S.anginosus, S.constellatus* and *S.intermedius*), because is able to cause abscesses by itself. *Streptococcus intermedius* produces hyaluronidase (an enzyme that breaks down host tissues) which seems to be associated with biofilm dispersion, and the cytotoxin intermedilysin which targets and lyses human cells (Masood *et al.*, 2016). *Streptococcus intermedius* is rarely implicated with colonic malignancy, unlike Group D Streptococci which includes the well-studied pathogen *Streptococcus gallolyticus,* referred in Introduction.

Reminding the recent literature concerning oral colonization of the gut, Komiya *et al* (2018) accurately proved that patients with CRC present identical strains of one species (*Fusobacterium nucleatum)* in their colorectal cancer and oral cavity. The present work, in addition, introduces previously unidentified strains within oral colonization of CRC tumour context. One *Fusobacterium nucleatum* isolate was recovered from the oral cavity of patient 1, but unfortunately no strains of this species were obtained from the paired tumour sample to prove or not their equal identity. *Porphyromonas gingivalis*, a periodontal pathogen like *Fusobacterium nucleatum*, and also usually captured by metagenomic analysis in association with CRC (Gallimidi *et al.*, 2015), was not isolated from the cultured samples.

We should remind that no selective isolation process targeting one specific species was performed, but rather a high number of different conditions in order to capture a high diversity of isolates, which is even more difficult if we are trying to find equal strains within paired plates filled of many diverse colonies. Indeed, isolation work is very randomized in picking different colonies – maybe *Porphyromonas gingivalis* wasn't picked and was in the plate; maybe it wasn't in the plate, because the culture conditions tested were not the preferential for the growth of that strain; or maybe it wasn't in the plate, because it wasn't in the oral sample. This can be demystified by looking into patient 1 and 2 metagenomic profiles. Since *Fusobacterium nucleatum and Porphyromonas gingivalis* are considered bacterial drivers and since the analysed biopsies are already fully developed CRC tumours, these bacteria were maybe in low abundance in the tumour by the time of the surgery (which is true mainly in oral swab of patient 1 and in both samples of patient 2 – see supplementary Figure E.1.A analysed in the next section).

3.4 Microbiota profile of CRC patients and success of the isolation process

In this last section, it is important to remind that additionally to the broad-range culture approach applied to the oral swab and tumour samples of CRC patients 1 and 2, a metagenomic approach (V3/V4 variable region of 16S rRNA gene - Illumina sequencing) was also performed in these 2 patients, in 39 CRC patients and 5 healthy controls, previously referred. Note again that these 39 CRC patients were divided into 2 groups for distinct analysis. 20 of them and the 5 healthy controls were used for microbiota profile comparison to validate that a dysbiotic microbiota is a feature of CRC (section 3.4.1). Remind also that the 7 CRC patients used in the histologic analysis also came from these group of CRC patients. The other 19 CRC patients were used for microbiota profile comparison with patient 1 and 2 (section 3.4.4). The success of the isolation protocol was evaluated through comparison with the Human Microbiome Project's 'Most Wanted' list of microbes (section 3.4.2), through analysis of the microbiota profile of oral swab and tumour samples of patient 1 and 2 (section 3.4.3), and through analysis of the selective power of the culture medium used in oral swab and tumour samples of patient 1 (also microbiota profiling) (section 3.4.5).

3.4.1 Microbiota profile of CRC patients *vs* healthy individuals

16S rRNA amplicon sequencing was performed in 5 healthy controls, corresponding to APC065, APC066, APC070 and APC071 patients; and performed in 20 CRC patients, including CRC038, CRC042, CRC044, CRC045, CRC056 and CRC057 patients, in order to obtain their microbiota profiles (see supplementary Figure E.1.B). In terms of phylum classification and relative abundance, healthy controls present similar results between ASC, DESC and TRANS samples within each patient and also a similar distribution and category patterns between patients. When we look to CRC patient profiles (faecal, Tumour, OFF and UD), there is more taxonomy variance between patients, although a high abundance of Firmicutes and Bacteroides is usually verified like in healthy controls. The absence of Fusobacteria in healthy controls and its high abundance in Tumour samples of CRC patients is in accordance with literature. No conserved CRC dysbiotic signature is verified across the CRC multiple samples.

3.4.2 Positive matches with the Human Microbiome Project's 'Most Wanted' list

A search was performed in order to detect if some isolated species in this project were present in the Human Microbiome Project's 'Most Wanted' list (*NIH Human Microbiome Project - Most Wanted Genomes*, 2018) of previously uncultured and unsequenced microbes (Figure 3.11). Comparisons with the Human Microbiome Project were carried out using 100% (perfect matches) and 99% (probable matches) sequence similarity of the 16S rRNA gene sequence. Chimeric sequences of the 'Most Wanted' list were removed for this analysis. In total (and without excluding repeated OTU), the pure collections created in this project contain 32 isolates with

perfect matches, 159 isolates with probable matches and 88 equal species (sequences that most likely belong to the same species) with this list. This is resumed to 10 perfect different matches, 35 probable different matches and 31 different equal species with the Human Microbiome Project's 'Most Wanted' list (see supplementary Figure E.1.C).

LIGHTrun barcode	Match	Priority	HMP_OTU	Taxonomy
20BE90	Probable match	LOW_PRIORITY	otu_451	Bacteria(100);Firmicutes(100);Bacilli(100);Lactobacillales(100);Lactobacillaceae(100);Lactobacillus(100)
20BE91	Probable match	LOW_PRIORITY	otu_26	Bacteria(100);Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Bacteroidaceae(100);Bacteroides(100)
20BE93	Perfect match	LOW_PRIORITY	otu_451	Bacteria(100);Firmicutes(100);Bacilli(100);Lactobacillales(100);Lactobacillaceae(100);Lactobacillus(100)
20BE94	Perfect match	LOW_PRIORITY	otu_167	Bacteria(100);Proteobacteria(100);Gammaproteobacteria(100);Enterobacteriales(100);Enterobacteriaceae(100);Escherichia/Shigella(92)
20BE95	Probable match	LOW_PRIORITY	otu_18	Bacteria(100);Firmicutes(100);Bacilli(100);Lactobacillales(100);Lactobacillaceae(100);Lactobacillus(100)
20BG06	Same species	LOW_PRIORITY	otu_486	Bacteria(100);Firmicutes(100);Bacilli(100);Lactobacillales(100);Streptococcaceae(100);Streptococcus(100)

Figure 3.11 – Example of local Blast results for Human Microbiome Project's 'Most Wanted' sequences. Isolates whose sequence presented a positive match with a most wanted sequence (identified by an OTU number- HMP_OTU) are indicated by their respective sequencing LIGHTrun barcodes. Type of match, Priority of isolation of the uncultured and unsequenced most wanted bacteria, and Taxonomy (taxonomic classification of the matching OTU from the most wanted list and which should be in accordance with the classification of the positively matched pure culture) are also presented in the table.

OTU numbers of the most wanted sequences that have perfectly matched with the isolates are: 7, 15, 167, 211, 253, 301, 374, 451, 638 and 683 (supplementary Figure E.1.C) - which correspond to the isolates *Veillonella* sp., *Gemella* sp., *Escherichia coli, Gemella* sp., *Enterococcus faecalis, Enterococcus faecalis, Fusobacterium necrophorum* subsp. *funduliforme, Lactobacillus salivarius, Lactobacillus fermentum* and *Lactobacillus rhamnosus,* respectively. Only low and medium priority OTU presented positive matches. Medium priority includes OTU numbers 62, 331, 522, 528, and 1150 - which matches with the isolates *Actinomyces* sp., *Bacterium (Bacteroides* sp.*), Bacteroides fragilis, Peptoniphilus coxii* and *Streptococcus parasanguinis*, respectively. The taxonomic classification of the matching OTU from the 'Most Wanted' list is coherent with the taxonomy previously attributed to the isolates by NCBI and RDP.

3.4.3 Metagenomics *vs* Culture at family level (patient 1 and 2)

The culture approach allowed to obtain isolates from 28 different families in total (from oral swab and tumour biopsy of patient 1 and 2), while metagenomic profiling of the same original samples only detected 22 different families (Figure 3.12.A). 7 families were identified by the metagenomic profile, but not isolated in the culture process – Acidaminococcaceae, Flavobacteriaceae, Peptostreptococcaceae, Porphyromonadaceae, Rikenellaceae, Ruminococcaceae and Sutterellaceae. The inverse – families recovered from oral and tumour specimens, but which do not feature in Illumina results – represents 12 families: Actinomycetaceae, Atopobiaceae, Bacillaceae, Dermacoccaceae, Enterococcaceae, Micrococcaceae, Peptoniphilaceae, Propionibacteriaceae, Staphylococcaceae, Streptomycetaceae, Tannerellaceae and Yersiniaceae (Figure 3.12.A).

From oral swab of patient 1 and 2, 13 and 12 families were recovered respectively (Figure 3.9.A). The metagenomic results for the same samples identified 12 families (7 of them with enough relative abundance – bigger than 1% - to appear in the metagenomic microbiota profile bar chart) and 13 families (5 of them with representative abundances), respectively (Figure 3.12.B and supplementary Figure E.1.A). From tumour samples

of patient 1 and 2, 12 families were isolated in each (Figure 3.9.A), in comparison with 19 families obtained with Illumina for both specimens (supplementary Figure E.1A). Note that even the lower relative abundant families (lower than 1%), indicated in the metagenomic analysis, were successfully isolated. For instance, in patient 1 oral swab, Prevotellaceae, Fusobacteriaceae and Lactobacillaceae were recovered, although their metagenomic percentages are 0.96%, 0.20% and 0.01%, respectively (supplementary Figure E.1.A).

A

Different Families from Pat 1 and Pat 2 (Oral swab and Tumour samples)

Recovered from Culture	Identified by Metagenomic analysis
Actinomycetaceae	Acidaminococcaceae
Atopobiaceae	Bacillales_Incertae_Sedis_XI
Bacillaceae	Bacteroidaceae
Bacillales Family XI. Incertae Sedis	Bifidobacteriaceae
Bacteroidaceae	Clostridiales_Incertae_Sedis_XI
Bifidobacteriaceae	Coriobacteriaceae
Clostridiaceae	Enterobacteriaceae
Coriobacteriaceae	Erysipelotrichaceae
Dermacoccaceae	Flavobacteriaceae
Enterobacteriaceae	Fusobacteriaceae
Enterococcaceae	Lachnospiraceae
Erysipelotrichaceae	Lactobacillaceae
Fusobacteriaceae	Neisseriaceae
Lachnospiraceae	Pasteurellaceae
Lactobacillaceae	Peptostreptococcaceae
Micrococcaceae	Porphyromonadaceae
Neisseriaceae	Prevotellaceae
Pasteurellaceae	Rikenellaceae
Peptoniphilaceae	Ruminococcaceae
Prevotellaceae	Streptococcaceae
Propionibacteriaceae	Sutterellaceae
Ruminococcaceae	Veillonellaceae
Staphylococcaceae	
Streptococcaceae	**Total:** 22
Streptomycetaceae	
Tannerellaceae	
Unclassified	
Veillonellaceae	
Yersiniaceae	
Total: 29	

B

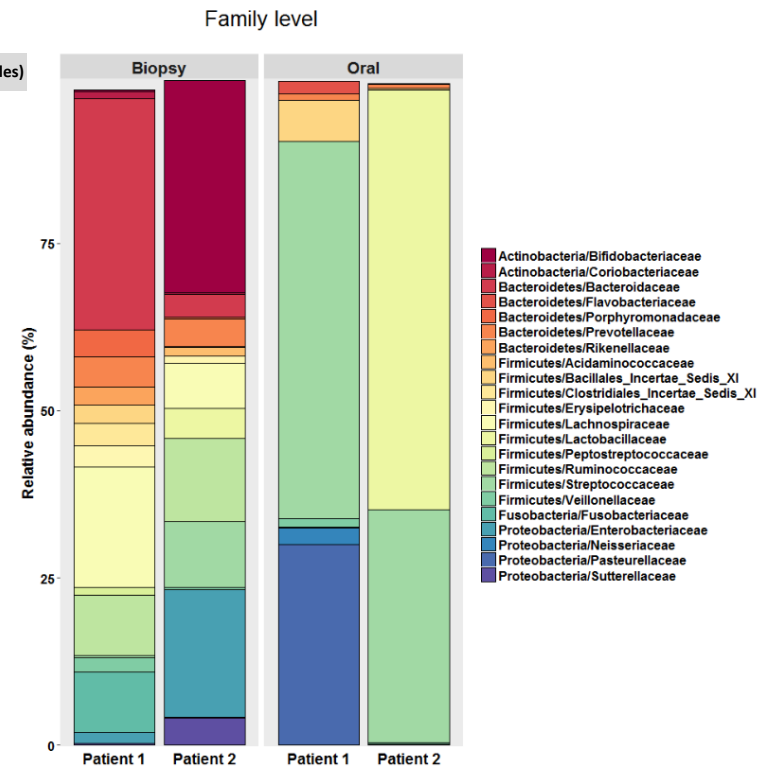


Figure 3.12 – Microbiota profile at the family level *vs* Isolated families of patient 1 and patient 2. A- List of different families obtained from culture (in orange are highlighted the taxa not detected by metagenomic analysis) *vs* families identified by Illumina sequencing results (in green are highlighted the taxa not isolated by the culture approach), from patient 1 and patient 2. B - Microbiota profile at the family level for patient 1 and 2 (original oral swab and tumour biopsy samples). Only dominant bacterial families (relative abundance ≥ 1%) are displayed.

The isolated phyla match the phyla identified by metagenomic profiling (5 in total) in patient 1 and 2 – Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria and Proteobacteria (Figure 3.12.B and Figure 3.8.A). In addition, unclassified phyla were detected in isolates of the tumour samples of both patients, but their phylogenetic trees allowed to associate them to Firmicutes and Bacteroidetes phyla (Figure 3.6.A and 3.7.A).

Continuing the analysis of the families identified in metagenomic vs culture approaches, is interesting to note that Fusobacteriaceae and Porphyromonadaceae were present in representative abundance in patient 1 tumour sample – 9% and 4% respectively (Figure 3.12.B and supplementary Figure E.1.A). However, members of these families were not isolated from this sample. Only *Fusobacterium nucleatum* was recovered from the oral cavity of patient 1.

3.4.4 Microbiota profile at family level of several CRC patients including patient 1 and 2

The microbiota profile of 19 CRC patients (ON Tumour, OFF and UDP samples) and of patient 1 and 2 (only ON Tumour sample) varies substantially in terms of relative abundance and taxonomic distribution between patients at the family level (Figure 3.13). Similar microbiota profile is verified in OFF and UDP samples, when comparing the same patients. A general pattern is visualized, with samples containing in great abundance Bacteroidetes and Firmicutes, and in lower abundance Actinobacteria, Fusobacteria and Proteobacteria (Figure 3.13). This distribution is in accordance with general gut microbiota composition which is dominated by Firmicutes (30 to 50%) and Bacteroidetes (20% to 40%) predominantly, and Actinobacteria in a lower percentage (1% to 10%) (Gagnière *et al.*, 2016). In colon, Bacteroidetes and Firmicutes are the most representative bacteria phyla (90% in total), which is verified in this population of CRC patients (Gagnière *et al.*, 2016).

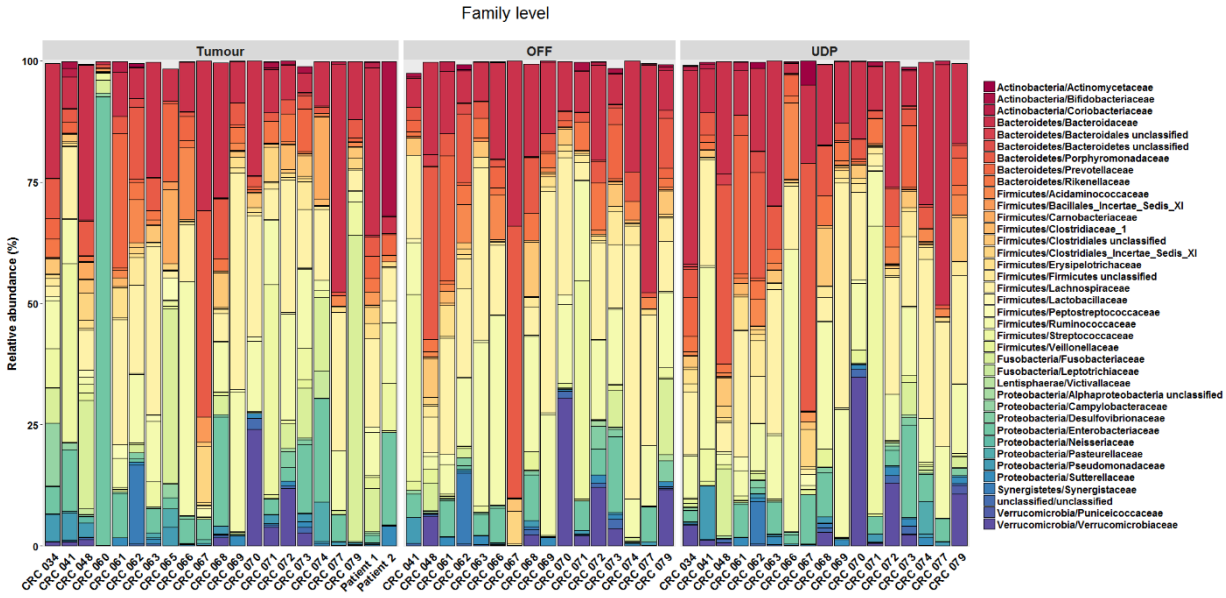


Figure 3.13 – Microbiota profile at family level for 19 CRC patients (ON Tumour, OFF and UDP samples). The microbiota profile presented by the tumour biopsies of patient 1 and patient 2 was also put side by side with the tumour profile of the rest of the CRC patients to compare differences in profile distributions. Only dominant bacterial families (relative abundance ≥ 1%) are displayed.

The microbiota pattern verified in the tumour samples of patient 1 and 2 is slightly different from the other patients (Figure 3.13). Since patient 1 had chemotherapy, a more irregular family distribution was expected when compared to the other individuals (didn't have chemotherapy). However, the only significant difference is the lower abundance of Proteobacteria phylum. In relation to patient 2, it should be highlighted the high percentage of Bifidobacteriaceae, never verified in the microbiota profile of the other CRC patients. Note that tumour sample of patient CRC060 is an outsider in terms of the regular CRC pattern verified in these 18 patients – it presents more than 92% of Enterobacteriaceae family members.

3.4.5 Selective power of tested culture media

The selective power of a selection of culture media used in this project was evaluated by comparing the metagenomic microbiota profile of bacteria grown in each liquid medium with the metagenomic profile of the original tumour and oral samples of patient 1 (Figure 3.14). It can be concluded that there is no excellent representative medium able to grow all the genera present in the original sample. In the biopsy case, Brain Heart infusion, Tryptic Soy, YCFA and YCFA + microaerophilic conditions are the ones that achieve more different genera. However, other media (more selective) allowed to obtain different genera not covered by these more general media. *Bifidobacterium* genus was grown mainly in Tomato Juice broth. *Bacteroides* genus was obtained in all media. *Anaerococcus* in M17 culture medium. *Clostridium* XIVa mainly in YCFA and YCFA 5% CO_2, and Todd Hewitt broth. *Clostridium* XVIII grew in YCFA, YCFA supplemented with kanamycin, and MRS medium. *Enterococcus* in MRS. *Finegoldia* in YCFA and BHI. *Fusobacterium* in TS and BHI. And finally, *Escherichia/Shigella* grew mainly in Tomato Juice and SAB medium (Figure 3.14).

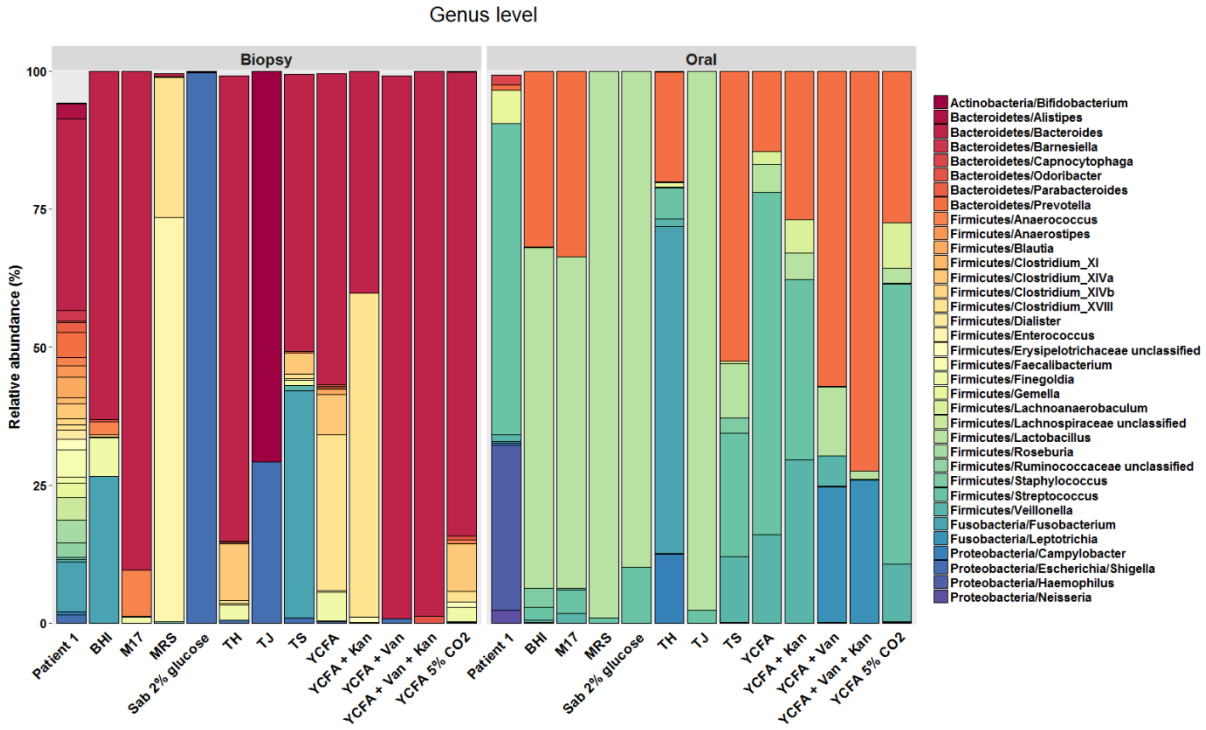


Figure 3.14 – Selective power of each medium at genus level. Microbiota profiles of patient 1 (tumour and oral samples) were gathered together with the profiles of each medium in order to compare easily the composition of the original samples and the captured composition in culture. Only dominant bacterial genera (relative abundance ≥ 1%) are displayed. BHI – Brain Heart Infusion. Sab – Sabouraud. TH – Todd Hewitt. TJ – Tomato Juice. TS – Tryptic Soy. Kan -Kanamycin. Van – Vancomycin.

Looking at the oral genera captured by each medium, we can see again that there is no medium capable of covering all the genera present in the oral cavity of patient 1. Here, Todd Hewitt, Tryptic Soy and YCFA (and YCFA variants) are the ones with more represented genera. *Prevotella, Lactobacillus and Streptococcus* genera grew in all media. *Staphylococcus* was captured mainly by BHI and TS. *Veillonella* in YCFA (and variants excepting YCFA with both antibiotics) and in TS. *Fusobacterium* and *Campylobacter* mainly in Todd Hewitt broth (Figure 3.14).

Looking now only to the genera presented in the microbiota profile of tumour and oral samples of patient 1 (first bars of both Biopsy and Oral bar chart groups – Figure 3.14), the main represented genera are *Bacteroides* (35%)*, Fusobacterium, Faecalibacterium, Prevotella, Roseburia, Lachnospiraceae* unclassified, *Blautia, Clostridum* XIVa and *Gemella* (by descending order of abundance) in tumour sample; and *Streptococcus* (56%)*, Haemophilus, Gemella, Neisseria* and *Veillonella* in oral sample (by descending order of abundance too). Indeed, *Bacteroides* and *Streptococcus* were the genera most recovered from tumour and oral samples in the isolation process, respectively (supplementary Figure C.1).

Mainly in biopsy, several genera indicated in its metagenomic profile were not isolated. However, it should also be highlighted that many other genera represented in low abundance by the metagenomic profile of the original samples, or not even detected, were isolated successfully. For instance, in the oral swab, *Lactobacillus, Fusobacterium*, *Prevotella* and *Veillonella* were recovered (supplementary Table C.3), although they present relative metagenomic abundances of 0.01%, 0.20%, 0.91% and 1.33%, respectively. From the same sample, *Atopobium, Bacillus, Dermacoccus, Actinomyces* and *Serratia* were isolated (supplementary Table C.3), although they weren't detected in the metagenomic analysis. In tumour sample, *Bifidobacterium, Streptococcus, Finegoldia* and *Escherichia/Shigella* with relative metagenomic abundances of 0.22%, 0.41%, 1.06% and 1.55%, respectively, were also isolated (supplementary Table C.4). Other genera not identified in the metagenomic analysis were recovered from the same sample – *Enterococcus, Collinsella, Kocuria, Micrococcus, Cutibacterium, Holdemanella* and *Peptoniphilus* (supplementary Table C.4).

Chapter 4

Conclusions and Future Work

Advanced CRC has a high mortality rate, being consistently the third leading cause of cancer death around the globe. In industrialized nations, approximately 5% of people will be diagnosed with CRC at some point during their lifetime (*Colorectal Cancer - Cancer Stat Facts*, 2018). Although FOBT and colonoscopy (the most used CRC screening methods) efficiently detect advanced adenomas and carcinomas, they present very low sensitivity in detecting early lesions, which leads to the necessity of studying more reliable biomarkers (Flemer, Warren, *et al.*, 2017). Sporadic CRC occurs in individuals who have accumulated a burden of mutations, the last of which may be triggered by environmental factors including specific bacteria.

Recent studies have reported changes in the faecal and colonic mucosal microbiota in CRC patients, being dysbiosis an already known feature/hall-mark of CRC (Flemer, Lynch, *et al.*, 2017). Several single bacterial species have been associated with higher CRC risk and data from several animal models has been able to characterize the mechanisms through they which modulate CRC carcinogenesis (Gagnière *et al.*, 2016). Part of this altered community of bacteria seems to be derived partly from the oral microbiota since high abundance of putative oral bacteria has been found on CRC tumours and polyps by metagenomic profiling approaches (Flemer, Warren, *et al.*, 2017). Given that the oral cavity is at the beginning of the digestive tract and that in average a person generates and ingests around 1.5 litres of saliva per day (Edgar *et al.*, 2004), the theory of CRC gut mucosal colonization by oral bacteria seems to be reasonable to investigate. Additionally, we do not currently know how oral bacteria might accelerate tumorigenesis.

To help advance this theory, there is the need to grow these CRC-associated bacteria in pure culture, to prove that they are truly derived from the oral cavity and to study their ability to interact with the innate immune system and with cultured cells. In this project, several objectives were established in order to investigate oral colonization in CRC tumours – culture and isolate bacteria from colonic tumours and oral swabs of 2 CRC patients; identify bacteria by sequencing an amplicon from the full length 16S rRNA gene by Sanger sequencing; and perform RAPD to compare the genomic structure of oral and colonic isolates of the same species from the same CRC patient.

In this work, CRC patients were recruited and a broad range approach was developed that includes histology and immunology techniques (H&E staining and IF, Gram staining and FISH); microbiota metagenomic profiling by using Illumina sequencing targeting the V3/V4 variable region of 16S rRNA gene; and a complete culture, isolation and identification methodology exploring multiple combinations of conditions that targeted faecal, oral and tumour bacteria of four CRC patients (patient CRC038, patient CRC057, patient 1 and patient 2). These conditions included anaerobic, microaerophilic and aerobic incubations in a panel of 11 different media (where YCFA

medium is included) at different temperatures (being 37 ºC the most used), pre-incubation in blood bottles and addition of blood to culture media, heat shock, selective filtration and pre-incubation in liquid medium prior to plating. All conditions were used simultaneously (twice) in oral swabs and tumour samples of CRC patients 1 and 2. Media suitable for the isolation of general oral bacteria and bacteria associated with periodontal disease and caries were investigated and selected.

Gram staining allowed to validate bacteria presence in all samples and to establish a distribution pattern presented by the seven CRC patients analysed. In ON Tumour samples, a co-occurrence of bacteria is visualized inside the tumoral mass; while in OFF samples, a dense community of bacteria which resembles an intricated bacterial biofilm is detected in the mucus layer, in close contact with epithelial tissue and sometimes even infiltrated in it and invading lamina propria. This is interesting since the mucus zone close to mucosa is germ-free in healthy individuals (Swidsinski et al., 2007). Gram and FISH together allowed to confirm the presence of invasive bacterial biofilms – findings that are in accordance with a recent study that showed for the first time that these biofilms can be intrinsically associated with CRCs (Dejea *et al.*, 2014). In future works, since IF and FISH protocols were optimized, these can be used together in order to identify local immune responses associated with specific groups of bacteria.

Within the scope of oral bacterial colonization in the gut, two articles seem to show substantiated evidence of this premise until now (Atarashi *et al.*, 2017 and Komiya *et al.*, 2018), with the latter proving the presence of identical strains of *Fusobacterium nucleatum* (one species) in colorectal cancer and oral cavity of CRC patients. The culture-based approach developed in this project allowed to achieve new findings. To our knowledge, this is the first time that oral colonization in CRC tumours was proven with a considerable range of isolated species (with new strains within this context) - 15 species in total belonging to 6 different families and 2 different phyla.

Cronobacter sakazaki, Enterococcus faecalis, Escherichia coli, Lactobacillus crispatus, Lactobacillus delbrueckii subsp. *bulgaricus, Lactobacillus fermentum, Lactobacillus gasseri, Lactobacillus oris, Lactobacillus rhamnosus, Lactobacillus salivarius, Serratia marcescens, Staphylococcus hominis, Streptococcus mutans, Streptococcus salivarius* and *Streptococcus intermedius* (Pat 1) were the species that presented at least one equal strain in paired oral swab and tumour samples of the same CRC patient. Only one equal strain was recovered from patient 1 (all the other 14 are derived from patient 2), probably due to the effect of chemotherapy treatments that this patient was subjected to, during the last 2 years, able to cause abrupt alterations in the host gut and oral microbiota (Montassier *et al.*, 2015).

Enterococcus faecalis and *E. coli* are single pathogens already well associated with CRC. Previous characterisation of the other isolated species by literature includes potential pathogenic and virulence mechanisms that can be evaluated in CRC tumorigenesis context. The most interesting finding is that 9 of the 15 species isolated (*Streptococcus mutans, Streptococcus salivarius* and more seven *Lactobacillus* spp.) are strongly linked with dental caries and plaque formation. They are strong producers of lactic acid and are associated with biofilm formation, characteristics that enable these oral bacteria to colonize the dental surface and cause damage in the

rigid tooth structure ((Becker *et al.*, 2002), (Aas *et al.*, 2008)). This can suggest a synergetic role on CRC tumorigenesis of these species, by similar pathogenic mechanisms already verified in the oral cavity.

More than proving that CRC patients have identical strains in their oral cavity and CRC tumour, a defined, phylogenetic diverse artificial faecal, oral and tumour community was obtained from the 4 CRC patients. In total, 500 pure cultures were archived belonging to 95 different species, 28 families and 5 phyla. In total, the pure collections created in this project contain 10 perfect matches (100% similar sequences), 35 probable matches (99% similar sequences) and 31 equal species (sequences that most likely belong to the same species) with the Human Microbiome Project's 'Most Wanted' list of previously uncultured and unsequenced microbes. It can, therefore, be concluded that the broad-range culture-based approach explored in here was a success - it led to bacterial discovery and contests the notion that the majority of the oral and gut microbiota is unculturable.

The metagenomic microbiota profiling of CRC samples allowed to confirm that dysbiosis is a feature of CRC, but as several other studies, no conserved CRC microbiota signature was verified (Flemer, Lynch, *et al.*, 2017). It also allowed to determine the selective power of each medium used and to analyse the efficacy of the isolation protocol. Comparing metagenomic and culture results from oral swab and tumour samples of patient 1 and 2, the culture approach allowed to obtain isolates from 28 different families in total, while the metagenomic microbiota profiling only detected 22 different families. Although several genera/families indicated in the metagenomic profile were not isolated, genera represented in low abundance by the metagenomic profile (lower than 1%) or not even detected by it were successfully isolated.

The representative pure collections obtained in this project can be used as a consortium to study microbial interaction networks, allowing to clarify cross-feeding interactions, targeted antibiotic mechanisms and identification of strains capable of competing in CRC environment (this would be interesting to test by culturing together tumour and oral isolated communities, for instance in a bioreactor or in a germ-free mouse model).

Since *Staphylococcus hominis* was the only strain susceptible of AP-PCR pattern similarity doubt, it is recommended to perform another AP-PCR using the gDNA of this isolate. Future work includes sequencing the entire genome of oral isolates that were able to colonize CRC tumours and search for virulence factors/plasmids. These single species can be tested for virulence and genotoxicity potential, being cultured *in vitro* - cells or organoids – or *in vivo* – in a mouse model - in order to investigate cytokine production, cellular proliferation and inflammation. This way, causal relationships with CRC development or even with CRC prevention can be explored. Metabolomic studies are also another possible approach to be explored.

Since CRC gut mucosal colonization by oral bacteria was only tested in two patients, future works can recruit more patients (a more representative sample) and use the optimized culture conditions and methodology explored in here to find new species within this context. If possible, it will be recommendable to perform first the microbiota profile of oral swab and tumour samples to confirm that there are at least some positive families or species matches in paired samples.

References

Aas, J. A. *et al.* (2008) 'Bacteria of dental caries in primary and permanent teeth in children and young adults', *Journal of Clinical Microbiology*, 46(4), pp. 1407–1417. doi: 10.1128/JCM.01410-07.

Akopyanz, N. *et al.* (1992) 'DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting', *Nucleic Acids Research,* 20(19), pp. 5137–5142. doi: 10.1093/nar/20.19.5137.

Allard, G. *et al.* (2015) 'SPINGO: a rapid species-classifier for microbial amplicon sequences.', *BMC bioinformatics*, 16(1), p. 324. doi: 10.1186/s12859-015-0747-1.

Amaro, A. *et al.* (2016) 'Molecular evolution of colorectal cancer: from multistep carcinogenesis to the big bang', *Cancer and Metastasis Reviews*, 35(1), pp. 63–74. doi: 10.1007/s10555-016-9606-4.

Armaghany, T. *et al.* (2012) 'Genetic alterations in colorectal cancer.', *Gastrointestinal cancer research : GCR*, 5(1), pp. 19–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22574233 (Accessed: 27 August 2018).

Arnold, M. *et al.* (2017) 'Global patterns and trends in colorectal cancer incidence and mortality', *Gut*, 66(4), pp. 683–691. doi: 10.1136/gutjnl-2015-310912.

Atarashi, K. *et al.* (2017) 'Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation', *Science*, 358(6361), pp. 359–365. doi: 10.1126/science.aan4526.

Badet, C. and Thebaud, N. B. (2008) 'Ecology of lactobacilli in the oral cavity: a review of literature.', *The open microbiology journal*, 2, pp. 38–48. doi: 10.2174/1874285800802010038.

Becker, M. R. *et al.* (2002) 'Molecular analysis of bacterial species associated with childhood caries.', *Journal of clinical microbiology*, 40(3), pp. 1001–9. doi: 10.1128/JCM.40.3.1001-1009.2002.

Berg, D. E. *et al.* (1994) 'Fingerprinting microbial genomes using the RAPD or AP-PCR method', *Methods in molecular and cellular biology,* 5(1), pp 13-24. Available at: https://www.researchgate.net/publication/279554115_Fingerprinting_microbial_genomes_using_the_RAPD_or_AP-PCR_method (Accessed: 16 September 2018).

Browne, H. P. *et al.* (2016) 'Culturing of "unculturable" human microbiota reveals novel taxa and extensive sporulation', *Nature*, 533(7604), pp. 543–546. doi: 10.1038/nature17645.

Bullman, S. *et al.* (2017) 'Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer', *Science*, 358(6369), pp. 1443–1448. doi: 10.1126/science.aal5240.

Cao, X. (2017) 'Intestinal inflammation induced by oral bacteria.', *Science*, 358(6361), pp. 308–309. doi: 10.1126/science.aap9298.

Caporaso, J. G. *et al.* (2010) 'QIIME allows analysis of high-throughput community sequencing data', *Nature Methods*, 7(5), pp. 335–336. doi: 10.1038/nmeth.f.303.

Caufield, P. W. *et al.* (2015) 'Oral lactobacilli and dental caries: a model for niche adaptation in humans.', *Journal of dental research*, 94(9 Suppl), p. 110S–8S. doi: 10.1177/0022034515576052.

Chaves, F. *et al.* (2005) 'Nosocomial spread of a *Staphylococcus hominis* subsp. *novobiosepticus* strain causing sepsis in a neonatal intensive care unit.', *Journal of clinical microbiology*, 43(9), pp. 4877–9. doi: 10.1128/JCM.43.9.4877-4879.2005.

Chen, T. *et al.* (2010) 'The human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information', *Database : the journal of biological databases and curation*, 2010(0), p. baq013-baq013. doi: 10.1093/database/baq013.

Colon Anatomy - National Cancer Institute (2018). Available at: https://visualsonline.cancer.gov/details.cfm?imageid=9380 (Accessed: 12 September 2018).

Colon Cancer Treatment - National Cancer Institute (2018) *PDQ Cancer Information Summaries*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26389319 (Accessed: 13 September 2018).

Colorectal Cancer - Cancer Stat Facts (2018) *National Cancer Institute*. Available at: https://seer.cancer.gov/statfacts/html/colorect.html (Accessed: 12 September 2018).

Darveau, R. P. (2010) 'Periodontitis: a polymicrobial disruption of host homeostasis', *Nature Reviews Microbiology*, 8(7), pp. 481–490. doi: 10.1038/nrmicro2337.

Dejea, C. M. *et al.* (2014) 'Microbiota organization is a distinct feature of proximal colorectal cancers', *Proceedings of the National Academy of Sciences*, 111(51), pp. 18321–18326. doi: 10.1073/pnas.1406199111.

Donnenberg, M. S. (2014) 'Enterobacteriaceae', *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 2, pp. 2503–2517. doi: 10.1016/B978-1-4557-4801-3.00220-4.

Dutilh, B. E. *et al.* (2013) 'Screening metatranscriptomes for toxin genes as functional drivers of human colorectal cancer', *Best Practice & Research Clinical Gastroenterology*, 27(1), pp. 85–99. doi: 10.1016/j.bpg.2013.03.008.

Eckburg, P. B. (2005) 'Diversity of the human intestinal microbial flora', *Science*, 308(5728), pp. 1635–1638. doi: 10.1126/science.1110591.

Edgar, R. C. (2010) 'Search and clustering orders of magnitude faster than BLAST', *Bioinformatics*, 26(19), pp. 2460–2461. doi: 10.1093/bioinformatics/btq461.

Edgar, W. M. *et al.* (2004) 'Saliva and oral health', in *London: British Dental Association (3rd edition)*. Available at: http://www.worldcat.org/title/saliva-and-oral-health/oclc/475498429 (Accessed: 18 September 2018).

Flemer, B., Warren, R. D., *et al.* (2017) 'The oral microbiota in colorectal cancer is distinctive and predictive', *Gut*, 67(8), pp. 1454–1463. doi: 10.1136/gutjnl-2017-314814.

Flemer, B., Lynch, D. B., *et al.* (2017) 'Tumour-associated and non-tumour-associated microbiota in colorectal cancer', *Gut*, 66(4), pp. 633–643. doi: 10.1136/gutjnl-2015-309595.

Flynn, K. J. *et al.* (2016) 'Metabolic and community synergy of oral bacteria in colorectal cancer', *mSphere*, 1(3), pp. e00102-16. doi: 10.1128/mSphere.00102-16.

Forssten, S. D. *et al.* (2010) '*Streptococcus mutans*, caries and simulation models.', *Nutrients*, 2(3), pp. 290–8. doi: 10.3390/nu2030290.

Gagnière, J. *et al.* (2016) 'Gut microbiota imbalance and colorectal cancer', *World Journal of Gastroenterology*, 22(2), pp. 501–518. doi: 10.3748/wjg.v22.i2.501.

Gallimidi, A. B. *et al.* (2015) 'Periodontal pathogens *Porphyromonas gingivalis* and *Fusobacterium nucleatum* promote tumor progression in an oral-specific chemical carcinogenesis model.', *Oncotarget*, 6(26), pp. 22613–23. doi: 10.18632/oncotarget.4209.

Gill, S. R. *et al.* (2006) 'Metagenomic analysis of the human distal gut microbiome.', *Science*, 312(5778), pp. 1355–9. doi: 10.1126/science.1124234.

Hajishengallis, G. (2015) 'Periodontitis: from microbial immune subversion to systemic inflammation', *Nature Reviews Immunology*, 15(1), pp. 30–44. doi: 10.1038/nri3785.

Hall-Stoodley, L. *et al.* (2004) 'Bacterial biofilms: from the natural environment to infectious diseases', *Nature Reviews Microbiology*, 2(2), pp. 95–108. doi: 10.1038/nrmicro821.

Hall, T. A. (1999) 'BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT', *Nucleic Acids Symposium Series*, 41, pp. 95–98. doi: citeulike-article-id:691774.

Han, Y. W. (2014) 'Oral bacteria and colorectal cancer : etiology and mechanism', in *AACR Education book*, pp. 61–64. doi: 10.1158/AACR.EDB-14-7995.

Han, Y. W. (2015) '*Fusobacterium nucleatum*: a commensal-turned pathogen', *Current Opinion in Microbiology*, 23, pp. 141–147. doi: 10.1016/j.mib.2014.11.013.

Han, Y. W. and Wang, X. (2013) 'Mobile microbiome: oral bacteria in extra-oral infections and inflammation', *Journal of Dental Research*, 92(6), pp. 485–491. doi: 10.1177/0022034513487559.

Handa, O. *et al.* (2010) '*Helicobacter pylori*: a ROS-inducing bacterial species in the stomach', *Inflammation Research*, 59(12), pp. 997–1003. doi: 10.1007/s00011-010-0245-x.

Hugon, P. *et al.* (2013) 'Molecular studies neglect apparently gram-negative populations in the human gut microbiota', *Journal of Clinical Microbiology*, 51(10), pp. 3286–3293. doi: 10.1128/JCM.00473-13.

Immunotherapy - CRC (2018) *National Cancer Institute*. Available at: https://www.cancer.gov/news-events/cancer-currents-blog/2018/fda-ipilimumab-nivolumab-colorectal-dna-repair (Accessed: 10 October 2018).

Johansson, M. E. V. and Hansson, G. C. (2012) 'Preservation of mucus in histological sections, immunostaining of mucins in fixed tissue, and localization of bacteria with FISH', in *Methods in molecular biology*, pp. 229–235. doi: 10.1007/978-1-61779-513-8_13.

Johansson, M. E. V. *et al.* (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*, 108(Supplement_1), pp. 4659–4665. doi: 10.1073/pnas.1006451107.

Jorth, P. *et al.* (2014) 'Metatranscriptomics of the human oral microbiome during health', *mBio*, 5(2), pp. 1–10. doi: 10.1128/mBio.01012-14.Editor.

Kato, I. *et al.* (2016) 'Oral microbiome and history of smoking and colorectal cancer', *Journal of Epidemiological Research*, 2(2), pp. 92–101. doi: 10.5430/jer.v2n2p92.

Komiya, Y. *et al.* (2018) 'Patients with colorectal cancer have identical strains of *Fusobacterium nucleatum* in their colorectal cancer and oral cavity.', *Gut*, 0, pp. 1–3. doi:10.1136/gutjnl-2018-316661.

Komiyama, E. Y. *et al.* (2016) '*Enterococcus* species in the oral cavity: prevalence, virulence factors and antimicrobial susceptibility.', *PloS one*, 11(9), p. e0163001. doi: 10.1371/journal.pone.0163001.

Kostic, A. D. *et al.* (2013) '*Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment', *Cell Host and Microbe*, 14(2), pp. 207–215. doi: 10.1016/j.chom.2013.07.007.

Kumar, S. *et al.* (2018) 'MEGA X: Molecular evolutionary genetics analysis across computing platforms', *Molecular Biology and Evolution*, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.

Lagier, J. C. *et al.* (2012) 'Microbial culturomics: paradigm shift in the human gut microbiome study', *Clinical Microbiology and Infection*, 18, pp. 1185–1193. doi: 10.1111/1469-0691.12023.

Lagier, J. C. *et al.* (2016) 'Culture of previously uncultured members of the human gut microbiota by culturomics', *Nature Microbiology*, 1(12), p. 16203. doi: 10.1038/nmicrobiol.2016.203.

Lagkouvardos, I. *et al.* (2017) 'Cultured microbes represent a substantial fraction of the human and mouse gut microbiota', *Gut Microbes*, 8(5), pp. 493–503. doi: 10.1080/19490976.2017.1320468.

Lau, J. T. *et al.* (2016) 'Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling', *Genome Medicine*, 8(1), pp. 1–10. doi: 10.1186/s13073-016-0327-7.

Letunic, I. and Bork, P. (2011) 'Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy', *Nucleic Acids Research*, 39(suppl), pp. W475–W478. doi: 10.1093/nar/gkr201.

Li, H. *et al.* (2015) 'The outer mucus layer hosts a distinct intestinal microbial niche', *Nature Communications*, 6(1), p. 8292. doi: 10.1038/ncomms9292.

LIGHTRUN Tube sample requirements (2018). Available at: https://www.eurofinsgenomics.eu/en/custom-dna-sequencing/gatc-services/lightrun-tube/ (Accessed: 19 August 2018).

Louis, P. *et al.* (2014) 'The gut microbiota, bacterial metabolites and colorectal cancer', *Nature Reviews Microbiology*, 12(10), pp. 661–672. doi: 10.1038/nrmicro3344.

Magoc, T. and Salzberg, S. L. (2011) 'FLASH: Fast length adjustment of short reads to improve genome assemblies', *Bioinformatics*, 27(21), pp. 2957–2963. doi: 10.1093/bioinformatics/btr507.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads',

EMBnet.journal, 17(1), p. 10. doi: 10.14806/ej.17.1.200.

Masood, U. *et al.* (2016) 'Colorectal cancer associated with *Streptococcus anginosus* bacteremia and liver abscesses.', *Case reports in gastroenterology*, 10(3), pp. 769–774. doi: 10.1159/000452757.

Mescher, A. (2013) 'Junqueira's Basic Histology Text and Atlas, 15th edition', in *McGraw-Hill Medical*. Available at: https://accessmedicine.mhmedical.com/content.aspx?bookid=2430&sectionid=190283267 (Accessed: 23 September 2018).

Miranda, N. and Portugal, C. (2016) 'Doenças oncológicas em números 2015 - Programa nacional para as doenças oncológicas', *Dgs*, pp. 5–65. doi: ISSN: 2183-0746.

Montassier, E. *et al.* (2015) 'Chemotherapy-driven dysbiosis in the intestinal microbiome', *Alimentary Pharmacology & Therapeutics*, 42(5), pp. 515–528. doi: 10.1111/apt.13302.

Moore, W. E. and Holdeman, L. V (1974) 'Human fecal flora: the normal flora of 20 Japanese-Hawaiians', *Applied Microbiology*, 27(5), pp. 961–979. doi: Human.Moore.1974.

Nakajima, M. *et al.* (2015) 'Oral administration of *P. gingivalis* induces dysbiosis of gut microbiota and impaired barrier function leading to dissemination of enterobacteria to the liver', *PLoS ONE*, 10(7), pp. 1–15. doi: 10.1371/journal.pone.0134234.

Nakatsu, G. *et al.* (2015) 'Gut mucosal microbiome across stages of colorectal carcinogenesis', *Nature Communications*, 6(1), p. 8727. doi: 10.1038/ncomms9727.

Nešić, D. *et al.*(2004) 'Assembly and function of a bacterial genotoxin', *Nature*, 429(6990), pp. 429–433. doi: 10.1038/nature02532.

NIH Human Microbiome Project - Most Wanted Genomes (2018). Available at: https://www.hmpdacc.org/most_wanted/ (Accessed: 6 November 2018).

Nucleotide BLAST: Search nucleotide databases using a nucleotide query (2018). Available at: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch (Accessed: 20 August 2018).

O'Keefe, S. J. D. (2016) 'Diet, microorganisms and their metabolites, and colon cancer', *Nature Reviews Gastroenterology and Hepatology*, 13(12), pp. 691–706. doi: 10.1038/nrgastro.2016.165.

O'Toole, P. W. *et al.* (2017) 'Next-generation probiotics: the spectrum from probiotics to live biotherapeutics', *Nature Microbiology*, 2(5), pp. 1–6. doi: 10.1038/nmicrobiol.2017.57.

Ochieng, J. B. *et al.* (2014) '*Serratia marcescens* is injurious to intestinal epithelial cells.', *Gut microbes*, 5(6), pp. 729–36. doi: 10.4161/19490976.2014.972223.

Pédron, T. *et al.* (2012) 'A crypt-specific core microbiota resides in the mouse colon.', *mBio*, 3(3), pp. e00116-12. doi: 10.1128/mBio.00116-12.

Perera, M. *et al.* (2016) 'Emerging role of bacteria in oral carcinogenesis: a review with special reference to perio-pathogenic bacteria', *Journal of Oral Microbiology*, 8(1), pp. 1–10. doi: 10.3402/jom.v8.32762.

RDP Sequence Match (2018). Available at: https://rdp.cme.msu.edu/seqmatch/seqmatch_intro.jsp (Accessed: 21 August 2018).

Rettedal, E. A. *et al.* (2014) 'Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria', *Nature Communications*, 5, pp. 1–9. doi: 10.1038/ncomms5714.

Rooks, M. G. and Garrett, W. S. (2016) 'Gut microbiota, metabolites and host immunity', *Nature Reviews Immunology*, 16(6), pp. 341–352. doi: 10.1038/nri.2016.42.

Sansonetti, P. J. (2018) 'Editorial: Editorial for the virtual issue on microbiome', *FEMS Microbiology Reviews*, 42(2), pp. 113–115. doi: 10.1093/femsre/fux058.

Schloss, P. D. *et al.* (2009) 'Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities', *Applied and Environmental Microbiology*, 75(23), pp. 7537–7541. doi: 10.1128/AEM.01541-09.

Schwabe, R. F. and Jobin, C. (2013) 'The microbiome and cancer.', *Nature reviews. Cancer*, 13(11), pp. 800–12. doi: 10.1038/nrc3610.

Sears, C. L. and Garrett, W. S. (2014) 'Microbes, microbiota, and colon cancer', *Cell Host and Microbe*, 15(3), pp. 317–328. doi: 10.1016/j.chom.2014.02.007.

Smith, A. J. *et al.* (2001) 'The ecology of *Staphylococcus* species in the oral cavity', *Journal of Medical Microbiology*, 50(11), pp. 940–946. doi: 10.1099/0022-1317-50-11-940.

Smith, L. and Garcia, J. (2011) 'Enteral nutrition', *Pediatric Gastrointestinal and Liver Disease*, p. 978–1001.e4. doi: 10.1016/B978-1-4377-0774-8.10089-2.

Sommer, M. O. A. (2015) 'Advancing gut microbiome research using cultivation', *Current Opinion in Microbiology*, 27, pp. 127–132. doi: 10.1016/j.mib.2015.08.004.

Srinivasan, L. and Evans, J. R. (2018) 'Health care-associated infections', *Avery's Diseases of the Newborn*, p. 566–580.e6. doi: 10.1016/B978-0-323-40139-5.00040-1.

Staley, J. (1985) 'Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats', *Annual Review of Microbiology*, 39(1), pp. 321–346. doi: 10.1146/annurev.micro.39.1.321.

Stenlid, J. *et al.* (1994) 'Intraspecific genetic variation in *Heterobasidion annosum* revealed by amplification of minisatellite DNA', *Mycological Research*, 98(1), pp. 57–63. doi: 10.1016/S0953-7562(09)80337-7.

Swidsinski, A. *et al.* (2007) 'Comparative study of the intestinal mucus barrier in normal and inflamed colon.', *Gut*, 56(3), pp. 343–50. doi: 10.1136/gut.2006.098160.

Takahashi, N. (2005) 'Microbial ecosystem in the oral cavity: metabolic diversity in an ecological niche and its relationship with oral diseases', *International Congress Series*, 1284, pp. 103–112. doi: 10.1016/j.ics.2005.06.071.

Thompson, S. L. *et al.* (2010) 'Mechanisms of chromosomal instability.', *Current biology : CB*, 20(6), pp. R285-95. doi: 10.1016/j.cub.2010.01.034.

Tilg, H. *et al.* (2018) 'The intestinal microbiota in colorectal cancer', *Cancer Cell*, pp. 1–11. doi: 10.1016/j.ccell.2018.03.004.

Tjalsma, H. *et al.* (2012) 'A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects', *Nature Reviews Microbiology*, 10(8), pp. 575–582. doi: 10.1038/nrmicro2819.

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

Vogelstein, B. *et al.* (2013) 'Cancer genome landscapes.', *Science*, 339(6127), pp. 1546–58. doi: 10.1126/science.1235122.

Wade, W. G. (2013) 'The oral microbiome in health and disease', *Pharmacological Research*, 69(1), pp. 137–143. doi: 10.1016/j.phrs.2012.11.006.

Warren, R. L. *et al.* (2013) 'Co-occurrence of anaerobic bacteria in colorectal carcinomas', *Microbiome*, 1(1). doi: 10.1186/2049-2618-1-16.

What Is Cancer? - National Cancer Institute (2018). Available at: https://www.cancer.gov/about-cancer/understanding/what-is-cancer (Accessed: 12 September 2018).

Whitmore, S. E. and Lamont, R. J. (2014) 'Oral bacteria and cancer', *PLoS Pathogens*, 10(3), pp. 1–3. doi: 10.1371/journal.ppat.1003933.

World Health Organization – Cancer Stats (2018). Available at: https://www.who.int/news-room/fact-sheets/detail/cancer (Accessed: 14 November 2018).

Zhou, Y.-J. *et al.* (2017) 'Cancer killers in the human gut microbiota: diverse phylogeny and broad spectra', *Oncotarget*, 8(30), pp. 49574–49591. doi: 10.18632/oncotarget.17319.

Appendix A

YCFA medium composition

Table A.1 – Detailed composition of solutions used in the preparation of YCFA medium.

Mineral Solution 1 (1 L)	
K2HPO4	3 g
H2Od	up to 1 L
Mineral Solution 2 (1 L)	
KH2PO4	3 g
(NH4)2SO4	6 g
NaCl	6 g
MgSO4 7H2O	0.6 g
CaCl2	0.6 g
H2Od	up to 1 L
VFA Solution (260 mL)	
Acetic Acid	170 mL
Propionic Acid	60 mL
n-Valeric Acid	10 mL
Isovaleric Acid	10 mL
Isobutyric Acid	10 mL
Vitamin Solution 1 (100 mL)	
Biotin	1 mg
Cobalemin	1 mg
PABA	3 mg
Pyridoxide	15 mg
Folic acid	5 mg
H2Od	up to 100 mL
Filter-sterilised	
Vitamin Solution 2 (100 mL)	
Thiamine	5 mg
Riboflavin	5 mg
H2Od	up to 100 mL
Filter-sterilised	
Haemin Solution (200 mL)	
KOH	0.56 g
EtOH	25 mL
Haemin	0.2 g
H2Od	Up to 200 mL
Resazurin Solution (100 mL)	
Resazurin	0.1 g
H2Od	Up to 100 mL

Appendix B

Accessing the quality of the PCR products (full length 16S rRNA gene)

Figure B.1 - 2% agarose electrophoresis gels of PCR products amplified with 27F2 and 1492R3 primers. This step was performed to access the quality of the DNA, before proceeding to purification and Sanger sequencing. The majority of the PCR products was amplified successfully without smearing for the four patients and their gel position corresponds to the 1500 bp band of the ladder, which is the full size of the amplified 16S rRNA gene.

Appendix C

Qualitative and quantitative summary of the different species isolated from the four patients

Table C.1 – Summary of the different phyla, families, species, apparent strains and number of pure isolates recovered from patient CRC038

Phylum	Family	Species	#	Apparent Strains
Bacteroidetes	Bacteroidaceae	*Bacteroides vulgatus*	1	mpk
Firmicutes	Lachnospiraceae	*Blautia sp.*	1	MC_32
		Lachnospiraceae bacterium	2	MC_36
	Clostridiaceae	*Clostridium perfringens*	5	**SD14-02** 18115
	Enterococcaceae	*Enterococcus avium*	6	**HCD36-3** **HCD36-2**
	Lactobacillaceae	*Lactobacillus casei*	2	**NWAFU1574** **MasaLam7**
		Lactobacillus paracasei	4	PN6 **BIOUAL009**
	Ruminococcaceae	*Ruminococcus sp.*	1	K-1
	Streptococcaceae	*Streptococcus equinus*	1	**HCD42-2**
		Streptococcus lutetiensis	2	**HCD23-5**
Fusobacteria	Fusobacteriaceae	*Fusobacteriaceae bacterium*	1	DJF_B254*
Proteobacteria	Enterobacteriaceae	*Escherichia coli*	6	50 **ISO3** INTA8 **E6** **TU-6**
Unclassified	Unclassified	*Bacterium*	2	NLAE-zl-P727 H-9137

*96% of identity

Note: Apparent strains in bold present 100% of identity with their closest relatives in database, while all the others present 99% of sequence similarity. This applies to all tables presented in the Appendix C (Table C.1 to Table C.6), excepting strains highlighted by the *symbol which present different percentages indicated in each table when applicable.

Table C.2 - Summary of the different phyla, families, species, apparent strains and number of isolates recovered from patient CRC057

Phylum	Family	Species	#	Apparent Strains
Bacteroidetes	Bacteroidaceae	*Bacteroides vulgatus*	1	JCM 5826
	Tannerellaceae	*Parabacteroides distasonis*	1	JCM 1294
Firmicutes	Bacillaceae	*Bacillus circulans*	1	PAM11B
	Erysipelotrichaceae	*Clostridium innocuum*	1	I46
	Clostridiaceae	*Clostridium perfringens*	1	18115
	Lachnospiraceae	*Blautia wexlerae*	1	AUH-JLD56
		Coprococcus comes	1	ATCC 27758
		Dorea longicatena	1	JCM 11232
	Lactobacillaceae	*Lactobacillus paracasei*	1	**HBUAS53106**
	Streptococcaceae	*Streptococcus anginosus*	1	ChDC B311
Fusobacteria	Fusobacteriaceae	*Fusobacterium necrophorum subsp. funduliforme*	1	1_1_36S
Proteobacteria	Enterobacteriaceae	*Escherichia coli*	26	50 NBRC 102203-96 **13A** E16 **E6** NBRC 102203-88 **TU-6** UCC 163
Unclassified	Unclassified	*Bacterium*	4	NLAE-zl-P727 NLAE-zl-P778 NLAE-zl-H411 **NLAE-zl-P911**

Table C.3 - Summary of the different phyla, families, species, apparent strains and number of isolates recovered from the oral swab sample of patient 1

Phylum	Family	Species	#	Apparent Strains
Actinobacteria	Actinomycetaceae	*Actinomyces sp.*	5	ICM54
		Actinomyces odontolyticus	3	14-211 GHTE7 CCUG 33767
	Atopobiaceae	*Atopobium parvulum*	3	ChDC B703 DSM 20469
	Dermacoccaceae	*Dermacoccus sp.*	1	Bma12
	Streptomycetaceae	*Streptomyces sp.*	1	MSP01SKC
Bacteroidetes	Prevotellaceae	*Prevotella melaninogenica*	1	**isolate 379N_14155**
Firmicutes	Bacillaceae	*Bacillus circulans*	1	LA396
	Bacillales Family XI. Incertae Sedis	*Gemella sp.*	3	ChDC B524 oral strain C24KA
	Lactobacillaceae	*Lactobacillus paracasei*	1	NWAFU1570
		Lactobacillus fermentum	2	DQ-4 **CAU:237**
	Streptococcaceae	*Streptococcus anginosus*	1	**clone CMC1237**
		Streptococcus intermedius	3	FDAARGOS_233
		Streptococcus mitis	4	Ejeff B6 MG3
		Streptococcus mutans	6	**clone QP057** H52 **KCOM 1054**
		Streptococcus parasanguinis	4	GIFU8099 C86 clone KZ028 ChDC B356
		Streptococcus salivarius	12	**isolate 278** isolate 263
		Streptococcus sp.	6	C2 10aMclG2 M23-2014 **FDAARGOS_192**
		Uncultured Streptococcus sp.	1	clone 030-42
	Veillonellaceae	*Veillonella dispar*	2	**isolate W297T_11223** isolate 377N_9424
		Veillonella parvula	10	1010-16111 **1004-13187** UTDB1-3 ChDC B271
		Veillonella sp.	1	OK8
Fusobacteria	Fusobacteriaceae	*Fusobacterium nucleatum subsp. Animalis*	1	KCOM 1279
Proteobacteria	Pasteurellaceae	*Haemophilus parainfluenzae*	1	isolate 387N_16932*
	Yersiniaceae	*Serratia nematodiphila*	1	**MU1**

*98% of identity

Table C.4 - Summary of the different phyla, families, species, apparent strains and number of isolates recovered from the tumour sample of patient 1

Phylum	Family	Species	#	Apparent Strains
Actinobacteria	Bifidobacteriaceae	*Bifidobacterium longum*	1	DMF1
	Coriobacteriaceae	*Collinsella aerofaciens*	2	JCM 10790
	Micrococcaceae	*Kocuria rhizophila*	1	**SL-08**
		Micrococcus sp.	1	**ZJ3005**
	Propionibacteriaceae	*Cutibacterium acnes*	1	**1202-14014**
Bacteroidetes	Bacteroidaceae	*Bacteroides fragilis*	56	W299-5 1303-01870 clone M40.8 W299-4
Firmicutes	Clostridiaceae	*Clostridium hathewayi*	2	1313
	Enterococcaceae	*Enterococcus faecalis*	4	**NR-IF7** **HCD42-1** **H25**
	Erysipelotrichaceae	*Erysipelatoclostridium ramosum*	8	JCM 1298 SRB509-5-F-B
		Holdemanella biformis	1	DSM 3989*
	Peptoniphilaceae	*Finegoldia magna*	1	HF3
		Finegoldia sp. feline oral taxon 341	1	7209
		Peptoniphilus coxii	2	RMA 16757*
	Streptococcaceae	*Streptococcus intermedius*	1	FDAARGOS_233
Proteobacteria	Enterobacteriaceae	*Escherichia coli*	4	ISO3 **NBRC 102203-87** **HT073016**
Unclassified	Unclassified	*Bacterium*	4	**NLAE-zl-H437** IARFR2395 **NLAE-zl-P911**

**Holdemanella biformis* – 98% of identity

**Peptoniphilus coxii* – 97% of identity

Table C.5 - Summary of the different phyla, families, species, apparent strains and number of isolates recovered from the oral swab sample of patient 2

Phylum	Family	Species	#	Apparent Strains
Actinobacteria	Actinomycetaceae	*Actinomyces sp.*	2	ICM39 A11
	Micrococcaceae	*Kocuria rhizophila*	1	KP135f
		Micrococcus sp.	1	CLC-F21
		Rothia dentocariosa	7	Inje L1 PG1B TRO214
		Rothia mucilaginosa	1	isolate 183T_2929
	Propionibacteriaceae	*Cutibacterium avidum*	3	**AN18** 44067
		Propionibacterium sp.	1	HKG209
Firmicutes	Bacillaceae	*Bacillus sp.*	1	PS*
	Enterococcaceae	*Enterococcus faecalis*	1	HCD42-1
	Bacillales Family XI. Incertae Sedis	*Gemella sanguinis*	1	NTUH_8428
		Gemella sp. oral	1	C24KA
	Lactobacillaceae	*Lactobacillus antri*	2	Kx146A4
		Lactobacillus casei	1	**NWAFU1574**
		Lactobacillus crispatus	5	**G-OK** **HBUAS53138** K3-1-3
		Lactobacillus delbrueckii subsp. bulgaricus	2	LB2
		Lactobacillus fermentum	8	FQ002 **SM39** **CAU2163** Y227D **B35** **PLc35A** DRZ87
		Lactobacillus gasseri	8	**HCD34-5** JCM 5343 BOV
		Lactobacillus harbinensis	4	MIV7MA1
		Lactobacillus oris	2	R12.19
		Lactobacillus paracasei	2	**BIOUAL009** KF83
		Lactobacillus rhamnosus	11	**HCD33-1** **CMGB-L10** Z5 **W120** **AF3G** LABIB1
		Lactobacillus salivarius	13	**358** **DJ-sa-01** C1-1-3 s4 HBUAS54082 DSPV011P
		Lactobacillus vaginalis	2	**C107**
	Staphylococcaceae	*Staphylococcus hominis*	2	**L0020-01R** **K23**
		Staphylococcus pasteuri	1	H23
		Staphylococcus saprophyticus	2	FDAARGOS_355 **HBUM07055**
		Staphylococcus sp.	2	**jxk1-6** ZSC
		Staphylococcus xylosus	1	B36
	Streptococcaceae	*Okadaella gastrococcus*	2	MT/JULY2010
		Streptococcus anginosus	1	isolate 522N_1587
		Streptococcus intermedius	2	F0395
		Streptococcus mutans	3	**P20** ChDC YM51
		Streptococcus parasanguinis	1	PTO10
		Streptococcus salivarius	31	NCTC7366 **B0023-01R** **isolate 278** isolate 263 B0023-01F B0021-04R **K12** YL456 MYZI-K2
		Streptococcus sp.	3	C150
Proteobacteria	Enterobacteriaceae	*Cronobacter sakazaki*	1	**PM468**
		Escherichia coli	1	RM9975
	Neisseriaceae	*Eikenella sp.*	1	NML130454
	Yersiniaceae	*Serratia marcescens*	1	MI17

*95% of identity

Table C.6 - Summary of the different phyla, families, species, apparent strains and number of isolates recovered from the tumour sample of patient 2

Phylum	Family	Species	#	Apparent Strains
Bacteroidetes	Tannerellaceae	*Parabacteroides distasonis*	1	JCM 13400
	Bacteroidaceae	*Bacteroides vulgatus*	1	mpk
Actinobacteria	Bifidobacteriaceae	*Bifidobacterium adolescentis*	6	**S52** 1-11 DMF2
		Bifidobacterium angulatum	1	GT102
		Bifidobacterium breve	2	**TPY2-3**
		Bifidobacterium longum	4	**HCD15-7**
		Bifidobacterium scardovii	1	JCM 12489
		Bifidobacterium sp.	4	S35 isolate ASS Bif
	Coriobacteriaceae	*Collinsella aerofaciens*	2	JCM 10803
	Micrococcaceae	*Kocuria rhizophila*	1	SL-08
Firmicutes	Enterococcaceae	*Enterococcus avium*	3	HCD36-3 HCD19-2
		Enterococcus casseliflavus	2	EC1
		Enterococcus faecalis	2	SP038 HCD42-1
		Enterococcus faecium	4	URL L2 **G3**
		Enterococcus gallinarum	4	HCD7-6
		Enterococcus sp.	2	**feacium URLL2**
	Lactobacillaceae	*Lactobacillus acidophilus*	1	**MB422**
		Lactobacillus crispatus	4	HBUAS53138 K3-1-3
		Lactobacillus delbrueckii subsp. bulgaricus	1	SM52
		Lactobacillus fermentum	6	**CAU2163** DRZ87 HBUAS54091 B35
		Lactobacillus gasseri	3	**JCM 5343** HCD34-5 BOV
		Lactobacillus oris	1	**J-1**
		Lactobacillus rhamnosus	7	**HCD33-1** **LABIB1** isolate hsryfm 1301 CMGB-L10
		Lactobacillus salivarius	15	**HBUAS54082** **2-5** **358** **C1-1-3**
		Lactobacillus sp.	1	**UVAS:RY2**
	Staphylococcaceae	*Staphylococcus hominis*	1	CLC-M5
	Streptococcaceae	*Streptococcus mutans*	3	**P20** LAR01
		Streptococcus pasteurianus	14	**HCD36-6**
		Streptococcus salivarius	1	B0023-01R
Proteobacteria	Enterobacteriaceae	*Cronobacter sakazakii*	2	G220 PM468
		Enterobacter sp.	1	**TA11**
		Escherichia coli	16	**RM9975** 2014C-3075 **AR_452** CCUG 70745 E2 **E23** VGC69*
		Escherichia fergusonii	2	LWIS 6 CGS25
		Klebsiella pneumoniae	4	MS1.5 11219
		Klebsiella sp.	1	CCFM8376
	Yersiniaceae	*Serratia marcescens subsp. sakuensis*	1	RY21
Unclassified	Unclassified	*Bacterium*	1	NLAE-zl-H479

*98% of identity

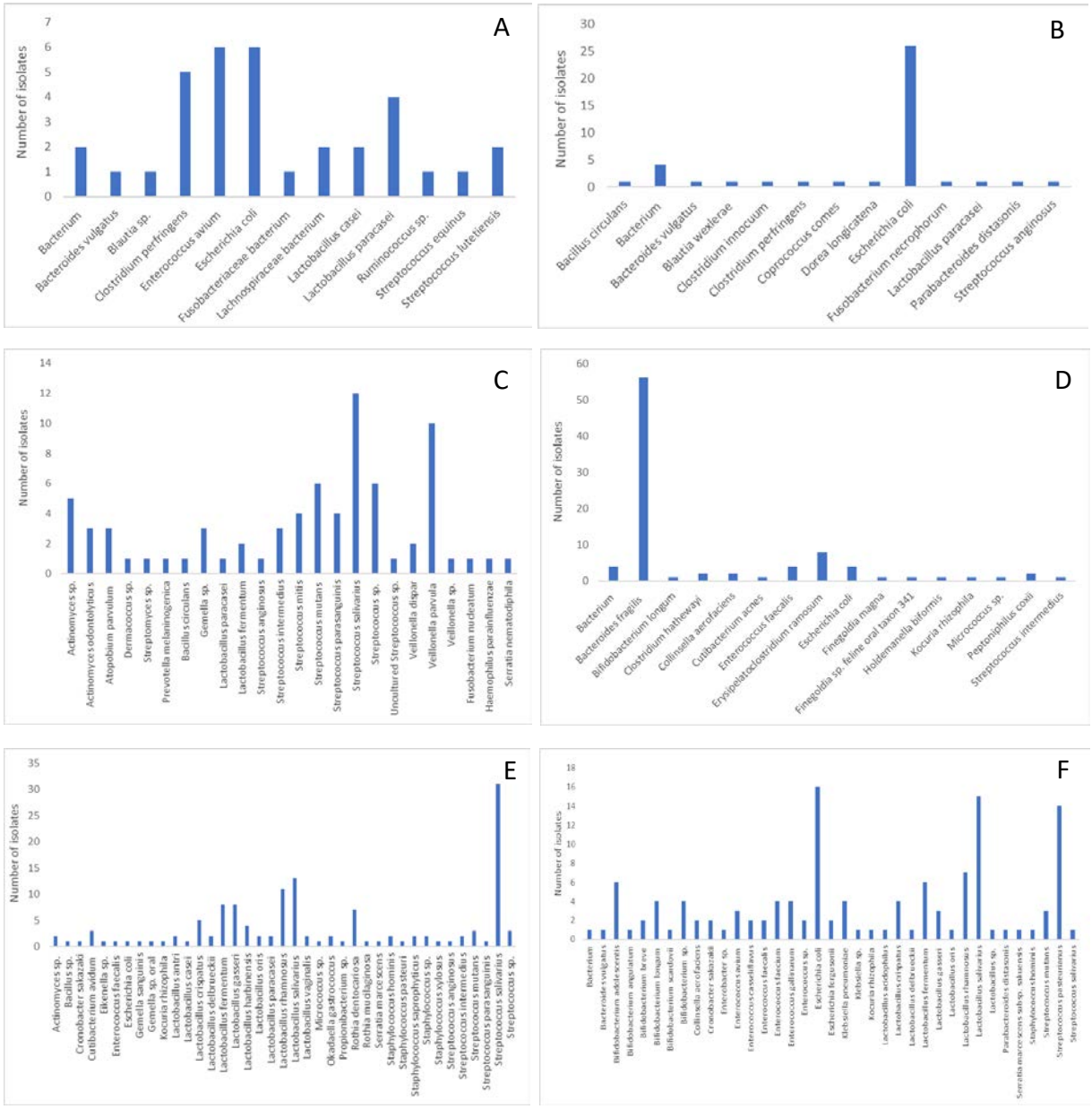


Figure C.1 – Quantitative analysis of isolates recovered from the four CRC patients. Isolates and respective species from: A - Faecal sample of patient CRC038; B – Faecal sample of patient CRC057; C – Oral swab sample of patient 1. D – Tumour sample of patient 1; E – Oral swab sample of patient 2; and F – Tumour sample of patient 2.

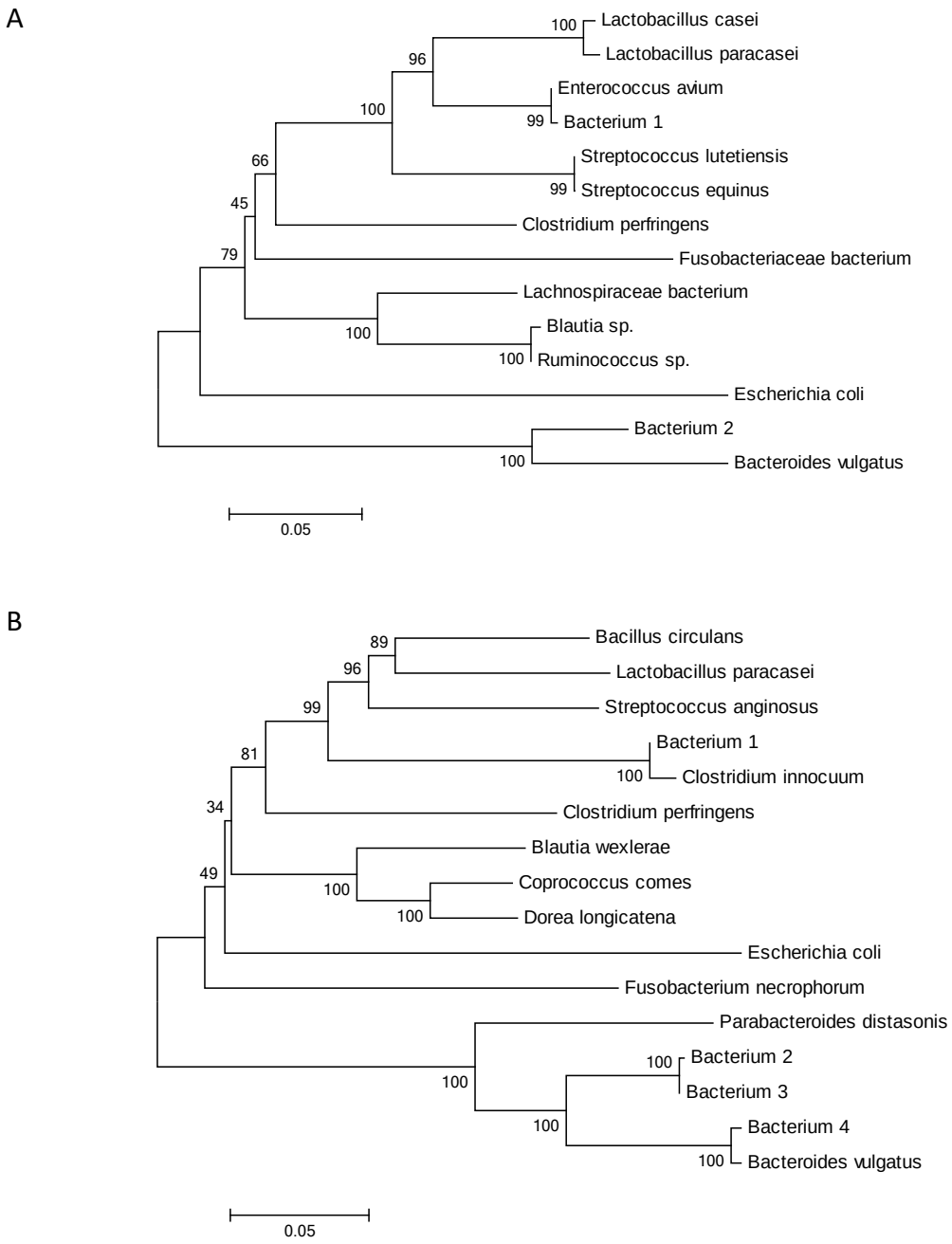
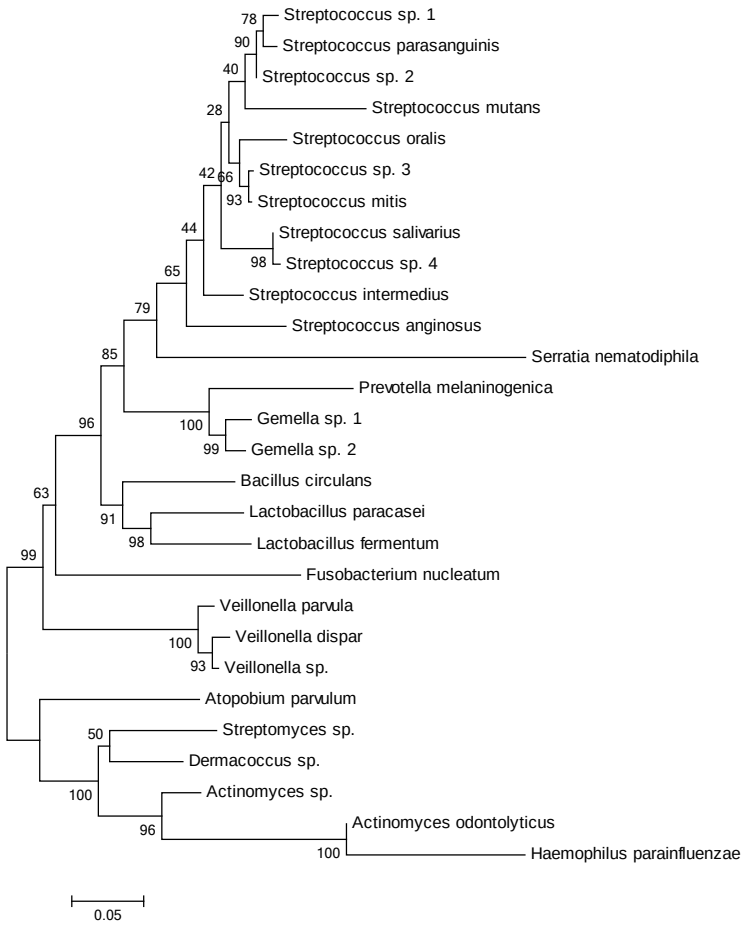


Figure C.2 – Evolutionary relationships of taxa (patients CRC038 and CRC057). Original phylogenetic trees plotted by species with node statistics and original branch lengths, from faecal samples of patient CRC038 (A) and patient CRC057 (B).

Note: The statistical method used was the Neighbor-Joining method, the test of phylogeny selected was the Bootstrap method with 1000 Bootstrap replications, and the substitution model used was the Maximum Composite Likelihood. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Evolutionary analyses were conducted in MEGA 6 and MEGA X. This applies to all phylogenetic trees presented in the Appendix C (Figure C.2 to Figure C.6).

A



B

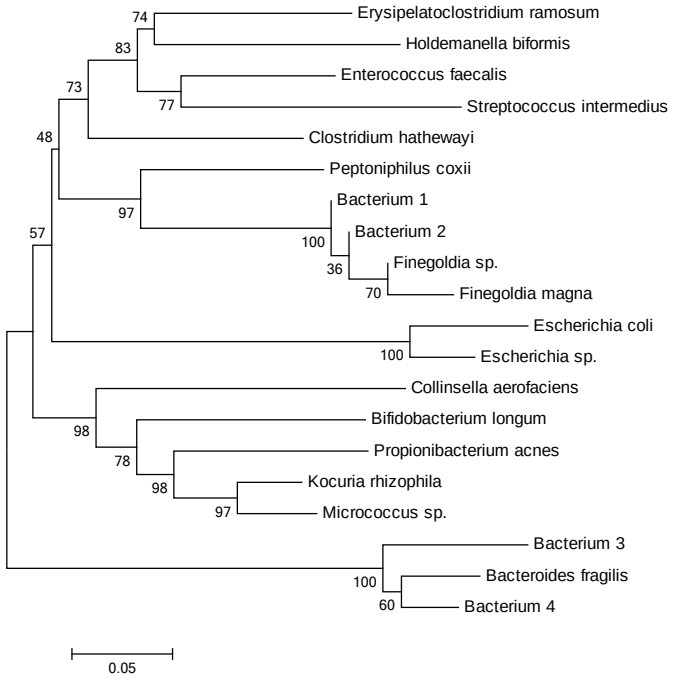


Figure C.3 – Evolutionary relationships of taxa (patient 1). Original phylogenetic trees plotted by species with node statistics and original branch lengths, from oral swab sample (A) and tumour sample (B) of patient 1.

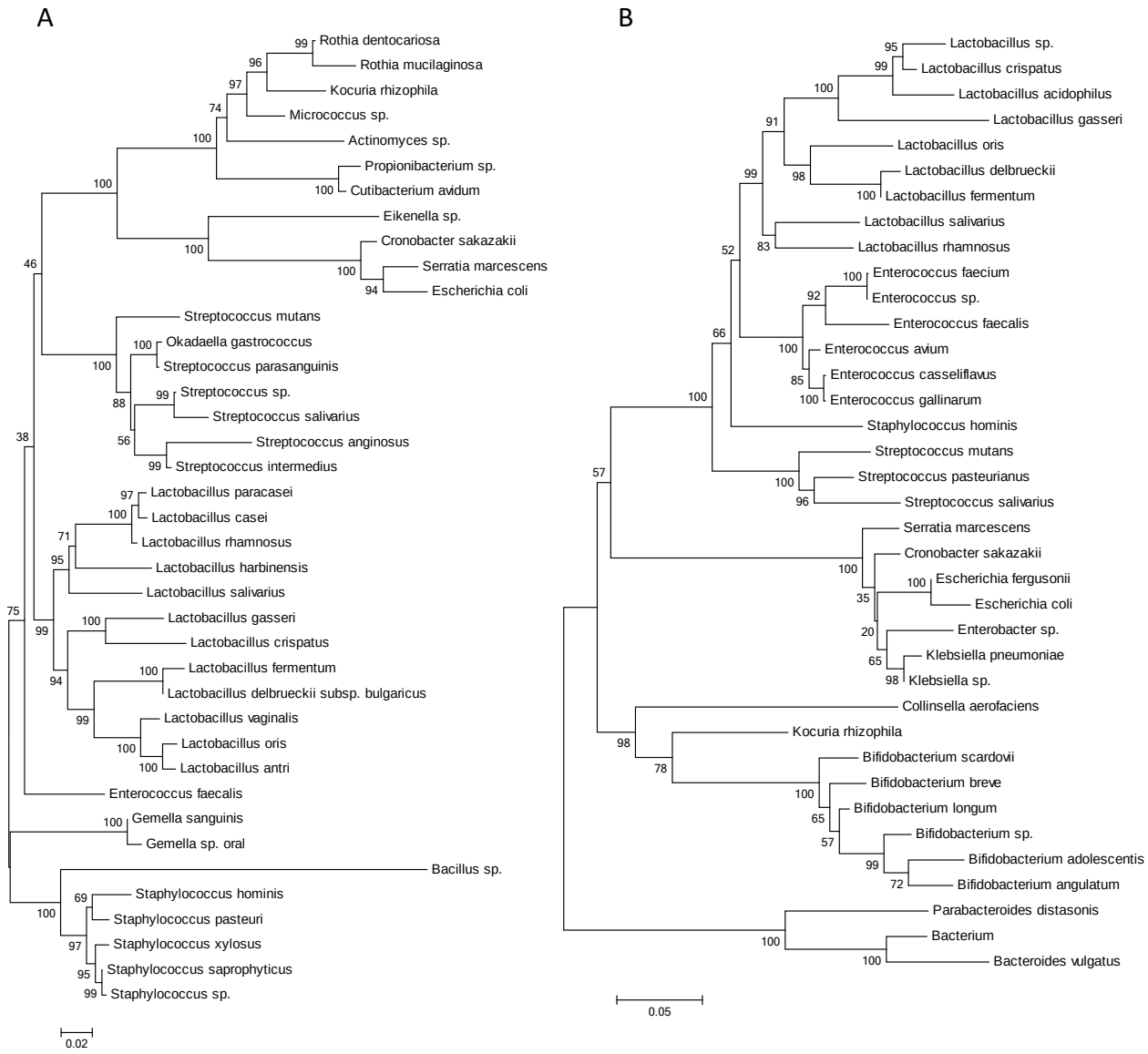
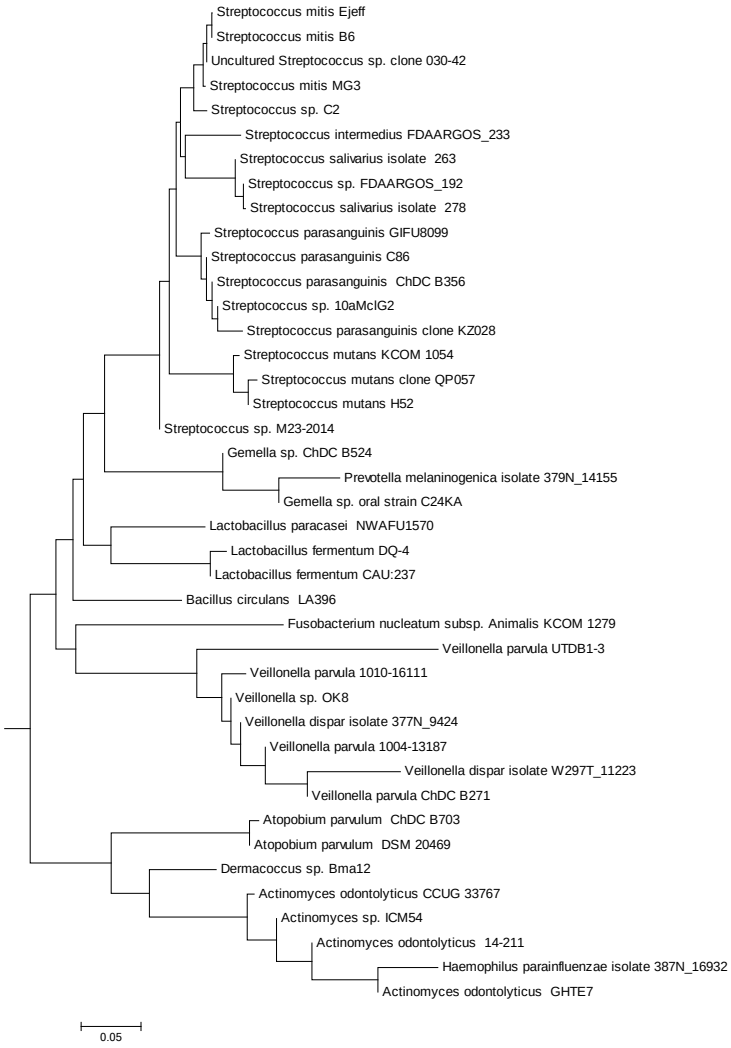


Figure C.4 – Evolutionary relationships of taxa (patient 2). Original phylogenetic trees plotted by species with node statistics and original branch lengths, from oral swab sample (A) and tumour sample (B) of patient 2.

A



B

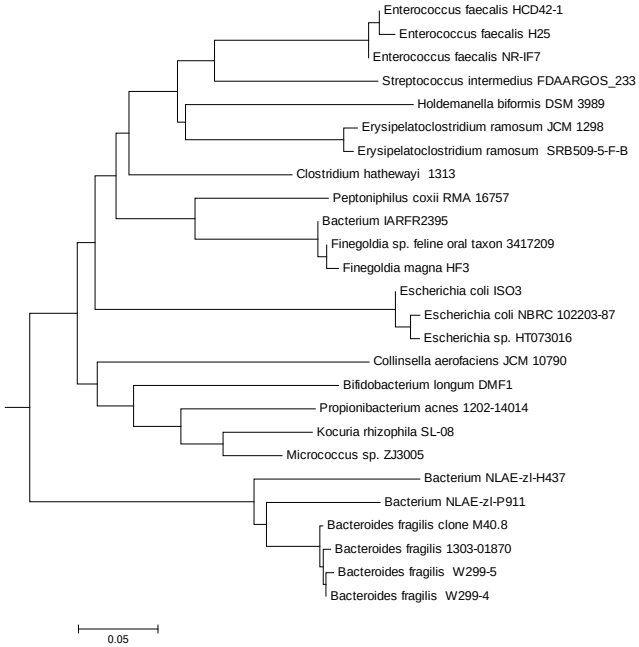


Figure C.5 – Evolutionary relationships of taxa (different strains of patient 1). Original phylogenetic trees plotted by apparent strain level with original branch lengths, from oral swab sample (A) and tumour sample (B) of patient 1. Only one of the many isolates for each apparent strain is presented here.

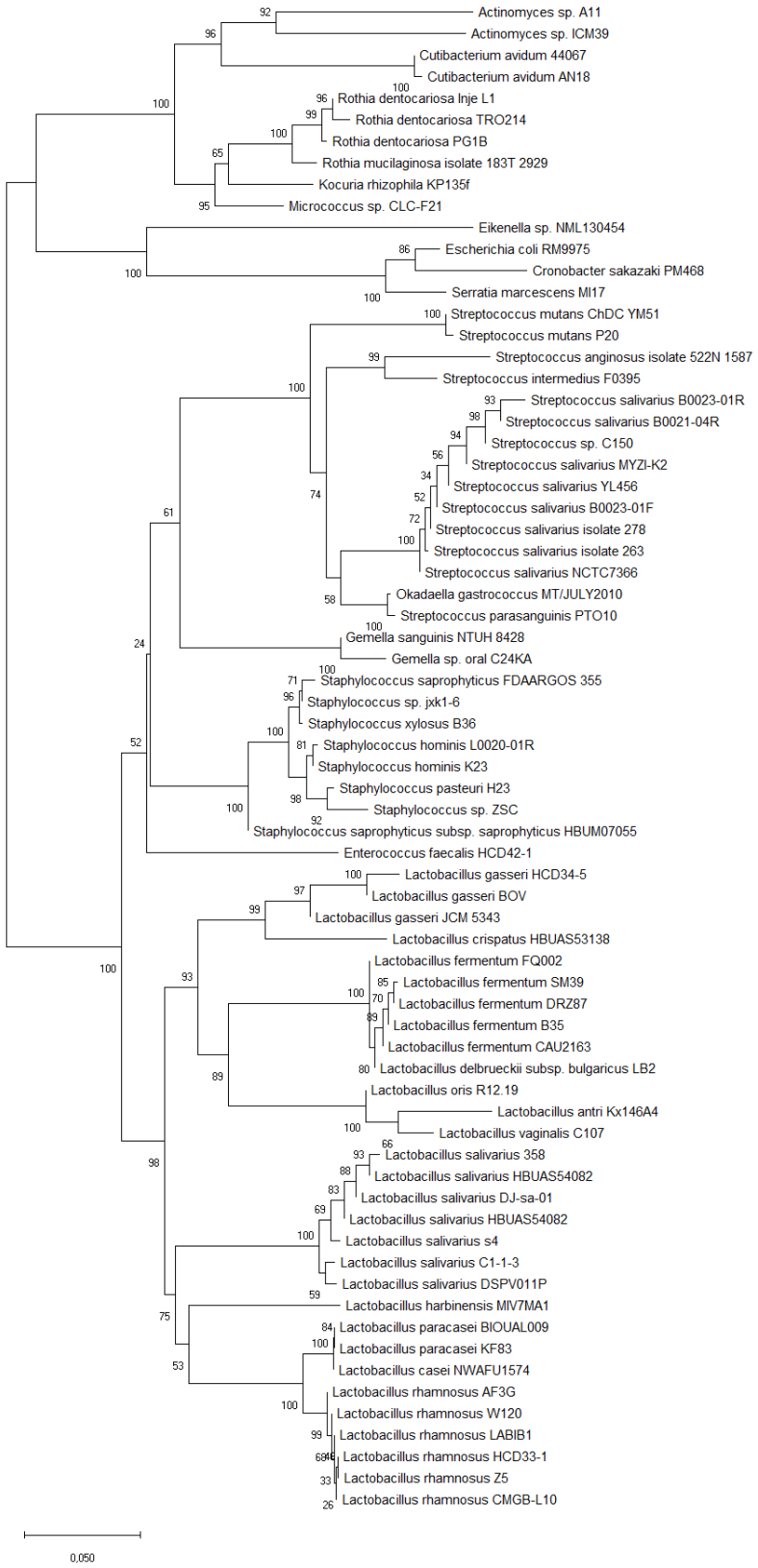


Figure C.6 – Evolutionary relationships of taxa (different oral strains from patient 2). Original phylogenetic trees plotted by apparent strain level with node statistics and original branch lengths, from oral swab sample of patient 2. Only one of the many isolates for each apparent strain is presented here.

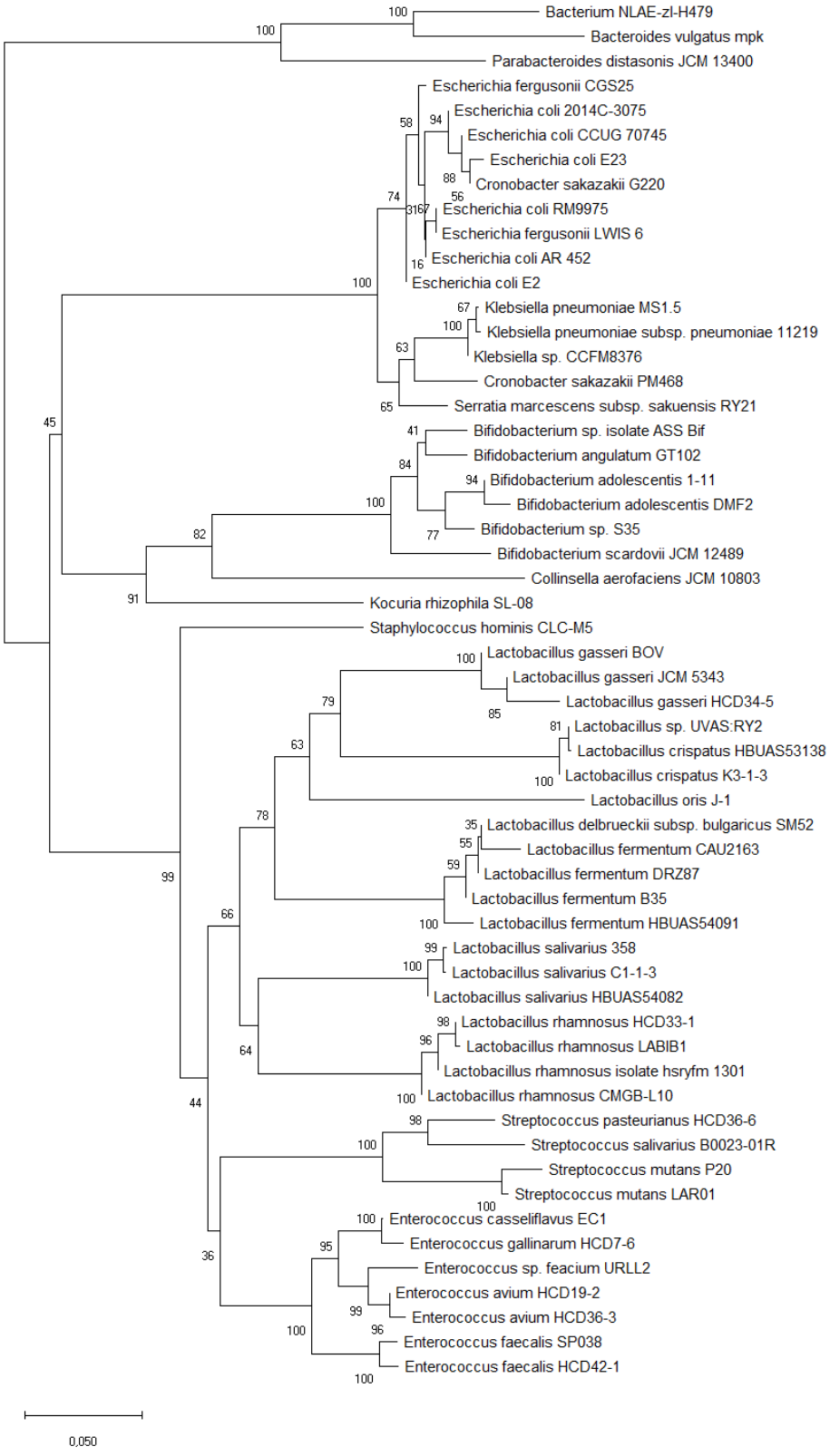


Figure C.7 – Evolutionary relationships of taxa (different tumoral strains from patient 2). Original phylogenetic trees plotted by apparent strain level with node statistics and original branch lengths, from tumour sample of patient 2. Only one of the many isolates for each apparent strain is presented here.

Appendix D

Closeup of CRC Gram-stained slides

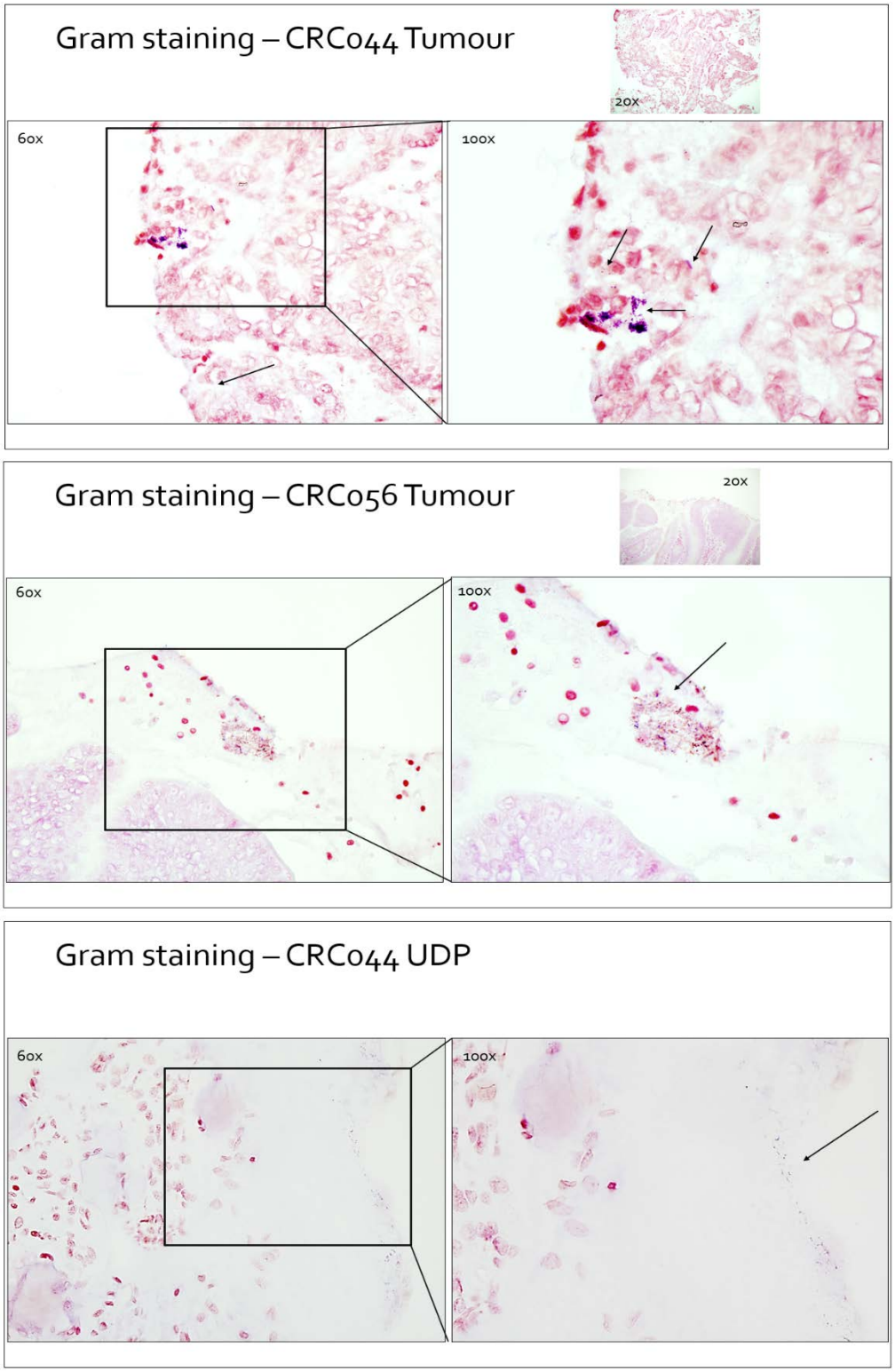


Figure D.1 – Extra Gram-stained slides of patients CRC044 Tumour, CRC056 Tumour and CRC044 UDP.

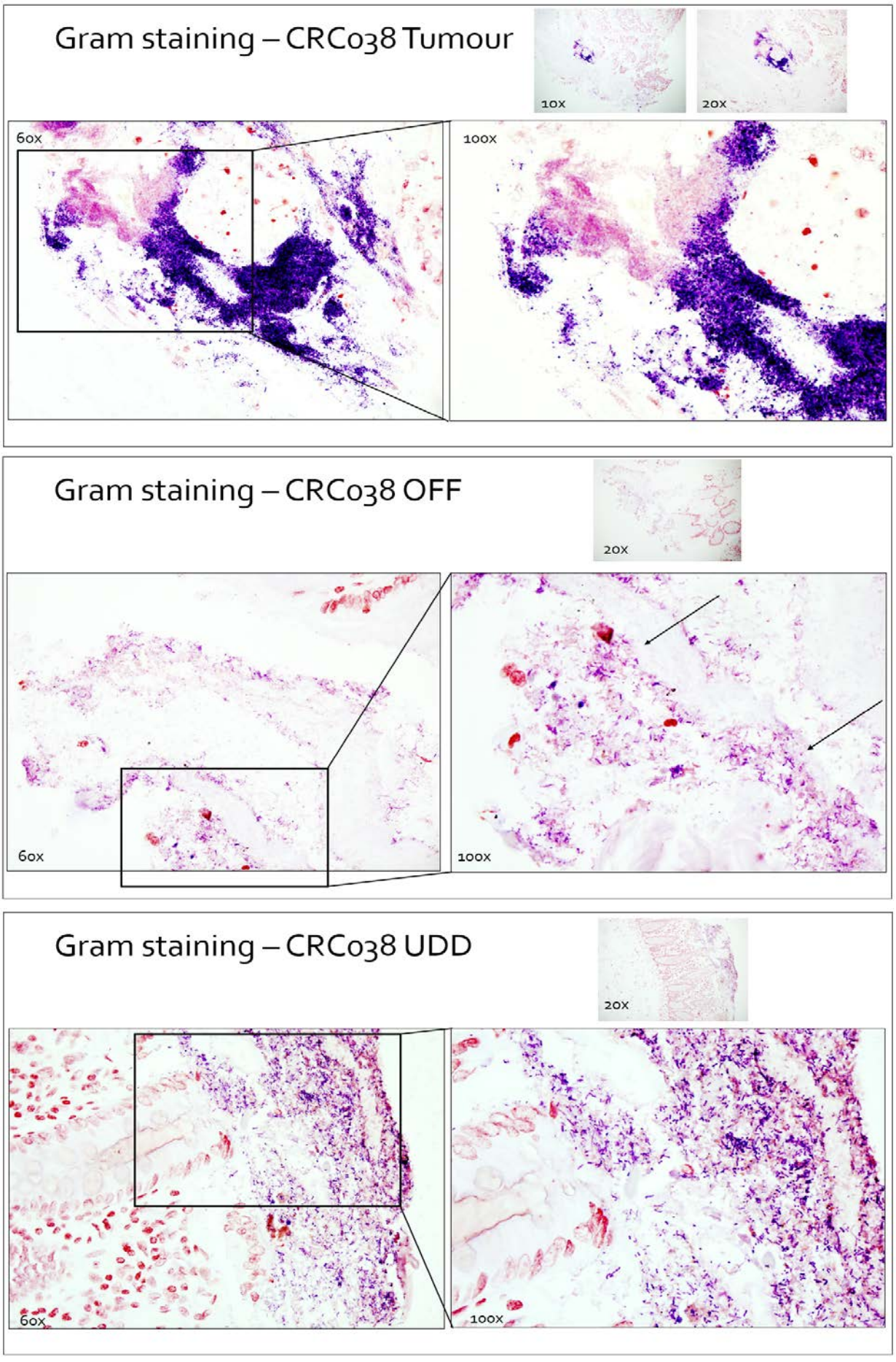


Figure D.2 – Gram-stained slides of patient CRC038 (Tumour, OFF and UDD samples).

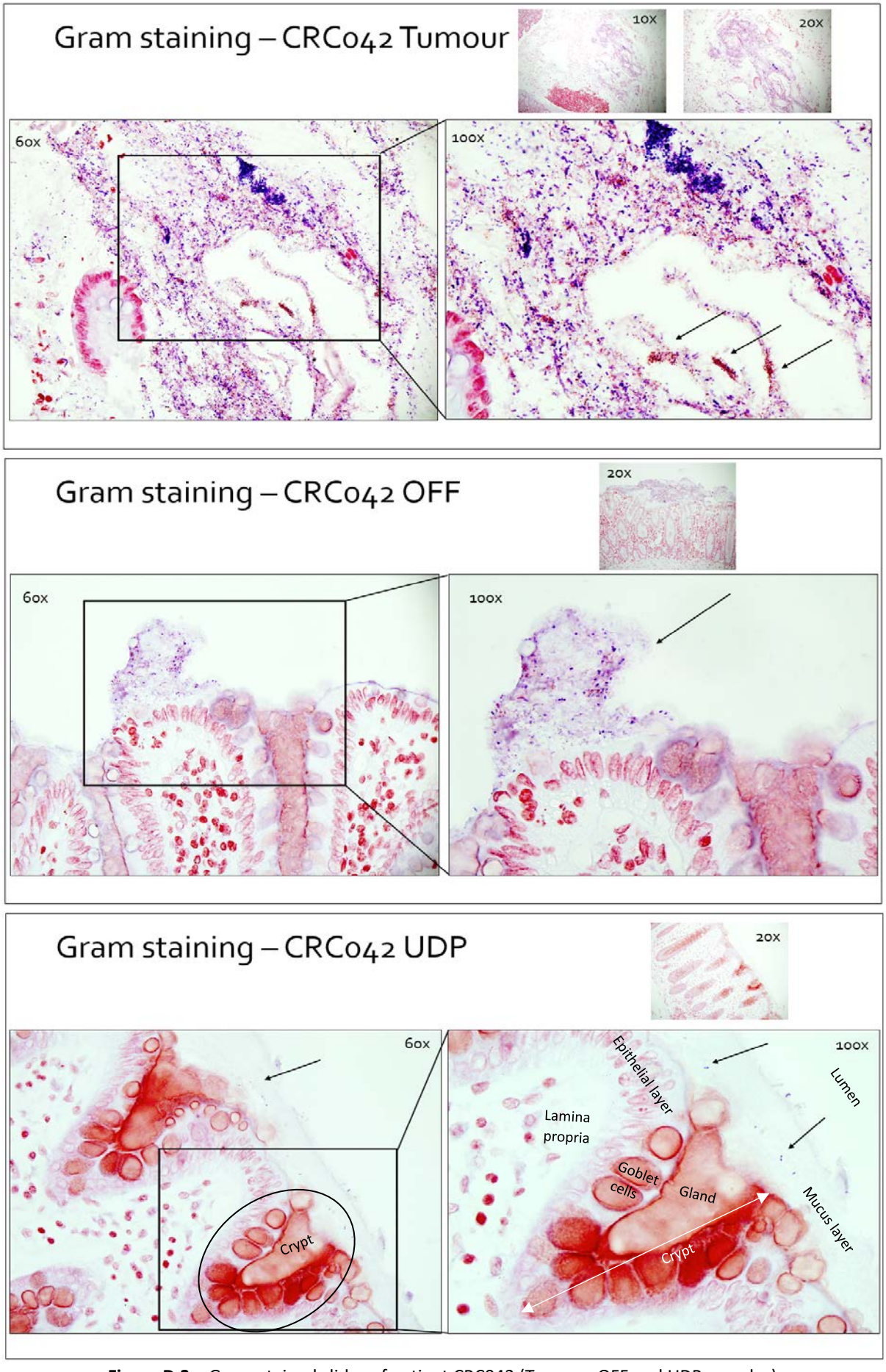


Figure D.3 – Gram-stained slides of patient CRC042 (Tumour, OFF and UDP samples).

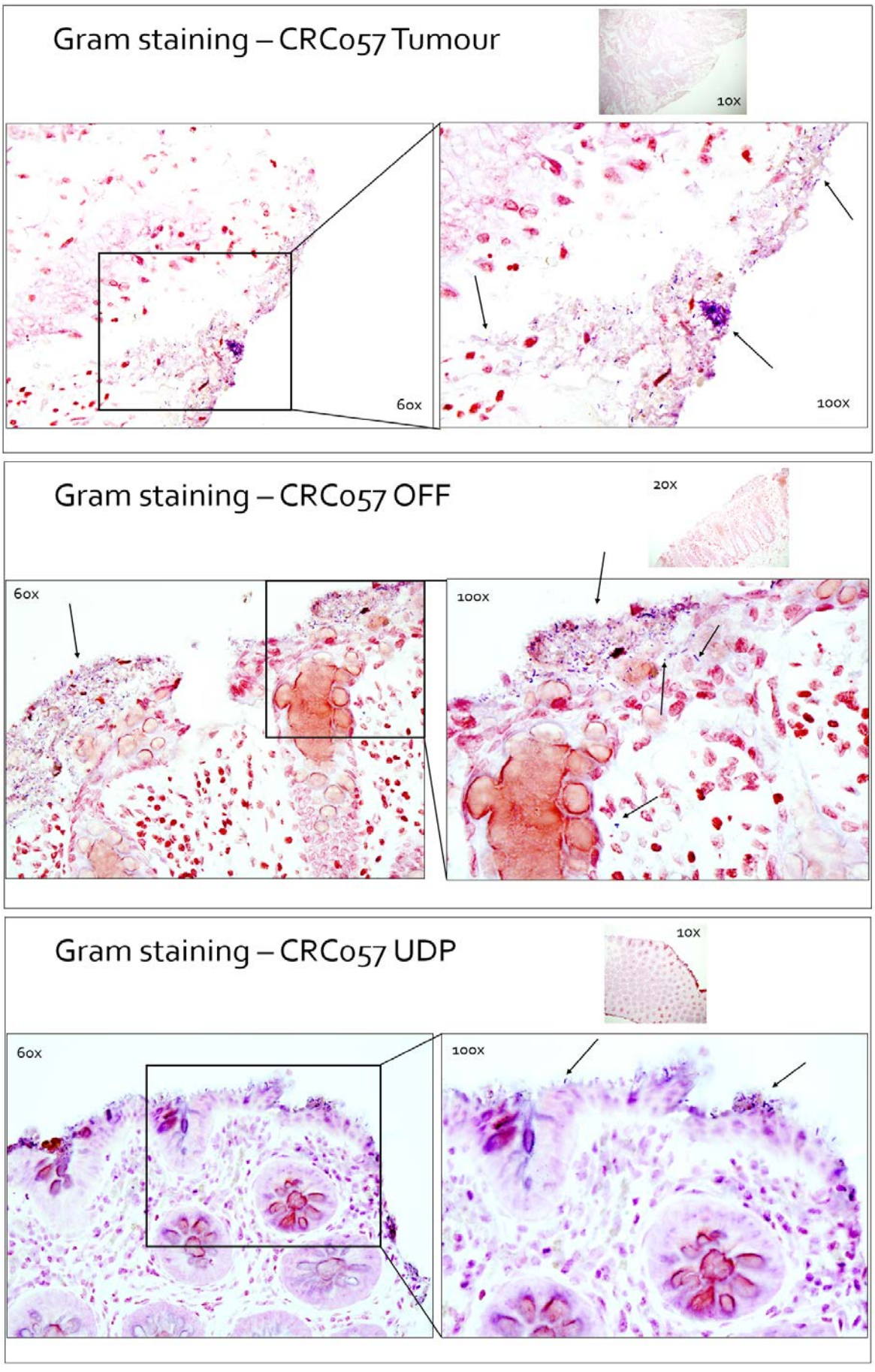


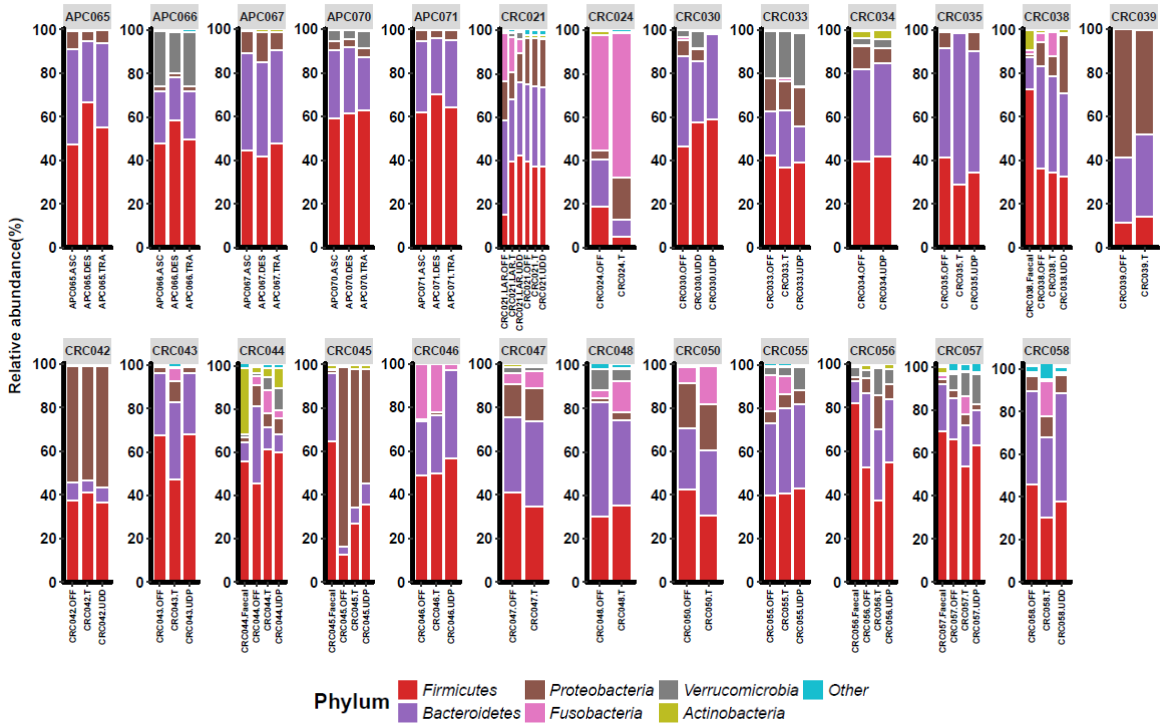
Figure D.4 – Gram-stained slides of patient CRC057 (Tumour, OFF and UDP samples).

Appendix E

A

Phylum/Family	Pat1 Tumour	Pat1 Oral	Pat2 Tumour	Pat2 Oral
Actinobacteria/Bifidobacteriaceae	0,216	0,000	31,705	0,000
Actinobacteria/Coriobacteriaceae	1,029	0,000	0,219	0,000
Bacteroidetes/Bacteroidaceae	34,651	0,000	3,420	0,010
Bacteroidetes/Flavobacteriaceae	0,000	1,955	0,000	0,162
Bacteroidetes/Porphyromonadaceae	3,928	0,000	0,307	0,022
Bacteroidetes/Prevotellaceae	4,551	0,963	4,172	0,604
Bacteroidetes/Rikenellaceae	2,763	0,000	0,092	0,000
Firmicutes/Acidaminococcaceae	0,000	0,000	1,300	0,000
Firmicutes/Bacillales_Incertae_Sedis_XI	2,667	6,067	0,002	0,164
Firmicutes/Clostridiales_Incertae_Sedis_XI	3,280	0,012	0,000	0,002
Firmicutes/Erysipelotrichaceae	3,183	0,000	0,988	0,000
Firmicutes/Lachnospiraceae	18,064	0,007	6,785	0,020
Firmicutes/Lactobacillaceae	0,009	0,012	4,502	62,699
Firmicutes/Peptostreptococcaceae	1,170	0,000	0,035	0,000
Firmicutes/Ruminococcaceae	8,935	0,000	12,446	0,000
Firmicutes/Streptococcaceae	0,411	56,315	9,797	34,813
Firmicutes/Veillonellaceae	2,171	1,331	0,354	0,313
Fusobacteria/Fusobacteriaceae	8,991	0,200	0,019	0,030
Proteobacteria/Enterobacteriaceae	1,595	0,002	19,019	0,000
Proteobacteria/Neisseriaceae	0,068	2,463	0,000	0,067
Proteobacteria/Pasteurellaceae	0,000	29,945	0,108	0,010
Proteobacteria/Sutterellaceae	0,218	0,000	4,080	0,000
Total of different families	19	12	19	13

B



C

Perfect match	**Human**	7	15	167	211	253	301	374	451	638	683		**Total: 10**																																					
Probable match	**Microbiome**	3	5	8	9	13	15	18	23	26	64	65	74	108	112	119	167	172	222	245	251	253	261	301	350	398	404	406	430	451	528	611	638	683	837	933		**Total: 35**									
Same species	**Project_OTU**	5	9	29	62	65	68	108	112	120	152	162	211	222	230	241	260	280	301	331	377	408	486	509	522	638	746	837	974	1148	1150	1223		**Total: 31**													

Figure E.1 – Resume of results from metagenomic analysis and positive matches with Human Microbiome Project's 'Most Wanted' list. A – Relative abundance (in%) of the taxa (family level) obtained from the metagenomic analysis of patient 1 and 2, and used to plot the bar charts in Figure 3.12.B – where only relative abundance ≥ 1% was displayed. B – Microbiota profile (phylum level) of 5 healthy individuals and 20 CRC patients. C – OTU numbers of the most wanted sequence that has matched with cultured isolates. Total of different perfect and probable matches, and same species is indicated. The OTU highlighted in orange correspond to medium priority uncultured and unsequenced microbes, while all the others correspond to low priority.