Exploring the gut microbiota of breast cancer patients

Date of submission: 30 June 2023

Author: Nancy MY Teng

University of East Anglia

A thesis submitted for the degree of Doctor of Philosophy

Abstract

Host-associated microbial communities play a key role in health and disease, and more recently there has been a growing appreciation for how particular microbes and microbial 'signatures' are associated with different cancers. However, breast cancer remains an understudied cancer type, and there is a pressing need to define, if and how, the gut microbiota maybe be linked to disease progression and treatment outcomes.

To investigate the gut microbiota and breast cancer, two clinical cohorts were profiled (using a range of sequencing and bioinformatics approaches) and additional mechanistic *in vitro* and *in vivo* studies were also undertaken. First, a local Norfolk cohort was established – BEAM, with the aim of longitudinally profiling newly diagnosed breast cancer patients (1 control and 35 breast cancer patients, as of 30 June 2023), however study recruitment was severely impacted due to the SARS-Cov-2 pandemic. My initial analysis indicated no significant shifts in microbiome profiles in the limited number of patients profiled, however I was able to establish a large culture collection through untargeted culturing. I obtained 298 strains from 50 different species which were whole genome sequenced and phylogenetically characterised. This work also led to the discovery and detailed description of one novel genus and one novel species - *Allocoprobacillus halotolerans* gen. nov., sp. nov and *Coprobacter tertius* sp. nov.

Concurrent to BEAM, the oral and gut microbiota samples from a phase 2a clinical trial (KELLY) that had been completed were processed, sequenced, and analysed which led to the creation of the CALADRIO study. The KELLY trial had one arm where all patients received treatment, a chemotherapeutic and immunotherapeutic. Overall, treatment did not cause significant gut or oral microbiota perturbations, which is usually indicative of drug-related microbiota toxicity. Differential analysis indicated that clinical benefit was driven, in part, by gut-associated *Bacteroides fragilis*. Further *in vitro* studies indicated a product present in the cell-free supernatant of *B. fragilis* led to greater cellular stress in breast cancer cells, but it did not result in complete cell death.

Bifidobacterium, generally considered a beneficial gut-associated bacterium, was consistently in the top ten most abundant genera of the gut microbiota in the BEAM and CALADRIO study. Thus, to define if *Bifidobacterium* was mechanistically associated with breast cancer outcomes, a *Bifidobacterium longum* subsp. *longum* isolate was selected and used as a live oral supplementation in a murine breast cancer model that was also treated

with chemotherapy (cyclophosphamide). Oral supplementation resulted in larger primary tumours than cyclophosphamide alone suggesting that oral supplementation interfered with treatment efficacy. Genomic screening of the isolate showed that it possessed aldehyde dehydrogenase which is known to inactivate cyclophosphamide.

These data allowed me to explore how the gut microbiota of breast cancer patients may link to treatment outcomes and indicated both positive (e.g., *B. fragilis*) and negative (e.g., *B. longum* subsp. *longum*) impacts. Translating it into the clinic, such findings could provide avenues for improving efficacy of anti-cancer therapeutics. To test these further *in vivo* studies could be conducted to determine how candidate bacterial strains could influence the immune system in the context of breast cancer and building on those findings *in vitro* studies would investigate the intricacies of the gut-immune axis. Overall, my thesis outputs highlight the complex interactions between the microbiota and their host, and suggest new avenues for biomarker and therapy development, particularly in breast cancer.

Access Condition and Agreement

Declaration

The work presented in the thesis has been produced solely for the purpose of this PhD, therefore has not been presented prior to the undertaking of this thesis. All work performed in this thesis has been done by me and any collaborators have been appropriately acknowledged.

Bioinformatic analysis regarding 16S rRNA amplicon sequencing and whole genome sequencing was guided by Dr. Raymond Kiu (Quadram Institute Bioscience, QIB). Dr. Raymond Kiu has performed the shotgun metagenomic pipeline which was used for both the BEAM and CALADRIO project. Data visualisation was guided by Dr. Matthew Dalby (QIB) who provided R scripts for microbiota visualisation plots where upon I edited the script to suit the purposes of my thesis. Whole genome sequencing, 16S rRNA gene amplicon sequencing and library preparation for shotgun metagenomics was performed by the QIB sequencing team: Dr. David Baker, Ms. Rhiannon Evans and Mr. Steven Rudder.

Clinical samples as part of BEAM were collected by the clinical team at NNUH (Dr. Simon Pain, Dr. Katalin Zechmeister, Ms. Tracey Parker), JPUH (Dr. Ibrahim Sallam, Dr. Sandy Leeper, Ms. Wendy Harrison, Prof. Sue Downs) and the Biorepository (Dr. Mark Wilkinson, Dr. Rachael Stanley, Ms. Roxanne Brunton-Sim, Dr. Louise Jones, Mr. Joel Wood, Ms. Cheryl Prior). Clinical samples as part of the CALADRIO study were collected by the clinical team in MedSir, where the clinical team is acknowledged in [1].

Dr. Ibrahim Sallam (JPUH) helped in DNA extraction for the JPUH clinical samples under my instruction. The Robinson group (Dr. Stephen Robinson, Dr. Wesley Fowler, Dr. Sally Dreger, Mr. Christopher Price, Mr. Christopher Benwell, Mr. James Taylor, Ms. Alicia Nicklin, Mr. Luke Mitchell) kindly performed the *in vivo* experiment and handling of animals. Ms. Rhianna-lily Smith assisted with rounds faecal culturing under my supervision.

All major works including *in vitro* work i.e., microbiological assays, cell-culture, DNA extractions and bioinformatic analysis, unless otherwise stated, was done by myself as part of this thesis.

Acknowledgements

The past four years and a few months have been a journey. I am incredibly fortunate to have been under the mentorship and tutelage of my supervisors: Lindsay Hall and Stephen Robinson. When I first decided to undertake a PhD, one piece of advice I was told was that your supervisors will make or break your entire PhD experience. And I can say, without fail, that I won the lottery. Thank you both for sharing your insights and advice on how to navigate the confusing but exciting process of undertaking a PhD. Words cannot express how grateful I am for having the best mentors I could ask for. To both of you, I'm honoured to have been your PhD student.

Lindsay, you are an inspiration to me and aspiring female scientists around the world. Your mentorship in raising the next generation of scientists is your legacy and I'm honoured to be part of this. Thank you for your patience, for your wisdom and for showing me my value and maintaining my own integrity throughout the PhD. Whenever I felt lost throughout this process you were there to push me in the right direction, for this I am grateful. To Stephen, I know you from my undergraduate days. Back then, you had a reputation of being a supportive, fair and knowledgeable supervisor. Being adopted into your group, I can say that this is all true. You've been a source of calmness when I was second guessing my progress, the voice of reason when diplomatic relations got hairy and reassured me when my cells supposedly got infected (rest in peace MCF-7s).

Thanks to the Hall and Robinson lab members for reminding me there is a life outside of the PhD and making my time at Quadram something to remember. Raymond, thank you for training me in the wonderful world of bioinformatics, and for your patience with my many questions. Wes, thanks for letting me bounce off ideas and concerns, as well as mentoring me how to be one of the Robinsonians. Matthew for helping me navigate the world of R because it isn't fun doing this alone. Lastly, all the current and past members of the Robinson and Hall labs for keeping things light-hearted and lively when times were rough. It's been a ride especially with the pandemic, but we made it.

Thank you BigC for funding this PhD and for providing me with this amazing opportunity where I started my science journey. Thank you to QIB who funded my last 6 months of my PhD due to the COVID disruption. Thank you to the other members of QIB who have taken time to help me with my experiments, including the QIB Sequencing team and the amazing lab technicians. I am grateful to the collaborators that gave rise to the completion of this

Thank you to my family for the support to pursue my fascination of the sciences. My sister, for letting me complain about why anaerobic bacteria act like moody teenagers. My mom, for keeping me grounded when I was being irrational like my bacteria. And my dad, for reminding me not to work too hard- and comparing cells to porridge, yum…

Lastly, thank you Jan for being there with the wine and takeout pick-me-ups. I will never admit to this in person, but you were right to say that looking after yourself is the best thing you can do to succeed. Thank you for your patience with my inevitable romps and stomps when I was learning R. Despite my tantrums and woes, you were there giving support and celebrating my achievements no matter how small. For this I can't thank you enough.

Table of Contents

List of abbreviations

BrCa	Breast cancer
PoM	Post-menopausal
PeM	Pre-menopausal
YLL	Years of lives lost
TME	Tumour microenvironment
ER	Oestrogen receptor
PR	Progesterone receptor
HER2	Human epidermal growth factor receptor 2
HR	Hormone receptor
NICE	National institute for healthcare excellence
UK	United Kingdom
MDT	Multi-disciplinary team
qPCR	Quantitative polymerase chain reaction
PCR	Polymerase chain reaction
NMR	Nuclear magnetic resonance
M/S	Mass spectrometry
TLR	Toll-like receptor
IFN	Interferon
rRNA	Ribosomal ribonucleic acid
SCFA	Short-chain fatty acid
HMO	Human milk oligosaccharides
QC	Quality control
OTU	Operational taxonomic unit
ML	Machine learning
GF	Germ free
SPF	Specific pathogen free
FMT	Faecal microbiota transplant
NSCLC	Non-small cell lung carcinoma
CRC	Colorectal cancer
IBD	Inflammatory bowel disease
BEAM	Breast health and microbiota
NNUH	Norfolk and Norwich University Hospital
JPUH	James Paget University Hospital

HR	Hormone receptor
NRES	National Research Ethics Service
BR	Biorepository
BCN	Breast Cancer Now
PBS	Phosphate buffered saline
FBS	Foetal bovine serum
NBI	Norwich Bioscience Institute
HPC	High performance computing cluster
LEfSe	Linear discriminant analysis of effect size
CEG	Centre for Genomic Epidemiology
MAG	Metagenome assembled genomes
BHI	Brain heart infusion
YCFA	Yeast casitone fatty acids
LDH	Lactate dehydrogenase
CFU	Colony forming units
REC	Research ethics committee
HRA	Human Research Authority 2004
NHS	National Healthcare Service
GP	General practitioner
DPU	Day Procedure Unit
HTA	Human tissue act
PIS	Patient information sheet
ILC	Invasive lobular carcinoma
IDC	Invasive ductal carcinoma
NMDS	Non-metric distance scaling
ANI	Average nucleotide identity
dDDH	Digital DNA-DNA hybridisation
TYGS	Type strain genome server
PoCP	Percentage of conserved proteins
MIDI	Sherlock microbial identification system
FAME	Faty acid methyl esters
mBC	Metastatic breast cancer
PD-1	Programmed cell death protein 1
PFS	Progression free survival
CB	Clinical benefit

EoT	End of treatment
NLR	Neutrophil to lymphocyte ratio
CTCAE	Common terminology for adverse events
PSA	Capsular polysaccharide A
BBE	Bile esculin agar
CFS	Cell free supernatant
EV	Extracellular vesicles
IP	Intraperitoneal
ALD	Aldehyde dehydrogenase

List of Figures

List of Tables

List of peer-reviewed articles

- Teng, NMY, Price, CA, McKee, AM, Hall, LJ, Robinson, SD. Exploring the impact of gut microbiota and diet on breast cancer risk and progression. *Int. J. Cancer*. 2021; 1– 11. https://doi.org/10.1002/ijc.33496.

- Teng, N.M.Y., et al., Allocoprobacillus halotolerans gen. nov., sp. nov and Coprobacter tertius sp. nov., isolated from human gut microbiota. International Journal of Systematic and Evolutionary Microbiology, 2023. 73(7).

1   General Introduction

This chapter focuses on the general background of this PhD project, and my thesis. It will introduce breast cancer (BrCa) and the gut microbiota, followed by keystone studies that led to the initial hypothesis that the microbiota can influence cancer aetiology. Section 1.1.6 is part of the review I co-authored with Dr. Alaistar Mckee and Mr. Christopher Price, which was published in the *International Journal of Cancer* [2]. For this review I wrote section 2 and 3 titled: "BrCa risk factors: is the microbiota the missing link" and "Growing evidence linking the gut microbiota and BrCa" respectively.

1.1 Breast Cancer

This first part of the introduction will briefly provide statistics on the incidence of BrCa in the UK and globally. Following this I will introduce BrCa pathology in a clinical setting and conclude by introducing risk factors related to the development of BrCa.

1.1.1   Epidemiology

BrCa is a disease which primarily affects, but is not exclusive to, post-menopausal (PoM) women [3]. The outlook of a BrCa diagnosis is often associated with the stage at which diagnosis is made and despite the prognosis generally being quite good, it is also dependent on the subtype of BrCa [4]. Due to the heterogenous nature of BrCa, treatment and therapies can be ineffective or induce unwanted side effects e.g., neutropenia or intestinal mucositis. Research into BrCa is driven by improving the diagnostic tools as well as exploring new less-toxic therapies to treat the disease. Despite the overall positive prognosis of BrCa, health forecasting reports expect an increase in BrCa incidences in the next twenty years [5]. As such research into the field of BrCa is as important as ever to develop new innovative diagnostic and therapeutic tools to reduce the burden of disease.

1.1.1.1   *Disease burden*

BrCa is the most commonly diagnosed cancer type among women, and it is the leading cause of cancer-related mortalities in females [6]. Every year it is estimated that 8.6 million new cases of cancer are reported in women, of which 25% is BrCa alone [7]. In the UK, roughly 55,000 new cases were reported, in 2013-2015, with 11400 cases of BrCa-related mortality reported in 2014-2016 [4] . Early diagnosis often correlates with a good prognosis; individuals diagnosed at a pre-metastatic stage have more than 80% five-year survival rate. However, this figure drops dramatically to less than 20% if diagnosed at a metastatic stage (Figure 1:1) [8]. Crucially, the average cost of BrCa treatment increases relative to the stage

of diagnosis with stage 3 and 4 treatment cost nearly double of that at stage 1 (Table 1:1). Thus, for patient care and overall treatment costs there is a pressing need to develop new therapeutics to decrease disease incidence, especially past stage 2.
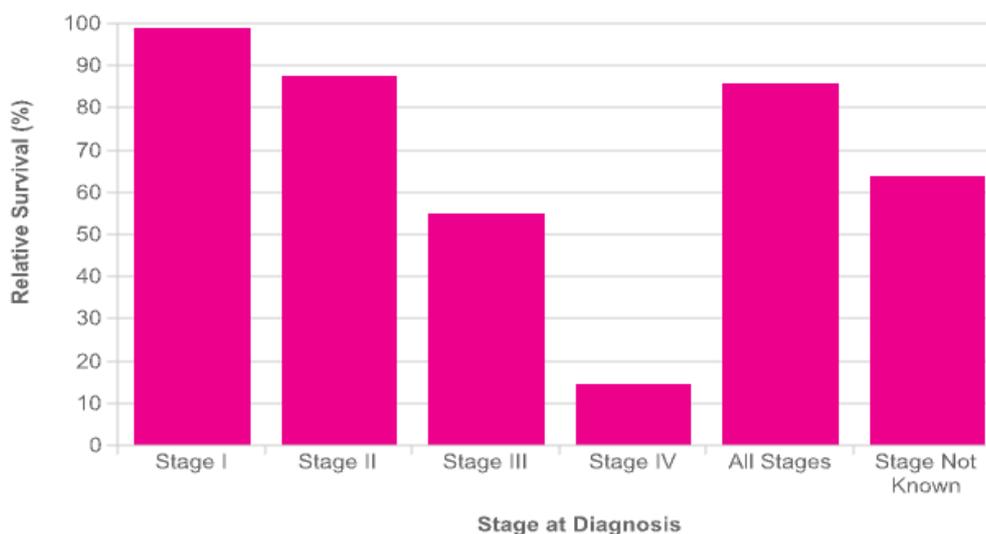
*Figure 1:1: **Five-year relative survival by stage in women (aged 51-99).** Data from Former Anglia Cancer network and taken from Cancer Research UK [4]. Statistics have been determined based on national cancer statistics across the UK and stratified according to primary cancer diagnosis, stage and mortality to name a few. "Stage not known" is cancer that has been found in a secondary site, but clinicians cannot determine the primary cancer type.*

Stage of diagnosis	Average cost (USD)	Average cost (GBP)	Increase (%) relative to stage 1
1	29 724	10 578	-
2	39 322	30 932	32
3	57 827	45 490	95
4	62 108	48 857	109

*Table 1:1 **Breast cancer treatment cost relative to stage of diagnosis.** Data was collected and reviewed by [9]. The figures were estimated using the FIGO staging system.*

1.1.1.2   *Global forecasts of breast cancer*

In 2018, the Lancet published an extensive study forecasting disease burden of several diseases, including BrCa. It is expected that BrCa cases will increase over the next 20 years [6, 10, 11] yet mortality will not. The expectation of increased incidence is largely driven by an ageing population. As BrCa is predominantly diagnosed in PoM women [3, 6] it is expected that as more women live longer and reach menopause, the incidence of BrCa should increase as well. Mortality is not expected to increase as much as incidence and this is largely due to improvements of diagnostics and therapies [10, 12]. Unfortunately, with the declaration of the COVID-19 pandemic many crucial diagnostic and screening processes were temporarily halted. Concerns arose over how this would impact cancer screenings and

ultimately cancer-related complications. As demonstrated in Figure 1:1, an earlier diagnosis is generally associated with a better prognosis [8]. The delays in routine appointments translate to a later diagnosis and consequently, possibly a worse prognosis [13-15]. During the national lockdown, only patients presenting with life-threatening symptoms related to BrCa would be referred to a clinician for further investigation [15]. Due to this, early symptoms of BrCa cases would have been missed and health care professionals project an increase in the number of preventable deaths due to BrCa in the future because of COVID-19. One study projects the number of years of lives lost (YLL) from excess deaths due to the delay in diagnosis as a result of the COVID-19 pandemic to increase by roughly 9%, compared to before the pandemic [15]. The pandemic has put clinicians under additional pressure to diagnose cases efficiently and accurately as they present itself. Therefore, having additional therapeutic or diagnostic tools at hand would benefit the healthcare system as well as patient care.

Epidemiological studies have observed an increased risk of BrCa associated with heightened socio-economic status [6, 10, 12]. BrCa has generally been prevalent in high-income "westernised" countries. With globalisation and modernisation of low-income countries, we observe a shift towards a "westernised" lifestyle including changes in diet and reproductive trends [6, 16]. This "western" lifestyle presents as a trend towards women having less children, having children later in life, experiencing a later menopause and spending less time breast feeding their children [6, 10, 12, 17, 18]. These are recognised risk factors for BrCa that will be discussed in further detail in section 1.1.6. A case-study example included researchers investigating Asian-American women and their offspring migrating from Asia to America and subsequent BrCa incidence. They observed that BrCa risk for future generations was determined by whether the individual came from rural or urban backgrounds and how long they lived in high-income countries. Although this study did not detail the reproductive information of these subjects, it is one of the earlier studies which demonstrated that BrCa risk is largely environmental as opposed to genetic [19], as the risk of BrCa increased generationally in this cohort. Should it be genetically driven, the risk should have been similar regardless of whether or not the individual came from a rural or urban background. A similar trend is currently being reported in urban Africa [16]. Women are moving towards having less children later, less breast-feeding time, later menopause and experiencing earlier menarche [16, 20]. All are known factors to influence BrCa risk in the future [3], and could explain the general trend of increasing incidence of BrCa globally. One thing to note is that technology has improved considerably which is not considered in these

studies. Improved diagnostics can increase the reported incidence of breast cancer which could influence the current trend being reported.

In the current climate there is also an economic incentive to reduce cost that healthcare services are currently experiencing, an example being the National Healthcare Service (NHS) for the UK. Therefore, research into BrCa is necessary to understand the aetiology and contributing factors to disease development that will affect so many women globally.

1.1.2   Pathophysiology of breast cancer

1.1.2.1   Hallmarks of cancer

Initiation of cancer involves a multi-step process; resisting cell death, sustained proliferative signalling, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality and inducing angiogenesis [21]. Generally, the process of cancer development includes a mutation in the cell allowing it to achieve the six criteria previously mentioned. Consequently, this cell will produce many clones of itself i.e., tumour clonality. Within this tumour community some cells will have a selective advantage over others as they have gained or lost some characteristics allowing enhanced immunosurveillance evasion and/or improved nutrient uptake. These cells will continue to grow and shape the tumour microenvironment [22].

1.1.2.2   Tumour microenvironment

The tumour microenvironment (TME) refers to the interface between cancer cells and the surrounding tissue including blood vessels, extracellular matrix, and immune cells. More recently, the extracellular matrix has been shown to be involved in cellular processes including growth, migration, and signalling. It has been noted that the extracellular matrix of a TME tends to be altered due to cancer-associated fibroblasts. Instead of a well organised meshwork, the extracellular matrix becomes disorganised and facilitates tumour growth [23, 24], and these changes can lead to abnormal cell signalling.

In BrCa, the TME is complex and specialised imaging revealed that the TME can exert selective pressures on malignant cells. The study identified ten structures associated with certain protein enrichments which can influence leucocyte infiltration, stromal quiescence or activation and vasculature infiltration. Depending on the TME structures and expression of proteins e.g., immune checkpoint over-expression, it can influence how cells organise and preferentially select for malignant cells [25, 26]. Normally, the myoepithelial cells are

located below the basement membrane and on top of luminal epithelial cells in the breast duct lobular unit. These cells are important for polarisation of luminal epithelial cells and correct orientation of the breast tissue [27, 28]. Abnormal myoepithelial cells by a compromised stroma can move into the physical barrier, allowing cancerous cells to invade neighbouring ducts resulting in invasive carcinoma (Figure 1:2A) [23, 29]. Consequently, the cells continue to grow and eventually become a tumour. If malignant cells grow inside of a duct but have not breached the basement membrane it is known as ductal carcinoma in-situ. This is an early stage of disease and can show up on mammograms due to the formation of microcalcifications within the duct. Once it has breached the basement membrane it is known as invasive ductal carcinoma, or ductal carcinoma, which accounts for 70-80% of diagnosed breast cancers [30]. For malignant cells growing in the lobe it is known as invasive lobular carcinoma, and accounts for 10% of diagnosed cases (Figure 1:2B) [31]. Should malignant cells grow along lymph vessels present in the breast tissue it is known as inflammatory breast cancer, which is extremely rare.

(A)

(B)
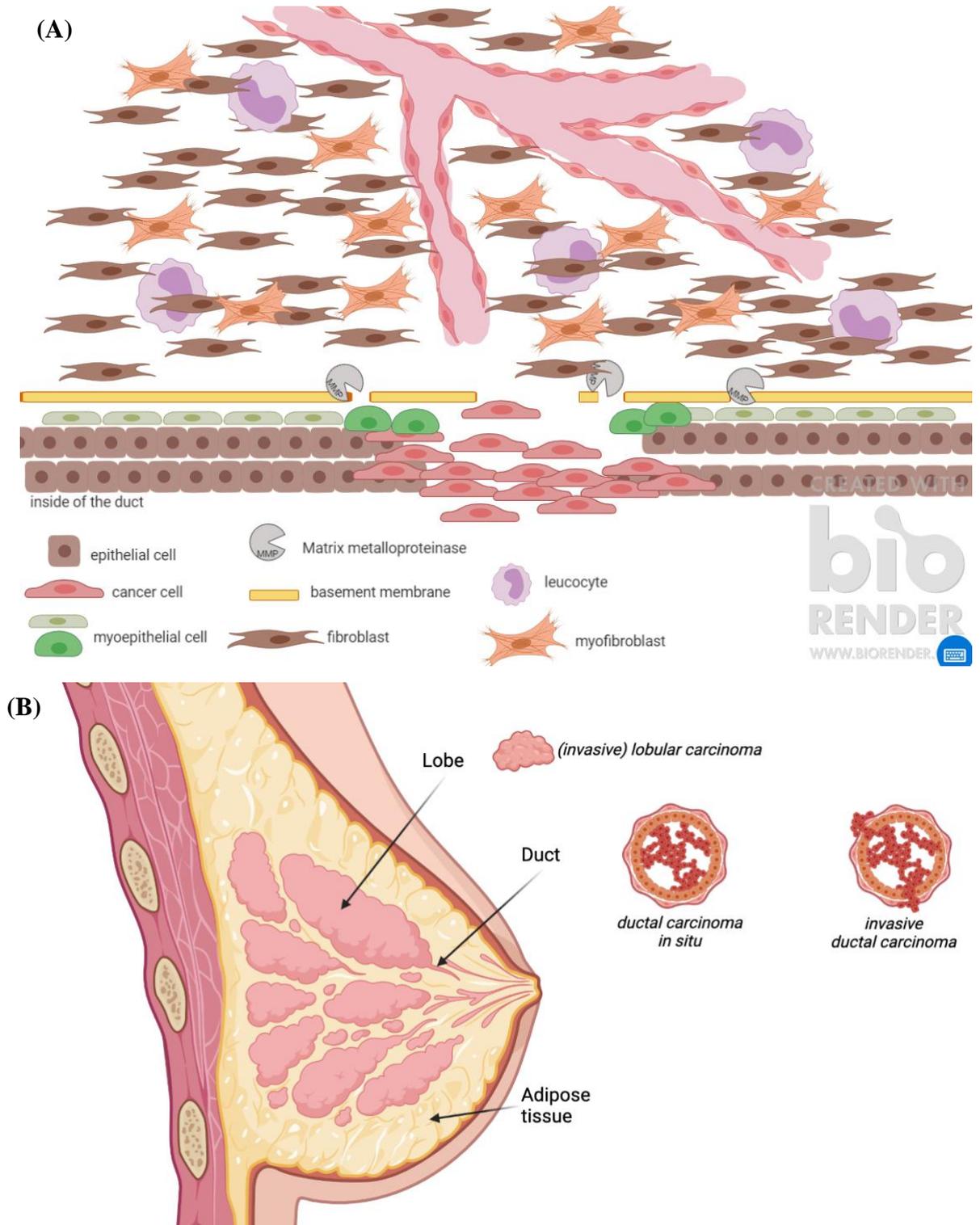
*Figure 1:2: **Structure of breast tissue to breast cancer.** Disruption of the basement membrane allows cancerous cells to migrate out of the duct which can result in the development of invasive ductal carcinoma. Disruption of the stroma by metalloproteinases remodelling the extracellular matrix (ECM), is characterised by leucocyte infiltration and increased presence of myofibroblast. Increased vascularisation into the TME allows the cancer to grow and invade further (A) [29]. The three most common types of invasive breast cancer are lobular carcinoma, where cells grow in a breast lobe. Ductal carcinoma in situ, where cells grow in a duct and once these cells invade through the basement membrane it becomes invasive ductal carcinoma. Cross-sections of the duct for ductal carcinoma in situ and invasive ductal carcinoma is shown on the right (B). Both images were made using Biorender.*

1.1.3   Clinical diagnosis of breast cancer

1.1.3.1   Molecular subtypes of breast cancer

BrCa as a disease has several subtypes that were previously mentioned. The three most common ones being ductal carcinoma in-situ, invasive ductal and invasive lobular carcinoma. Despite having three common BrCa types when diving deeper and assessing molecular markers these groups segregate into further molecular subtypes (Table 1:2). The presence or absence of key markers ultimately determines the grade of BrCa, its prognosis, and the ultimate course of treatment [32]. Luminal BrCa tumours are the most diagnosed, where luminal A exceeds luminal B in how often it is diagnosed. Molecularly, luminal A and B are nearly identical however luminal A tends to express more oestrogen receptors (ER) and the luminal B tends to be a higher grade than its counterpart [33]. Although luminal A and normal-like BrCa are similar in terms of marker status, the prognosis varies. Patients with normal-like BrCa do not show similar histopathology as luminal A which could result in delayed diagnosis or altered treatment which leads to a less desirable outcome [33]. Roughly 30% of diagnosed BrCa cases are due to genetic factors. The most well-known BrCa-related genes are BRCA 1 and BRCA 2 and have high penetrance in society [34].

Molecular subtype	Marker status	Grade	Clinical outcome
Luminal A	ER+PR+ HER2- Ki67-	1/2	Good
Luminal B	ER+PR+ HER2+/- Ki67+	2/3	Intermediate/Poor
HER2	ER+ PR+ HER2+ Ki67+	2/3	Poor
Basal /Triple negative	ER- PR- HER2-	3	Poor
Normal-like	ER+ PR+ HER2- Ki67-	1, 2, 3	Intermediate

*Table 1:2: **Summary of BrCa molecular subtypes.** Table is adapted from [33]. ER, oestrogen receptor; PR, progesterone receptor; HER2, human epidermal receptor 2; Ki67, cell proliferation marker. Normal-like is similar to luminal A in terms of molecular markers however patients will usually have negative lymph node infiltration and show a normal breast tissue profile [35].*

1.1.4   Breast cancer diagnosis

The initial symptom of BrCa is usually a lump in or around the breast tissue. Presentation of any indicators suggestive of BrCa warrants a physical examination by a certified physician. If concerned, they will refer the patient to the hospital for further examinations. Referral to hospital can also be done when the patient has a high familial risk of developing BrCa.

According to NICE (National Institute for Health and Care Excellence) guidelines, followed in the United Kingdom (UK), clinicians do a triple diagnostic assessment for a rapid diagnosis. This includes a physical examination where the physician reports a Z-score and if it exceeds the score of 3, the patient is referred to further breast screenings which can either be a mammogram or ultrasound [36]. The Z-score is a scale of 1-5 which is a score the radiologist assigns to scans and concludes if it is normal (Z-score of 1) or highly suspicious of malignancy (Z-score of 5). If during the screening it is suspected to be a tumour, needle biopsies are taken and sent for pathology [37]. The pathology report would confirm the molecular subtype as well as the stage at which the tumour is. The staging process is universal and follows the TNM method. T indicates the primary tumour, and it ranges from T1 to T4 where the number is determined by the tumour diameter and invasion into neighbouring tissue. N refers to the lymph node and reflects the extent of tumour invasion into the lymphatic system. Lastly, M stands for metastasis where there are only two scores; presence or absence of metastasis which scores a 1 and 0 respectively [38].

Once a case has been confirmed it is usually reviewed with computational tools and a care pathway discussed with the patient. Validated tools like PREDICT or Nottingham Prognostic Index [39] are used to guide the multi-disciplinary teams (MDT) what neoadjuvant treatment to recommend, if applicable [40], and the patients' prognosis relative to the suggested treatments.

1.1.5 Breast cancer treatment

As aforementioned, the recommended course of treatment is often based on the absence or presence of molecular markers. In current society, technology is crucial in the determination of the correct treatment. Clever algorithms like PREDICT have provided MDTs a realistic and accurate model of the prognosis of patients with or without the administration of adjuvant therapies [39]. The initial model was primarily based on oestrogen receptor (ER) status, and subsequent versions since have included more molecular markers like human epidermal growth factor receptor 2 (HER2) [41] status and Ki67 [42]. Since these updates, the model has been refitted and improved based on new 15-year follow up cases [43]. Tools like PREDICT are a valuable resource to assess the prognosis of the patient as well as recommend treatment options.

Classically there are three treatment options patients can take: hormone therapy, radiotherapy, and chemotherapy [44]. Chemotherapy has existed since the mid-1950s, where empirical observations during World War II suggested that nitrogen mustards acted as

effective anti-cancer substances [45]. Prior to the first approved use of chemotherapy for BrCa the only recommended procedure was conservative breast surgery i.e., resect as much of the tumour as possible and adjacent tissue. Only in 1976 did the first BrCa chemotherapy regime show that it was effective in inducing remission in combination with surgery [46, 47]. Since then, more chemotherapy reagents have been approved for the use of treating BrCa, see Table 1:3.

Drug name (e.g., Brand name)	Mechanism of Action	Approval	Study
Cyclophosphamide (Cytoxan)	Crosslinks with DNA inducing apoptosis.	1959	[48, 49]
5-FU (Capecitabine)	Blocks DNA synthesis by inhibiting the production of thymidine monophosphate.	1962	[50, 51]
Doxorubicin (Adriamycin)	Interferes with topoisomerase II thereby inhibiting DNA synthesis and replication.	1993	[52, 53]
Docetaxel (Taxotere)	Inhibits the depolymerisation of microtubules, necessary for mitosis.	1995	[54]
Epirubicin (Ellence)	Causes DNA damage by intercalating with the DNA/RNA strands thereby triggering cell death.	1999	[55]
Trastuzumab (Herceptin)	Monoclonal antibody that binds to the HER2 receptor thereby slowing tumour growth.	2000	[56]

*Table 1:3: **List of approved BrCa chemotherapy agents, their mechanism of action, year of approval and associated publications.***

Prior to staining for molecular markers, individuals were likely overtreated with chemotherapeutics as these were seen as the novel cancer-beating drug to a fatal disease [57]. There is a need for fewer novel toxic treatments to be made available for patients, which includes targeted immunotherapy drugs. With the emerging information of markers e.g., ER and progesterone receptor (PR) markers clinicians soon realised that not all chemotherapy therapies are necessary, and in doing so introducing the beginnings of precision medicine, targeting the cancer based on the markers it has.

Alternative therapies to chemotherapy include radiotherapy, hormone therapy and for HER2+ BrCa specifically, immunotherapy. As HER2 is an oncogene, targeting the product

of this gene with immunotherapy can halt proliferation of tumour cells [56]. Hormone therapy, by supressing ovarian function, is recommended for women who are diagnosed with ER+ or PR+ BrCa. The therapy can be extended up to five years to prevent the recurrence of HR+ BrCa [58]. Radiotherapy after breast surgery has shown high local control rates up to a decade after the treatment [44], and could, in a hypothetical setting, reduce one in four deaths when a local recurrence occurs within 15 years [59]. Current therapies are allowing patients to live longer with cancer and is heavily driven by seeking novel therapies. For example, exemestane and ovarian suppression in premenopausal (PeM) women after adjuvant therapy was shown to be more effective in reducing occurrence in the SOFT trial [58], versus tamoxifen and ovarian suppression alone which was standard care previously. Trials like these will become crucial in finding novel therapies to treat BrCa cost-effectively and with less side effects for the patient.

1.1.6   The gut microbiota, a new risk factor for breast cancer?

All cancers are considered to have an 'environmental' element associated with the risk of developing the cancer and at what rate the disease may progress. These elements are often observed in large human epidemiological studies that correlate lifestyle factors (e.g., smoking, drug exposure and diet) with cancer onset and clinical outcomes [20, 60, 61]. Studying the mechanism of how lifestyle factors influence health in relation to diseases like BrCa has also introduced an additional risk factor. Residential microbial communities i.e., the microbiota (discussed in more detail in section 1.2), have been identified to be involved with the pathogenesis in certain diseases including cancer. For example, *Fusobacterium nucleatum* in colorectal cancer or *Helicobacter pylori* in gastric cancer [62, 63]. This has led to a new era where studies are investigating the potential link of cancer and residential microbial communities including the gut microbiota. Often, lifestyle factors, further discussed in section 1.2, e.g., high-fat diets which can lead to obesity, are also recognised as cancer risk factors which influence the gut microbiota. As there is more interest in how the microbiota may influence health, more population-based microbiota studies have taken place and are showing that the gut microbiota can influence cancer outcomes [60, 64, 65]. The gut-tumour axis includes locations known to have direct crosstalk between the host and the gut microbiota (e.g., colorectal cancer), but also in sites further from the gut (e.g., the skin, liver, and breast).

It is likely that what applies in one cancer setting is by no means universal, and BrCa by its extreme heterogeneity and relative low incidence of genetic predisposition, is particularly unique. Consequently, there are many large studies focussing on understanding how

different environmental factors influence BrCa, and how each of these factors influence (and are influenced by) the microbiota. Risk factors influencing risk of BrCa include:

1. **Diet**. In the 1990s several groups investigated the association between diet and BrCa risk. For example, a low-fat diet elicited a lower risk of relapse after tumour resection [66]. Recent meta-analyses of cohort studies continue to correlate dietary patterns with BrCa risk [67]. No studies published to date have assessed risk of incidence and specific diets or food groups. However, there is a consensus to high-fat diet increasing risk of BrCa though this lends itself to obesity which is discussed in the next point.

2. **Obesity**. Complementary to a high-fat diet, obesity is associated with increased risk of developing PoM BrCa with a worse clinical outcome. Meta-analysis of 9 studies showed increased BrCa risk with increased body-mass index [68]. Associations between obesity and PoM BrCa may be due to adipose tissue catalysing the formation of oestrogen after menopause, thereby increasing circulating oestrogen levels [68, 69]. See point (4) below.

3. **Alcohol consumption**. Excessive alcohol intake is also recognised as a risk factor for BrCa [70]. Whilst the specific molecular mechanisms driving this effect remain unknown, ethanol may: (a) induce molecular damage in mammary cells; (b) inhibit oestrogen-metabolising enzymes in the liver; and (c) increase aromatase activity in the liver, which has been reported to facilitate the conversion of testosterone to oestrogen [71]. See point (4) below.

4. **Changes in circulating hormonal levels**. Alongside uterine, ovarian, and prostate cancers, some forms of BrCa are oestrogen driven. Both a late menarche and an early menopause decrease the risk of developing BrCa [20]. For a recent review on this subject see [72].

5. **Antibiotic exposure**. Use of antibiotics is becoming increasingly controversial, with unexpected adverse effects being reported in several disease contexts [73-75]. In 2004, Velicer *et al*., concluded that cumulative days of antibiotic exposure were associated with increased risk of BrCa [76]. A follow-up study also showed that antibiotic use may be associated with less favourable tumour features [77].

One commonality between each of these risk factors is that they significantly alter the profile of the gut microbiota (see Table 1:4), suggesting a strong link between microbiota make-up and BrCa development. Although the gut microbiota could influence BrCa development, the

aforementioned risk factors show that these factors influence each other one cannot determine if the gut microbiota influences BrCa risk and conversely if BrCa diagnosis influences gut microbiota profiles. There is a need for observational studies to be set up to investigate this association and collect data. Then any observation should be followed up in more controlled settings i.e., *in vitro* or *in vivo* experimentation to investigate the direction of the association i.e., is it the gut microbiota influencing BrCa or BrCa influencing the gut microbiota profile.

Factor	Influence on the gut microbiota

Diet	● Members of the microbiota can digest otherwise indigestible components of our diet (e.g., dietary fibre).
	
	● Dietary fibre constituents can: (1) boost nutritional intake, (2) act as a substrate for other microbiota members to colonise and (3) act as a metabolite [78].
	
	   ○ Short-chain fatty acids (SCFA), a constituent of metabolised dietary fibre, can module host immune responses.
	
	   ○ Bioactive compounds, a constituent of metabolised polyphenols, encourage growth of beneficial bacteria e.g., *Bifidobacterium* and *Lactobacillus* and production of SCFA [79, 80].

Obesity	● Gut microbiota profiles differ amongst obese and lean patients and between those with metabolic syndrome [81]. ● In mice, studies showed that an obese microbiota profile had a greater nutritional intake capacity [82]. ○ Members of an obese microbiota profile encoded enzymes that could more efficiently degrade polysaccharides.

Alcohol	● Perturbations of the gut microbiota profile was observed in alcoholics vs. non-alcoholics [83].
	
	   ○ This resulted in lower abundance of Bacteroidota and higher abundance of Proteobacteriota.
	
	   ○ Alcoholics also had higher levels of serum endotoxin.
	
	● Alcoholics tended to have greater gut permeability, which could lead to a local inflammatory state and disease e.g., alcohol-related liver disease [84].
	
	● It can be hypothesized that changes in microbiota members due to alcoholism alter the metabolites available by the host to use for other physiological processes including gut barrier function.

Hormones	● In 1998 a group observed that germ-free mice, which do not have a gut microbiota, regained normal oestrous levels upon accidental bacterial contamination. ○ This suggested a link between gut bacteria and reproductive capacity [85], where bacteria seem to influence hormone cycles in mice. ● Microbiota members possess β-glucuronidase, which can deconjugate already metabolised oestrogen.

	o Thereby increasing levels of systemic oestrogen, increasing the risk of ER+ breast cancer [86].
	● A population-based study demonstrated an association between oestrogen metabolism and phylogenetic diversity of the gut microbiota, suggesting a link between the gut bacteria influencing circulating reproductive hormones [85].

Antibiotics	● Antibiotics severely impact the gut microbiota, most notably they reduce microbial diversity.
	    o After depletion due to antibiotics, it became easier for pathogenic bacteria e.g., *Salmonella* to colonise due to lack of competitive exclusion [87].
	    o The change in microbiota members consequently influenced the availability of metabolites used by the host, which could influence e.g., host immune responses [87].

*Table 1:4: **Reported BrCa risk factors can also influence the gut microbiota.***

1.2   The gut microbiota

Both on and within humans there are resident microbial communities (bacterial, fungi, viruses, and archaea) termed the microbiota. The gut represents the most densely colonised microbiota site (over 10^{14} microbes), which has in recent years gained significant attention as researchers have observed its importance in our health [88]. These microbes have been shown to be involved in metabolising certain drugs, as well as influencing our immune system [89-91]. In a "healthy" state, our gut microbiota is in a stable community, however, when disturbances occur from e.g., taking antibiotics, this can significantly impact ecosystem diversity. The consequences of these microbiota disturbances can be as minor as diarrhoea, or more severe manifesting as inflammatory bowel disease [92]. This section will introduce the gut microbiota and why it is important for health and disease. This will be followed by discussing methodologies on researching and analysing the gut microbiota. Lastly, this section ends with a summary of keystone studies showing the potential that the gut microbiota has in a clinical setting including in BrCa.

1.2.1   Importance of the gut microbiota

It has long been known that what we consume influences our health and in turn our gut microbiota. These bacteria obtain their nutrients from what we eat, as such our diet shapes our microbiota composition and now, we appreciate that our gut microbiota in turn influences our health. As we age, our diets change. Epidemiological studies have shown that lifestyle choices, largely diet, can increase risk of diseases [19]. A monotonous diet limits

the nutrients available to the microbiota. This can result in lower gut microbial diversity, which translates to an increased risk of developing either obesity, type 2 diabetes or metabolic syndrome, systemic low-grade inflammation and weakened intestinal barrier integrity [93]. It should be noted though that as more research into metabolic syndromes and the gut microbiota is being published it is suggested that this relationship is bi-directional, where the gut microbiota can worsen disease phenotype and where the disease phenotype can cause a perturbed microbiota which exacerbates the disease. Nevertheless, these metabolic conditions can increase the risk of cancer. Other factors that influence our gut microbiota include antibiotics and genetics. Antibiotics have become an asset in the treating of bacterial diseases. Unfortunately, their widespread use has resulted in microbial resistance. Our understanding in how antibiotic usage impacts health is increasing, specifically in relation to the impact it has on our gut microbiota and how this may influence our health [94, 95]. To minimise relative risk of disease due to a disturbed microbiota we would need to establish what a healthy microbiota is.

Microbiota studies allow researchers to associate species, strains, or consortia with specific host functions, and start to address the question; what is a healthy microbiota? Some may associate 'good' bacteria like probiotics being present at high levels as a 'healthy' microbiota. *Lactobacillus* is a commensal that has been implicated in several health benefits e.g., lactose metabolism and anticarcinogenic properties [96]. *Bifidobacterium* is another important commensal in early life. It has been shown that this species ferments human milk oligosaccharides (HMOs) into various metabolites that 'cross-feed' other commensals [97]. *Bifidobacterium* make up roughly 15% of an adult gut [98] and *Lactobacillus* <1% autochthonously [99], it seems unlikely that probiotic bacteria alone determines a 'healthy' gut microbiota. Other microbiota members, although insignificant individually, can collectively constitute a 'healthy' microbiota. For example, *Prevotella* and *Xylanibacter* digest plant components and the metabolites include short-chain fatty acids (SCFA), which have been shown to be anti-inflammatory and anticarcinogenic in several studies [100, 101]. A healthy microbiota may instead be a sum of its individual parts, and the products they produce giving benefits to its host.

1.2.2   Researching the gut microbiota

Microbiota profiling methods and analytical tools are varied and there are important points to be considered when designing or interpreting these types of studies (for further details see Table 1:5). There are pros and cons to each of these techniques and usually factors like sequencing depth, cost, and the research aim, determines what platform is utilised. As a

result, each study has unique study protocols and methods of analysing sequencing reads. This has several consequences including huge variability in the results of similar studies with the same sequencing platform. Therefore, efforts should be made to agree on a standardised method of analysing microbiota data.

Microbiota profiling method	Advantages	Disadvantages
16S rRNA amplicon sequencing	• Cost effective. • Downstream analysis is less computationally intensive. • Comprehensive databases Analysis pipelines robust and easy to use.	• Only allows bacterial genus determination. • Issues with contamination (from low biomass samples i.e., tissues). • Amplification (PCR) biases; based on 16S rRNA region targeted and primers used.
Shotgun metagenomics	• Species and strain level profiling (bacteria) and relative abundance. • Also captures total microbial community (e.g., viruses and fungi) • Infer functional potential of metagenome. • Can 'assemble' novel taxa and gene families.	• More expensive than 16S rRNA amplicon sequencing. • Requires skilled bioinformaticians for in-depth analysis. • Databases for non-bacterial taxa less well curated.
Cultureomics	• In-depth genomic analysis (via whole genome sequencing). • Phenotypic assays (e.g., growth studies on different dietary components). • Isolates can be used in model systems to probe mechanisms (e.g., *in vivo* models).	• Time consuming and labour intensive. • Low abundance taxa may be missed.

	• Novel and beneficial microbes could be used as probiotics or live biotherapeutic products in clinical studies.	
(Meta)transcriptomics	• Allows determination of active transcripts, rather than just functional potential. • Can infer 'live' vs. dormant/dead microbes. • Can be used to determine how communities respond and adapt to different conditions.	• (Most) expensive. • Prokaryotic RNA has short half-life, so samples need to be processed very rapidly. • High level bioinformatics required for data processing.
Metabolomics	• Profiling small molecules i.e., metabolites which allows more direct assessment for potential microbe and host. • Absolute quantification possible (with standards) • Assess interactions between the gut microbiota and host tissue. • Important for determining microbiota-diet interactions.	• Careful preparation of samples is needed (and choice of platform i.e., NMR vs. M/S). • Can be difficult to delineate between microbe and host-derived metabolites (unless known e.g., SCFAs). • Untargeted analysis is difficult as many metabolites are unknown, therefore limiting downstream analysis and insights.

*Table 1:5: **Summary of the advantages and disadvantages of various techniques to investigate microbiota compositions.** Adapted from: [102, 103].PCR: polymerase chain reaction, NMR: nuclear magnetic resonance, M/S: mass spectrometry, rRNA: ribosomal ribonucleic acid.*

There are various methods and software available to analyse microbiome data. Each software has its own method to do quality control (QC) and prepare the raw reads for taxonomic assignment. As demonstrated by Nearing *et al*., the same microbiome dataset with different testing methods resulted in variable results [104]. 16S rRNA gene amplicon

sequencing targets a conservative area in the prokaryotic genome and amplifies this by PCR with rarefying often used to normalise the quantity of amplicons. Rarefying refers to the removing of reads from the dataset relative to the selected threshold. This includes removing libraries that do not meet this threshold, and down-sampling libraries to meet the minimum threshold [105]. There is discussion if this method of normalising is correct as low abundance taxa, or low amplified taxa due to PCR bias, are discarded incorrectly [106]. Alternative methods include using operational taxonomic unit (OTU) assignment. OTU assignment is grouping microbes based on DNA sequence similarity. OTU-clustering and then rarefying has decreased run time for analysis but it is argued that this can create artificial features [105]. Fortunately, technology has improved significantly to the point where rarefying and normalising methods that used to create artificial features and noise have become minimal. The study by Nearing *et al*., showed that methods that included rarefying the datasets did not perform significantly worse than other methods. However, this is not always the case which is why critical thinking of the sequencing data post-processing should always take place.

With advancements in technology especially in machine learning (ML) approaches we can learn a lot more about the microbiome's influence on health and disease. ML is a powerful tool that can help disentangle trends from noise. Studies have used ML for e.g., metabolomics [107] and microbiota signatures [108] associated with cancers. However, as Giwahi *et al*., point out there are some precautions researchers should have when undertaking microbiota analysis. More often than not, issues in microbiota studies are due to experimental design rather than the analysis [109]. The paper by Giwahi *et al.,* demonstrates the pitfalls that Whalen discusses in their review about using ML techniques in genomic studies [110]. As Giwahi *et al*., Nearing *et al.,* and the results of the CALADRIO study (discussed in The CALADRIO Study) show multi-variate analysis and ML approaches can show false-positive associations. These tools should not be discounted for their possibility to report false positives but should be used cautiously in the experimental design. A researcher that does have significant results should ask the question: "Do these results make sense? Is this the native environment of the microbe? How many samples is this association reported in?". In doing so it provides more context to the results of exploratory microbiota studies and hopefully reduces false positive conclusions.

1.3   A link between breast cancer and the gut microbiota

Metagenomic profiling of the microbiota of cancer patients belonging to the Twins UK cohort showed that some cancers have a 'distinct' profile. One of those cancers is BrCa.

Jackson *et al.*, used 16S rRNA gene amplicon microbiota profiling data to probe for correlations between certain microbiota traits (e.g., diversity and presence or absence of bacterial taxa) with health outcomes. One of the interesting findings was that BrCa did appear to strongly correlate with a "disrupted" or non-healthy microbiota signature [111], although this cohort was comprised of solely PoM women. Whilst BrCa was not the primary focus of this publication, rather general health was, the study suggests that the gut microbiota is a useful biomarker or target for prevention or treatment of BrCa patients [111].

1.3.1   Human observational studies of the gut microbiota and BrCa

BrCa is incredibly heterogenous on a molecular level. As such disentangling the relationship between the gut microbiota and disease can be a difficult to navigate. To narrow down the scope, one can focus on specific types of BrCa. Most of the diagnosed BrCa cases occur in PoM women, as such most studies focus on PoM cohorts. An early study in 2015 by Goedert *et al.*, observed a significantly lower alpha diversity in PoM BrCa patients than in control patients. Looking at the microbiota composition in detail the group reported an increase of relative abundance in the classes *Clostridiaceae, Faecalibacterium* and *Ruminococcaceae,* and a relative decrease in abundance of *Dorea* and *Lachnospiraceae* in PoM BrCa patients compared to controls. For this study, controls were patients invited for mammogram screening with a negative diagnosis. This study also looked at urinary oestrogen and, although not significant, observed that case patients had higher urinary oestrogen levels than control patients [112]. In a follow-up study by the same group, the researchers again observed a lower diversity in case patients compared to age-matched controls [113]. Another study by Zhu *et al.,* looked at PeM and PoM BrCa patients and observed an altered gut microbiota in PoM patients. This study used shotgun metagenomic sequencing as opposed to 16S rRNA gene amplicon sequencing as the two studies previously mentioned. When determining differences between PoM control and case patients using the Shannon diversity index it was not significant, but it was significant for PeM control vs. case patients where PeM case patients had a higher Shannon diversity. Unlike the studies reported by Goedert *et al.*, this study reported a high species number in BrCa patients compared to controls. They did however observe several species differences in PoM cohorts. They reported a positive but a weak association between certain species and oestrogen metabolism i.e., *Shewanella putrefaciens* and *Erwinia amylovora*, and the butyrate producing species i.e., *Roseburia inulinivorans* [114]. When assessing the functional capacity of the microbiota they observed that PoM case patients had enriched genes in e.g., lipopolysaccharide biosynthesis and beta oxidation compared to PoM controls.

A study in 2017 found that *Blautia* spp. is associated with early grade BrCa histopathology. The study also reported significant taxa associations, but this was more likely due to BMI differences as opposed to a relation with BrCa [115][1]. In 2020, a group in Ghana studied the microbiota profiles of BrCa and non-malignant breast disease patients. They consented 379 BrCa patients, 102 non-malignant cases and 414 population-based controls in a large population-based study. They utilised the V4 16S rRNA gene region to compare microbiota profiles. Unfortunately, they could not find a statistically significant difference between odds ratio of BrCa and non-malignant cases, the P-trend was not significant (P-trend = 0.66). However, they did observe an inverse association with alpha diversity for BrCa versus non-malignant cases [11].

The studies published by Goedert *et al.*, reported a lower microbiota diversity in BrCa patients compared to controls whereas Zhu *et al.*, reported no difference in Shannon diversity PoM case vs. PoM controls, although this could be due to the small sample size and high variance within the group. The contrasting conclusions could be due to the country where these studies were conducted. Zhu's study was based in China while Goedert's study was based in USA. Fundamentally their diets are different, and this is a recognised confounding factor for gut microbiota composition. The change in diet is a factor the Ghana study reported. A shift to a more refined "westernized" diet may be associated with the increased incidence of BrCa cases, which could be reflected in their gut microbiotas. Despite not finding a significant result they did observe a strong trend, where alpha diversity was strongly inversely associated with BrCa cases than none. One factor to appreciate in epidemiological studies is that should a consistent trend be found across different populations, even if it does not reach statistical significance, that differ in a recognised influencing effect e.g., diet, then perhaps that trend is a true observation that should be investigated. Human observational studies, like the aforementioned, are crucial in elucidating the association between microbes and human health and disease. Either the study size must be large enough to capture a true representative cohort or validation studies need to be done to understand how microbes can influence BrCa development. The former is difficult to achieve in terms of cost and time but achievable. The latter is achievable using animal models, which will be discussed below.

---

[1] The study design was based on qPCR of representative strains for bacterial groups or species. *Blautia coccoides DSM935* was the representative of Bacillota as such the conclusion can be an artefact of the study design.

1.3.2 Immunotherapy animal studies of gut microbiota and cancer

Large cohort and population studies provide potential associations between the gut microbiota and BrCa patients. As briefly discussed before, there are many confounders that cannot be controlled for in a large population-based study but can be adjusted should a potential confounder be reported. Another issue is if an association is present, it would be difficult to conclude if e.g., A influences B or B influences A, which is known as reverse causation. Therefore, further research into mechanistic pathways of how candidate microbiota members may influence the tumour microenvironment are necessary. Within the last few years several studies have been published suggesting mechanistic pathways how microbes are capable of increasing BrCa progression. In 2013 Iida *et al*., showed that germ free (GF)-mice or antibiotic-treated mice had impaired anti-tumour responses in CpG-oligonucleotide immunotherapy and platinum-based chemotherapy. The authors observed that animals with a perturbed gut microbiota had a decrease in expression of genes controlling anti-inflammatory pathways which is usually activated for this anti-cancer therapy [116]. Thereby demonstrating that certain bacteria are necessary to stimulate the correct immunomodulatory pathways required in anti-cancer therapies.

Two noteworthy studies were published showing that the previous observation stated by Iilda *et al*., could be narrowed down to certain bacteria. Sivan *et al*., showed that *Bifidobacterium* was responsible for a spontaneous anti-tumour response in mice. The group took genetically similar mice from two housing facilities and noticed one group showed anti-tumour responses while the other did not. Further investigations showed that the anti-tumour group harboured a distinct microbiota profile, where *Bifidobacterium* showed the strongest association in mice with anti-tumour responses. Dendritic cells isolated from the mice gavaged with *Bifidobacterium* could be stimulated at lower antigen levels compared to naïve cells [117]. It is postulated that these cells have become primed by the gavaging of *Bifidobacterium* thereby being more effective in anti-tumour responses. The other paper was from Vétizou *et al*. CTLA-4 is a negative regulator of T-cell activity. Cancer cells can interact with CTLA-4 by expressing CD80 or CD86 to switch off activated T-cells and blocking consequent cytotoxic reactions induced by its binding [118]. Blocking the binding site of CTLA-4 will allow the cytotoxic T-cell to remain active and kill off cancer cells despite the cancer cell expressing CTLA-4 ligands. A serious immune related adverse event related to CTLA-4 blockade is inflammation at mucosal sites usually interacting with microorganisms e.g., the gut or lungs. The group observed a difference in specific pathogen free (SPF) and GF mice and noticed that GF-treated mice could not control tumour progression with CTLA-4 blockade. Administration of antibiotics could replicate the

previous observation reported in GF mice. These results implicate the gut microbiota as a player in the effectiveness of CTLA-4 blockade. Random forest classification of human melanoma gut microbiota profiles showed three distinct clusters. Longitudinal analysis showed that during CTLA-4 blockade treatment patients were moving into one distinct cluster. Faecal microbiota transfers (FMTs) into GF mice from patients in these distinct clusters showed that mice receiving a FMT from those in cluster C responded the best to CTLA-4 blockade. Further in-depth analysis by qPCR showed this was driven by the expansion of *Bacteroides fragilis*. As an added benefit, intestinal reconstitution with immunogenic *Bacteroides* spp. reduced CTLA-4 induced colitis a common side effect of CTLA-4 therapies [119]. In summary, this in-depth study demonstrated that immunogenic bacteria present in the gut can influence the efficacy of immunotherapy.

Lee *et al*., demonstrated the anti-cancer potential of *Bifidobacterium bifidum* in non-small cell lung carcinoma (NSCLC). Comparing 16s ribosomal ribonucleic acid (rRNA) gene sequencing from 96 NSCLC patients to healthy control patients they found a significant difference in both alpha and beta diversity, where non-responders had a higher diversity than responders. Further analysis segregating microbiota profiles according to response and no response to anti-cancer treatment showed that *B. bifidum* was enriched in responders. This was also confirmed using quantitative polymerase chain reaction (qPCR) methods. Despite not colonising, the transient passing of *B. bifidum* was enough to reduce tumour growth in the syngeneic murine models. Mice treated with anti-PD-1 and the synergistic *B. bifidum* strain showed a significantly smaller tumour volume compared to anti-PD-1 treated animals alone. Intra-tumoural cytokine profiling showed an increase in interferon (IFN)-ɣ and toll-like receptor (TLR)-2. Blocking the IFN-ɣ receptor abrogated the observation of a reduced tumour volume in anti-PD-1 and anti-PD-1 with *B. bifidum* administration. Furthermore, they used TLR2 knock-out mice and the effective anti-cancer synergistic effect induced by *B. bifidum* disappeared. This mechanistic work suggests that *B. bifidum* exerts its anti-cancer effect by TLR-2 which stimulates an IFN-ɣ cell mediated pathway [120]. This study demonstrated a possible mechanism of *B. bifidum*'s anti-cancer effect in NSCLC, which could provide novel therapeutics to improve the efficacy of anti-cancer therapies.

The previous studies demonstrated how gut microbes could possibly influence anti-cancer immunotherapy outcomes. There are some postulations about the cellular mechanism, where immunomodulatory factors of bacteria stimulate the gut mucosal immune system i.e., metabolites [121]. Research and development teams strive to find a biomarker for response to anti-cancer therapies or as a diagnostic tool for cancer. There have been several studies

showing an association between cancers and metabolites derived from the gut microbiota [107, 122-126]. A large Mendelian randomisation study by Liu *et al*., is an attempt linking blood metabolites with specific gut bacteria. In doing so, it can provide a reference for other researchers as a potential tool for exploratory microbiome studies [122]. Inosine is a metabolite found to be significantly abundant in the serum of colorectal cancer (CRC) mono-colonised mouse models with *Bifidobacterium pseudolongum*. The group showed that inosine sustained T-cell activation by increased IFN-ɣ production necessary for anti-CTLA-4 therapy [126]. Choline degradation was associated with CRC using ML approaches [107], thus demonstrating a potential biomarker for CRC. These are just some of the studies linking the changes in microbiota profiles with possible mechanisms, specifically metabolites in these examples. It is inferred that metabolites can influence intermediary cells, including T-cells and dendritic cells, which in turn can improve the efficacy of anti-cancer treatments.

1.3.3 Current mechanistic hypothesis related to the gut microbiota and BrCa

Current hypothesis of mechanistic pathways that may link the gut microbiota with BrCa includes the oestrobolome and metabolite theory. As aforementioned, age of menarche and first pregnancy to obesity and high alcohol intake can all influence the level of circulating oestrogen. It has become apparent that the gut microbiota, and some members of it possess an enzyme: β-glucuronidase which can deconjugate already metabolised oestrogen. In doing so, the oestrogen is reabsorbed into the bloodstream [86]. Another theory is that the microbiota can release metabolites, by-products of digestion, which influence our immune system. GF mice or antibiotic-treated mice treated with immunotherapy or chemotherapy showed poor immune infiltration into tumours. Antibiotic treatment also ablated the ability of tumour necrosis factor (TNF)-producing immune cells to secrete TNF. Gavage of a strain, *Alistipes shahii*, that seems to be involved with TLR-4 in priming TNF production, to antibiotic-treated mice restored ability to of immune cells to produce TNF. These results suggested that commensal strains are necessary to induce proper immune response brought by certain therapies [116].

In 2021, Mckee *et al., [127]* and Buchta Rosean *et al., [128]* demonstrated a link between antibiotic-induced perturbations of the gut microbiota and BrCa. Mckee *et al.,* reported larger breast tumours in mouse models when their gut microbiotas were ablated with antibiotics. They observed the same outcome in mice when they used a clinically relevant antibiotic regime. Further research showed a higher proportion of myeloid-derived tumour-infiltrating cells [129]. Buchta Rosean *et al*., performed the same experiment, and although they did not observe different primary tumour sizes, they did observe greater metastasis.

However the exact mechanism of this remains unknown [130]. Further investigations by researchers showed that these myeloid-derived cells in the tumour may in fact be mast cells. Blocking mast cells resulted in slower tumour progression in antibiotic-treated mice but not control mice. Thereby suggesting that a perturbed microbiota encourages infiltration of mast cells into the breast tumour [131]. As it stands, current literature assessing the functional mechanism between the gut microbiota and BrCa does not exist. There are preliminary results suggesting that microbiota-derived metabolites may influence immune cell behaviour as reported by these studies, however more investigations into the mechanism are necessary to fully understand this relationship.

Studies in the last decade have shown the potential that microbiota profiling can have for the use of immunotherapy. Most papers assessing gut microbiota and immunotherapy responses have been observational in nature with validation in murine models. Only recently have phase 1a clinical trials approached the power of FMT to influence responses in anti-PD-1 therapy in refractory melanoma [132, 133]. These studies tended to show an increased infiltration of CD8+ T cells resulting in an improved pathological response. In comparison to other cancers i.e., melanoma or CRC, studies into BrCa are lacking. Observational studies of the gut microbiota in BrCa patients (e.g., BEAM study) are the first step to understanding the interaction between breast tissue and the gut microbiota. In addition, undertaking profiling studies in BrCa related clinical trials will also further elucidate the delicate relationship the gut microbiota may have on the efficacy of anti-cancer therapies (e.g., CALADRIO).

1.4   Project introduction

Research has shown that the gut microbiota holds potential for many aspects regarding health and disease. Most of the gut microbiota profiling studies have taken place for gut related illnesses i.e., CRC, Crohn's, or inflammatory bowel disease (IBD). A literature search via Pubmed[2] using the keywords "gut microbiota" AND "breast cancer" yielded 137 results (January 2023), when narrowed down to anything but systematic reviews or reviews, only 3 remain. It should be noted that this is not an exhaustive literature search, as not all Boolean operators were utilised, but this is merely to demonstrate that research studies specifically looking at the gut microbiota and breast cancer are limited. Of these three only one of them is an observational, case-control study based in Spain (NCT03885648) [134]. With the two remaining studies, one is a human interventional trial assessing the impact of

---

[2] Search performed April 2023

a Mediterranean diet with probiotics on BrCa survivors, and the other is an interventional trial progressing into an *in vitro* experiment assessing polyphenols and urinary metabolites on BrCa cells. There exists a gap in the number of studies assessing baseline microbiota profiles of individuals at risk of developing BrCa. The literature that has been reported by Jackson *et al.,* Goedert *et al*., and Zhu *et al.,* suggest there may be a perturbed gut microbiota profile associated with BrCa in humans. Possible mechanisms were alluded to by McKee *et al.,* and Buchta Rosean *et al*., indicating that there may be a mechanistic foundation to support the concept that the gut microbiota may play an active role in BrCa. This PhD project attempts to address this knowledge gap by setting up an observational trial at the local hospital (Norfolk and Norwich University Hospital, NNUH) to profile the gut microbiota of first-time diagnosed BrCa patients, leverage a completed BrCa clinical trial and associated sample set, and use *in vitro* and *in vivo* models to probe underlying mechanisms.

1.4.1   Research questions, project aims and objectives

The project was originally focussed on profiling the gut microbiota of first-time diagnosed BrCa patients in the Norfolk region to elucidate if there is a baseline profile associated with BrCa, and then expanding on this by examining how treatment impacts the gut microbiota. Due to the SARS-CoV-2 pandemic the original aim of the project was changed and was broadened to investigate the relationship between the gut microbiota and breast health.

The main research question of this project was: in women who experienced a diagnosis of non-hereditary BrCa, how different is their gut microbiota after treatment compared to before the commencement of treatment i.e., at baseline?

The main goals of this project were to:
1. Characterise the gut microbiota of first-time diagnosed BrCa patients using 16S rRNA gene amplicon sequencing and cultureomics approaches.
2. Identify associations between the gut microbiota and clinical metadata e.g., grade of tumour, remission status, blood circulating cytokines.
3. Validate any associations through *in vitro* or *in vivo* experimentation.

This thesis with its respective goals is presented in the following chapters:

Chapter 3: The BEAM study.

This chapter focusses on the establishment of the Breast hEalth And Microbiota (BEAM) Study at the NNUH and James Paget University Hospital (JPUH). The chapter will also

describe the microbiota profiles based on 16S rRNA gene amplicon sequencing, shotgun sequencing and cultureomics efforts. This chapter addresses the main research question mentioned in 1.4.1. I will describe establishing and operationalising the BEAM study to address the current knowledge gap in evaluating baseline microbiota profiles of individuals at risk of developing BrCa. The research hypothesis driving this endeavour was to assess if a perturbed gut microbiota driven by antibiotics is associated with adverse clinical outcomes.

Chapter 4: The CALADRIO study.

This chapter is part of a collaborative effort with MedSir, an institute based in Madrid. MedSir set up the KELLY trial (NCT03222856) and the CALADRIO study was a spin-off to assess the oral and gut microbiota of metastatic BrCa human epidermal growth factor receptor 2 (HER)-/hormone receptor (HR)+ patients undergoing a novel combination therapy of pembrolizumab (PD-1 inhibitor) and eribulin (mitotic inhibitor). The KELLY trial only had one arm where all patients enrolled received the therapeutics. Here, I evaluate the potential of microbiota profiling as a predictive marker in a clinical setting. The research hypothesis driving this was discussed in 1.3.2. Specifically, I want to determine if metastatic HER2-/HR+ BrCa patients that experienced a clinical benefit had an altered gut or oral microbiota at baseline compared to patients that did not experience a clinical benefit.

Chapter 5: The *Bifidobacterium longum project*.

This chapter builds on my observations while culturing samples from the BEAM study as well as literature reporting protective properties of bifidobacteria (see section 1.3.2). I, with help from the Robinson group, performed and designed the experiment with the hypothesis being that bifidobacteria will work synergistically with cyclophosphamide chemotherapy to reduce tumour burden in murine BrCa models.

2   Materials and Methods

All methodologies used to produce the data reported are described in this thesis. Should there be any methods that build on the work of others, the method will be referenced accordingly. Unless otherwise stated all experiments took place at room temperature and pressure. When referring to anaerobic conditions these conditions were 37°C in an atmosphere containing N_2, CO_2, H_2 (85%, 5% and 10% respectively).

2.1   BEAM study pathway, a brief summary

The BEAM (Breast hEalth And Microbiota) study gained favourable ethical approval from the University of East Anglia (UEA) Faculty of Medical Health Sciences Ethical committee (FMH 201819-092HT). Sample collection was in accordance with the protocols approved by the National Research Ethics Service (NRES) for the NRP Biorepository tissue bank (Human Tissue Authority license number: 11208). Upon identification of possible eligible patients, research staff would notify a member of staff of the Biorepository who would then contact the patient to determine if they were eligible. Once deemed eligible, the patient would be consented into the BEAM study and receive a faecal collection kit to their home address. Three different pathways were established for BEAM patients, which was dependent on the clinical care that the patient would receive (see, Appendix: BEAM Human Tissue Protocol). Blood and resected tissue would be obtained by the clinical team at the NNUH and passed to the researcher. Faecal samples would be collected by the participant at home using a collection kit which had all the necessary items to collect a sample. Upon collection, the sample would be posted back to Quadram Institute Bioscience via a pre-paid label where the researcher would collect and process the sample within 24h. Once aliquoted, any remaining sample would be destroyed via incineration and the aliquots returned to the Biorepository (BR) for storage until further use. For the disease cohort, patients would receive a minimum of four kits. The baseline kit would be given when consent was taken. The second kit would coincide closely after their surgery or chemo/radiotherapy. The third kit would occur six months after initial diagnosis at an appointment called "moving forward" or shipped to their home address if "moving forward" would not occur (i.e., due to the pandemic), and the last kit would be one year after recruitment into the BEAM study. The pathway for BEAM at JPUH was the same as NNUH but without resected tissue or blood.

2.1.1   BEAM faecal collection kit

Eligible patients were given or posted the faecal kit. This kit included: gloves, pre-paid safe box, envelopes, Fe-coll paper, bio-hazard safe specimen bag with absorbent sheet, an ice pack and a 30mL universal with a spoon. The individual components are listed below.

Item	Catalogue number	Supplier
Safebox 1st Class	SB1CB48	Royal Mail UK
Peal and Seal White C4 Polythene Envelopes	Vow_2815771	Royal Mail UK
Faeces Collection Paper Fe-coll	FC-2010	Alpha laboratories
Pouch 95kPA with absorbent sheet	FS95-C5A-100	Alpha laboratories
30mL Universal with spoon	CD3810	Alpha laboratories
Gloves nitrile starguard protect, medium	SG-P-M	Amazon.co.uk
Small ice packs	-	Amazon.co.uk

All components of the faecal collection kit were labelled with the BR number. A BR number was assigned to a patient once consented under the Biorepository ethics. Everything was labelled with this BR number to minimise identifiable information given by the patient as well as to minimise human error when processing. As more samples would arrive there were increased chances of mis-placing samples by labelling everything with the BR number this would minimise this mistake. The safebox 1st Class also had a return address labelled on it: "BEAM Study c/o [investigator's name and lab], Quadram Institute Bioscience, Norwich Research Park, Rosalind Franklin Road, NR4 7UQ, Norwich, United Kingdom".

2.1.2   BEAM study research team

BEAM faecal samples were collected through the collaboration with the following individuals and organisations:

Collaborator	Organisation	Location
Dr. Simon Pain	NNUH, Consultant (General surgery)	Norwich, UK
Dr. Katalin Zechmeister	NNUH, Consultant (General surgery)	Norwich, UK
Ms. Tracey Parker	NNUH, Deputy sister (Pre-op Assessment DPU)	Norwich, UK
Dr Rachael Stanley	NRP Biorepository	Norwich, UK
Dr. Mark Wilkinson	NRP Biorepository	Norwich, UK

Dr Louise Jones	NRP Biorepository and Breast Cancer Now	Norwich, UK
Ms. Roxanne Brunton-Sim	NRP Biorepository	Norwich, UK
Mr. Joel Wood	NRP Biorepository	Norwich, UK
Ms. Cheryl Prior	NRP Biorepository	Norwich, UK
Dr. Ibrahim Sallam	JPUH, Breast care	Great Yarmouth, UK
Dr. Alexander Sandy Leeper	JPUH, Breast care	Great Yarmouth, UK
Ms. Wendy Harrison	JPUH, Clinical research nurse (cancer)	Great Yarmouth, UK
Prof. Sue Downs	JPUH, Consultant oncoplastic breast surgeon and clinical lead for breast surgery	Great Yarmouth, UK

Table 2:1: Collaborators and organisations who have helped in the collection of BEAM samples.

2.1.2.1 Sample processing and storage

Materials, equipment and reagents used for processing BEAM samples were:

Item	Catalogue/model number	Supplier
Lysing Matrix E, 2mL tube	116914050-CF	MP Biomedicals
RNAlater Stabilisation Solution (100mL)	AM7020	Thermo Fischer Scientific Biosciences GmbH
Cryovial self-standing 2mL	479-1220	VWR International GmbH
Glycerol ANALAR 500mL	10579570	Fisher Scientific UK Ltd
MSC II Safety Cabinet	Microbiological Safety Cabinet	Walker Safety Cabinets
Autoclave (portable)	Classic 2100 Standard (9-litre)	Prestige Medical

When a safety box would arrive at QIB, the researcher would collect the box and disinfect the external surface before opening it in a MSC class II Safety cabinet. The sample record sheet would be taken, and information copied onto the spreadsheet kept by the researcher. A BEAM study number would be allocated to the sample. The BEAM study numbers followed the format: P,[XXX: number of the patient e.g., 001], [A-Z: time point of sample e.g., A is baseline], [v1: denoted the vial including preservation medium from this time point]. Thus, P001A/v1 was the baseline sample of patient 1 in the Lysing matrix tube E. As the BEAM study ran in parallel with the Breast Cancer Now (BCN) tissue bank, one aliquot would always be designated for the BCN tissue bank. Therefore, the label would be prefixed with "BCN". Of each sample the following aliquots were taken and stored in a cryovial apart from vial 1, which was aliquoted into a lysing matrix tube E:

- V1: 200mg of faecal sample in lysing matrix tube E.
- V2: 1g of faecal sample in 1mL of sterile PBS+20% glycerol solution.

- V3: 1g of faecal sample in no preservation medium.
- V4: 1g of faecal sample in 1mL of sterile PBS+20% glycerol solution.
  - This is the designated BCN aliquot.
- V5: 200mg of faecal sample into 1mL of RNAlater solution.

Should there be inadequate sample provided by the participant, the order of the vials denote priority for aliquoting. For consequent samples from the same patient, the vial numbers continue in blocks of five. For example, a patient only provided enough sample to fill two vials e.g., P001A/v1 and P001A/v2. The second sample would start from vial 6 onwards e.g., P001B/v6. The third sample would start from vial 11 e.g., P001C/v11, then the final sample would start from vial 16 e.g., P001D/v16.

Sample was processed into different vials with different preservation medias depending on their downstream application to ensure correct preservation for future use. For retrospective culturing I chose a preservation media that could protect cells from the ice crystals over a long period of time at -80°C. I opted to go for 20% glycerol in phosphate buffered saline (PBS) as it was a cheap preservation media to make in large volumes. Several papers have previously stated that glycerol in a range of 10-25% (v/v) acted as a good cryopreserving media [135, 136]. Although faecal samples can be kept 'pure' in a -80°C, changes in relative abundances of phyla have been reported when extended past three months and after multiple freeze-thaw cycles[3]. As such it was decided to aliquot samples into the Lysing matrix tube E immediately to expedite DNA extraction and prevent freeze-thaw cycles for aliquoting and one kept pure as a last resort. Lastly, RNAlater was used to stabilise RNA should RNA-seq be done later.

2.2   DNA extraction of bacterial or faecal samples

2.2.1   FastDNA SPIN kit materials

Item	Catalogue/model number	Supplier
FastDNA 2mL SPIN kit for Soil (50preps)	11492400	Fischer Scientific (MP Biomedicals)
FastPrep-24 Classic bead beating grinder and lysis system	116004500	MP Biomedicals
Lysing Matrix E, 2mL tube	116914050-CF	MP Biomedicals

---

[3] Reported by previous lab members.

Ethanol	-	Sigma-Aldrich, Dorset, UK
Eppendorf tubes 5.0mL, PCR Clean	0030 119.460	Eppendorf
Eppendorf tubes 2.0mL, PCR-clean, safe lock	0030 123.344	Eppendorf
Microcentrifuge	Prism	Labnet, New Jersey, USA

2.2.1.1 Short read fragmentation using the fastDNA SPIN kit

Manufacturer's instructions of the fastDNA SPIN kit were followed to obtain short-read DNA fragments but amended as described in the protocol [137]. Briefly, 200mg of faecal sample or 200µL of bacterial sample was aliquoted into a Lysing matrix E tube with 980µL of sodium phosphate buffer and 120µl of MT buffer. Sample homogenisation using the FastPrep-24 Classic bead beating system was done for 3min at speed setting 6.0m/s. DNA extraction protocol was then followed accordingly with a final elution volume of 60µL. Eluted DNA was stored at -20˚C after quantification until further use.

2.2.1.2 Long read fragmentation using the fastDNA SPIN kit

This protocol was an amended protocol using the same reagents from the fastDNA SPIN kit. 200µL of sample, faecal or bacterial, was aliquoted into Lysing Matrix E tubes. 980µL of Sodium Phosphate Buffer and 120µL of MT buffer was added to the sample. Samples were mixed by inverting before vortexed for 10 minutes at speed 5. Consequently, samples were centrifuged at 10,000 g for 15 minutes to pellet debris. Supernatant was transferred to a clean 2mL Eppendorf with 500µL of Protein Precipitation Solution and mixed by inverting 10 times before being centrifuged at 10,000 g for 10 minutes. Supernatant was transferred to 5mL Eppendorf tubes with 2mL of Binding Matrix. The tube was mixed by inverting for 2 minutes by hand before letting the silica matrix settle for at least 3 minutes. 1mL of supernatant was discarded and the remaining solution resuspended gently. 750µL of the resuspended matrix was added to a SPIN filter column and centrifuged at 10,000 g for 4 minutes. The catch tube was emptied and the step until all the resuspended matrix has spun through the column. The pellet was resuspended with 500µL of prepared SEWS-M and centrifuged at 10,000 g for 10 minutes. The catch tube was emptied, and a dry spin was done before allowing the column to air dry, this is done to remove residual ethanol. Appropriate volume of DES i.e., 65µL was added to the matrix and DNA eluted through the column into a new clean catch tube. Eluted DNA was kept at -20˚C after quantification for further analysis.

2.2.2 Maxwell Promega DNA extraction

Item	Catalogue/model number	Supplier
Lysing Matrix E, 2mL tube	116914050-CF	MP Biomedicals
Promega Maxwell RSC PureFood GMO and Authentication kit	AS1600	Promega
FastPrep-24 Classic bead beating grinder and lysis system	116004500	MP Biomedicals
Block heater		VWR International
MSC II Safety Cabinet	Microbiological Safety Cabinet	Walker Safety Cabinets
Vortex	Vortex Genie 2	Cole-Palmer

This protocol was followed for the processing of the CALADRIO study samples, as it was more efficient to process all samples this method than the method described in 2.2.1.1. This method had been validated by other lab members in the group to be used for large-scale microbiota studies to accurately reflect the microbial community using mock samples[4]. For the protocol the reagents and materials were part of the Promega Maxwell RSC PureFood GMO and Authentication kit. 200µL of faecal sample was aliquoted into Lysing Matrix E tubes. 1mL of CTAB was added to the sample and vortexed for 30 seconds before being incubated on a heat block at 95˚C for 5 minutes and then vortexed for a minute. Samples were homogenised in the FastPrep-24 machine for 45 seconds at a speed of 6.0m/s. To each sample 40µL of Proteinase K and 20µL of RNase A was added and vortexed to mix. Samples were heated to 70˚C for 10 minutes while the cartridges were prepared for the Maxwell robot according to manufacturer's instructions.

2.2.3 DNA quantification

After extraction of genomic DNA, the DNA was quantified using the Qubit 2.0 dsDNA BR assay kit. Protocol was as per manufacturer's kit. Briefly, a volume up to 5µL was used for sample DNA and the working solution was made up using a 1:200 dilution of DNA reagent in dsDNA BR buffer.

Item	Catalogue/model number	Supplier
Axygen 0.6mL MaxyClear Snaplock Microcentrifuge tube	MCT-060-C	Corning Life Sciences
Qubit Fluorometer 2.0	Q32866	ThermoFisher Scientific
Qubit dsDNA BR reagents	Q32850	ThermoFisher Scientific

---

[4] This protocol was used by the PEARL and MOTION study which validated this methodology to process a high number of samples efficiently and accurately reflect the microbial community.

2.3 16S rRNA gene amplification

This protocol described amplifying the 16S rRNA gene from a pure culture of bacteria.

Item	Catalogue/model number	Supplier
KAPA 2G Robust PCR kit	KK5004	Sigma-Aldrich
DNA ladder (1kb)	N3232S	New England Biolabs Ltd
SYBR Safe DNA Gel stain	S33102	ThermoFisher Scientific
PCR Thermal cycler	Applied Biosystems Veriti 96-well thermal cycler	Thermo Fisher Scientific
Gel electrophoresis unit	ENDURO Gel XL	Labnet

2.3.1 16S rRNA Sanger sequencing

Full-length 16S rRNA gene PCR was performed on bacterial isolates to determine preliminary bacterial identity. KAPA2G Robust PCR reagents were used to prepare the PCR master mix, to which 5µL of bacterial DNA (10-30ng/µL) was added and amplified with a thermal cycler, conditions explained below in Table 2:2 and Table 2:3. The PCR master mix consisted out of: 10µM primers (1µL per sample), 10mM of dNTPs mix (1µL per sample), 5X GC buffer with $MgCl_2$ (10µL per sample), molecular H_2O (31.6µL per sample) and Taq DNA Polymerase (0.4µL per sample) [138].

2.3.1.1 PCR primers, conditions, and product confirmation

Primers	Sequences (5' to 3')
fD1	AGA GTT TGA TCC TGG CTC AG
fD2	AGA GTT TGA TCA TGG CTC AG
rP1	ACG GTT ACC TTG TTA CGA CTT

*Table 2:2: **Sequence of primers used for full-length 16S rRNA PCR reactions.** Primers were prepared to 100µM. Primers were taken from: [139]. Use of these primers was based on data by previous users demonstrating it could provide a nearly complete 16S rRNA gene PCR product.*

PCR Step	Temperature (°C)	Duration	Cycles
Initial denaturation	94	5 min	1
Denaturation	94	1 min	
Annealing	43	1 min	35
Extension	72	2 min	
Final extension	72	2 min	1

*Table 2:3: **PCR conditions for full-length 16S rRNA reaction.***

PCR product amplified the 16S rRNA gene of bacteria, successful amplification was confirmed by running on 1% agarose gel stained with SYBR Safe with a 1kB ladder. The gel was visualised and a product of 1,500bp was expected. Once confirmed the PCR

amplicons were sent off according to the sample submission guide (Eurofins, Luxembourg or Source Bioscience, Cambridge) for Sanger sequencing. After quality trimming most PCR products were within the range of 900-1,000bp.

Once results came in, FASTA sequences were run through the NCBI Blastn[5] database, limited to rRNA/ITS databases. Sequence identity was confirmed with the similarity score and used for preliminary identification.

2.3.2 Genomic sequencing

The methodology described below was optimised by the QIB Sequencing team (David Baker, Rhiannon Evans and Steve Rudder). This is taken from their methodology protocol available on the intranet website to be used for publications.

Item	Catalogue number	Supplier
KAPA 2G Robust PCR kit	KK5004	Sigma-Aldrich
EB	19086	Qiagen
Qubit high sensitivity kit	Q32851	ThermoFisher Scientific
Illumina® DNA Prep, (M) Tagmentation (96 Samples, IPB	20060059	Illumina
D5000 ScreenTape	5067-5579	Agilent
NSQ® 500 Mid Output KT v2 (300 cycle)	FC-404-2003	Illumina
PhiX Control v3	FC-110-3001	Illumina
MiSeq Reagent Kit v3 (600 cycle)	FC- 102-3001	Illumina

2.3.2.1 *Whole genome sequencing, PE150*

Genomic DNA was normalised to 5ng/µL and submitted to the QIB sequencing team for whole genome sequencing. The libraries were prepared using a novel modified Illumina DNA prep tagmentation approach as described in [140], under CoronaHiT-Illumina library preparation. The final library pool was double-SPRI size selected between 0.5 and 0.7X bead volumes using sample purification beads from the Illumina DNA prep kit. The final pool was quantified on a Qubit 3.0 instrument and run on a D5000 ScreenTape using the Agilent Tapestation 4200 to calculate the final library pool molarity.

---

[5] http://blast.ncbi.nlm.nih.gov/

At a final concentration of 1.5ρM the pool was run on an Illumina Nextseq500 instrument using a Mid Output Flowcell (NSQ® 500 Mid Output KT v2 (300 cycle), following the Illumina recommended denaturation and loading recommendations which included a 1% PhiX spike in (PhiX Control v3 Illumina Catalogue FC-110-3001).

2.3.2.2 Nanopore MinIon sequencing
Genomic DNA was extracted using the materials and method previously described in: Long read fragmentation. The genomic DNA was sequenced using the Oxford Nanopore native barcoding genomic DNA protocol (SQK-LSK109).

2.3.3 16S rRNA gene amplicon sequencing

Primers	Sequences (5' to 3')
16S rRNA V1-V2 forward	AGM GTT YGA TYM TGG CTC AG
16S rRNA V1-V2 reverse	GCT GCC TCC CGT AGG AGT

Table 2:4: Primer sequences for 16S rRNA gene targeting the V1 and V2 hypervariable region.

PCR Step	Temperature (˚C)	Duration	Cycles
Initial denaturation	95	5 min	1
Denaturation	95	30 sec	
Annealing	55	30 sec	30
Extension	72	30 sec	
Final extension	72	5 min	1

Table 2:5: PCR reaction conditions for the first step amplifying the 16S rRNA V1-V2 hypervariable region.

PCR Step	Temperature (˚C)	Duration	Cycles
Initial denaturation	95	5 min	1
Denaturation	95	30 sec	
Annealing	55	30 sec	10
Extension	72	30 sec	
Final extension	72	5 min	1

Table 2:6: PCR reaction conditions to prepare library for 16S rRNA gene amplicon sequencing.

Genomic DNA was extracted using the materials and method previously described in: Short read fragmentation using the fastDNA SPIN kit. DNA was quantified using Qubit, as previously described in: DNA quantification. The genomic DNA was diluted to 5ng/µL and submitted to the QIB sequencing team. A PCR master mix was prepared consisting out of 10µL of KAPA2G Fast Hot Start ready Mix. 0.1µL of both 100µM primers (sequence listed in Table 2:4), was mixed with 7.8µL PCR grade per sample, to which 2µL of DNA was added and mixed. PCR conditions for the first reaction is listed in Table 2:5. PCR products

were purified using a 0.7X SPRI clean-up with Illumina Sample purification beads. DNA was eluted with 20uL of water. The second PCR was prepared with 10µL of KAPA2G Fast Hot Start Ready Mix added to each well of the 96-well microplate where 2µL of 10µM 8bp Unique Dual Indexes were added. To each well, 8µL of clean specific PCR was added. The second PCR reaction conditions is listed in Table 2:6. The final library was quantified using Qubit and once equimolar, samples were pooled together. The equimolar pooled product underwent a single 0.7X SPRI clean-up. Final Qubit and sizing using D5000 Screen Tape with the Agilent Tapestation 4200 was used to calculate the final library pool molarity. The pool was run with a final concentration of 12ρM on Illumina MiSeq instrument with the MiSeq Reagent Kit v3 (600 cycle). Illumina recommendations were followed regarding denaturation and loading suggestions, this included a 20% PhiX spike in. The raw data was analysed locally on the MiSeq using MiSeq reporter. The data was consequently uploaded on the NBI IRIDA platform for further analysis.

2.4   Bioinformatics

All of the bioinformatics described in this thesis, unless otherwise stated was provided by Dr. Raymond Kiu. All scripts and notes can be found on his Github repository[6]. All bioinformatic tools were done on the Norwich Bioscience Institute (NBI), High-Performance Computing cluster (HPC). The NBI HPC at the time of writing this utilised the SLURM workload manager (v16.05.8) to submit jobs to the cluster. Dr. Raymond Kiu ran the metagenomic shotgun pipeline on the HPC while I did the remaining pipelines using the scripts Dr. Kiu provided, making adjustments where necessary.

2.4.1   16S rRNA gene amplicon sequencing taxonomic analysis

QIIME v1.9.1 (Quantitative Insights Into Microbial Ecology) was used to analyse the 16S rRNA gene amplicon sequencing data, Illumina 300bp paired-end reads. Briefly, FASTQ paired-end reads were merged using PEAR v0.9.6 [141], with a Phred quality score of 33. Quality filtering was done by: --phred_quality_threshold 29 and --phred-offset 33 (split_libraries_fastq.py).

USEARCH v6.1 [142] was used to remove chimeras (identify_chimeric_seqs.py). OTU assignment was done (pick_open_reference_otus.py) based on SILVA rRNA database [143] (released in September 2016). USEARCH utilises UCLUST clustering algorithm at a 97% sequence identity to identify chimeras. Once OTUs were assigned BIOM files were

---

[6] https://github.com/raymondkiu/informatics-tools

generated and data visualised on MEGAN v6.20.19 [144] or on R [145], after normalising data to relative abundance (%).

2.4.2   Whole genome *de novo* assembly

2.4.2.1   *Short-read assembly, PE300*

FASTQ files were downloaded from the QIB IRIDA platform and uploaded onto the NBI HPC. To remove bad reads (phred value less than 20) and primers the sequences were run through fastp v0.20.0 [146] with a phred quality cut-off of 20. *De novo* assembly was done using Spades v3.11 with the parameters: --careful [147]. Any contig less than 500 base pairs was removed using the script "filter-contig.pl", found on Dr. Kiu's GitHub page. BactspeciesID v1.2 was used to check for contamination and provided a preliminary identity, this was achieved using the parameters: --m TRUE [148]. Using the output of BactspeciesID, if available, the type strain was downloaded and fastANI v1.33 used to compare the query whole genome to the type strain whole genome to confirm identity [149]. For isolates that did not provide a BactspeciesID output, the 16S rRNA gene sequence was extracted *in silico* using BactspeciesID parameter --r FALSE, and run through Blastn [150] and type strain whole genomes downloaded to use for final confirmation using fastANI v1.33.

2.4.2.2   *Long-read assembly using Flye*

This software was used for the description of the novel genus *Allocoprobacillus halotolerans* LH1062. The sequence reads were initially filtered through Filtlong v0.2.1 [151], with only reads larger than 1,000bp and the top 90% quality of reads remaining for subsequent genome assembly. Consequently, the genome was assembled using Flye v2.9 specifying five polisign iterations [152]. Once assembled the FASTA file was compared with the 'short-read' assembly of *A. halotolerans* LH1062 to confirm if it was successful and contamination was checked using checkm, described in 2.4.4.1.

2.4.3   Metagenome shotgun sequencing

2.4.3.1   *Taxonomic assignment of reads*

For metagenome shotgun sequencing, fastp v0.20.0 [146] was used to run a quality check on the FASTQ files with a phred score of 20. Bowtie2 v2.3.4.1 [153] via KneadData v0.10.0 [154] was used to remove host-associated reads i.e. human reads, using the database "GrCh38_noalt_decoy_as" using --bypass-trim, --reorder and --bypass-trf. Metagenome co-assembly was achieved using Spades v3.14.1 [147] and Megahit v1.2.9 [155], run via MetaWrap v1.3.2 [156]. To assign taxonomy, Kraken2 v2.1.2 [157] was used with a

confidence level of 0.1 and the output run through Bracken v2.6.2 [158] to estimate relative abundances with a threshold of 10 reads minimum.

2.4.3.2   *Metagenome assembled genome assembly (MAGs)*

Binning for metagenome assembled genomes (MAGs) was done via MetaWrap v1.3.2 with "metawrap binning" and using Checkm v1.1.3 [159-161]. Maxbin2 [162] and MetaBAT2 [163] was used for bin refinement with the parameters: >80% completeness and <10% contamination. Salmon v1.5.1, part of the metawrap pipelines, [164] was used to quantify the bins by "metawrap quant_bins" and GTDB-tk v1.5.1 [165] to get the closest average nucleotide identity (ANI) genome.

2.4.3.3   *Functional assignment using Humann3*

For functional assignment Humann3 v3.0.0 [154] via Metaphlan v3.0.13 [154] pipeline, was used with Chocophlan as the database. Prior to this all paired-end reads were concatenated into a single file. Default parameters were used with the paths to the databases, reads specified. The data of the file, "humann_all_pathabundance_cpm.tsv" was normalised then stratified by species or genus contribution before using for data analysis.

2.4.4   Strain level quality control and screening

2.4.4.1   *Contamination check using checkm*

To confirm the purity of strains I utilised checkm v.1.1.3. I generated the lineage marker according to the manual i.e., "checkm lineage_wf", followed by performing a quality assessment. The output, in a table format, was used to assess contamination and genome quality.

2.4.4.2   *Circular chromosome check*

To check if a bacterial chromosome was circular, the software circulator v1.5.5 was used [166]. The specified options to do so was: "--merge_min_id 85, --merge_breaklen 1000". This was used to determine if *Allocoprobacillus halotolerans* LH1062 was circular when submitting the genome to NCBI.

2.4.4.3   *Taxonomy assignment with GTDB-TK*

Taxonomy assignment using GTDB-Tk was done using software gtdbtk v.1.5.1. I used "classify_wf" to do taxonomy assignment for our isolates. The final "gtdbtk.bac120.summary.tsv" file was used to check the taxon of each genome.

2.4.4.4   *Screening using Abricate*

Abricate was used to screen for resistance genes using the provided database ResFinder. I also used Abricate v1.0.1 to screen for Cytochrome P450 genes in the gut microbiota FASTA files of CALADRIO patients, and to screen for the capsular polysaccharide, and fragilysin proteins in *B. fragilis* MAGs. A database had to be curated for cytochrome P450 (list found in 8.3.1), capsular polysaccharide (specified in 4.2.4.1) and fragilysin proteins (specified in 8.3.2) before screening could take place. To do so, the nucleotide FASTA sequences were downloaded from Genbank and compiled into one file and indexed according to the Abricate manual. Next the query genomes were screened with the default parameters of Abricate.

2.4.5   Analytical tools

2.4.5.1   *LEfSe on Galaxy*

Linear discriminant analysis of effect size (LEfSe) was performed on the sequencing reads against the clinical parameter 'clinical benefit' status for the CALADRIO study. This was done using the Galaxy module provided by the Huttenhower lab, available at: http://huttenhower.sph.harvard.edu/galaxy.

2.4.5.2   *CEG Tools: PathFinder and ResFinder-FG*

I used the Centre for Genomic Epidemiology (CEG) to screen the *B. fragilis* metagenome assembled genome (MAGs) for a resistance phenotype. This differed from the normal Resfinder database as it was based on functional metagenomic antimicrobial resistance determinants. This web tool is available at: https://cge.food.dtu.dk/services/ResFinderFG/. I also used PathFinder available at: https://cge.food.dtu.dk/services/PathogenFinder/. This tool predicts its potential pathogenicity of a bacteria based on its genome to humans.

2.4.6   Phylogenetic trees

2.4.6.1   *Kmer based phylogenetic trees- Mashtree*

I utilised Mashtree to create a phylogenetic tree of the cultured BEAM isolates. As it is a diverse set of strains, it was suggested to use a Kmer based approach which the software mashtree v.1.2.0 utilises. To create the phylogenetic tree, I used the outputs: "--outmatrix mash" and utilised bootstrapping by: "mashtree_bootstrap.pl --reps 100, --min_depth 0". The mashtree file was then visualised using iTOL v6.

2.4.6.2 *Phylogenetic trees using 16S rRNA genes*

To create a phylogenetic tree based on the 16S rRNA gene sequences I first utilised muscle v3.8.31 to create an alignment file. The consequent alignment file was used to create a 16S rRNA phylogenetic tree using IQ-Tree v2.0.5 with the arguments: "-m TEST, -alrt 1000, -B 1000, -T AUTO". This specifies the tree to be built based on the best-fit model determined by IQ-Tree, specifies 1000 replicates, the use of 1000 bootstraps, and automatic allocation of the number of CPU cores to use. This generated a tree file that was then used in iTOL v6 to visualise the tree.

2.4.6.3 *Phylogenomic trees using PhyloPhlAn*

A phylogenomic tree was constructed using PhyloPhlan v3.0.51. FASTA files, both 16S rRNA and whole genome sequences downloaded from NCBI, of the isolates to be included in the tree were downloaded from NCBI. PhyloPhlan has an automated pipeline which I utilised. Briefly, for PhyloPhlan to run a configuration file needs to be created. This file specifies the software and settings used to run the automated pipeline. In our case, the configuration file specified the use of diamond v0.9.19 and mafft v7.515 as the aligner. Sequences were trimmed using trimal v2.4.rev15 and the tree constructed using iqtree v2.1.4. The tree was constructed with the PhyloPhlAn options –diversity medium and – accurate. iTOL v6 was used to visualise the generated tree file.

2.4.7 Graphing and visualisation

Figures were made using Rstudio v2022.01.1+554 for Mac, "Spotted Wakerobin" release and RGui 4.0.0 "Arbor Day" and 4.2.2 "Innocent and Trusting", dependent if the package was supported by the updated RGui. Microsoft PowerPoint 2022 was used to edit and annotate figures and graphs where necessary. The scripts and pipelines used in this PhD is available at: https://github.com/nteng22/PhD-Thesis.

2.4.8 Statistical analysis

All statistics were performed in Rstudio, versions and RGui as previously described. Whilst a p-value of 0.05 was used to indicate an association, the magnitude and precision of estimates were interpreted on a continuous scale of evidence i.e., due to the low powered study trends were taken into consideration when drawing conclusions. Data expression and the statistical tests were stated in the figure legend or test where applicable. For microbiota data pertaining to comparing relative abundances of certain strains I checked for normality by plotting a Q-Q plot and performing a Shapiro test. I performed a Levene's test to check

for variance. When it was not normally distributed, I transformed the data logarithmically, and repeated the tests. Should the transformed data not be normally distributed or be homoscedastic I then performed non-parametric tests which is specified in the figure legends.

2.5   Faecal culturing materials and reagents

Item	Catalogue/model number	Supplier
Yeast casitone fatty acid (YCFA)	-	-
Brain heart infusion (BHI)	CM1135	Oxoid
Glycerol	10579570	Fisher Scientific UK Ltd
Axygen 1.7mL MaxyClear Snaplock Microcentrifuge tube	MCT-175-C	Corning Life Sciences
Phosphate buffered saline (PBS), sterile and reduced	18912014	Thermo Fisher Scientific
10ul Plastic Inoculation Loops 1000 - in Grip Seal Bags of 20 (Aseptic Manufacture)	A3	Microspec Limited
Wedge Shaped Spreaders – in Grip Seal Bags of 20 (Aseptic Manufacture)	D3	Microspec Limited
140mm petri dish triple vent sterile	501V	Slaughter Ltd, R&L
9mm petri dish triple vent sterile	101VR20	Slaughter Ltd, R&L
15mL Centrifuge tubes	430790	Corning B.V. Life Sciences
Cryovial self-standing 2mL	479-1220	VWR International GmbH
MSC II Safety Cabinet	Microbiological Safety Cabinet	Walker Safety Cabinets
Autoclave (portable)	Classic 2100 Standard (9-litre)	Prestige Medical
pH meter	Martini MI 151	Rocky Mount
Anaerobic cabinet	Ruskinn Concept Plus	Baker Ruskinn

BHI and YCFA media was used for untargeted faecal culturing described in section 2.5.3.

2.5.1   Sample collection from the NRP Biorepository

An email request would be sent to the NRP Biorepository at least 24 hours before the sample would be used. The sample would be transferred using a double-contained biohazard container before being processed in a MSC II cabinet. If depleted the Biorepository would

be notified. After taking the appropriate aliquot necessary the sample was kept on dry ice until transferred back to the NRP Biorepository.

A sample stored in the lysing matrix tube E would be depleted when requested as it would undergo DNA extraction. A sample stored with PBS+20% glycerol will be used for faecal culturing, and therefore would need to be returned to the Biorepository after the appropriate amount has been taken.

2.5.2 Culture media preparation

To prepare brain heart infusion (BHI) agar and broth, manufacturer's instructions were followed.

For yeast casitone fatty acids (YCFA) the recipe was followed as stated in [167], and printed below in Figure 2:1.

Ingredient	Amount		Components of solutions and mixes:	
Before Autoclaving			**Resazurin Solution**	
Agar (optional)	8 g		Resazurin	0.1 g
Tryptone	5.0 g		d. H₂O	100 ml
Yeast extract	1.25 g			
NaHCO3	2.0 g		**Mineral Solution I:**	
(D)+Glucose	1.0 g		K₂HPO₄	3 g
(D)+Maltose	1.0 g		d.H₂O	1 L
(D)+Cellobiose	1.0 g			
L-cysteine	0.5 g		**Mineral Solution II:**	
Mineral Solution I	75 ml		KH₂PO₄	3 g
Mineral Solution II	75 ml		(NH₄)2SO₄	6 g
Resazurin Solution	0.5 ml		NaCl	6 g
Haemin Solution	5 ml		MgSO₄	0.6 g
Vitamin solution I	0.5 ml		CaCl₂ (dry)	0.6 g
d.H2O	up to 500 ml		d.H₂O	1 L
VFA mix	3.1 ml			
NaOH	pH to 7.45		**VFA mix:**	
			Acetic acid	17 ml
After Autoclaving			Propionic acid	6 ml
Vitamin solution II	0.5 ml		n-Valeric acid	1 ml
			Isovaleric acid	1 ml
			Isobutyric acid	1 ml
			Haemin Solution:	
			KOH	0.28 g
			Ethanol 95 %	25 ml
			Haemin	0.1 g
			d.H₂O	up to 100 ml
			Vitamin Solution I:	
			Biotin	5 mg
			Cobalamin (Vitamin B12)	5 mg
			PABA (4-Aminobenzoic Acid)	15 mg
			Folic acid	25 mg
			Pyridoxine	75 mg
			d.H₂O	up to 500 ml
			Vitamin Solution II:	
			Thiamine hydrochloride	25 mg
			Riboflavin	25 mg
			d.H₂O	up to 500 ml

*Figure 2:1: **Recipe for YCFA supplemented with carbohydrates (glucose, maltose and cellobiose).** Recipe was taken from [167].*

To prepare the three different media types i.e., media broth, agar and 20% glycerol the following was done. For media broth, solution was stirred until media powder was completely dissolved. Media was autoclaved to sterilise, then left to cool. For agar solution, 8g of agar was added to 500mL of media broth. The solution was stirred until the media powder was completely dissolved and consequently the agar solution sterilised by autoclaving. Once the media completed the sterilisation process and cooled down enough to handle, 25mL of agar broth was dispensed per 9cm petri dish plate and 125mL of agar broth per 14cm petri dish plate. For glycerol, 20% of glycerol was added (v/v) to the media

solution prior to autoclaving. The solution was then sterilised by autoclaving and left to cool until further use.

2.5.3   Faecal culturing process

The work was carried out in a MSC II safety cabinet where work with human faecal samples can take place. Samples (PBS+20% glycerol) were defrosted on ice before 100mg was weighed out into a sterile microcentrifuge tube. Remaining sample was frozen again at -80°C while preparing the faecal slurry. To the 100mg faecal aliquot, 1mL of sterile reduced PBS was added and the sample vortexed until homogenised. A dilution series down to 10^{-6} was prepared and 250μL of the 10^{-4} dilution was plated out on a 14cm plate for both BHI agar and YCFA agar. The diluted faecal slurry was spread using a disposable spreader until dry. This was repeated for the dilution 10^{-5} and 10^{-6}. Plates were labelled with the sample ID and dilution and left to grow anaerobically for 72 hours. After the plates were placed in the anaerobic cabinet the faecal samples were returned to the Biorepository.

After 72 hours, plates were inspected to look for morphologically distinct colonies. Should a distinct colony be found it was isolated using an inoculation loop and streaked, using quadrant streaking, onto a new 9cm petri dish with the respective media it was isolated from. For example, a colony that was isolated from a 14cm BHI plate would be streaked onto a 9cm BHI plate. Should the same looking colony be present on two different plates, streaking it from one plate would be sufficient. The colonies were left to grow under anaerobic conditions for two days and the process of isolating and streaking onto a new plate was repeated, this was done to ensure purity of the isolate. This was done a total of three times. 10mL of the respective broth was added to a 15mL centrifuge tube. Once the isolate was confirmed to be pure, an inoculation loop was used to inoculate 10mL of the respective broth with the pure colony. The inoculum was left to grow anaerobically for 2 days. After which, the inoculum was spun down at 2,600rpm for 10min to create a cell pellet. The supernatant was removed, and the pellet resuspended in 1mL of the glycerol media broth. The resuspended pellet in glycerol broth was transferred to a sterile cryovial and kept at -80˚C until further use.

2.6   *In vitro* co-culture

Item	Catalogue/model number	Supplier
DMEM high glucose, no phenol red	D1145-500mL	Sigma

L-glutamine (200mM)	G7513-100ML	Sigma
Sodium pyruvate (100mM)	S8636-100mL	Sigma
Penicillin-Streptomycin (10,000 U/mL)	15140122	Gibco
Foetal Bovine Serum		Thermo Fisher Scientific
Trypsin EDTA (100mL)	25200056	Thermo Fisher Scientific
0.1% Porcine gelatin		
Nunc™ EasYFlask™ Cell Culture Flasks	156472	Thermo Scientific

2.6.1 Cell housekeeping protocols

2.6.1.1 Complete and minimal media

Complete media for MCF-7 cells included DMEM high glucose, no phenol red. To which 5% FBS, 1% penicillin/streptomycin, 1% sodium pyruvate and 1% L-glutamate was added to create the complete media which was used. Minimal media had the same supplements apart from 5% FBS, which was changed to 2% FBS.

2.6.1.2 STR profile check

I received the MCF-7 and MDA-MB-231 cell aliquots from the Robinson group. To confirm their identity cell pellets were sent off to Eurofins for cell line authentication checking.

2.6.1.3   Mycoplasma check

Item	Catalogue/model number	Supplier
MegaMix	2MMB	Clent Life Sciences
Mycoplasma positive control		DSMZ
Mycoplasma internal control		DSMZ
Axygen 0.6mL MaxyClear Snaplock Microcentrifuge tube	MCT-060-C	Corning Life Sciences
Gelatin from porcine skin	G2500	Sigma
PCR Thermal cycler	Applied Biosystems Veriti 96-well thermal cycler	Thermo Fisher Scientific
96-well PCR plates (Axygen® 96 Well Polypropylene Segmented PCR Microplate, Clear, Nonsterile)	PCR-96-SG-C	Thistle Scientific

Primers	Sequences (5' to 3')
Mycoplasma 1	GGG AGC AAA CAG GAT TAG ATA CCC T
Mycoplasma 1	TGC ACC ATC TGT CAC TCT GTT AAC CTC

*Table 2:7: **Primer sequences for mycoplasma reactions. Primers were prepared to 100µM.***

PCR Step	Temperature (°C)	Duration	Cycles
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	40
Annealing	52	20 sec	
Extension	72	30 sec	
Final extension	72	2 min	1

*Table 2:8: **PCR reaction mixture to detect mycoplasma in cell supernatant.***

Mycoplasma testing was done every month for both cell lines to ensure cells were free of infection and behaving as the cells should behave. 100µL of cell supernatant that was on the cells for at least 2 days was collected. The supernatant was boiled at 95°C for 5min in a heating block, and then spun down quickly with a microcentrifuge to collect the liquid from the lid. The reaction mixture consisted out of 12.9µL of MegaMix, 0.8µL of each primer and 1.5µL of the cell supernatant or control. Primer sequences are listed in Table 2:7. The PCR conditions for this reaction mixture is listed in Table 2:8.

After the PCR reaction, the product was run on a 1% agarose gel as previously described in section 2.3.1.1. A band present at 500-520bp would be indicative of a mycoplasma infection, provided the internal control, present at around 975bp, was also amplified.

2.6.1.4   *Human breast cancer cell culture*

Incubation of cell cultures were performed with the following conditions: 37˚C, 5% CO_2 with a 95% humidity. Cells were seeded into T-75 flasks pre-coated with 0.1% porcine gelatine. Cell detachment for sub-culturing was achieved with 0.25% trypsin-EDTA and seeded as per recommended by the literature.

2.6.2   Co-culture setup

Item	Catalogue/model number	Supplier
AlamarBlue Cell Viability Reagents	DAL110	Thermo Fisher
MTS Assay	Ab197010	Abcam
CytoTox 96® Non-Radioactive Cytotoxicity Assay (LDH assay)	G1780	Promega
Nunc™ EasYFlask™ Cell Culture Flasks	156472	Thermo Scientific
Corning™ Falcon™ 96-Well, Non-Treated, Flat-Bottom Microplate	10748692	Fisher Scientific
Microplate reader	VersaMax	Molecular Devices
Costar 24-well Clear TC-treated Multiple Well Plates	3526	Corning Costar
RNeasy Micro kit (50)	74004	Qiagen
DNA LoBind® Tubes 2.0mL	0030108078	Eppendorf UK Ltd

Experiment was planned to ensure splitting coincided with the start day of the co-culture. After splitting the cells, the left-over cell suspension was counted, and diluted to get 9,000 cells/mL. Cells were seeded at a density of 3,600 cells per well of a 24-well microplate and left to adhere overnight in the incubator. The total volume per well was 400µL. Once confirmed that cells were fully adhered by light microscope, media was replaced with the cell free supernatant of *B. fragilis* NCTC 9343 from different growth stages, as described in: CFU growth curves, diluted with BHI. The concentration of this mixture was 20% CFS with 80% BHI and a total volume of 200µL per well.

Positive control for the lactate dehydrogenase (LDH) assay was 10% cell lysis solution, from the kit, was added to cells, this confirms cell death and LDH release. Negative control for the LDH and MTS assay was cells in 2% FBS i.e., minimal media, which was determined to keep cells alive but not proliferating limiting LDH release and MTS metabolism. Positive control for the MTS assay was cells in 5% FBS i.e., complete media, to stimulate cell proliferation resulting in greater overall cell metabolism. Background control for these tests were media mixed with test reagents without cells.

Cells were left in the incubator for two hours, after which the supernatant was transferred to a microplate to be used later for the LDH assay, which was done following manufacturer's instructions. Relative cell viability was calculated against maximum cell lysis and background control values were removed from treated cell values. Cell media was replenished with 10% MTS with minimal media and left to incubate for one hour. After the hour, supernatant was transferred to a microplate and absorbance read using a spectrophotometer at 490nm. Relative cell viability was calculated against cells in minimal media and background values were removed from treated cell values. With the remaining cells, room temperature sterile PBS was used to wash the cells. After which 250μL of RLT lysis buffer, part of the RNeasy Micro kit, was added to each well. Lysed cells were collected in a 2.0 mL Eppendorf and frozen at -20°C for further extraction in the future.

To thaw cells, tubes were left to defrost at room temperature until completely thawed and salts have dissolved. If any insoluble material remains, pellet can be centrifuged for 5 minutes at 3,000-5,000g. Supernatant was transferred to a new tube, which will be the QiaShredder (part of the RNeasy Micro kit) for subsequent homogenisation. After which manufacturer's instructions were followed.

2.6.3 *B. fragilis* culturing and growth curve

Item	Catalogue/model number	Supplier
Brain heart infusion (BHI)	CM1135	Oxoid
Agar	AGA03	Formedium
Stratus plate reader		Cerillo
Anaerobic cabinet	Ruskinn Concept Plus	Baker Ruskinn

Corning™ Falcon™ 96-Well, Non-Treated, Flat-Bottom Microplate	10748692	Fisher Scientific
50mL Falcon Conical centrifuge tubes	352070	Falcon
Centrifuge	Eppendorf 5810R	Eppendorf
pH Indicator strips	WHA10362000	Sigma
Cuvettes	67.742	Starstedt
Spectrophotometer	WPA CO8000	Biochrom

2.6.3.1   Media preparation for the CFU growth curve

Please refer to section 2.5.2. Culture media preparation, for BHI broth, agar and glycerol solutions.

2.6.4   *B. fragilis* inoculation

B. fragilis NCTC 9343 was revived in 10mL of BHI broth and left to grow overnight anaerobically until confluent, usually 24h. 1mL of the inoculum was moved to new 9mL of BHI and this was done at least three times before used for any experiment.

To check morphology 100µL of the inoculum was spread on BHI agar plate after the first sub-culture. Plates were left to incubate anaerobically for 48h before being assessed. *B. fragilis* colonies on BHI are usually small and flat with a rounded edge. Should the morphology look different 16S rRNA Sanger sequencing was performed.

2.6.4.1 OD_{600} growth curves

To characterise the growth of *B. fragilis* in BHI broth the stratus plate reader was used. A total volume of 200µL per well was used. The microplate was prepared and inoculated with a 1:10 ratio with an actively growing *B. fragilis* inoculum. The plate was started to read over a period of 72 hours. The data was imported on MS Excel and then visualised using R.

2.6.4.2   CFU growth curves

To characterise the growth of *B. fragilis* by colony forming units (CFU), an inoculation was prepared with a 1:10 inoculum. 1mL of the inoculum was taken, from which 100µL was used to prepare the serial dilution. With the left over 1mL aliquot the OD_{600} was measured using the spectrophotometer and then spun down to be sterile filtered (0.2µm) to be used in downstream experiments. The OD_{600} informed me to what serial dilution to prepare. Three

10µL spots of each dilution were dropped to a BHI agar plate and left to dry before incubating anaerobically for 48h. To calculate the CFU/mL the following calculation was applied:

$$\frac{CFU}{mL} = (average\ number\ of\ colonies) * \left(\frac{1000\mu L}{10\mu L}\right) * \left(\frac{1}{dilution\ factor}\right)$$

To maintain similar population growth, the OD_{600} was taken and diluted to ensure the same starting OD_{600} of 0.05nm.

2.7 *In vivo* murine work

2.7.1 Orthotopic mammary gland tumour assay

Murine breast carcinoma cells B6BO1 were prepared at 10^5 per 50µl in a 1:1 mixture of PBS to Matrigel. The cells were injected into the left number four abdominal mammary fat pad of age matched (8-12 weeks) female mice under anaesthesia. Mice were subjected to the treatment regime, which was gavage with live *B. longum* subsp. longum or water every other day until sacrifice and if applicable, injection of cyclophosphamide (100mg/kg) or saline. Tumour growth was continued until fifteen days post-implantation at which point animals were sacrificed by cervical dislocation and tumours were harvested. Collected tumours were weighed (g) and measured (width and length in mm), tumour volume was calculated using the formula: (length * width^2) * 0.52, where length is the larger of the two measurements recorded.

2.7.1.1 *Tissue fixation and sectioning*

Tumours were photographed and then bisected. One half was snap frozen in liquid nitrogen for subsequent analysis, and the other half was placed into histological cassettes and submerged into 4% paraformaldehyde overnight at 4°C. After overnight fixation specimens were washed twice in PBS for 30 minutes each, before being moved to 70% ethanol until further tissue processing in the Leica Tissue Processor ASP-300S. For tissue processing the specimens underwent washes in ethanol of increasing concentrations i.e., 70% for an hour, 80% for 90 minutes, 90% for two hours, 100% for an hour, 100% for 90 minutes, 100% for two hours and then undergoing three xylene washes for 30 minutes, one hour and 90 minutes respectively. Cassettes were then placed into three paraffin wax washes for two hours, two hours and one hour respectively. Once finished, specimens were embedded in paraffin using the Leica EF-1150H Paraffin Embedding Station before being sectioned using a HM-355S Microtome. Tissue was sectioned at a thickness of 5µm and mounted onto positively charged slides (Thermofisher) before being baked overnight at 37˚C.

2.7.2 Administration of *B. longum*

Animals were treated up to three times weekly until day of sacrifice with a gavage of live *B. longum* subsp. *longum* in 200µL PBS. On the day of gavage, inoculums (starting culture a 1:10 ratio) were spun down by centrifugation 4,000rpm for 10 minutes at 4˚C. The supernatant was discarded, and the bacterial pellet washed with sterile cold PBS three times before being resuspended in the final volume necessary for gavage. A 100µL aliquot is taken from the final resuspension to determine CFU/mL retrospectively.

2.7.3 qPCR reaction targeting *B. longum*

Item	Catalogue/model number	Supplier
LightCycler 480 SYBR Green I Master	04887352001	Roche
LightCycler 480 Compatible PCR Plates, 96-well plate	AXYPCR96LC480WNF	Sigma
B. longum groEL forward primer (10µM):	5'-CTGAGGCTCTGGACAAGGTCG-3'	
B. longum groEL reverse primer (10µM):	5'-GGTGCCACGGATGTTGTTCAGG-3'	
RNAase-free water		
Roche Lightcycler	LC480II	Roche Diagnostics
Clear PCR sealing foil		

Genomic DNA from *B. longum* subsp. *longum* YP5B-G and *B. bifidum* subsp. *bifidum* 8809 was quantified using Qubit measurements, described in 2.2.3. Genomic DNA was extracted from the faecal pellets of the sacrificed mice using the protocol described in 2.2.1.1. To prepare the standard curve, DNA at a concentration of 100ng/µL was prepared and serially diluted down to 0.001ng/µL with a ratio of 1:10. Of these volumes 0.2µL was added to each well of the microplate. RNase-free water alone acted as the negative control.

For the final reaction, genomic DNA was diluted using RNA-ase free water to a concentration of 480ng/µL, where 0.2µL is the maximum volume to be taken into each well resulting in 96ng of DNA per well. To the 0.2µL of genomic DNA in each well the following master mix was added: 12.5µL of LightCycler 480 SYBR Green I Master mix, 2.5µL of forward and reverse primer (10µM) and 7.3µL of RNase-free water giving a final volume of: 25µL. Plate was sealed using clear foil and briefly spun down before being loaded onto the LightCycler. Conditions of the reaction was followed as described in the paper [168]. Using the standard curve and the calculations described in the paper the number of *Bifidobacteria* was calculated in the sample. Briefly, the DNA concentration was extrapolated using the standard curve and the following was calculated using the information provided in the paper, where N_A, is Avogadro's constant.

$$Concentration_{amplicon} \ (molecules.mol^{-1}) = \frac{Concentration_{amplicon}(g.L^{-1}) * N_A}{Molecular \ mass_{amplicon}(g.mol^{-1})}$$

3   BEAM Study

This chapter describes the set-up and running of the BEAM study at the Norfolk and Norwich University Hospital (NNUH, Norwich) and James Paget University Hospital (JPUH, Great Yarmouth). This will be followed by a description of the approaches used and downstream results pertaining to the gut microbiota of the patients enrolled into the BEAM study: 16S rRNA gene amplicon sequencing, metagenome shotgun sequencing and faecal culturing. The background will also introduce current literature describing the gut microbiota of breast cancer patients and describe why observational studies like BEAM are needed to address the knowledge gap. The results section summarised the trends and differences observed between patients for the described methods. Parts of this chapter were also included in the IJSEM manuscript where I am first author, in which we described two novel isolates, Novel isolate 1- *Allocoprobacillus halotolerans* LH1062 and Novel isolate 2- *Coprobacter tertrius* LH1063.

The clinical team (mentioned in Table 2:1) at NNUH were responsible for identifying eligible patients. The clinical team at JPUH, specifically Dr. Ibrahim Sallam and Ms. Wendy Harrison, were responsible for the identification, consenting of patients and sending the faecal kits. The NRP Biorepository team were responsible for consenting and sending the faecal kits at the NNUH site as well as storing samples and providing the clinical metadata. Dr. Ibrahim Sallam was also involved in the aliquoting, processing and DNA extraction of faecal samples collected at JPUH. I operationalised the BEAM study, this included setting up the pathway to determine when it would be best to intercept potential eligible patients. I aliquoted the samples once it arrived at QIB and processed the samples for sequencing. I performed the post-sequencing analysis for 16S rRNA gene amplicon sequencing using scripts provided by Dr. Raymond Kiu. Raymond did the post-sequencing analysis for shotgun metagenomic sequencing, for both gut microbiota taxonomy abundances and functional annotations. Dr. Matthew Dalby provided me with initial R scripts, which I used as a foundation to get initial plots and later added my own lines to create the plots shown in this thesis. I performed all the faecal culturing from BEAM samples and characterised the strains.

3.1   Breast cancer and the gut microbiota

As previous highlighted in the introduction, evidence exists to suggest a link between the gut microbiota and breast cancer. Briefly, the study performed by Jackson *et al.,* using the gut microbiota data from the Twins UK cohort demonstrated that BrCa patients had a

perturbed profile compared to controls [111]. Other observational studies focussing on PoM and PeM BrCa patients noted a difference in the gut microbiota profiles between the two groups [113, 114, 169]. Furthermore McKee *et al.*, demonstrated a mechanistic link where a perturbed gut microbiota induced by antibiotic administration resulted in a larger primary mammary tumours in mice [127]. These are just a few of the key studies published in the last decade that suggested a potential link between the gut microbiota and BrCa. Probing into this relationship could improve diagnostic or treatment outcomes for a disease that affects many women globally.

3.1.1   Clinical trials in the UK

Clinical trials, interventional or observational, are important for the development and validation of novel therapeutics. Observational studies are studies where clinical samples are collected from patients but no intervention i.e., a drug or treatment, is given. Any study working with human patients would need favourable ethics prior to starting. This is given by a recognised research ethics committees (REC) that are across the UK. UK universities often have internal RECs that adhere to the governance set out by the NRES, but they are not part of the NRES. University projects that require ethical approval can apply for approval from their university's REC. This was the case for the BEAM study, where ethical approval was given by FMH UEA research ethics sub-committee (reference: 201819-092 HT). However, as the BEAM study involves a partnership with the NHS, approval from the Human Research Authority (HRA) is required. The NRP Biorepository has been approved by the HRA (license number 11208) to work in partnership with the NHS, and also has appropriate NHS Research Ethics Committee approval. This allows the NRP Biorepository[7] to provide access to human tissue of studies that have ethical approval, which was also provided for the BEAM study (reference: BAC.001.A1). Both the FMH UEA research ethics application and the NRP Biorepository ethics application can be read in appendix: FMH REC application and BEAM Human Tissue Protocol, respectively. The Biorepository ethics application is different than the start date of the study as the NRP Biorepository introduced a new way of obtaining research ethics which led to the updated submission of documents with the new date.

---

[7] More information found at: https://biorepository.org.uk/about-biorepository/working-with-us/

3.1.2 Breast screening pathways in the UK

The National Healthcare Service follows guidance provided by NICE. For this thesis and in relation to this project, only the recommendation for: "suspected cancer: recognition and referral" will be discussed. According to the guidance, patients suspected of having breast cancer are referred by their general practitioner (GP) or referred to the clinic after attending the NHS Breast Screening Programme. Upon arrival to the clinic some or all of the following diagnostic tests are done: physical examination, mammogram, ultrasound and core biopsy [36, 37]. Having spoken to consultants and nurses at the breast care clinic at the NNUH I learnt that radiologists usually score the mammogram, and dependent on these, the patient is referred for further tests or not. Patients who are referred to the breast clinic for assessment will usually undergo some or all these tests in one morning at the NNUH.

3.1.3 Methods of investigating the gut microbiota

Sequencing technologies have allowed researchers to delve into different microbial niches and define ecosystem signatures, without the need to culture the bacteria [170]. Large sequencing studies utilising 16S rRNA gene amplicon sequencing or shotgun metagenomic sequencing can be used to determine trends and associations between gut microbiota profiles of healthy participants and compare them to disease cohorts. However, as mentioned in "Researching the gut microbiota" section, researchers cannot determine a causal relationship in exploratory observational studies like BEAM. Thus, to provide further insights that may help define potential causal relationships, microbes of interest can be isolated and subjected to whole genome sequencing and appropriate downstream mechanistic studies. However, most gut microbes are anaerobic and fastidious [103], and providing a physiological environment *ex vivo* can be difficult. There are also complex cross-feeding relationships which may make culturing individual microbes from faecal slurries problematic. Moreover, culturing can take place retrospectively after initial analysis, and appropriate storage of clinical samples is crucial in preserving the microbes. Inadequate storage conditions can bias microbes that are to be cultivated, where studies reported a decline in diversity in Bacteriodota over time at -80˚C [171, 172]. For our study we utilised glycerol (20% in PBS) to protect microbes from ice crystals produced by freezing [136, 172]. I compared the recovery between two faecal samples preserved with and without glycerol and noted that I could recover more colonies from glycerol preserved faecal samples[8], data not shown.

---

[8] We aliquoted 200mg of faecal sample from the same patient and stored it at -80°C for 2 months. We performed faecal culturing on BHI plates and observed more morphologically distinct colonies on the glycerol preserved sample (n = 14) vs. the non-glycerol preserved sample (n = 4).

As discussed previously, methods of investigating the gut microbiota each have their advantages and disadvantages. With sequencing becoming more affordable, shotgun metagenomics is preferred over 16S rRNA gene amplicon sequencing. Shotgun metagenomics sequences all genomic DNA in a sample thereby providing insights into e.g., potential functionality and metagenome assembled genomes, while 16S rRNA gene amplicon sequencing can only resolve to the genus level (although it is still cheaper than shotgun) [170]. However, one consideration for shotgun metagenomics is the necessity of additional computing 'power' given the large amount of data to be processed.

In contrast, 16S rRNA gene amplicon sequencing is a lot more affordable than metagenomics. This method uses the highly conserved 16S rRNA genome region to identify bacteria, accurate down to a genus level. 16S rRNA sequencing uses primers that flank one of the nine hypervariable regions of the 16S rRNA gene, with amplification allowing identification based on reference database alignments. However, a caveat is that this technique focusses on conserved regions, and two closely related species can share up to 99% sequence similarity. In addition, depending on which hypervariable region the researcher selects it can give different results [103, 173, 174]. 16S rRNA gene amplicon sequencing also does not require extensive computational tools or specialised bioinformatic pipelines that shotgun metagenomics demands. Nevertheless, if analysis is limited by cost, 16S rRNA gene amplicon sequencing can be an effective tool to achieve results.

3.1.4   The BEAM Study

A combination of tools i.e., 16S rRNA gene amplicon sequencing, shotgun metagenomics and cultureomics, were used to investigate the relationship between the gut microbiota and breast health. We had the pathway in place when the COVID-19 pandemic was declared by the WHO. Consequently, this caused an immense pressure and backlog of mammogram screenings and cancer diagnosis. Ultimately, the clinical team and I, mutually agreed that we could not actively recruit patients without compromising the standard of care given to patients. Thus, active BEAM recruitment at the NNUH stopped in January 2021. Recruitment at an alternative site (JPUH) started in August 2022 and is expected to end June 2024.

The BEAM study was a collaboration between the Day Procedure Unit (DPU), breast care clinic at the Norfolk Norwich university Hospital (NNUH, Norwich) and James Paget University Hospital (JPUH, Great Yarmouth) and the local tissue bank: NRP Biorepository. I will outline the study protocol and address the premise behind these decisions. Due to the

nature of the COVID-19 pandemic, there were some adjustments made in the recruitment pathway, which will be outlined below, to ensure risk was minimized for all BEAM participants. Following the operationalisation of the BEAM study we introduced methods of analysing and researching the gut microbiota including: 16S rRNA gene amplicon sequencing, whole genome shotgun sequencing and cultureomics approaches.

3.1.4.1   Ethics statement

This study gained favourable ethical approval by the Faculty of Medicine and Health Ethics board at the University of East Anglia (FMH 201819-092HT). The patients involved in the study provided signed and informed consent to participate in this study. The protocol for faecal collection was laid out by the NRP Biorepository (Norwich, UK), and was in accordance with the terms of the Human Tissue Act (HTA) 2004 and approved with license number 11208 by the HTA. The most up-to-date ethics application is provided in the appendices under: BEAM Human Tissue Protocol and FMH REC application.

3.1.5   Aims and hypothesis

Gut microbiota profiles will change due to treatment pathways that first-time diagnosed breast cancer patients experience. I hypothesised that a baseline gut microbiota becoming perturbed, driven by prophylactic antibiotics, translated to adverse clinical outcomes compared to those who did not have a perturbed gut microbiota at baseline. Following this, I aimed to investigate:

- The impact of prophylactic antibiotics and/or chemotherapy on gut microbiota profiles of patients.
- Explore possible associations with the gut microbiota and clinical outcomes.
- Investigate different methods of profiling the gut microbiota including: 16S rRNA gene amplicon, metagenomic sequencing and faecal culturing.

3.2   Results

3.2.1   Operationalising the BEAM study at NNUH

The recruitment target is described in Appendices: BEAM Human Tissue Protocol. Briefly, power calculations using microbiota diversity at 5% significance level suggested 72 subjects each samples thrice over a year would be enough to detect a true time-averaged chance, if that exists, at a 95% power. This was based off of the Goedert *et al*., [112], database using PASS software v8 with 100 simulations. Speaking to a BrCa oncologist, Dr. Susanna

Alexander, at the NNUH she was optimistic that this target could be achieved within one year as she sees over 700 patients a year, see Appendix: BEAM Human Tissue Protocol and FMH REC application for more information on the power calculations.

Upon starting the PhD project, the priority was to establish the recruitment pathway. During planning stages I met with the nurses at DPU and the breast care clinic to understand the pathway patients would have to undertake when they visit NNUH for their screening. I joined MDT meetings with the breast consultants and research staff to share study aims and obtain their input into the design of the patient information sheet (PIS). For the recruitment pathway it was agreed to be as minimally disruptive to the current NHS clinical pathways. In doing so I believed compliance and agreement with the clinical staff would be high and would have the best chance of success in achieving the recruitment target.

The team and I eventually decided on the following recruitment pathway. We would include the PIS in the letter that would invite patients for their mammogram. This would allow patients to read and be aware of the study prior to their appointment at NNUH. A member of the Biorepository team would sit in the breast care clinic for a morning during the one-stop screening clinics. Anyone who would be undergoing a mammogram would be asked if they were interested in consenting into the study, and if consented would be given their first faecal kit to collect a baseline sample at home. Patients that were later diagnosed with BrCa would undergo screenings at the DPU for their pre-operative assessment. At this screening, they would be asked for a blood sample to allow researchers to investigate metabolites. After surgery, where applicable, patients would receive a faecal kit for at-home collection, which would be returned in a bio-safety postal box. When patients visited the NNUH at the "Moving forward" appointment they would receive their third kit (usually 6 months after diagnosis). Finally, the fourth kit would be sent a year after consent into the study, Figure 3:1. For control patients, the original plan was to only request one sample at baseline, which was the point of consent. This would be representative of the control group. However, it was soon recognised that microbiota profiles changed, and we amended the study protocol to reflect that up to four kits could be sent should the participant confirm participation. The sampling would have been done every three months.
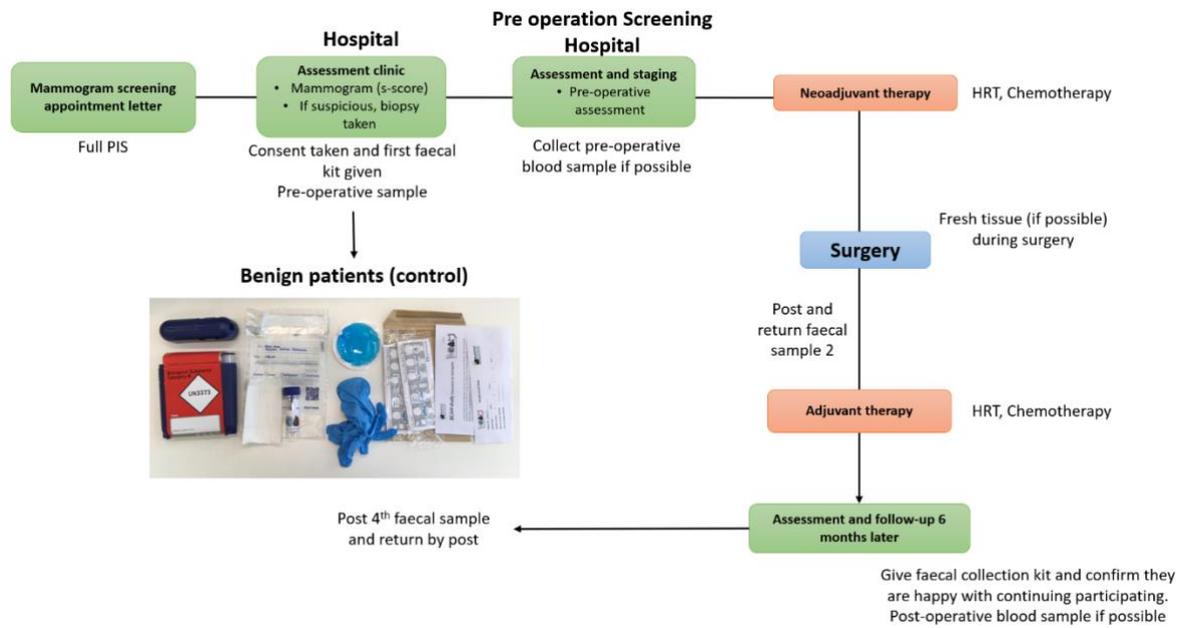
*Figure 3:1: **Proposed pathway for collecting human samples of patients consented into the BEAM study.** The main priority were faecal samples. At the NNUH patients did receive prophylactic antibiotics. The faecal kit that patients receive is shown in the image above this includes: a bio-safety box, absorbent sheet and leak-proof bag, ice pack, 20mL spoon Universal tube, nitrile gloves, faeces Catcher and documentation related to the study.*

Our main eligibility criteria for the BEAM study were as follows:

1. Females (or males) aged between 30-60 years of age.

I made the decision to exclude patients below the age of 30 as I believed BrCa diagnosis may be driven primarily by genetic factors, thus by excluding this age group I hoped to control a potential confounding factor. In the study protocol (Appendix: BEAM Human Tissue Protocol), the exclusion criteria related to minimum age is the same at JPUH and NNUH i.e., anyone below the age of 30 will be excluded. Despite BrCa being more common in patients above 60-year-old, it is well known that microbiota profiles change with age, thus having a 'younger' cohort would minimise age-associated changes [175]. I did not exclude males as male breast cancer cases are rare and would thus represent a potentially important and novel sample set. Dietary information through an online survey was also requested. We did change the age criteria for JPUH as clinicians and I noticed that the demographic of patients who were diagnosed with BrCa were often above 60-year-old. I decided to change the eligibility to "above the age of 30" and allowed the clinician consenting the patient determine if they were within capacity to consent. No dietary information was requested from patients who were consented at JPUH.

2. No antibiotics of any kind, 3 months prior being recruited into the study.

Antibiotics can severely reduce the diversity of the microbiota, which may mask any baseline trends [176]. As I was interested in following microbiota profiles as patients undergo treatment, I needed to establish their initial 'baseline'. This criterion was based on patients disclosing this information at the point of consent.

3. No previous primary cancer diagnosis of any kind

There are some studies that have reported that particular microbiota members link to certain cancer types. To adjust for this effect, I decided to exclude any patients who had a previous cancer diagnosis of any kind. Additionally, primary cancer diagnosis i.e., first-time diagnosed patients would be more representative of the general population and allow us to potentially find a signature that could aid in earlier diagnosis or better outcomes.

Although we tried to exclude several important influencing effects, some effects we could not adjust for e.g., diet or smoking habits, thus for these we relied on (optional) patient questionnaires.

3.2.2   BEAM patient characteristics

A total of 5 BrCa patients, and 1 control patient were recruited into the BEAM study from the NNUH. At the point of writing this chapter (May 2023), a total of 29 BrCa patients were recruited from the JPUH, with no controls. The summary of patients recruited, and their samples provided at either site is shown in Table 3:1, a summary of patient characteristics is found in Table 3:2. We followed up with the NNUH patients and were pleased to know all of them are in remission.

Patient ID	BrCa type	Age	Chemotherapy	Samples given			
				Baseline	Post-surgery	6-months	One-year
NNUH001	ILC	51	Cyclophosamide: 3x900mg; Epirubicin: 3x180mg doses; Flurouracil: 3x900mg; Docetaxel: 3x180mg	x	x		x
NNUH002	ILC	60		x	x		x
NNUH003	IDC	46	Cyclophosamide: 3x800mg; Epirubicin: 3x160mg; Flurouracil: 3x800mg; Docetaxel: 2x160mg and1x180mg	x	x		x
NNUH004	ILC	48		x	x		x
NNUH005	IDC	58			x		x
NNUH006	Control	51		x	x	x	x
JPUH001	IDC	58		x	x	x	
JPUH002				x	x	x	
JPUH003	ILC	65		x	x	x	
JPUH004	IDC	58		x	x		
JPUH005	IDC	65		x	x		
JPUH006				x	x	x	
JPUH007				x	x	x	
JPUH008				x	x		
JPUH009				x	x		
JPUH010				x	x		
JPUH011				x	x		
JPUH012				x	x		
JPUH013				x	x		

JPUH014			x	x
JPUH015	IDC	51	x	x
JPUH016	ILC	60	x	x
JPUH017	IDC	65	x	x
JPUH018	IDC	56	x	x
JPUH019	IDC	57	x	
JPUH020	ILC	62	x	
JPUH021	NEUROENDOCRINE	65	x	
JPUH022			x	
JPUH023	IDC	65	x	x
JPUH024	IDC	65	x	x
JPUH025			x	x
JPUH026	IDC	54	x	x
JPUH027	IDC	64	x	x
JPUH028	IDC	68	x	
JPUH029	IDC	68	x	

*Table 3:1: **Summary of patient characteristics enrolled in the BEAM study**. NNUH: Norfolk and Norwich University Hospital, JPUH: James Paget University Hospital denotes where the patient was enrolled. IDC: invasive ductal carcinoma, ILC: invasive lobular carcinoma. Last updated as of 4 May 2023 and recruitment is ongoing until April 2024 which is reflected in missing samples. Missing metadata reflects unknowns due to the information not being recorded into the database in time for thesis submission. In green are samples submitted for 16S rRNA gene amplicon sequencing only, and in yellow are samples submitted for both 16S rRNA gene amplicon (n = 54) and shotgun sequencing (n = 34).*

Characteristic	Overall (n = 35)
Age: median (IQR), years	60 (55-65)
BrCa status: IDC | ILC | Control | Unknown/Others, %	42.9 % | 17.1% | 2.9% | 37.1%
Site: NNUH | JPUH, %	22% | 78%
Female, %	100%
Clinical outcome NNUH patients: remission or recurrence, %	Remission (100%)
Chemotherapy: yes | no | unknown, %	11.4% | 54.3% | 34.2%
ER Status: positive | negative | unknown, %	45.7% | NA | 54.3%
HER2 Status: positive | negative | unknown, %	5.7% | 40% | 54.3%

*Table 3:2: **Summary of patient characteristics in the BEAM study.***

3.2.3   16S rRNA gene amplicon sequencing

All samples coloured in Table 3:1 were submitted for 16S rRNA gene amplicon sequencing, targeting the V1 and V2 hypervariable region of the 16S rRNA gene. At the time of submitting this thesis, recruitment was still ongoing until April 2024 as such not all samples were submitted for sequencing and not all metadata was shared in time of thesis submission. Samples from 6-months and one-year are missing from JPUH patients as recruitment for this group started in August 2022. My thesis submission was June 2023 resulting in only a small number of patients being able to provide their 6-month sample which could have been submitted for sequencing and thus included in this thesis.

A total of 54 samples were submitted, two of which did not pass QC, and one had less than 2000 reads. Therefore, a total of 51 samples were available for downstream analysis. Sequencing was performed on the NextSeq PE300 with an average read depth of 440Mbp, and a median read number of 128,175. With the help of Dr. Ibrahim Sallam, I obtained the metadata of most patients recruited into the BEAM study. Metadata collected included: age, type of BrCa (limited to invasive ductal carcinoma (IDC) or invasive lobular carcinoma (ILC), chemotherapy (if applicable, type and regime), and radiotherapy (if applicable, regime).

3.2.3.1   *Site is an influencing factor to the gut microbiota profile of patients*

I initially intended to pool NNUH and JPUH samples, however random selection of 18 samples from JPUH patients indicated a (minimal) deviation between the sites (non-metric distance scaling (NMDS) plots, Figure 3:2), Additional Permanova analysis suggested site was a significant factor in explaining sample differences, where $p = 0.001$. Thus, I decided to analyse the samples two ways: pooled and JPUH only. Although only 18/51 samples were from NNUH, given the significant recruitment efforts I wanted to include them with the expectation that any significant observation would likely be driven by JPUH due to the larger sample size.
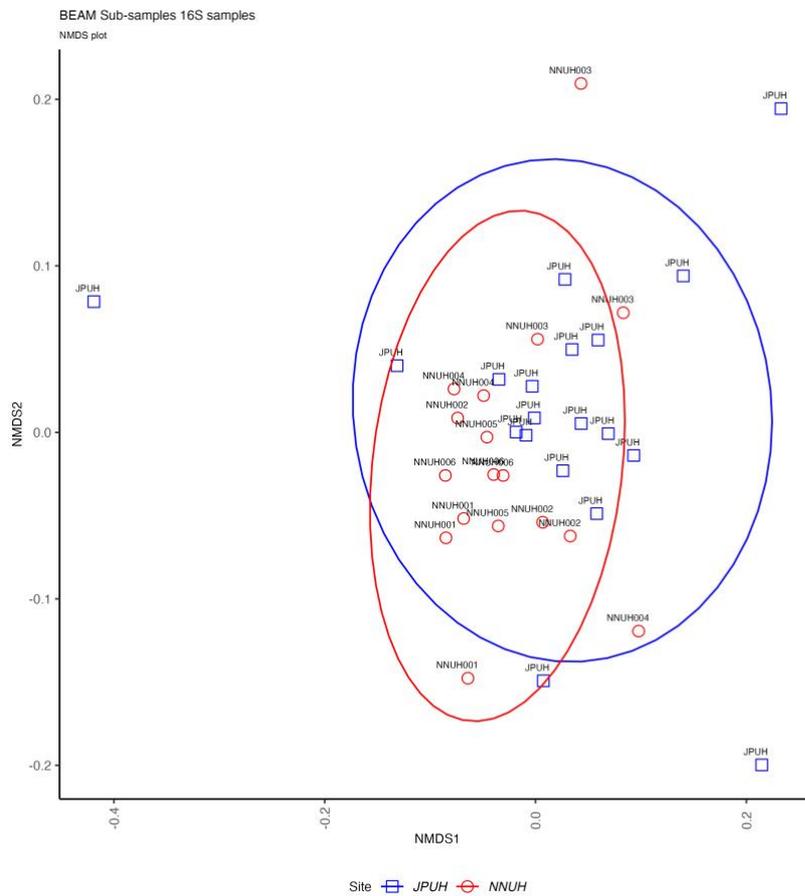
*Figure 3:2: **NMDS plot by Bray-Curtis dissimilarity, of a sub-sample of BEAM 16S rRNA gene amplicon sequencing at genus level.** Samples annotated by Site, where blue is JPUH and red is NNUH. Permanova of the samples confirmed that site was significant in explaining the spread of samples, where p = 0.001. Time point was not significant, p = 0.55.*

3.2.4   Initial analysis of all BEAM samples by time point

Initial assessment of all BEAM samples (JPUH and NNUH) involved assessing alpha and beta diversity. I limited the taxonomy level to genus level as this is more accurate than species for 16S rRNA gene amplicon sequencing [177]. Across time points the trend was a decreasing Shannon diversity (Figure 3:3A), but at the time of submitting this thesis only three samples were submitted for 16S rRNA gene amplicon sequencing, two were from NNUH and one from JPUH where unfortunately one NNUH sample did not pass sequencing QC and was dropped from analysis. For one-year we only had samples from NNUH patients submitted for sequencing, and none from any JPUH patients. As I had determined previously that the gut microbiota from JPUH and NNUH differed, the decreasing could also have been influenced by the fact that the last time point was primarily only NNUH patients and not JPUH. There were no significant differences in alpha diversity across time points, determined by Kruskal-Wallis where $p = 0.61$ but this could be due to the small sample size of the study. Additionally, there were less samples at 6-months and one-year, and these samples came from NNUH patients, which may drive the trend that was observed since site was an influential factor as previously reported. Focussing only on baseline and post-

surgery, which have 21 and 26 samples respectively, there was a clear trend of increasing diversity after surgery despite not being statistically significant.

The NMDS plot, annotated by time point (Figure 3:3B), assessed beta diversity. I observed some separation between baseline and one-year. Post-hoc pairwise analysis on time points was done using pairwiseAdonis v-0.4 [178]. Bonferroni correction suggested that post-surgery and one-year was significant (p = 0.03). However, all six one-year samples are from NNUH, and site was a significant influencing factor. Top ten genera, determined by the sum of relative abundances by genera across all samples, of BEAM patients were: *Bacteroides, Blautia, Faeclibacterium, Collinsella, Bifidobacterium, Ruminococcus, Coprococcus, Eubacterium, Dialister* and *Oscillospira.* Visualised by time point there were no discernible genus changes (Figure 3:3C).

(A)

(B)

(C)
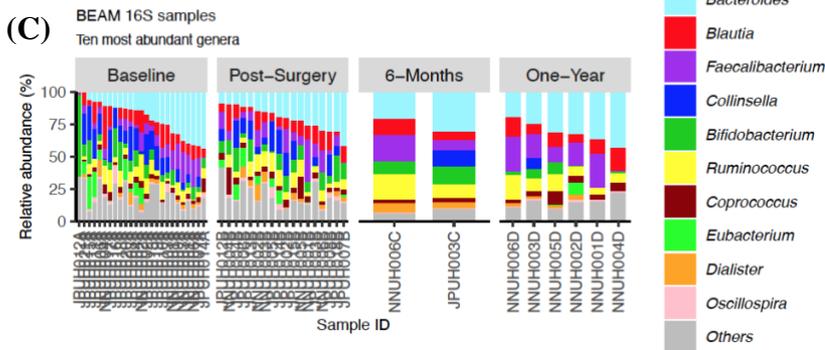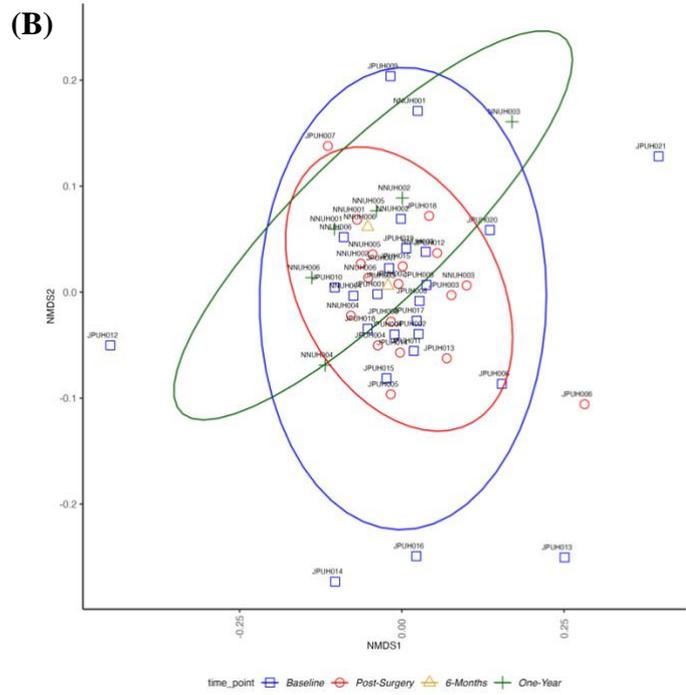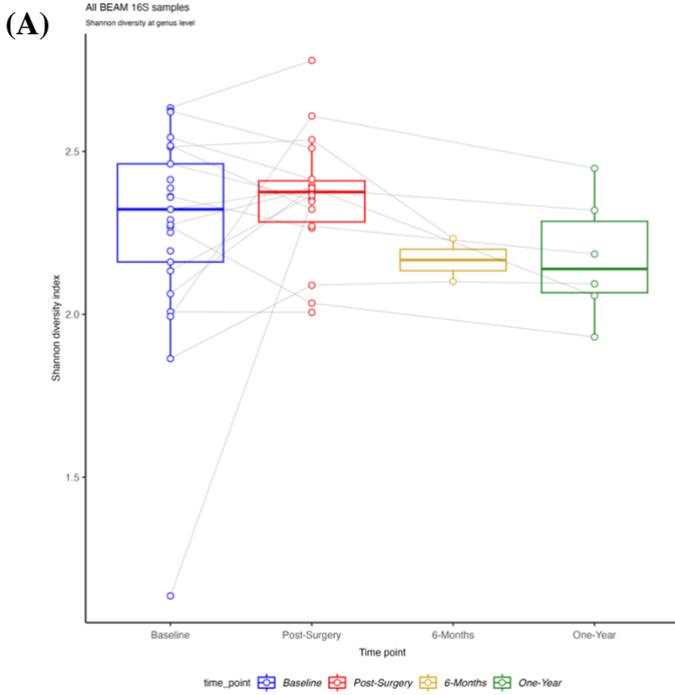
*Figure 3:3: **Initial analysis of diversity at genus level of all BEAM samples that underwent 16S gene amplicon sequencing, targeting the V1-V2 hypervariable region.** Shannon diversity index was used to assess alpha diversity at genus level, and showed no significant changes at any time point (A) while for beta diversity, (B), the time point post-surgery (red) and one-year (green) was significant, p=0.03, according to multilevel pairwise comparison with Bonferroni correction. However, the sample sizes were 26 and 6 for post-surgery and one-year and all one-year samples came from NNUH patients which was a significant influencing factor. Beta diversity was assessed by plotting a NMDS plot using Bray-Curtis dissimilarity matrix. General taxa analysis by top ten genera, determined by the total relative abundance across all samples did not show anything significant (C).*

3.2.5 JPUH BEAM sample analysis
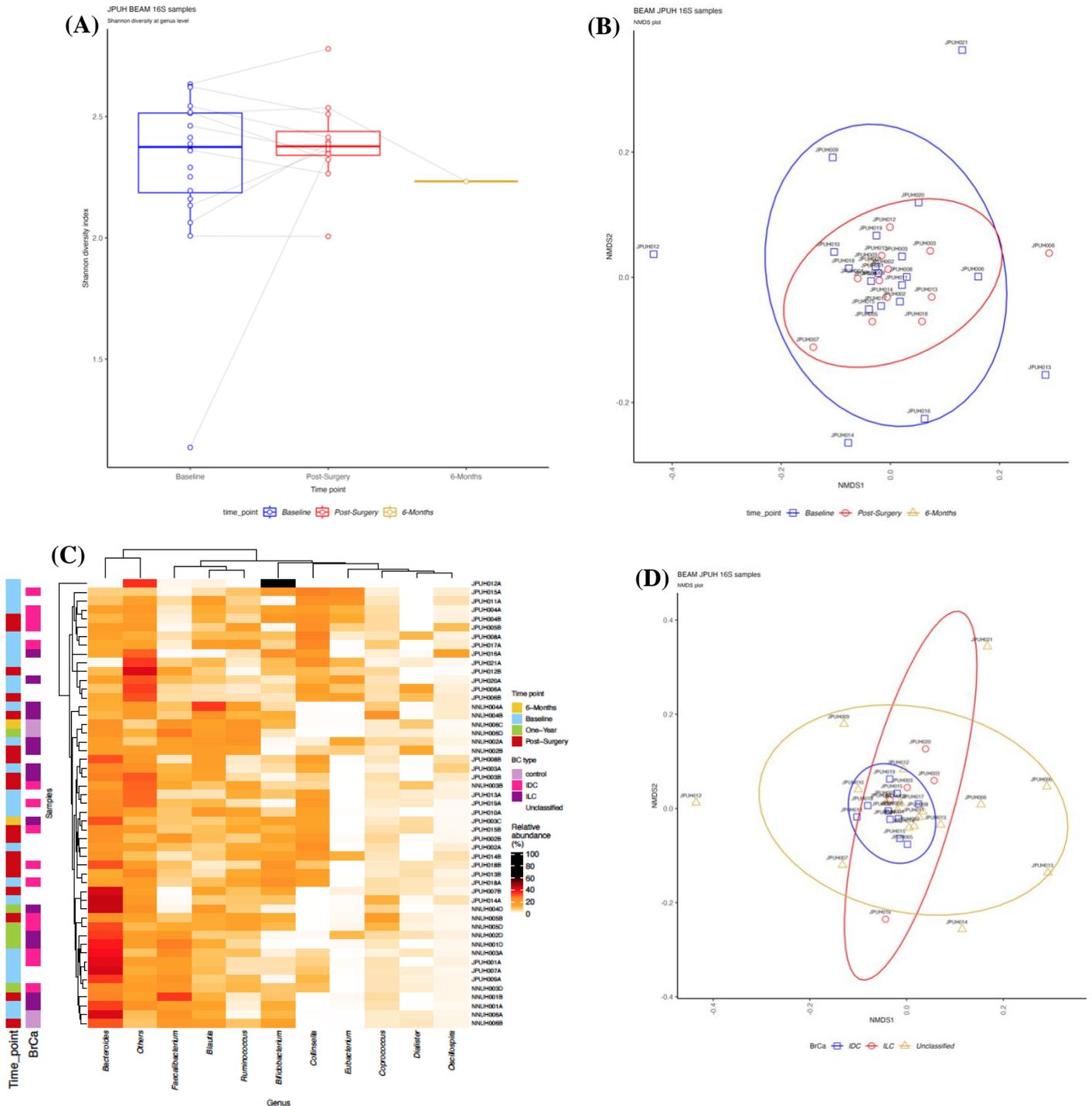
*Figure 3:4: **Analysis of BEAM 16S rRNA samples from JPUH only at genus level.** Alpha diversity was assessed by Shannon diversity index (A), Kruskal-Wallis confirmed limited changed were observed across the time points, p = 0.47. Beta diversity was assessed by a NMDS plot by Bray-Curtis dissimilarity matrix. Permanova confirmed no significant changes by time point, p =0.99. Heatmap by Bray-Dissimilarity matrix showed some clustering of a relative high abundance of Bacteroides (C), but this was not significant by Permanova (p=0.86). Annotating the NMDS plot by BrCa status showed different clustering for ILC compared to IDC (D).*

Re-analysis of the JPUH samples indicated no significant differences by time point for either alpha, Shannon diversity (Figure 3:4A), or beta diversity (Figure 3:4B) by NMDS (confirmed by Kruskal-Wallis and Permanova where p = 0.47 and 0.99 for alpha and beta diversity respectively). Visualising by heatmap did indicate some clustering based on

Bacteroides relative abundances, but this was not associated with BrCa status (Figure 3:4C). Annotating a NMDS plot (Figure 3:4D) by BrCa status showed different clusters had formed, however this was not significant according to Permanova, p = 0.86. The data shown suggests there is little difference in the gut microbiota profiles of ILC or IDC BrCa patients The trend however suggests there may be an association with *Bacteroides*, as indicated in Figure 3:4C, but due to the underpowered study significance may not have been detected.

3.2.6   Metagenomic shotgun sequencing

A total of 34 samples, as these were available at the time of submitting for sequencing, underwent metagenomic shotgun sequencing. Reads had an average depth of 3.63Gbp and median read numbers of 18,235,166. This included all NNUH samples, 18 samples: 5 baseline, 6 post-surgery, one 6-months and 6 one-year, and 16 JPUH samples: 11 baseline, 6 post-surgery, making up a total of: 15 baseline samples, 12 post-surgery samples, one 6-months and 6 one-year samples. Top ten genera were: *Bacteroides, Faecalibacterium, Phocaeicola, Blautia, Coprococcus, Alistipes, Anaerostipes, Bifidobacterium, Roseburia and Mediterraneibacter* (Figure 3:5). The top ten species were: *Faecalibacterium prausnitzii, Akkermansia muciniphila, Blautia* spp.*, Phocaeicola vulgatus, Collinsella aerofaciens, Bacteroides uniformis, Escherichia coli, Bacteroides cellulosilyticus, Anaerostipes hadrus* and *Phocaeicola dorei.* As I did not observe any significant profiles from the 16S rRNA gene amplicon sequencing data on genus level, I focussed my analysis on species level for the shotgun metagenomics data given its better accuracy.
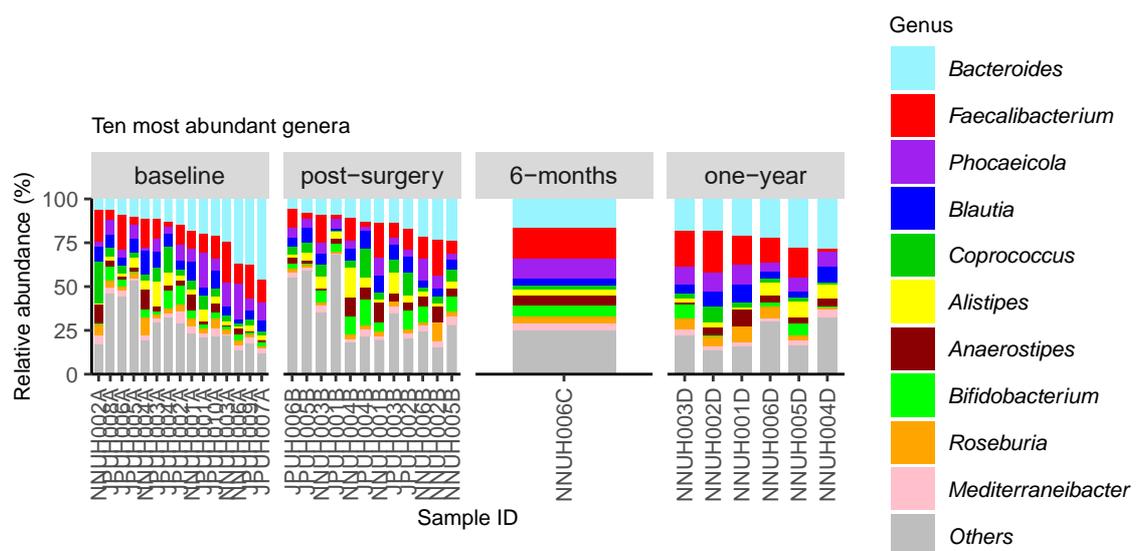
*Figure 3:5: **Top ten genera of BEAM shotgun metagenomic samples by time point.***

Alpha diversity was assessed using Shannon diversity. Initial visualisation indicated alpha diversity trended to decrease across time points, Figure 3:6A. However, the number of

patients at time points 6-months and one-year, is only one and six respectively. Kruskal-Wallis testing confirmed there were no differences in Shannon diversity at each time point, where p = 0.244 and 0.330 for genus and species respectively. No segregation based on time points was observed for beta diversity, Figure 3:6B, which was confirmed by Permanova, where p = 0.274 and 0.611 for genus and species respectively. Visualised as a heatmap no clustering of samples was observed (Figure 3:6C). However, Permanova by BrCa status indicated a shift (p = 0.001), however after post-hoc Bonferroni pairwise comparison this significance was lost (p = 0.1, Figure 3:6D). Further analysis indicated that of all ILC samples, n = 11, nine were from NNUH patients, which may link to site being a variable rather than BrCa status.

(A)

(B)

(C)

(D)

*Figure 3:6: **BEAM metagenomic shotgun samples analysis at species level.** Alpha diversity (A) was assessed by Shannon diversity index across time points. Beta diversity was visualised with a NMDS plot by Bray-Curtis dissimilarity matrix, plots were annotated by time point (B) and BrCa status (D). Permanova was used to determine if there are any significant shifts by time point or BrCa status, which was p = 0.64 and 0.001 respectively. Post-hoc pairwise comparison showed that it was a weak significance (p = 0.1) where the BrCa status was ILC or control, which is most likely an artefact of site. A heatmap by Bray-Curtis dissimilarity matrix and annotated by time point and BrCa status showed no clustering of samples (C).*

3.2.6.1 *Bifidobacterium and BEAM samples*

Previous literature, and Hall and Robinson lab studies, indicates the potent immunological effects of *Bifidobacterium* species in anti-cancer therapies. Thus, I decided to assess *Bifidobacterium* species profiles in our BEAM samples. Visualising as a stacked bar plot by BrCa status, the one control sample had a high relative abundance of *Bifidobacterium pseudocatenulatum* (Figure 3:7), with *Bifidobacterium longum* the most common species in patient samples.
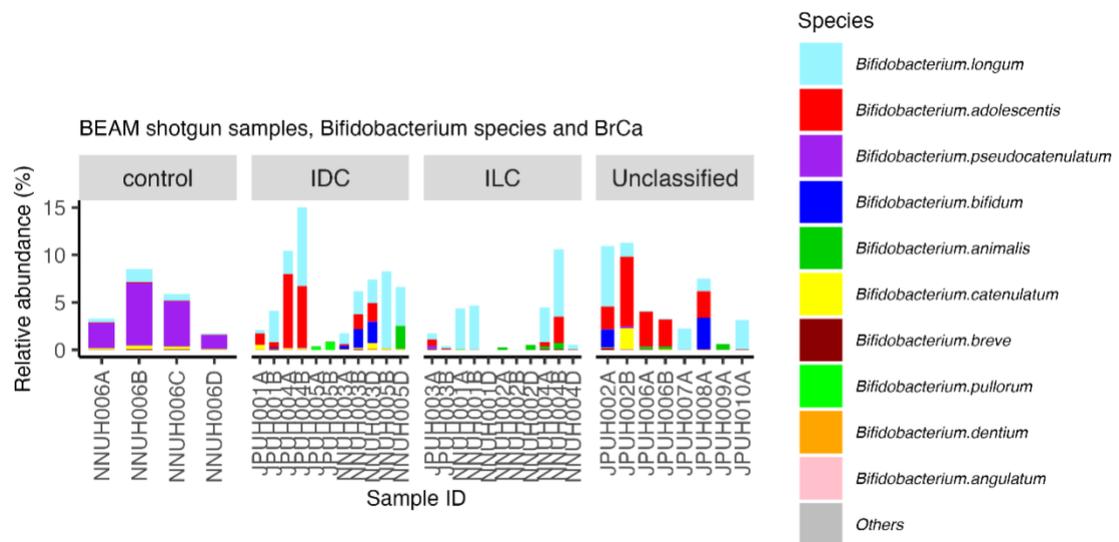
*Figure 3:7: **Bifidobacterium species in BEAM samples.***

3.2.7 LEfSe analysis

I also ran both 16S rRNA gene amplicon and shotgun metagenomic sequencing reads through linear discriminant analysis of effect size (LEfSe) with BrCa status (ILC or IDC) as the outcome. No significant determinant features could be determined. However, this could be due to the proportion of unclassified metadata.

3.2.8 Cultureomic approaches

The initial plan was to culture out candidate bacteria from the faecal sample based on downstream metagenomics and clinical metadata analysis. However, with the slow initial recruitment, later impacted by the COVID-19 pandemic, it was decided to do non-specific (i.e., non-targeted) faecal culturing to create a bank of isolates obtained from BrCa patients. The first round of culturing was performed with YCFA media supplemented with carbohydrates as specified in [179]. The initial round also included ethanol-shocking the bacterial cells to promote the isolation of spore-formers. In subsequent rounds of culturing BHI media was included, and ethanol-shocking stopped as I noticed most of the ethanol treated isolates were either *Turicibacter* or *Clostridium.* A total of ten rounds of culturing

took place, I isolated 298 strains of which 190 underwent QC and whole genome sequencing. I had a total of 138 confirmed pure cultures with an average nucleotide identity (ANI) of >95%, 15 isolates with an ANI <95% and >80%. Fifty-three species were isolated with two potential novel ones (described in more detail below, Table 3:3).

To investigate the phylogenies of these isolates I downloaded 5 reference strains per isolate and used this to create a phylogenetic tree using Mashtree v1.2.0 and visualised it using iTOL v6. The tree is visualised in Figure 3:8, to view at better resolution see: https://itol.embl.de/export_shared/149155192142892616801643I9/BEAM_2023.svg.

Of the 154 strains that went through the QC process and a match of ANI >80%, 38.4% were *Bacteroides* with 12 individual species being identified. I only isolated one of each *Blautia, Collinsella* and *Escherichia*, and two individual *Alistipes* isolates. 13% were *Bifidobacterium* across 4 different species. These isolates only cover six out of the ten top genera that was reported with metagenomic shotgun sequencing. The remaining four genera could not be cultured. Notably, two isolates could not be matched to the GTDB-Tk database, suggesting they represented novel taxa.
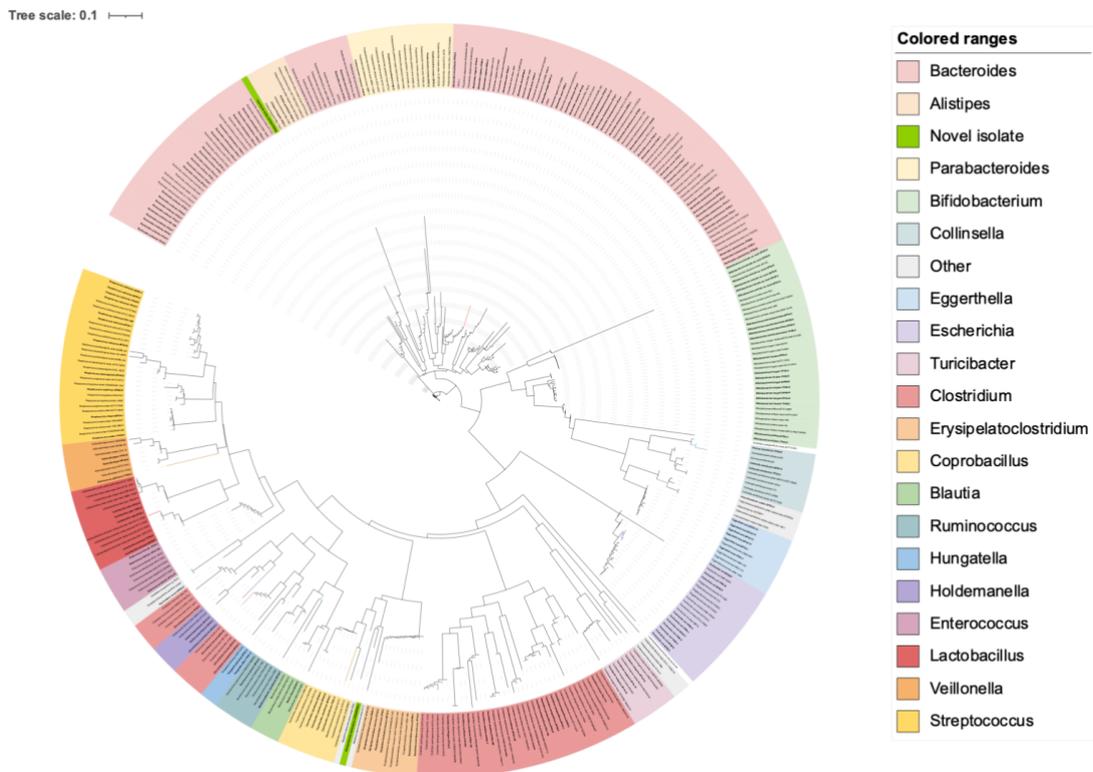
*Figure 3:8: **A phylogenomic tree made from the 167 pure strains, in bold are those isolated from BEAM patients.** The tree was generated using Mashtree v1.2.0 using 100 bootstrapping repetitions. The tree was visualised using iTOL v6. I downloaded 5 reference strains per isolate which was identified using GTDB-Tk v4.5.1. Coloured ranges denote genera and novel isolates are in solid green.*

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
				Bacteroidaceae			
YP3D-E	46	4567178	41.92	*Bacteroides caccae* JCM 9498	*Bacteroides caccae*	99.06	Obligate anaerobe
YP4D-C	34	5244670	42.86	*Bacteroides caccae* JCM 9498	*Bacteroides caccae*	99.36	Obligate anaerobe
YP4D-H	36	5249648	42.85	*Bacteroides caccae* JCM 9498	*Bacteroides caccae*	99.32	Obligate anaerobe
YP4A-D	162	7318751	43.18	*Bacteroides cellulosilyticus* JCM 15632	*Bacteroides cellulosilyticus*	97	Obligate anaerobe
YP4A-F	109	7317576	43.18	*Bacteroides cellulosilyticus* JCM 15632	*Bacteroides cellulosilyticus*	97	Obligate anaerobe
YP4B-A	142	7311413	43.19	*Bacteroides cellulosilyticus* JCM 15632	*Bacteroides cellulosilyticus*	97	Obligate anaerobe
YP6A-D	72	6634155	42.74	*Bacteroides cellulosilyticus* JCM 15632	*Bacteroides cellulosilyticus*	98.47	Obligate anaerobe
YP6A-E	75	6633803	42.74	*Bacteroides cellulosilyticus* JCM 15632	*Bacteroides cellulosilyticus*	98.47	Obligate anaerobe
YP6A-G	75	6636186	42.74	*Bacteroides cellulosilyticus* JCM 15632	*Bacteroides cellulosilyticus*	98.54	Obligate anaerobe
BP5B-A	72	5234730	41.9	*Bacteroides dorei* 175	*Bacteroides dorei*	99	Anaerobe
YP3D-C	55	5392824	41.71	*Bacteroides dorei* 175	*Bacteroides dorei*	99.15	Anaerobe
YP4D-F	94	5647888	42.12	*Bacteroides dorei* 175	*Bacteroides dorei*	98.98	Anaerobe
YP6B-C	45	4896739	43.2	*Bacteroides fragilis* ATCC 25285	*Bacteroides dorei*		Anaerobe
BP2A-13	64	5109508	43.48	*Bacteroides fragilis* ATCC 25285	*Bacteroides fragilis*	98.9	Anaerobe
YP6A-A	103	4243020	41.8	*Bacteroides massiliensis*	*Bacteroides massiliensis*	98.71	Anaerobe
BP2A-8	2057	8509356	46.47	*Bacteroides xylanisolvens* XB1A	*Bacteroides ovatus*	91	Obligate anaerobe
YP1D-E	64	6425065	41.7	*Bacteroides ovatus* JCM 5824	*Bacteroides ovatus*	96.45	Obligate anaerobe
YP2A-3	142	6628251	41.89	*Bacteroides xylanisolvens* XB1A	*Bacteroides ovatus*	91	Obligate anaerobe

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
YP2A-7	120	6634700	41.89	*Bacteroides xylanisolvens* XB1A	*Bacteroides ovatus*	91	Obligate anaerobe
YP5B-B	159	6342733	41.71	*Bacteroides xylanisolvens* XB1A	*Bacteroides ovatus*	91	Obligate anaerobe
YP6A-F	156	6799318	41.75	*Bacteroides koreensis* YS-aM39	*Bacteroides ovatus*	96.76	Obligate anaerobe
YP6A-H	161	6795423	41.74	*Bacteroides koreensis* YS-aM39	*Bacteroides ovatus*	96.74	Obligate anaerobe
YP1D-B	102	5164401	41.96	*Bacteroides salyseriae* JCM 12988	*Bacteroides salyersiae*	98.87	Anaerobe
YP1D-F	101	5108927	41.99	*Bacteroides salyseriae* JCM 12988	*Bacteroides salyersiae*	98.72	Anaerobe
YP1D-A	77	3707042	45.84	*Bacteroides stercoris* ATCC 43183	*Bacteroides stercoris*	98.24	Anaerobe
BP4D-C	104	6414207	43.2	*Bacteroides thetaiotaomicron* JCM 5827	*Bacteroides thetaiotaomicron*	97.76	Anaerobe
YP1B-E	165	6469099	42.97	*Bacteroides thetaiotaomicron* JCM 5827 16S ribosomal RNA, partial sequence	*Bacteroides thetaiotaomicron*	98	Anaerobe
YP2A-1	143	6592204	43.15	*Bacteroides thetaiotaomicron* JCM 5827	*Bacteroides thetaiotaomicron*	97	Anaerobe
YP2A-11F	116	6592566	43.14	*Bacteroides faecis* MAJ27	*Bacteroides thetaiotaomicron*	97	Anaerobe
YP2A-2	90	6597907	43.14	*Bacteroides thetaiotaomicron* JCM 5827	*Bacteroides thetaiotaomicron*	97	Anaerobe
YP2D-G	85	6621143	43.12	*Bacteroides thetaiotaomicron* JCM 5827	*Bacteroides thetaiotaomicron*	97.58	Anaerobe
YP6A-B	79	6094134	42.76	*Bacteroides thetaiotaomicron* JCM 5827	*Bacteroides thetaiotaomicron*	99.11	Anaerobe
BP3D-A	86	4762471	46.32	*Bacteroides uniformis* JCM 5828	*Bacteroides uniformis*	98.21	Microaerophile

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
BP3D-F	74	4894497	46.51	*Shigella flexneri* ATCC 29903	*Bacteroides uniformis*	99.05	Microaerophile
YP1B-C	77	4639941	46.53	*Bacteroides uniformis* JCM 5828 16S ribosomal RNA, partial sequence	*Bacteroides uniformis*	98	Microaerophile
YP2A-10	43	4919119	46.25	*Bacteroides thetaiotaomicron* JCM 5827	*Bacteroides uniformis*	97	Microaerophile
YP2A-16B	73	5393885	46.29	*Bacteroides uniformis* JCM 5828	*Bacteroides uniformis*	97	Microaerophile
YP2A-4	84	5418990	46.29	*Bacteroides uniformis* JCM 5828	*Bacteroides uniformis*	96	Microaerophile
YP2A-5	78	5380772	46.28	*Bacteroides thetaiotaomicron* JCM 5827	*Bacteroides uniformis*	96	Microaerophile
YP2A-6	75	5418013	46.29	*Bacteroides thetaiotaomicron* JCM 5827	*Bacteroides uniformis*	96	Microaerophile
YP2A-8	81	5420309	46.29	*Bacteroides thetaiotaomicron* JCM 5827	*Bacteroides uniformis*	96	Microaerophile
YP2D-D	79	5372482	46.29	*Bacteroides uniformis* JCM 5828	*Bacteroides uniformis*	97.35	Microaerophile
YP2D-F	84	5410000	46.31	*Bacteroides uniformis* JCM 5828	*Bacteroides uniformis*	96.86	Microaerophile
YP3D-F	61	4791320	46.56	*Bacteroides uniformis* JCM 5828	*Bacteroides uniformis*	98.97	Anaerobe, microaerophile
YP5B-I	44	4921592	46.25	*Bacteroides uniformis* JCM 5828	*Bacteroides uniformis*	97	Microaerophile
YP5B-J	41	4908890	46.26	*Bacteroides uniformis* strain JCM 5828	*Bacteroides uniformis*	97	Microaerophile
YP5D-D	43	4526562	46.47	*Bacteroides uniformis* JCM 5828	*Bacteroides uniformis*	97.15	Microaerophile

BEAM Study

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
BP2A-6	146	5027985	42.18	*Bacteroides vulgatus* ATCC 8482	*Bacteroides vulgatus*	99	Anaerobe
YP1D-K	224	5041572	42.18	*Bacteroides vulgatus* ATCC 8482	*Bacteroides vulgatus*	98.95	Anaerobe
YP2B-C	132	5037380	42.2	*Bacteroides vulgatus* ATCC 8482 16S ribosomal RNA, partial sequence	*Bacteroides vulgatus*	98	Anaerobe
YP2A-11	152	6630882	41.89	*Bacteroides xylanisolvens* XB1A	*Bacteroides xylanisolvens*	97	Obligate anaerobe
YP2B-A	147	6638792	41.89	*Bacteroides xylanisolvens* XB1A 16S ribosomal RNA, partial sequence	*Bacteroides xylanisolvens*	97	Obligate anaerobe
YP2D-C	388	6795874	41.89	*Bacteroides xylanisolvens* XB1A	*Bacteroides xylanisolvens*	97.63	Obligate anaerobe
YP5B-C	179	6337793	41.71	*Bacteroides xylanisolvens* XB1A	*Bacteroides xylanisolvens*	97	Obligate anaerobe
YP5B-M	156	6339943	41.7	*Bacteroides xylanisolvens* XB1A	*Bacteroides xylanisolvens*	98	Obligate anaerobe
Bifidobacteriaceae							
BP2D-E	62	1921384	60.48	?	*Bifidobacterium animalis*	99.89	Anaerobe
BP5D-A	24	1919909	60.5	*Bifidobacterium animalis* subsp. *lactis*	*Bifidobacterium animalis* subsp. *lactis*	99.97	Anaerobe
BP5D-B	29	1920179	60.51	*Bifidobacterium animalis* subsp. *lactis*	*Bifidobacterium animalis* subsp. *lactis*	99.97	Anaerobe
BP5D-C	28	1919341	60.5	*Bifidobacterium animalis* subsp. *lactis*	*Bifidobacterium animalis* subsp. *lactis*	99.97	Anaerobe
BP5D-D	26	1920567	60.49	*Bifidobacterium animalis* subsp. *lactis*	*Bifidobacterium animalis* subsp. *lactis*	99.97	Anaerobe

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
BP5D-E	59	1937737	60.22	*Bifidobacterium animalis* subsp. *lactis*	*Bifidobacterium animalis* subsp. *lactis*	99.97	Anaerobe
BP3D-C	57	2168388	62.93	*Bifidobacterium bifidum* NBRC 100015	*Bifidobacterium bifidum*	99.05	Anaerobe
YP3D-D	56	2166843	62.92	*Bifidobacterium bifidum* NBRC 100015	*Bifidobacterium bifidum*	99.01	Anaerobe
BP4D-D	60	2366964	60.09	*Parabacteroides merdae* JCM 9497	*Bifidobacterium longum*	98.65	Anaerobe
BP5B-B	71	2321282	60.09	*Bifidobacterium longum* subsp. *suillum* Su 851	*Bifidobacterium longum*	98	Anaerobe
BP6A-D	35	2382132	59.97	*Bifidobacterium longum* subsp. *suillum* Su 851	*Bifidobacterium longum*	96.29	Anaerobe
EP4A-D	37	2299957	59.98	*Bifidobacterium longum* subsp. *suillum* Su 851	*Bifidobacterium longum*	98	Anaerobe
YP3A-A	66	2418113	60.15	*Bifidobacterium longum* subsp. *suillum* Su 851	*Bifidobacterium longum*	98	Anaerobe
YP4B-C	44	2294132	60.05	*Bifidobacterium longum* subsp. *suillum* Su 851	*Bifidobacterium longum*	98	Anaerobe
YP5B-G	56	2312570	60.15	*Bifidobacterium longum* subsp. *suillum* Su 851	*Bifidobacterium longum*	98	Anaerobe
YP5D-E	45	2268753	60.07	*Bifidobacterium longum* subsp. *suillum*	*Bifidobacterium longum*	98.6	Anaerobe
BP6A-C	22	2163668	56.63	?	*Bifidobacterium pseudocatenulatum*	97.57	Anaerobe
BP6B-A	26	2164941	56.64	*Bifidobacterium pseudocatenulatum* B1279	*Bifidobacterium pseudocatenulatum*	97.55	Anaerobe
BP6B-C	23	2161425	56.63	*Bifidobacterium pseudocatenulatum* B1279	*Bifidobacterium pseudocatenulatum*	97.57	Anaerobe
YP6B-D	22	2161014	56.63	*Bifidobacterium pseudocatenulatum* B1279	*Bifidobacterium pseudocatenulatum*	97.54	Anaerobe
Clostridiaceae							

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
BP2A-9	163	7280801	48.36	*Hungatella hathewayi* 1313	?	98	Anaerobe
EP3A-17E	145	3547379	28.79	*Clostridium saudiense* JCC	?	79	Anaerobe
EP3A-19C	195	3767858	27.92	*Clostridium saudiense* JCC	?	99	Anaerobe
EP3A-20B	151	3896205	27.77	*Clostridium saudiense* JCC	?	82	Anaerobe
BP1D-B	86	6488519	49.36	*[Clostridium] citroniae* RMA 16102	*[Clostridium] citroniae*	99.32	Anaerobe
BP6A-B	155	6185836	49.52	?	*[Clostridium] citroniae*	99.24	Anaerobe
BP6A-A	63	6104883	49.35	*[Clostridium] bolteae* JCM 12243	*[Clostridium] clostridioforme*	91.36	Anaerobe
YP2D-E	136	4953630	44.28	*[Clostridium] innocuum* B-3	*[Clostridium] innocuum*	97.49	Anaerobe
EP3A-17B	379	3647183	30.07	*Clostridium paraputrificum* JCM 1293	*Clostridium paraputrificum*	95	Anaerobe
EP3A-H	70	3678133	29.98	*Clostridium paraputrificum* JCM 1293	*Clostridium paraputrificum*	95	Anaerobe
EP2A-G	812	3157665	28.57	*Clostridium perfringens* ATCC 13124 16S ribosomal RNA, complete sequence	*Clostridium perfringens*	98	Anaerobe
EP3A-17G	914	3787903	28.02	*Clostridium tertium* JCM 6289	*Clostridium tertium*	99	Anaerobe
YP1D-H	97	7171936	48.91	*Hungathella effluvi* UB-B.2	*Hungatella hathewayi/Hungathella effluvii*	<80/95.5	Anaerobe
EP2A-C	613	3710267	28.12	*Clostridium saudiense* JCC 16S ribosomal RNA, partial sequence	?	87	Anaerobe
EP2B-D	754	3505952	30.1	*Clostridium paraputrificum* JCM 1293 16S ribosomal RNA, partial sequence	?	95	Anaerobe
EP3A-18C	319	3541099	28.87	*Clostridium saudiense* JCC	?	80	Anaerobe

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
EP3A-C	67	3731523	29.83	*Clostridium paraputrificum* JCM 1293	?	95	Anaerobe
EP3B-E	69	3655149	29.83	*Clostridium paraputrificum* JCM 1293	?	95	Anaerobe
Coprobacillaceae							
BP4D-A	111	3871933	31.29	*Coprobacillus cateniformis* JCM 10604	*Coprobacillus cateniformis*	99.17	Anaerobe
BP4D-B	114	3882022	31.33	*Coprobacillus cateniformis* JCM 10604	*Coprobacillus cateniformis*	99.41	Anaerobe
BP4D-G	114	3870593	31.29	*Alistipes onderdonkii* WAL 8169	*Coprobacillus cateniformis*	99.05	Anaerobe
YP4D-A	116	3870887	31.29	*Coprobacillus cateniformis* JCM 10604	*Coprobacillus cateniformis*	99.18	Anaerobe
BP1D-C	55	3443630	31.45	*Erysipelatoclostridium ramosum* JCM 1298	*Erysipelatoclostridium ramosum*	99.57	Anaerobe
BP2A-4	100	3609275	31.26	*Erysipelatoclostridium ramosum* JCM 1298	*Erysipelatoclostridium ramosum*	99.74	Anaerobe
EP3B-C	156	3694196	31.51	*Erysipelatoclostridium ramosum* JCM 1298	*Erysipelatoclostridium ramosum*	99	Anaerobe
YP1D-D	167	3458531	31.54	*Erysipelatoclostridium ramosum* JCM 1298	*Erysipelatoclostridium ramosum*	99.98	Anaerobe
YP1D-G	36	3372663	31.34	*Erysipelatoclostridium ramosum* JCM 1298	*Erysipelatoclostridium ramosum*	99.98	Anaerobe
YP1D-I	46	3444198	31.45	*Erysipelatoclostridium ramosum* JCM 1298	*Erysipelatoclostridium ramosum*	99.58	Anaerobe
Coriobacteriaceae							
BP3D-G	121	2203907	59.75	*Collinsella aerofaciens* JCM 10188	*Collinsella aerofaciens*	94.77	Anaerobe
YP3D-A	73	2176679	59.97	*Collinsella aerofaciens* JCM 10188	*Collinsella aerofaciens*	94.75	Anaerobe

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
YP4D-D	56	2282526	62.22	*Enorma massiliensis* phI	*Enorma massiliensis*	97.36	Obligate anaerobe
Eggerthellaceae							
BP2A-10	150	3310300	64.22	*Eggerthella lenta* DSM 2243	*Eggerthella lenta*	98.27	Anaerobe
BP2A-12	205	3531457	64.35	*Eggerthella lenta* DSM 2243	*Eggerthella lenta*	97.94	Anaerobe
BP2A-14	160	3484583	64.21	*Eggerthella lenta* DSM 2243	*Eggerthella lenta*	98.24	Anaerobe
BP2A-3	167	3455425	64.21	*Eggerthella lenta* DSM 2243	*Eggerthella lenta*	98.24	Anaerobe
YP6A-C	201	3249684	64.23	*Eggerthella lenta* DSM 2243	*Eggerthella lenta*	98.41	Anaerobe
Enterobacteriaceae							
BP4D-F	62	4959962	50.34	*Bacteroides uniformis* JCM 5828	*Escherichia coli*	93.42	Facultative anaerobic
BP5B-D	59	4613277	50.73	*Escherichia fergusonii* ATCC 35469	*Escherichia coli*	96	Facultative anaerobic
BP5B-F	57	4613607	50.72	*Escherichia fergusonii* ATCC 35469	*Escherichia coli*	96	Facultative anaerobic
BP5D-F	79	4914312	50.42	*Escherichia fergusonii* ATCC 35469	*Escherichia coli*	99.12	Facultative anaerobic
BP6B-D	99	5157983	50.57	*Escherichia fergusonii* ATCC 35469	*Escherichia coli*	99.6	Facultative anaerobic
YP5B-E	65	4617454	50.72	*Escherichia fergusonii* ATCC 35469	*Escherichia coli*	96	Facultative anaerobic
YP5B-F	70	4607621	50.73	*Escherichia fergusonii* ATCC 35469	*Escherichia coli*	96	Facultative anaerobic
YP5B-K	152	4911388	50.67	*Escherichia fergusonii* ATCC 35469	*Escherichia coli*	96	Facultative anaerobic
YP5B-L	154	4914573	50.67	*Escherichia fergusonii* ATCC 35469	*Escherichia coli*	96	Facultative anaerobic
YP5D-B	75	4915230	50.42	*Escherichia fergusonii* ATCC 35469	*Escherichia coli*	99.27	Facultative anaerobic

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
Enterococcaceae							
YP1A-1A	24	2891514	37.29	*Enterococcus faecalis* ATCC 19433	*Enterococcus faecalis*	98	Microaerophile
YP5D-A	43	2892487	37.39	*Enterococcus faecalis* ATCC 19433	*Enterococcus faecalis*	98.92	Microaerophile
YP3A-19A				*Enterococcus faecium* NBRC 100486	*Enterococcus faecium*	94	Microaerophile
YP6B-A	69	3744822	41.98	*Enterococcus gallinarum* LMG 13129	*Enterococcus gallinarum*	80	Microaerophile
Erysipelotrichaceae							
BP5D-G	51	1839127	35.1	*Faecalitalea cylindroides* JCM 10261	*Faecalitalea cylindroides*	98.56	Anaerobe
BP2D-C	1319	3237307	35.41	*Holdemanella biformis* DSM 3989	*Holdemanella biformis*	90.27	Anaerobe
BP5B-C*	163	2728805	30.9	*Massiliomicrobiota timonensis* SN16	*Massiliomicrobiota timonensis*	<80	Anaerobe
Lachnospiraceae							
BP6B-B	30	3393378	44.22	?	*Anaerostipes caccae*	98.35	Anaerobe
EP2B-B	689	3090848	44.51	*Anaerostipes caccae* L1-92 16S ribosomal RNA, partial sequence	*Anaerostipes caccae*	98	Anaerobe
BP2D-D	34	3426956	41.89	*Blautia obeum* ATCC 29174	*Blautia obeum*	98.27	Anaerobe
BP2D-F	174	3086112	42.3	*[Clostridium] glycyrrhizinilyticum* ZM35	*Mediterraneibacter massiliensis*	<80	Anaerobe
Lactobacillaceae							
BP5B-E	88	3045877	46.23	*Lactobacillus paracasei* R094	*Lactobacillus paracasei*	98	Microaerophile
YP5B-A	86	3048442	46.23	*Lactobacillus paracasei* NBRC 15889	*Lactobacillus paracasei*	98	Microaerophile
EP2A-B	33	1844109	41.13	*Lactobacillus sakei* NBRC 15893 16S ribosomal RNA, partial sequence	*Lactobacillus sakei*	97	Microaerophile

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
EP2A-I	35	1842186	41.13	*Lactobacillus sakei* NBRC 15893	*Lactobacillus sakei*	97	Microaerophile
Oscillospiraceae							
BP1D-E	83	3915614	42.63	*[Ruminococcus] gnavus* ATCC 29149	*Ruminococus gnavus*	98.35	Anaerobe
EP4A-A	107	3449561	43.01	*[Ruminococcus] gnavus* ATCC 29149	*Ruminococus gnavus*	98.88	Anaerobe
Porphyromonadaceae							
YP6B-B*	25	3313166	39.23	*Gabonia massiliensis* GM3	*Gabonia massiliensis* GM3		Anaerobe
Rikenellaceae							
YP5D-C	47	3617262	57.07	*Alistipes finegoldii* JCM 16770	*Alistipes finegoldii*	98.83	Anaerobe
YP4D-G	72	4033001	57.71	*Coprobacillus cateniformis* JCM 10604	*Alistipes goldsteinii*	98.45	Anaerobe
Streptococcaceae							
BP2D-B	22	1881675	38.6	*Streptococcus anginosus*	*Streptococcus anginosus*	95.71	Microaerophile
BP2A-1	25	2029267	36.64	*Streptococcus mutans* NBRC 13955	*Streptococcus mutans*	99.2	Microaerophile
YP2D-A	15	2057322	36.82	*Streptococcus mutans* NCTC 10449	*Streptococcus mutans*	98.87	Microaerophile
BP2A-7	20	2112467	39.94	*Streptococcus salivarius* ATCC 7073	*Streptococcus salivarius*	96	Microaerophile
BP6A-E	51	2307808	39.44	*Streptococcus salivarius* ATCC 7073	*Streptococcus salivarius*	95.45	Microaerophile
BP6A-F	47	2310605	39.44	*Streptococcus salivarius* ATCC 7073	*Streptococcus salivarius*	95.61	Microaerophile
YP2A-12B	22	2107316	39.96	*Streptococcus vestibularis* ATCC 49124	*Streptococcus salivarius*	96	Microaerophile
YP2A-9	65	2296657	39.41	*Streptococcus salivarius* ATCC 7073	*Streptococcus salivarius*	95	Microaerophile

Sample ID	No. contigs	Genome Size (bp)	GC%	16S rRNA result	BactSpeciesID result	fastANI (%)	Anaerobe status
YP2D-B	66	2292412	39.38	*Streptococcus salivarius* ATCC 7073	*Streptococcus salivarius*	95.66	Microaerop hile
Tannerellaceae							
YP4D-B	114	5112504	45.3	*Parabacteroides distasonis* JCM 5825	*Parabacteroides distasonis*	98.87	Anaerobe
YP4D-E	86	6870689	43.54	*Parabacteroides goldsteinii* WAL 12034	*Parabacteroides goldsteinii*	95.82	Anaerobe
BP4D-E	169	4888325	45.38	*Parabacteroides merdae* JCM 9497	*Parabacteroides merdae*	98.62	Anaerobe
YBP4D-E	110	4850664	45.46	*Parabacteroides merdae* JCM 9497	*Parabacteroides merdae*	98.72	Anaerobe
Turicibacteraceae							
EP1A-1A	66	2953211	34.22	*Turicibacter sanguinis* MOL361	*Turicibacter sanguinis*	99	Anaerobe
EP1A-1C	109	3195805	34.04	*Turicibacter sanguinis* MOL361	*Turicibacter sanguinis*	99	Anaerobe
EP1A-1E	119	3137079	34.09	*Turicibacter sanguinis* MOL361	*Turicibacter sanguinis*	99	Anaerobe
Veillonellaceae							
BP3D-E	61	2257756	38.63	*Veillonella dispar* ATCC 17748	*Veillonella dispar*	93.45	Anaerobe
YP3D-B	36	2237651	38.65	*Veillonella dispar* ATCC 17748	*Veillonella dispar*	93.42	Anaerobe

Table 3:3: ***Summary of BEAM isolates that have been sequenced.*** *The 159 isolates in the table are those that had a minimum ANI of 79% and had an identity according to BactspeciesID. They represent 18 different families. Isolates have been ordered by family, 16S rRNA result refers to the Blastn identity when PCR of the 16S rRNA gene was done and then subjected to Sanger sequencing. The BactspeciesID result was the preliminary result suggested by BactspeciesID based on the 16S rRNA gene in silico from the whole genome sequence. Anaerobe status is reported what is available on BacDive (as part of DSM). Microaerophiles tolerate oxygen between 2-10%. Facultative anaerobes tolerate growth in and in absence of oxygen. Anaerobes do not require oxygen to grow but cannot grow optimally or not at all in the presence of oxygen. Obligate anaerobes will die in atmospheric oxygen (21%). * Denotes novel isolate.*
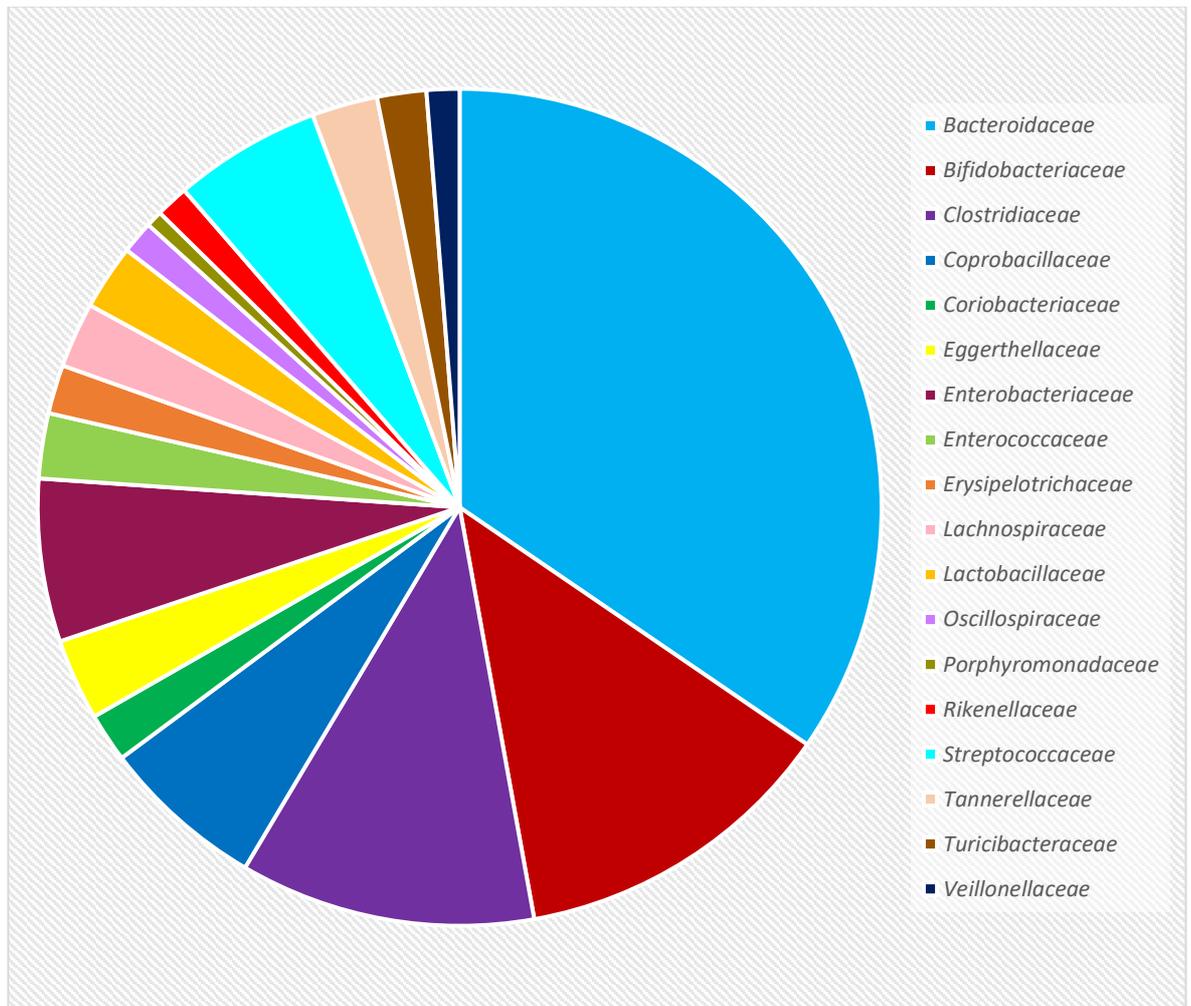
*Figure 3:9: **Relative percentages of BEAM isolates by Family listed in Table 3:3.***

Of the 160 isolates, a total of eighteen families were represented (Figure 3:9). The majority being member of the *Bacteroidaceae* family. This is followed by members of the *Bifidobacteriaceae* family and *Clostridiaceae*. The most abundant family members for 16S rRNA gene amplicon sequencing was: *Ruminococcaceae*, *Lachnospiraceae*, *Bacteroidaceae*, *Erysipelotrichaceae*, *Coriobacteriaceae*, *Bifidobacteriaceae*, *Clostridiaceae*, *Enterobacteriaceae*, *Rikenellaceae* and *Veillonellaceae*. A member of every family apart from *Ruminococcaceae* was cultured, but *Ruminococcaceae* has now been a recognised as a synonym for *Oscillospiraceae* [9].

3.2.8.1 The discovery of novel isolates

As part of the QC process for the faecal culturing, a PCR reaction targeting the 16S rRNA gene was performed and Sanger sequencing done on the PCR product. The sequencing results were blasted through NCBI Blastn to obtain a putative identity. The putative identity was used to download the type strain genome, which was then used as the reference genome

---

[9] According to LSPN (bacterio.net).

for fastANI comparison with the isolate's whole genome sequence. During this QC process, isolate BP5B-C and isolate YP6B-B, had an ANI <80% to any known genome available. BactspeciesID was used to detect any contamination *in silico* of the genomes. Once confirmed that the whole genome sequences did not have contamination, the genomes were further investigated. Submitting the genomes to type strain genome server (TYGS) indicated no known match, and suggested both isolates were novel, where BP5B-C was suggested to be a novel genus, while YP6B-B was suggested to be a novel species. I submitted the genomes to Protologger.de which is an automated Galaxy pipeline to characterise the genome of your candidate to determine if it is novel or not [180]. Once again, it suggested that BP5B-C was a novel genus while YP6B-B was a novel species. I renamed the strains to LH1062 and LH1063 for BP5B-C and YP6B-B respectively, and this nomenclature will be followed hereafter.

Stocks of the isolates were created; LH1062 in YCFA and LH1063 in BHI. LH1062 was also submitted for Oxford Nanopore sequencing and one contig was obtained. Both LH1062 and LH1063 official master and working stocks were confirmed to be free of contamination and then sent to DSMZ Services, Leibniz-Institute DSMZ- Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany for characterisation of their metabolism and growth.

3.2.8.2   Novel isolate 1- *Allocoprobacillus halotolerans LH1062*

Protologger.de suggested that this isolate belonged to the family *Erysipelotrichaceae*, therefore the full-length 16S rRNA gene sequences (1.5 Kb) of 35 species representing 35 genera within *Erysipelotrichaceae* family, were downloaded from the List of Prokaryotic names with Standing in Nomenlature (LSPN: June 2022) [181]. The 16S rRNA gene sequence for *Coprobacillus cateniformis* was also included after Protologger suggested it was the closest relative based on ANI results. The 16S rRNA gene sequences were aligned using Muscle v3.8.31 [182] prior to the reconstruction of a maximum-likelihood phylogenetic tree using IQ-tree v2.0.5 [183] with the TEST model at 1000 bootstrap replications and subsequent visualisation using iTOL v6 [184]. LH1062^T was placed next to *Massiliomicrobiota timonensis* SN16 (Figure 3:10A), nevertheless according to LSPN this has yet to be validated as an official new genus and species. The 16S rRNA gene sequence similarity between LH1062^T and *M. timonensis* SN16 was 96.49% [185]. According to Figure 3:10A, *I. porci* KCTC 15725 seems to be the closest relative to LH1062^T however based on 16S rRNA percentage identity *I. porci* KCTC 15725 is only 89.32% compared to

L. muris DSM 29487 which is 93.04%. I compared the 16S rRNA of LH1062[T] with *C. cateniformis* JCM 10604, after suggestion by Protologger, which had a nucleotide identity of 90.80%. I constructed a phylogenomic tree using PhyloPhlan v3.0.51 after downloading the genomes of the species used in Figure 3:10A. No whole genome sequences could be found for *Breznakia pachnodae* Pei061 and *Absiella argi* N6H1-5. The configuration file specified the use of diamond v0.9.19 and mafft v7.515 as the aligner. Sequences were trimmed using trimal v2.4.rev15 and the tree constructed using iqtree v2.1.4. The tree was constructed with the PhyloPhlAn options –diversity medium and –accurate. Figure 3:10B shows the genomic tree. As suggested by PhyloPhlAn, LH1062[T] is placed amongst the family *Coprobacillaceae* and is closely related to *L. muris* DSM 29487 and *C. cateniformis* JCM 10604. However, based on the 16S rRNA gene sequences it has a higher identity percentage to *L. muris* DSM 29487 than *C. cateniformis* JCM 10604. Further genomic investigation between LH1062[T] and *L. muris* DSMZ 29487 indicated digital DNA-DNA hybridisation (dDDH) was estimated at 21.7% (TYGS), whilst the ANI is 79.3% (fastANI v1.3) [149, 186]. The dDDH comparison between LH1062[T] and *Coprobacillus cateniformis* JCM 10604 is 21% and ANI 78.8%. With the highest ANI being 79.3% and dDDH 21.7% it was significantly below the intra-species threshold of 95% and 70% for ANI and dDDH respectively. I used EzAAI-v1.2.1 [187] to calculate average amino acid identity (AAI) using the genomes shown in Figure 3:10A. The highest percentage was matched to *Coprobacillus cateniformis* JCM 10604 at 73.25%. This was followed by *Longibaculum muris* DSM 29487 at 71.22% and *Intestinibaculum porci* SG0102/KCTC15725 at 62.29%. AAI to *M. timonensis* SN16 was only 50.65%. Using Protologger [180] the percentage of conserved proteins (PoCP) assigned LH1062[T] to *Clostridium* with a value of 50.08%, which is borderline to be suggestive of a novel genus. However, using Blastn and limiting the search to "Bacillus/Clostridium group", the highest 16S rRNA gene sequence from the genus *Clostridium* was 89% which is even lower than to *L. muris* DSM 29487 and to *C. cateniformis* JCM 10604. Taken together and given the inconsistencies within the genus *Clostridium*, a novel genus *Allocoprobacillus* is proposed where *Allocoprobacillus halotolerans* LH1062[T] represents the type strain.

(A)

Tree scale: 0.1

Breznakia pachnodae Pei061 (AJ629069)
Breznakia blatticola ErySL (KU043523)
Longicatena caecimuris PG-426-CC-2 (KR364771)
Amedibacterium intestinale 9CBEGH2 (LC509015)
Absiella argi N6H1-5 (KR185326)
Amedibacillus dolichus JCM 10413 (AB649277)
Merdibacter massiliensis Marseille-P3254 (LT598590)
Faecalitalea cylindroides JCM 10261 (AB558487)
Catenisphaera adipataccumulans GK12 (AB537978)
Absicoccus porci YH-PanP20 (MG931004)
Holdemanella biformis ATCC 27806 (M59230)
Dubosiella newyorkensis NYU-BL-A4 (KU744405)
Ileibacterium valens NYU-BL-A3 (KU744404)
Allobaculum stercoricanis DSM 13633 (AJ417075)
Traorella massiliensis Marseille-P3110 (LT615365)
Dielma fastidiosa JC13 (JF824807)
Tannockella kyphosi BP52G (MG827409)
Faecalibacillus intestinalis SNUG30099 (KY992931)
Faecalibacillus faecis SNUG30370 (KY992930)
Coprobacillus cateniformis JCM 10604 (AB030219)
Longibaculum muris MT10-315-CC-1.2-2 (KR364765)
Intestinibaculum porci SG0102 (MK159121)
Allocoprobacillus halotolerans LH1062 (ON553235.1)
Massiliomicrobiota timonensis SN16 (LN998062)
Erysipelothrix larvae LV19 (KJ670316)
Erysipelothrix inopinata MF-EP02 (AJ550617)
Erysipelothrix piscisicarius 15TAL0474 (CP034234)
Erysipelothrix tonsillarum ATCC 43339 (AB019248)
Erysipelothrix rhusiopathiae ATCC 19414 (AB019247)
Anaerorhabdus furcosa T-301-A2 (GU585668)
Holdemania massiliensis AP2 (JX101683)
Holdemania filiformis ATCC 51649 (Y11466)
Anaerolactibacter massiliensis Marseille-P4302 (LT960585)
Galactobacillus timonensis Marseille-P4641 (LT934541)
Lactimicrobium massiliense Marseille-P4301 (LT996090)
Solobacterium moorei JCM 10645 (AB540987)
Bulleidia extructa DSM 13220 (AF220064)

(B)

(C)

*Figure 3:10: **A mid-point rooted maximum likelihood phylogenetic tree of Allocoprobacillus halotolerans LH1062.** Isolate is relative to 16S rRNA gene sequences of the genera in the family Erysipelotrichaceae (A) and a genome tree constructed using PhyloPhlAn for LH1062ᵀ. In blue are outgroup genomes, which are the species in the genus Coprobacillus, family Coprobacillaceae (B). Comparison of analytical profile indexes of strain LH1062, M. timonensis SN16 and C. cateniformis JCM 10604. Two tests were used, API50CHB and Biolog GenIII Microplate to test for a broader range of substrates was used for A. halotolerans LH1062ᵀ. Blank cells indicated that the substrate was not present in the test (C).*

Phenotypic investigations were carried out by DSMZ Services. This involved: cell and colony morphology, salt and temperature tolerance, fermentation profiles of different carbohydrates, catalase activity, oxidase activity and fatty acid analysis. Despite the strain being isolated in BHI, it was found to grow better in PY+X medium (DSMZ: 104b), which was subsequently used for the phenotypic testing. LH1062ᵀ cells were found to grow in rods and in chains and are Gram-positive. They were negative for catalase and oxidase activity

but positive for haemolytic activity. The bacterium could grow in a relatively broad range of salt conditions (1-20%), with growth delayed between 7-20%. It failed to grow at temperatures below 25°C, grew normally up to a maximum of 40°C, however weak growth was observed at 45°C. Optimum growth was observed between 30˚C to 40˚C. Biochemical characteristics were observed using API50CHB strips. Weak activity for D-arabinose, L-arabinose, D-Xylose, L-Xylose, fructose, mannose, Sucrose, D-turanose and D-lyxose and positive activity for glucose, sorbose, esculin, gentiobiose, D-tagatose and 5-ketogluconate was observed. The isolate was also incubated in the Gen III Biolog Microplate using Medium A. Initial transmission of 93% after incubation anaerobically for 48h at 37°C was determined. The bacterium was positive for gentiobiose, D-fructose, D-fucose, L-fucose, L-rhamnose, D-serine, D-fructose-6-PO4, minocycline, L-galactonic acid lactone, D-glucuronic acid, glucuronamide and sodium butyrate. It is noted that were inconsistencies between API strips and the Gen III Biolog Microplate, it should be mentioned that API50CHB was incubated for 12 days at 37˚C aerobically (covered in paraffin) whereas the Biolog microplate was done anaerobically resulting in contrasting results. Comparing the reaction patterns with the not validly published isolate: *Massiliomicobiota timonensis* SN16, there is a distinct difference in the ability of LH1062[T] to react with these substrates (Figure 3:10C), the profiles were based on the previous publication [185]. Although the closest relative *L. muris* DSM 29487 was negative for the acidification of carbohydrates, LH1062[T] was not. *C. cateniformis* JCM 10604 was found to have acidification of glucose, mannose, galactose, fructose, sucrose, maltose, cellobiose, lactose and trehalose, following a similar profile to LH1062[T]. Cellular fatty acids were detected after converting them into fatty acid methyl esters (FAMEs) following a modified protocol [188]. The FAMEs mixture was separated by gas chromatography and detected by a flame ionization detector using Sherlock Microbial Identification System (MIDI) based on the TSBA6 database. $C_{16:0}$ was the most abundant fatty acid for LH1062 at 19.08%. This was also the major fatty acid for the not validly published isolate: *M. timonensis* SN16, at 41% and for *L. muris* DSM 29487 at 30.1%.

Based on the genomic and phenotypic results presented above LH1062[T] as the type strain of the proposed new genus *Allocoprobacillus* is proposed. Naming it like its closest genomic relative, based on 16S rRNA, *Coprobacillus*. Strain LH1062[T] is suggested to be the type strain named: *Allocoprobacillus halotolerans* LH1062[T].

3.2.8.3   Novel isolate 2- Coprobacter tertrius LH1063

Similarly, to *Allocoprobacillus halotolerans* LH1062, Protologger.de suggested that LH1063 belonged to the genus *Coprobacter*. Therefore, for *Coprobacter tertius* LH1063[T], the 16S rRNA gene sequences representative of 6 *Coprobacter* species and 2 *Coprobacter secundus* subspecies type strains were downloaded from LSPN (LSPN: June 2022) [181]. The maximum-likelihood phylogenetic tree, both 16S rRNA gene and phylogenomic tree, was generated as aforementioned for *A. halotolerans* LH1062[T]. *C. tertius* was placed next to *Coprobacter fastidiosus* and *Coprobacter secundus* (Figure 3:11A). From the phylogenomic tree (Figure 3:11B), *C.* tertius LH1063[T] is more closely related to *C. fastidiosus* NSB1 than *C. secundus* species. Based on 16S rRNA gene sequences comparison (Protologger), the closest relative was *Coprobacter secundus* with a nucleotide identity of 91.5%. *Coprobacter secundus* was also the closest match based on ANI at 77.81% (Table 3:4). It was noted that the ANI for *C. secundus* 177 reported by Protologger and OrthoANI [189] is different, being 77.81% and 72.8% respectively. This discrepancy is explained due to Protologger using fastANI while OrthoANI uses USearch. OrthoANI was used as opposed to fastANI as fastANI would not provide an output if the ANI < 80%, which it was for each species in the genus Coprobacter. The dDDH comparison between LH1063[T] and *Coprobacter fastidiosus* DSMZ 26242, *Coprobacter secundus* 177 and *Coprobacter secundus* subsp. *similis* 2CBH44 were 20.1%, 19.4% and 19.3% respectively. LH1063[T] had a genome size of 3.3Mbp and a G+C content of 39.23 mol% whilst the genome size and G+C content for *C. secundus* 177 is 4.1Mbp 37.8 mol%.

Species	dDDH$_4$	ANI
Coprobacter fastidiosus DSM 26242	20.1	71.9
Coprobacter secundus 177	19.4	72.8
Coprobacter secundus subsp. *similis* 2CBH44	19.3	72.0
Caldicoprobacter oshimai DSM 21659	18.7	63.3
Caldicoprobacter guelmensis DSM 24605	18.6	63.3
Caldicoprobacter faecalis DSM 20678	18.5	61.8
Caldicoprobacter algeriensis DSM 22661	18.2	60.6

*Table 3:4: **dDDH and ANI percentages of the species in Coprobacter compared to Coprobacter tertius LH1063^T**. dDDH was determined using TYGS [186]. ANI was determined using EzBio Cloud ANI calculator [189]*

(A)

(B)

(C)

(D)

(E)
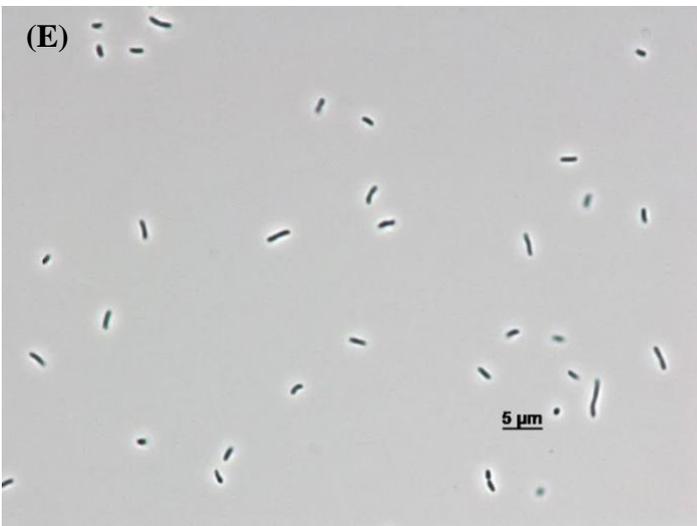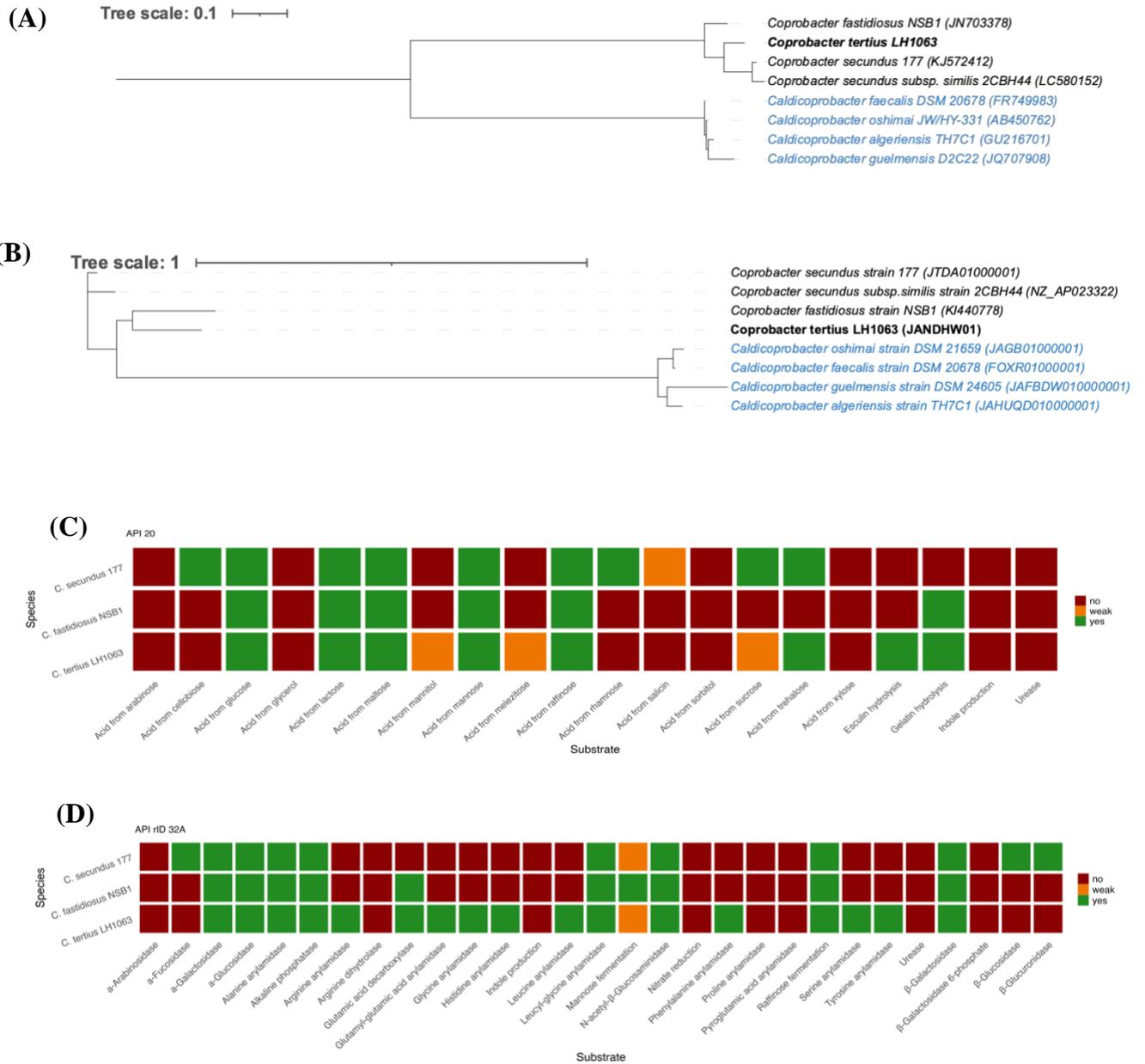
Figure 3:11: ***Mid-point rooted maximum likelihood phylogenetic tree of Coprobacter tertius LH1063^T.*** *Isolate was relative to the 16S rRNA gene sequences of the species in the genus Coprobacter. Outgroup were members of the genus Caldicoprobacter, belonging to the family Caldicoprobacteraeceae (A). Genome tree constructed using PhyloPhlAn for LH1063^T. In blue are outgroup genomes, which were the species belonging to the genus Caldicoprobacter, family Caldicoprobacteraceae (B). Comparison of analytical profile indexes of strain LH1063 and the other members of the Coprobacter genus. Three tests were used: API 20A (C), API rID 32A (D) and Biolog GenIII Microplate to test for a broader range of substrates was used for strain LH1063. The profiles of C. fastidiosus and C. secundus was based on previous published literature. Phase contract microscopy image of LH1063^T(E).*

Phenotypic investigations were also carried out by DSMZ Services. This involved: cell morphology, salt, bile and temperature tolerance, fermentation profiles of different carbohydrates and fatty acid analysis. *C. tertius* LH1063[T] cells grow as rods in pairs measuring roughly 3µm in length (Figure 3:11E). It is Gram-negative, and negative for catalase, oxidase and hemolytic activity. The bacterium tolerated up to 3% salinity and was shown to grow well between 30-40˚C, with weak growth at 25˚C. Unlike *C. fastidiosus* and *C. secundus*, LH1063[T] failed to grow in any concentration of ox gall [190, 191]. Biochemical characteristics were observed using API20A strips, the inoculation was grown anaerobically at 37˚C for 24h before the test was performed. The strain produced acid from: glucose, lactose, maltose, mannose, raffinose and trehalose. The strain hydrolysed gelatine and esculin, and was weakly positive for acid production from mannitol, sucrose and melezitose (Figure 3:11C). From APIrID32A, the strain was positive for: α-galactosidase, β-galactosidase, α-glucosidase, N-acetyl- β-glucosaminidase, raffinose and glutamic acid decarboxylase fermentation, alkaline phosphatase, arginine phosphatase, leucyl-glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, tyrosine arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl-glutamic acid arylamidase and serine arylamidase. The bacterium also had a weak reaction for mannose fermentation (Figure 3:11D). Profiles listed in Figure 3:11C and D was based on previous published literature: [190, 191]. The strain LH1063[T] was additionally run on the Gen III Biolog Microplate with the same conditions described previously for strain LH1062. It was positive after 48h incubation for: gentiobiose, D-melibiose, α-D-glucose, D-mannose, D-fructose, D-galactose, 3-methyl glucose, D-fucose, L-fucose, L-rhamnose, D-glucose 6-phosphate, D-fructose 6-phosphate, minocycline, D-galacturonic acid and D-glucuronic acid. The acidification of the various carbohydrates including α-glucose, D-mannose and D-fructose to name a few was also confirmed using the Gen III Biolog Microplate. The biochemical profile of strain LH1063 is quite similar to *C. fastidiosus* NSB1 and slightly different from *C. secundus* 177 (Figure 3:11C and D). Cellular fatty acids were detected using the same method for LH1062[T]. The major fatty acid produced by LH1603[T] was anteisoC$_{15:0}$ at just 25%, followed closely by isoC$_{15:0}$ at 20%. This was similar to *C. fastidiosus* NSB1 and *C. secundus* 177. For *C. fastidiosus* NSB1 it was 23-27% and 26-27% and for *C. secundus* 177 it was 0.24-0.34% and 0.59-0.70% for anteisoC$_{15:0}$ and isoC$_{15:0}$ respectively.

Based on the genomic and phenotypic results presented above, I propose LH1063[T] as a novel species within the genus *Coprobacter*. The epithet tertius, as this is the third species of this genus is proposed.

3.3   Discussion

In this chapter I sought to profile the gut microbiota of first-time diagnosed BrCa patients in Norfolk and associate how a perturbed gut microbiota, due to prophylactic antibiotics, may associate with clinical outcome. Unfortunately, due to the COVID-19 pandemic, recruitment numbers could not be reached. A total of 34 BrCa patients and one control patient were recruited at two sites: NNUH and JPUH. I used several microbiota profiling approaches including 16S rRNA gene amplicon sequencing, shotgun metagenomic and culturing.

Initial analysis of the gut microbiota using 16S rRNA amplicon and shotgun metagenomic sequencing platforms showed no differences in alpha or beta diversity by time point or BrCa status. Unfortunately, due to the issues in recruitment this study is underpowered and therefore led to the statistically non-significant conclusions reported. Trends in alpha diversity suggests that Shannon diversity decreases by time point, and there was some clustering observed according to gut-associated *Bactoides*. Additionally, any significant changes that were observed was possibly influenced due to site, as I had determined that the gut microbiotas of patients from NNUH were different than JPUH patients. 60% of all ILC samples and 100% of one-year samples were from NNUH patients. It was surprising to see a difference in site, as JPUH and NNUH are both in the county of Norfolk and it was expected that patient diets would be similar. The city of Norwich and Great Yarmouth are located less than an hour away from each other. However, the cities differ in terms of socioeconomical factors, which could translate to limited access to nutrition and thus shape their gut microbiotas which was reflected in this data. Unfortunately, I did not have dietary information to investigate if the diets of NNUH patients did differ from JPUH patients' diets and therefore account for this potential difference in microbiota by site. Another potential source explaining the difference between NNUH and JPUH gut microbiotas is the point of consent. All NNUH patients were consented prior to the start of the pandemic, while all JPUH patients were consented two years after the pandemic. Lifestyle habits changed during the pandemic [192], and there is some data indicating that prior SARS-Cov-2 infection is also associated with perturbations in the gut microbiota [193, 194].

The BEAM study was heavily impacted by the global pandemic. In total there were 34 BrCa patients recruited to the BEAM study. Due to the low sample size only, limited conclusions can be drawn. In addition, at the time of writing this chapter I had limited metadata available. This restricted any further analysis that could have been performed. Due to the nature of shotgun metagenomics, there is the potential to probe into functional potential of the microbiota. Moreover, when comparing methods, one study compared 16S rRNA gene amplicon sequencing to shotgun metagenomics in chicken metagenomes and found that shotgun metagenomics could report rarer genera over 16S rRNA [195]. Should these low abundant genera be significant in the outcome, an important observation may have been missed. Metagenomics have the advantage with sequencing all genomic content but only if the sequences can be assigned to all members and their respective functions present in the community. Unfortunately, I experienced issues with software dependencies and could not achieve this in the time frame of the thesis. This is something worth pursuing as functionally the microbiota can have an impact on health, as indicated by several studies [89, 196, 197]. I considered pooling studies, with a focus on 16S rRNA gene amplicon sequencing as more such studies are available than shotgun metagenomics. A literature search for gut microbiota studies of BrCa or healthy patients e.g., Twins UK cohort or Dutch microbiome project, showed that the vast majority are 16S rRNA gene amplicon sequencing targeting the V4-V6 region, while BEAM targeted V1-V2 region, thereby I was unable to pool data. I utilised the V1-V2 hypervariable region for this study as published studies, discussed later in Final discussion and future perspectives, indicated that in terms of reporting relative abundances there is no difference in using V1-V2 or V4-V6. As the protocol in our lab was established using V1-V2 I decided to continue using this region as we had established protocols and pipelines available to analyse this.

16S rRNA gene amplicon sequencing can be preferred for low biomass samples and for exploratory studies, as it is cost effective. However, it is recognised that 16S rRNA gene is not accurate in reflecting taxonomic distributions due to PCR amplification bias. One study showed that central areas of the 16S rRNA gene tend to be more conserved when it came to accurately reflecting the taxonomy distribution [198]. However, when reporting abundances in downstream analysis as relative abundances, the choice of hypervariable region became irrelevant [177]. These aspects should be carefully considered prior to pooling and also solely relying onto 16S rRNA gene taxonomic profiles from clinical samples. Should I have been able to pool, there would still be geographical differences, as demonstrated with Norwich and Great Yarmouth. Geographical differences could be adjusted with dietary or

lifestyle information, though this is not standard data collection in large scale microbiome studies. Lastly, I had hoped to include a representative cohort of control patients which unfortunately did not happen. No significant trends were observed in the BrCa cohort, but this could be different from control patients. As an example, I observed a relative high abundance of *B. pseudocatenulatum* in the BEAM control patient which was less in BrCa patients. This could be consistent at a population level but as I did not have enough cohort patients no conclusions can be drawn.

Culturing efforts of faecal samples is labour intensive and time consuming but could provide further insights and strains that can be used for mechanistic investigations (as highlighted in Chapter 5). Culturing also provides high quality genomes of individual bacteria which could allow for more in-depth *in silico* analysis including probing functionality. As I also had shotgun metagenomics data, MAGs could have been extracted and compared with the genomes that were sequenced from culturing to look more broadly at genomic diversity and probe possible features that guide microbe-host mechanisms. Of the isolates cultured I mostly recovered non-obligate anaerobes and microaerophiles, and a smaller number of obligate anaerobe species (which represented 6/10 genera as defined by sequencing). *Akkermansia* and *Faecalibacterium* were not recovered most likely due to their oxygen sensitivity [199, 200]. Due to the nature of how the faecal kits were transported, true anaerobic conditions could not be guaranteed which could explain why I did not recover any strict anaerobes. These culturing studies also led to the discovery of two isolates which could not be matched to any validly named bacterium in the LSPN database. Further characterisation resulted in validly naming a new genus, *Allocoprobacillus* with the representative type strain *Allocorprobacillus halotolerans* and a new species *Coprobacter tertius*.

Bacteroides were also the most abundant genera in both shotgun and 16S rRNA gene amplicon sequencing. As mentioned in the Introduction, *B. fragilis* has been indicated in several studies to be immunogenic, especially in anti-cancer therapies. I had attempted to investigate the anti-cancer potential of *B. fragilis* described later in '*In vitro* validation of *B. fragilis* NCTC 9343 with MCF-7 (HR+/HER2-) BrCa cells', using the strain I isolated from a BEAM patient but unfortunately this was found to be contaminated. Another potential anti-cancer bacteria mentioned in the Introduction was *Bifidobacteria*, which I showed was present in the BEAM study patients. Notably, *B. pseudocatenulatum* was present at higher abundances in the healthy control patient but not in the BrCa-diagnosed patients. This could

be chance as there is only one healthy control but would be interesting to investigate further if given the opportunity. *B. longum* however was present in all patients and was collective the highest abundant *Bifidobacterium* species in these patients. This species was also cultured readily from faecal samples and recognised as a probiotic strain. As such it is worth investigating if it has any influence on the anti-cancer responses *in vivo*, which is discussed further in the *Bifidobacterium longum* project.

Faecalibacterium was also present as the most abundant genera by sequencing data, but this genus was not recovered with culturing. *Faecalibacterium* belongs to the family *Oscillospiraceae* and members of *Oscillospiraceae* are obligate anaerobes [201]. Despite members of the *Oscillospiraceae* family being recovered, this was only *Ruminococcus gnavus* and not *Faecalibacterium*. *Clostridiaceae* was the second most recovered family but were not in the top genera regarding sequencing data. Members of *Clostridiaceae* that were isolated tended to be C*lostridium*, which are known spore-formers. This adaptation allowed the bacterium to tolerate oxygen contributing to its recovery in nutrient-rich media. Most of the isolated cultures are bacteria that can tolerate oxygen, but will not die from exposure i.e., microaerophiles or anaerobes. I only cultured five known obligate anaerobes from the fifty-three species available.

Although it is expected that purely sequencing based approaches provides accurate representation of taxonomic profiles, this is reliant on sequencing depth [202]. This is important as although culturing might not enable all taxa to be recovered, recent work indicates that culturing was able to 'uncover' 10% of *Bifidobacterium* species that were missed by metagenomics, therefore showing that culturing may provide greater sensitivity for certain low abundance members. Additionally, next generation sequencing approaches will be constrained by the databases used for taxonomy assignment, and *de novo* assembly of novel metagenomic isolates may have artefacts of sequencing present which is less likely to happen with a pure culture [179].

In a clinical setting, cultureomics is used for determining antibiotic resistance profiles, and this is key as it may uncover unknown AMR determinants that would be missed by sequencing (and mapping to databases) alone [203]. However, this approach takes time and efforts and there are efforts to develop next-generation sequencing to provide a rapid diagnosis. As one can appreciate there is much discussion to the advantages and limitations to culturing approaches.

3.4   Conclusion

A total of five BrCa patient and one control patient were recruited from the Norfolk and Norwich University Hospital with 29 BrCa patients from James Paget University Hospital. Downstream microbiota profiling (using both 16S rRNA and shotgun metagenomics) did not indicate any significant differences in alpha or beta diversity by time point or BrCa status. No structural microbial changes by species, genera or family were detected by time point either. However, hospital site was a significant differentiating factor, possibly due to geographical dietary or temporal differences. Whilst no statistically significant findings were concluded, this study was limited by size. We did report trends in decreasing alpha diversity across time points in this group, which could have clinical implications. As such, the findings drawn from this study should be interpreted with caution. Cultureomics did not recover any obligate anaerobes that were in high abundance i.e., *Faecalibacterium* or *Akkermansia*, possibly due to aerobic conditions during transport. However, the isolation studies did provide a large strain collection, which were used in *in vitro* and *in vivo* studies as described in later chapters and led to the discovery of a novel genus and species that are found in the human gut microbiota.

3.5   Future works and direction

The BEAM study is one of the first of its kind in the UK. Although initial recruitment was not as successful as anticipated there is capacity for it to continue at a national scale due to its ability to ship samples to the researcher. As the pathway is up and running, and has shown to be successful, the observational study could be continued. Although limited conclusions can be drawn with these samples alone, it would be possible to compare samples of age-matched control patients to determine if the BEAM cohort already had a perturbed microbiota signature at baseline. Ideally this should be individuals within the UK as diet is a known influencing factor, so limiting it to the same country would hopefully minimise dietary effect. Furthermore, due to software issues I could not probe into potential functionality nor extract metagenome assembled genomes from the BEAM metagenomes. Despite limited evidence for changes in microbiota structure being found in the cohort, functionally there may have been a signature which could be associated with an outcome. Extracting metagenome assembled genomes would allow further investigations with the genomic data from the cultured isolates I had extracted.

4 The CALADRIO Study

This chapter focuses on a collaboration we established with a team in Madrid, Spain: MedSir. It covers a spin-off study (CALADRIO) based on the phase 2a clinical trial: KELLY (NCT03222856). In this chapter I will start by introducing the need for studies like CALADRIO for design and development of anti-cancer therapies in relation to host microbiomes. This will be followed by the results of 16s rRNA gene amplicon sequencing and shotgun metagenomic sequencing of the oral and gut microbiota respectively. I will then progress to *in vitro* validation of the profiling results, before concluding how the results have contributed to our understanding of novel BrCa therapies and potential bi-directional impact of oral and gut microbial communities.

The team in Spain collected the clinical samples for this project. The Spanish team extracted the genomic DNA of the buccal samples and submitted the DNA for 16S rRNA gene amplicon sequencing. The faecal samples were shipped to QIB where I extracted the genomic DNA and submitted to Source Bioscience Cambridge for shotgun metagenomic sequencing. Dr. Raymond Kiu ran the metagenomic pipeline to provide me with taxonomic abundances, with Dr. Matthew Dalby helping by providing initial R scripts to visualise and analyse the sequencing reads. I used these scripts as a starting point to do the microbiota analyses and built upon them to make the plots and produce the results reported in this chapter. Dr. David Seki ran the METABOLIC pipeline on the *B. fragilis* MAGs. Galaxy modules provided by the Huttenhower lab were used for LEfSe analysis. MCF-7 BrCa cells used for the *in vitro* part of this chapter were kindly provided by the Robinson group (QIB).

4.1 Background

4.1.1 Anti-cancer therapies and the gut microbiota

HR+/HER2- BrCa is the most prevalent biological subtype of BrCa. Endocrine therapy resistant HR+/HER2- BrCa, has a median overall survival of 24.8 months in the advanced setting. Despite being sensitive to endocrine therapies, HR+/HER2- BrCa is only moderately sensitive to cytotoxic agents e.g., eribulin [204, 205]. New therapeutic strategies are needed for endocrine therapy resistant, HR+/HER2- metastatic breast cancer (mBC) with high efficacy to improve patient outcomes. Pembrolizumab is an immunotherapy drug that targets the programmed cell death protein 1 (PD-1) expressed on T-cells. PD-1 when bound to its ligand PDL1 acts as an immune checkpoint and downregulates T-cell responses. Tumours express PDL1 to downregulate T-cells to evade detection. Pembrolizumab blocks PD-1 on

active lymphocytes to prevent tumour cells from downregulating responses. Eribulin is a mitotic inhibitor approved for refractory HER2- mBC patients, Figure 4:1. The KELLY study, a phase 2a clinical trial (NCT02778685) assessed the safety and efficacy of the novel combination therapy of pembrolizumab and eribulin in the treatment of HR+/HER2- mBC. The study showed that the clinical benefit was 56.8% (95% CI: 41.0-71.7) and median progression free survival (PFS) was 6 months (95% CI:3.7-8.4 months), exceeding the activity observed with eribulin when used as monotherapy in a similar patient population [206].
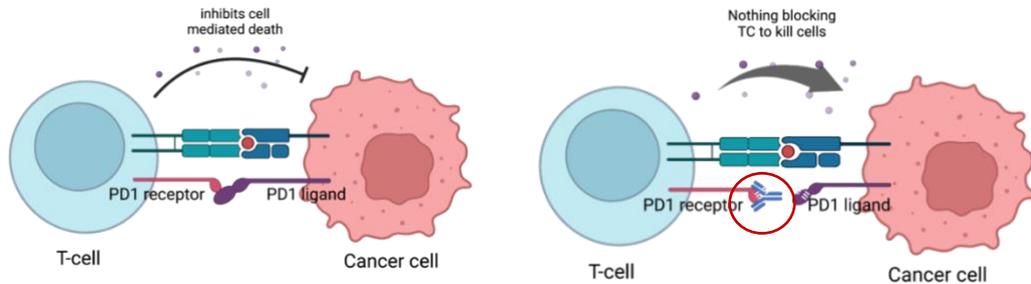
*Figure 4:1: **Mechanism of PD-1 immunotherapy.** PD-1 receptor when bound to its ligand, PDL1, inhibits cell mediated death by the T-cell. This prevents the host immune response attacking its own cells. Cancer cells have evolved to present PDL1 on its cell surface, thereby evading cell mediated death by the T-cell. When the PD-1 receptor is blocked, e.g., by pembrolizumab (red circle) on the T-cell, even when PDL1 is expressed on a cell surface will not inhibit cell mediated death. Image made using BioRender.*

In the last decade many studies have investigated the association of the gut microbiota and clinical outcomes regarding anti-cancer therapies [116, 117, 120, 207-210]. Most of these studies have been highlighted in the Introduction. Briefly though, Gopalakrishnan *et al.*, demonstrated that a higher gut diversity was present in individuals who responded to anti-PD-1 immunotherapy in metastatic melanoma [209]. Others have shown *Bifidobacterium* species and strains can promote anti-tumour activity in anti-PDL1 responses [117, 120, 132] or, that members of the *Bacteroides* genus improve CTLA-4 responses in melanoma and CRC mouse models [119]. Conceptually, faecal microbiota transplants to GF mice could alter the anti-tumour responses of recipients to responders [117, 207, 211]. Recently, a phase 1 trial in humans for refractory melanoma and FMT was conducted to influence clinical outcomes [132, 133]. In these trials, melanoma patients who were not responsive to anti-PD-1 therapy received a FMT from patients that experienced an anti-PD-1 response. Patients that did experience an anti-cancer response showed signs of increased tumour infiltrating lymphocytes, and gut microbiota analysis showed an increased relative abundance of species previously reported to be associated with a 'response' to immunotherapy e.g., *Bifidobacterium bifidum* and *Ruminococcus*. These studies demonstrate the potential of the gut microbiota predicting or influencing the outcome of anti-cancer immunotherapies.

Mechanistically it is hypothesised that the gut microbiota can influence anti-cancer responses by modulating immune responses including enhanced T-cell responses, which may be via production of certain microbial metabolites [117, 119, 121, 209, 211]. Notably, many of these studies have focused on strongly immunogenic cancers such as melanoma, with few exploring microbiota profiles in poorly immunogenic cancers such as BrCa [212]. Thus, there is a pressing need to define how novel therapies impact patient microbiota profiles and if there are any microbial biomarkers that correlate with anti-cancer responses in patients with BrCa. Though many studies looking at the microbiota and anti-cancer therapies tend to focus on the gut, we have also chosen to include the oral microbiota. There is some evidence to suggest a link between these two sites resulting in a perturbed microbiota leading to adverse health outcomes [213]. Here, I profile the oral and gut microbiota of mBC patients receiving pembrolizumab and eribulin. I have performed an exploratory retrospective analysis to explore associations between the gut or oral microbiota and clinical outcomes in the CALADRIO study.

4.1.2 Aims and hypothesis

To assess the oral and gut microbiota of CALADRIO patients in relation to their clinical outcome. The specific aims are:

- Assess the alpha and beta diversity of the gut and oral microbiota according to clinical parameters.

- Assess taxa level and taxa specific changes according to clinical parameters.

- Investigate species specific changes according to clinical parameters indicated by multivariate analysis.

- Validate any changes in the microbiota of patients experiencing a clinical benefit and those who did, if any, not by the use of *in vitro* methods.

4.2 Results

A total of 28 patients were recruited to the CALADRIO study, 17 of whom experienced a clinical benefit (CB, 60.7%) according to the RECISTv1.1 code [214]. The median age of the patients was 53.5-year-old and 48.1% of patients were PDL1 positive, see [1]. Histology of the tumour would be stained for PDL1 which can be expressed on cancer cells to suppress cell death by T-cells, PDL1 positive patients had tumours with PDL1 expressed on their cell surfaces. We collected 65 faecal samples and 70 buccal samples, where faecal samples collected were: baseline 28/28, after three treatment cycles (C4D1) 22/22 and end of treatment (EoT) 15/20 and all samples at all time points were collected for buccal samples. Four faecal samples failed to have sufficient sequencing depth and were removed from the dataset, and three samples did not have clinical data recorded and were also removed. Thus, the final number of faecal samples was 58, with all 70 buccal samples used for analysis. For shotgun sequencing, the average read depth was 4.14Gbp, the median number of reads of this dataset was 3,0197,291 reads and the average number of reads per sample was 27,592,595. For 16S rRNA gene amplicon sequencing, the average read depth was 37.2Mbp, the median number of reads in this dataset was 122,973 reads and the average number of reads per sample was 12,378. A positive control, which included a specific amount of genomic DNA from specific bacteria, was used to confirm the pipeline used to assign taxonomic abundances to the samples. As shown in Table 4:1, the percentages fall within ±5% of the theoretical composition. Two *Bacillus* species that were not part of the positive control were detected. However, *B. spizizenii* is closely related to *B. subtilis,* therefore it could represent a misclassification.

Name	Actual percentage of reads (%)	Theoretical Composition (%)
Staphylococcus aureus	16.442	12
Enterococcus faecalis	13.897	12
Listeria monocytogenes	13.484	12
Salmonella enterica	12.815	12
Escherichia coli	11.136	12
Limosilactobacillus fermentum	9.765	12
Pseudomonas aeruginosa	9.047	12
Bacillus spizizenii	7.184	0
Bacillus cereus	2.671	0
Bacillus subtilis	2.385	12
Others	1.175	4

Table 4:1: ***Theoretical vs. actual percentage of reads in the positive control submitted for shotgun metagenomic sequencing.*** *Others includes 2% each of Saccharomyces cerevisiae and Cryptococcus neoformans.*

4.2.1 Pembrolizumab and eribulin did not significantly alter the oral microbiota

4.2.1.1 *No significant differences were observed in alpha or beta diversity of the oral microbiota*

The primary objective of this study was to assess if the microbiota profile between patients who experienced a CB was different from patients who did not experience a CB. Initially, microbial alpha diversity was explored using Shannon's diversity index and inverse Simpson's diversity index. Based on diversity indexes alone little evidence for changes were observed as patients underwent treatment. Visualised as an alluvial plot, the oral genera stay relatively stable from baseline to EoT (Figure 4:2A). Looking at the alpha diversity by CB and no-CB, Shannon diversity index was similar (p= 0.97, Figure 4:2B). This was the same for other clinical metadata parameters (PDL1 status, neutrophil:lymphocyte ratio (NLR) and PFS (time to progression free survival; >6 months and <6months). Beta diversity was assessed by NMDS and permutational multivariate analysis of variance using distance matrices (adonis2 function from vegan package-v2.6-4, engine set to "monoMDS"). Permanova (from vegan package-v2.6-4, function "adonis2", method set to "bray" and permutations to "999") did not indicate that any of the clinical parameters could explain the variation observed in the oral microbiota as the remaining residual was 92%. NMDS scores were plotted and could not observe segregation by CB status or by PDL1 status as suggested by the Permanova calculations (p = 0.2, 0.45, 0.08, 0.52 for CB status, NLR, PDL1 and PFS respectively), as seen in Figure 4:2C.
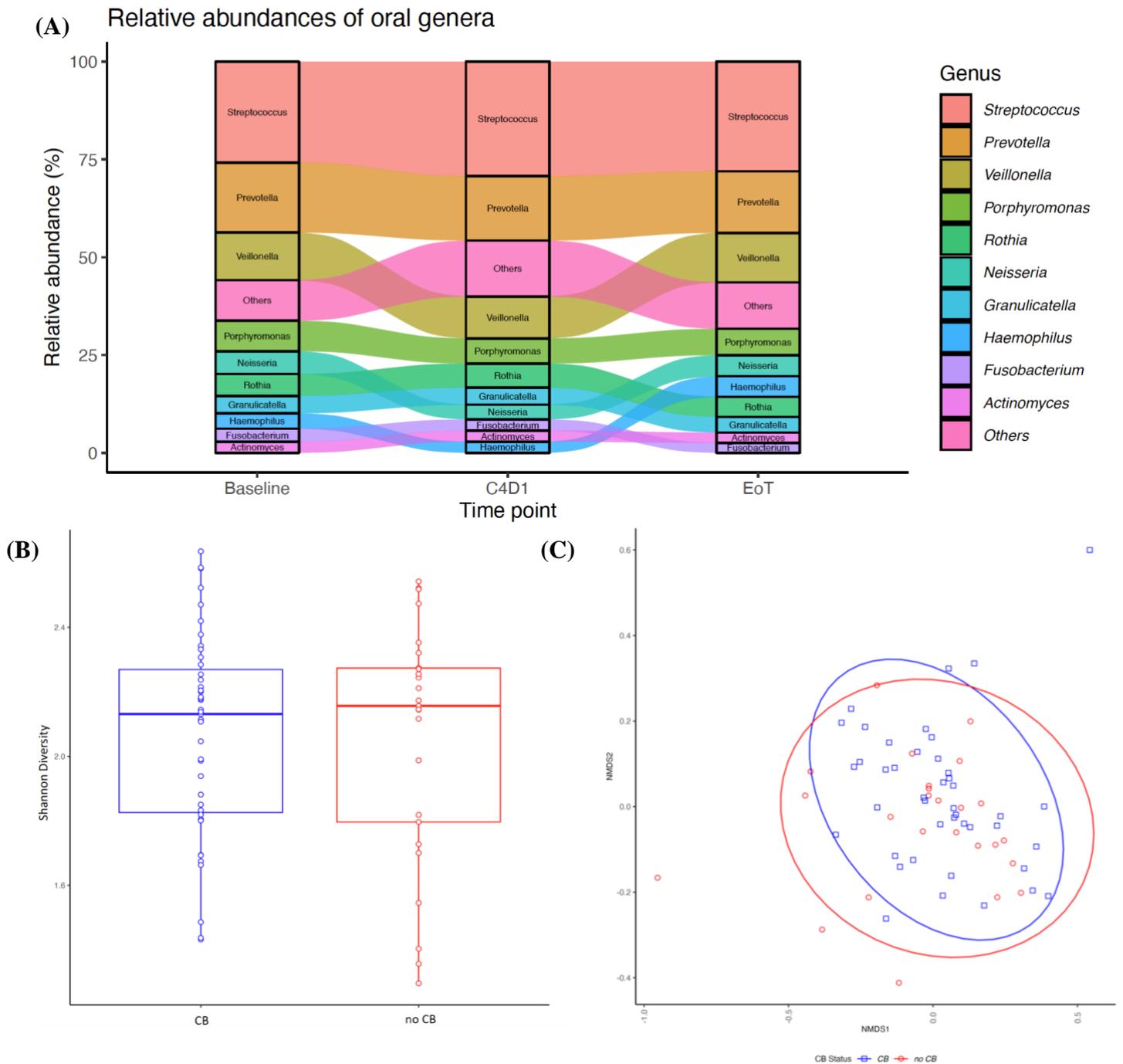
*Figure 4:2: **Exploratory analysis of the oral microbiota in CALADRIO patients.** Top ten genera of the oral microbiota presented as an alluvial plot, showed no significant changes at the three time points: baseline, C4D1 and EoT. (A). Shannon diversity index showed little evidence of changes between the oral genera of patients who experienced a CB (blue) vs. those who did not, no-CB (red). The same was observed for inverse Simpson diversity index, data not shown (B). Assessing the beta diversity of the oral microbiomes by NMDS showed no segregation between CB (blue) and no-CB (red) groups. The same was done for PDL1 presence and absence and observed no segregation between the groups, data not shown (C). This was also confirmed by Permanova.*

The top ten most abundant genera of the oral microbiota were: *Streptococcus*, *Prevotella*, *Veillonella*, *Poryphyromonas*, *Rothia*, *Neisseria*, *Granulicatella*, *Haemophilus*, *Fusobacterium* and *Actinomyces.* I noted a high relative abundance of *Streptococcus* at baseline CB samples. Visualising the results as a heatmap I noted a clustering of CB samples with relative abundance of 40% or more (Figure 4:3A). These suggest that oral *Streptococcus* at baseline could be a biomarker in clinical settings to predict responses to this therapy. However, when investigating this further, and performing a Mann-Whitney U, it was not significant at any time point (p = 0.09, 0.67 and 0.9 for Baseline, C4D1 and EoT respectively), as shown in Figure 4:3B. Control samples would be necessary to determine if *Streptococcus* could indeed be a biomarker. However, as CB patients tended to have higher *Streptococcus* abundances compared to no-CB it could be that *Streptococcus* was beneficial or alternatively that a lower *Streptococcus* abundance could be a biomarker for no-CB in this context.

4.2.1.2   *Capnocytophage, an oral opportunistic pathogen, could be associated with no-CB*

I next sought to see if any oral microbiota members could explain CB status. Using the Huttenhower lab galaxy pipeline, I ran the oral microbiota data through LEfSe. Six genera were reported to be discriminative features associated with either CB or no-CB (Figure 4:3C). These were *Capnocytophaga*, *Filifactor*, *Clostridium*, *Pyramidobacter* and *Sphaerochaeta* for no-CB and only *Atopobium* for CB. *Capnocytophaga*, *Filifactor* and *Clostridium* were also present in the gut, which will be further discussed in section 4.3 in the context of potential translocation. *Atopobium* is a genus usually reported in bacterial vaginosis [215], however there has been a report that *Atopobium* was in higher abundances in saliva of gingival squamous cell carcinoma compared to periodontitis patients. In the same study, the authors report that *Capnocytophaga* was highly abundant in gingival plaque compared to cancerous tissue [216]. Though in the study they did not elaborate on possible mechanisms. I observed higher abundances of *Capnocytophaga* in no-CB patients (Figure 4:3D) and higher abundances of *Atopobium* in CB patients (Figure 4:3E).

(A)
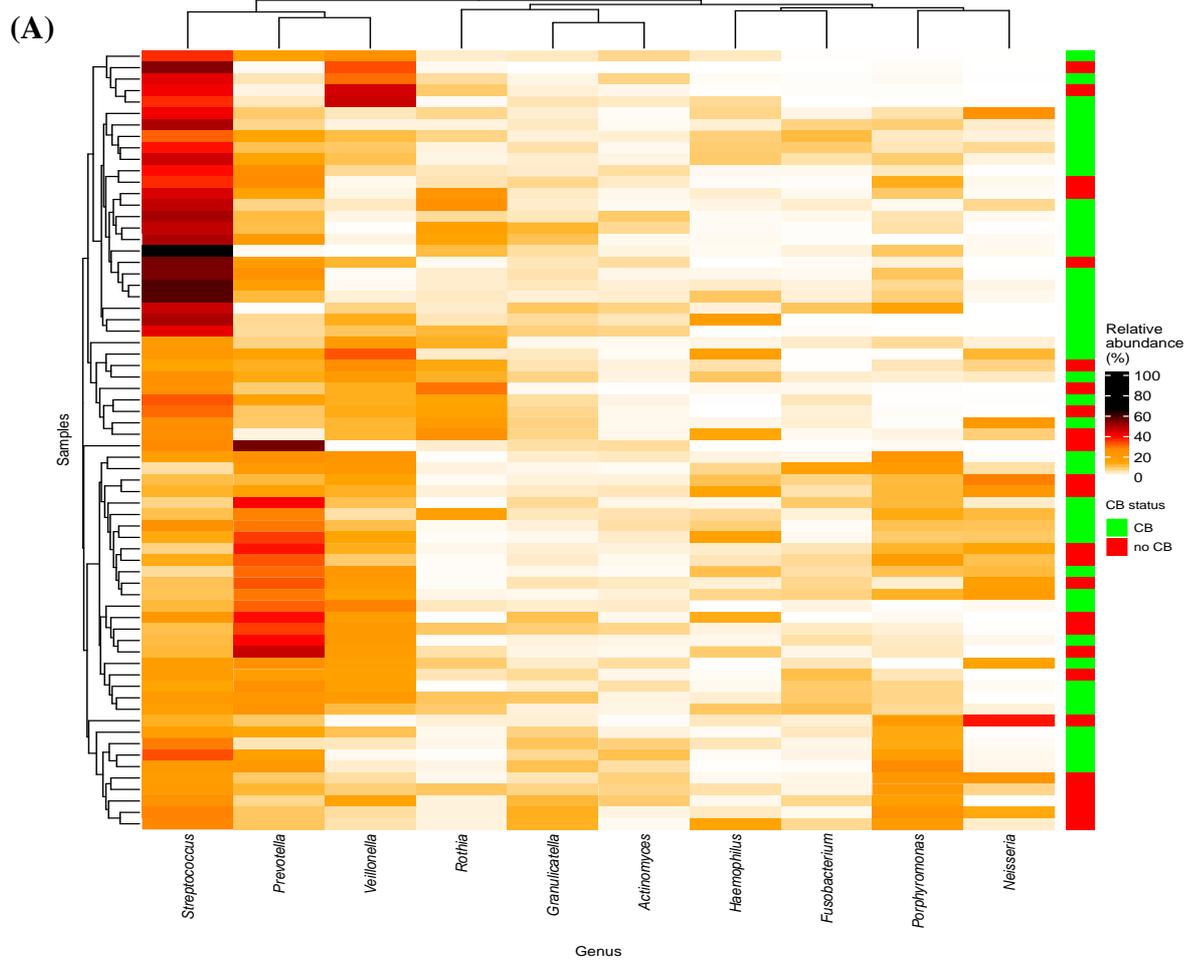
(B) Relative abundance of oral Streptococcus

(C)

(D)

(E)
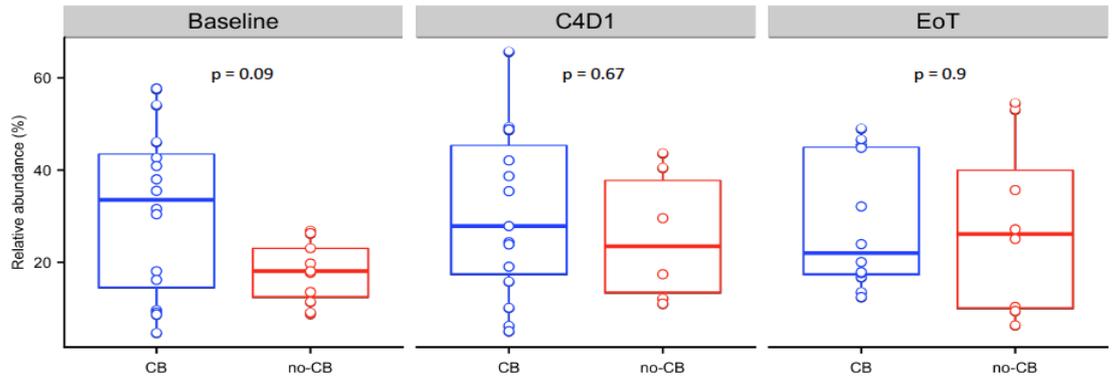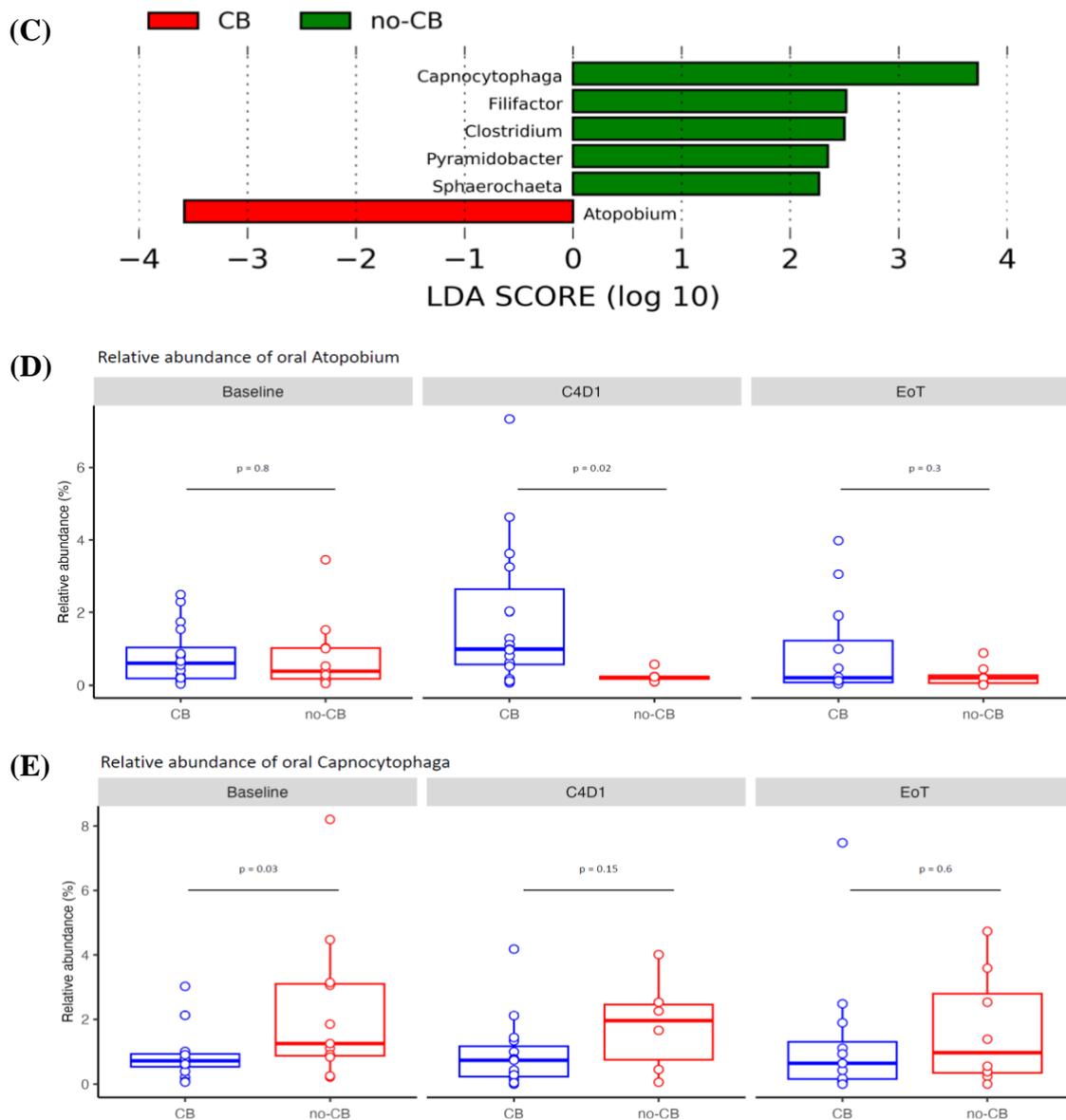
*Figure 4:3: **Oral Streptococcus, Capnocytophaga and Atopobium could be associated with CB status.** CB patients had higher relative abundances of oral Streptococcus at baseline (A). This observation was lost at consecutive time points and was not significant (B). LEfSe analysis suggested six genera, five associated with CB and one associated with no-CB (C). Relative abundances of Atopobium (D) and Capnocytophaga (E) by CB status. Statistical analysis was performed using Mann Whitney U. Linear Discriminant Analysis (LDA) is a method to find the maximum separability between categories. In the LEfSe pipeline it is used to estimate the order of magnitude of each differential abundant feature i.e., the LDA score indicates the effect size associated with the bacteria and clinical parameter.*

4.2.2 Pembrolizumab and eribulin did not significantly alter the gut microbiota

4.2.2.1 *No significant changes were observed in alpha or beta diversity of the gut microbiota*

Similar to the oral microbiota profile, alpha diversity was assessed by the Shannon and inverse Simpson index. No significant changes were observed in either index as patients underwent treatment, nor in CB status. This was also confirmed with a Mann Whitney U where p = 0.44 and 0.45 for inverse Simpson and Shannon index respectively when comparing by CB status (i.e., CB vs. no-CB). As observed in the alluvial plot, there are no significant changes in the gut genera (Figure 4:4A). An expansion in *Prevotella* in the no-CB group between baseline and C4D1 was noted. Further analysis indicated that the expansion of *Prevotella* was driven by three patients who had >20% of *Prevotella*, which is the likely reason why this is observed in the alluvial plot. When assessing beta diversity, none of the clinical parameters were significant in Permanova (p = 0.44, 0.20, 0.36 and 0.14 for PDL1, NLR, CB and PFS). Visualised as a NMDS plot, no clear segregation of any group was observed, suggesting no significant difference in beta diversity (Figure 4:4B). Notably, as no significant differences were observed (in either oral or gut microbiota profiles), this suggests that the combination therapy of pembrolizumab and eribulin does not lead to substantial microbiota perturbations via associated toxicity. I noted the presence of *Bifidobacteria* as being in the top ten most abundant genera in this cohort, which considering the previously reported literature is of interest as it has shown to influence anti-cancer outcomes. Probing into the mechanism of *Bifidobacteria* and anti-cancer therapies will be explored further in Chapter 5.

(A)

(B)
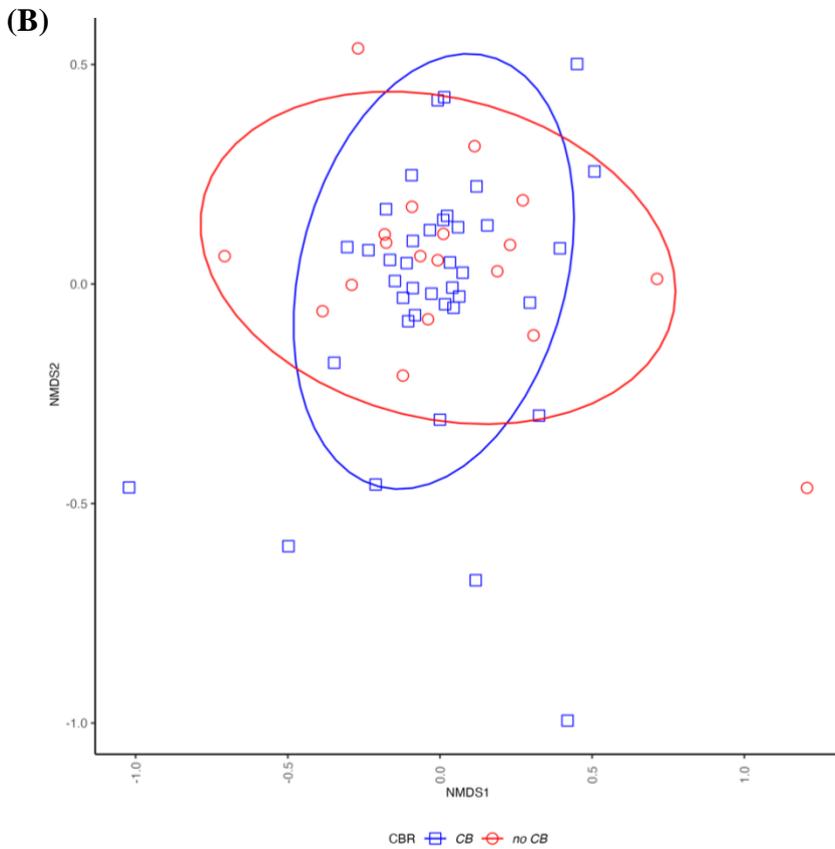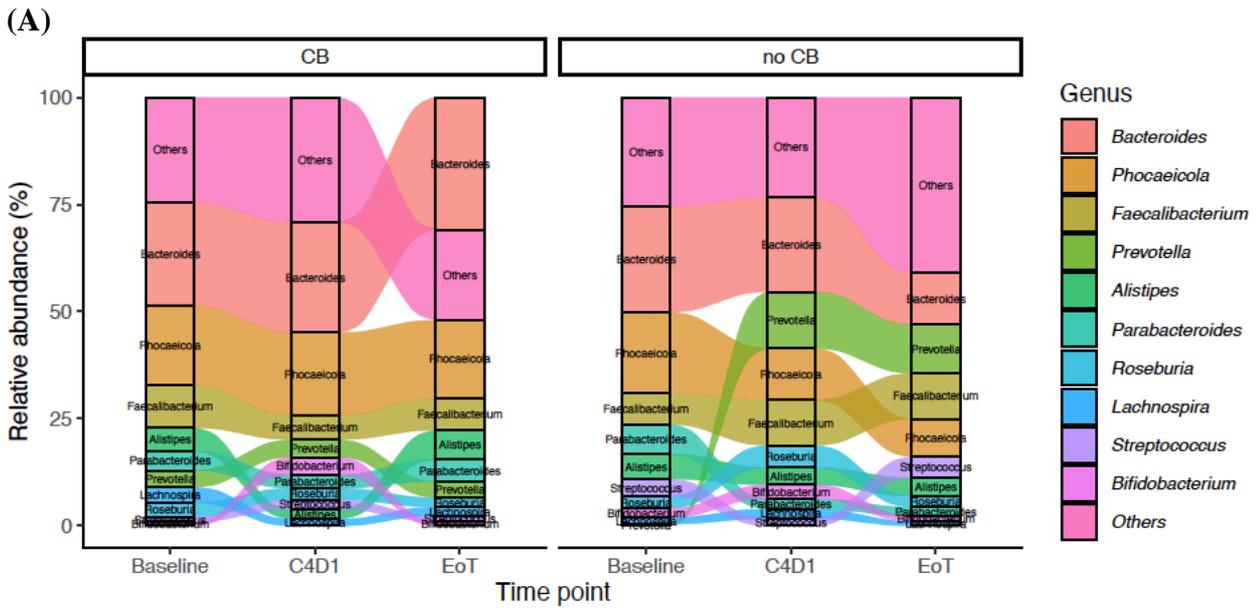
Figure 4:4: **Exploratory analysis of the gut microbiota of CALADRIO patients.** *No significant differences in the gut microbiota profile by CB status was observed (A). This was also the same on species level. Beta diversity was assessed by NMDS plotting (B). No segregation between groups, CB status, shown here, but also by other clinical parameters could be observed.*

4.2.2.2   Associations with adverse events and antibiotics

As part of the KELLY trial, any adverse events were recorded. I explored the associations of Grade 3 to 4 adverse events with microbiota profiles. By far the most common reported adverse event, possibly related to treatment, was a decreased neutrophil count i.e., neutropenia. The NMDS plot showed no segregation based on the common terminology criteria for adverse events (CTCAE) grades. This was not unexpected as no significant changes in either alpha or beta diversity was reported for either the oral or gut microbiota.

Antibiotics are known to severely perturb the microbiota, which was therefore of interest in our exploratory analysis. Concomitant drug information was requested and filtered to only include antibiotics. Simply by scrolling through the excel spreadsheet it was soon apparent that all patients had at least two different antibiotic classes by the second time point (C4D1), which complicated any additional analysis (Figure 4:5A).

4.2.2.3   Probing into the potential functionality of the gut microbiota

As I performed shotgun metagenomics on the faecal samples, the potential function encoded by the gut microbiota could be explored. The genomic data was run through the Humann3 database, the outputs of which were then put through LEfSe. LEfSe suggested sixteen potentially significant pathways, but only two pathways were different for less than or more than 6 months (Figure 4:5B). These were: "super pathway of purine nucleotides de novo biosynthesis II" and "super pathway of histidine, purine and pyrimidine biosynthesis". The former was a super pathway of purine nucleotides *de novo* biosynthesis II, and the latter was a super pathway of histidine, purine and pyrimidine biosynthesis. Mann Whitney U comparing the two pathways against "less than 6 months", or "more than 6 months" showed that the difference was significant (p = 0.02, for both pathways). However, caution should be taken with this result as the "<6months" value is 0, but it is unknown if no matches could be reported for this pathway by Humann3 or if there is genuinely no presence of this functional pathway.

Epidemiological studies have suggested that antibiotics influenced immune-checkpoint inhibitor success, which may be via modulation of host microbial communities [216]. As nearly all patients had at least two different classes of antibiotics by the second time point, I hypothesised that individuals who harboured greater antibiotic resistance experienced no-CB. FASTQ files of each sample was run through Resfinder to gauge if clinical parameters were correlated with higher 'hits' to resistance genes present in the gut microbiota. As seen

in Figure 4:5C, no distinct differences in antimicrobial resistance potential between CB and no-CB samples was observed. Given the treatment regimen, it was queried if microbial cytochrome P450 was differentially abundant between CB vs. no-CB, as it has previously been reported cytochromeP450 isoform 3A4 can metabolise eribulin [217-219]. Surprisingly, CB patients presented with a higher abundance of cytochromeP450 genes present.

(A)

(B)

(C)
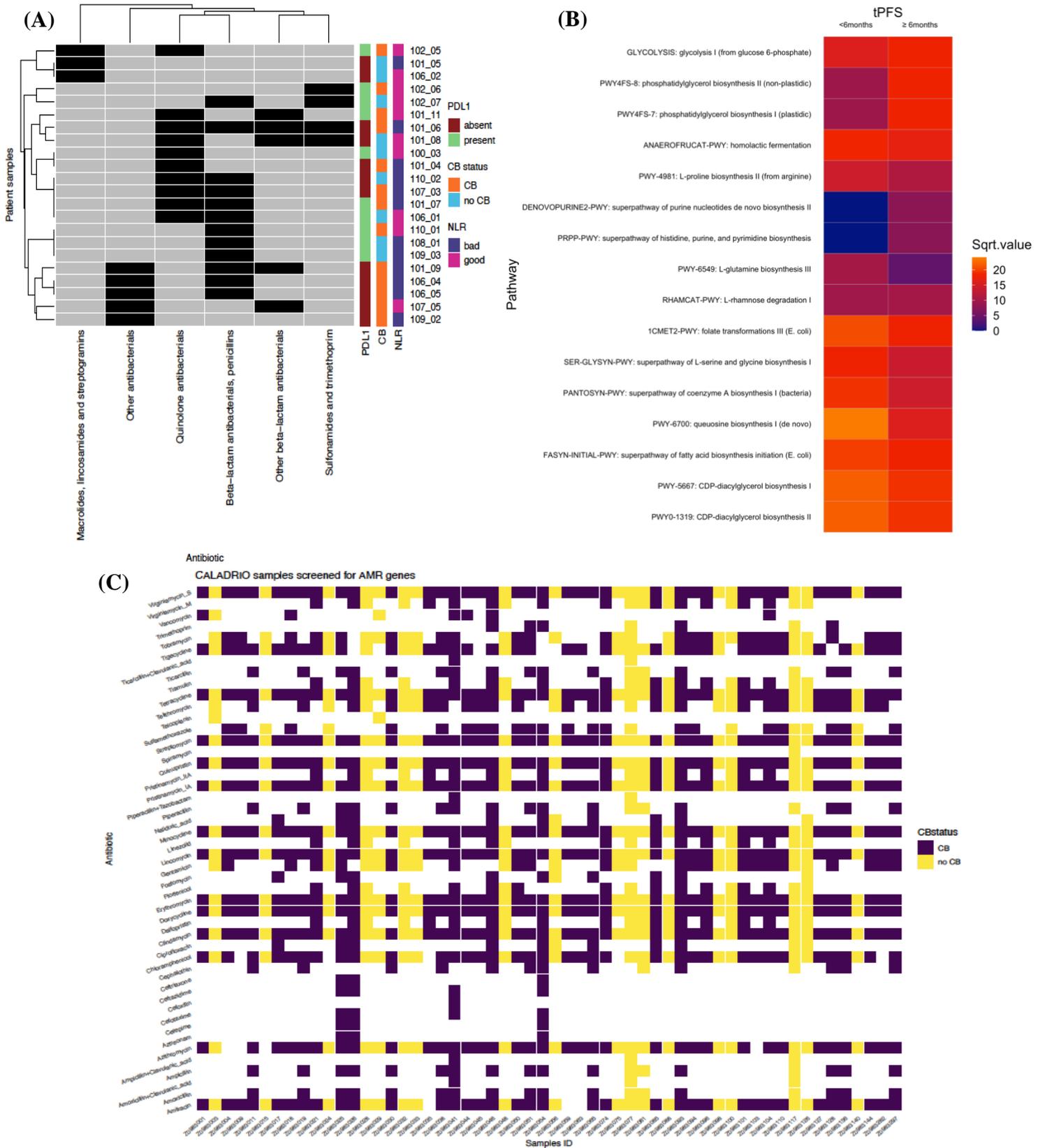
Figure 4:5: *Investigation into adverse events and potential function i.e., antimicrobial resistance in the gut microbiota. Nearly all patients enrolled in the study were administered antibiotics while undergoing their treatment of pembrolizumab and eribulin (A). The functional potential of the gut microbiota was assessed and there was only a difference in the clinical parameter: PFS. There was evidence to suggest that the pathways: "super pathway of purine nucleotides de novo biosynthesis II" and "super pathway of histidine, purine and pyrimidine biosynthesis" were present at less copies per million in PFS less than 6 months compared to more than 6 months (Mann Whitney U, p = 0.02 for both pathways). (B). No discernible differences could be observed in antimicrobial resistance potential between gut microbiota profiles of CB patients compared to no-CB patients (C). Figure A was made by Dr. Matthew Dalby.*

4.2.3   Translocation between oral and gut microbiotas

As both oral and gut microbiota profiles were obtained from patients, I was interested in investigating the potential of translocation. Assessing the significant oral LEfSe features, there were three common genera: *Capnocytophaga*, *Filifactor* and *Clostridium* within the gut. *Capnocytophaga* and *Filifactor* are common periodontal microbes, while *Clostridium* is usually associated with the intestinal tract of mammals. The presence of *Clostridium* in the oral samples was investigated. Only four samples with relative high abundances of this genus were reported. Similar for *Capnocytophaga,* there were two stool samples with high relative abundances. *Filifactor* was not significant in the LEfSe analysis of the gut microbiota. It is likely that this was a false positive driven by those samples having a high relative abundance compared to other samples [220]. In total there were 57 common genera between the oral and gut microbiota as shown in Figure 4:6. These 57 genera, listed in Table 4:2, could be reflective of oral-gut translocation. However, as the oral microbiota was 16S rRNA gene amplicon sequencing, while the gut microbiota was shotgun sequencing, I cannot confirm if these are the same strains.
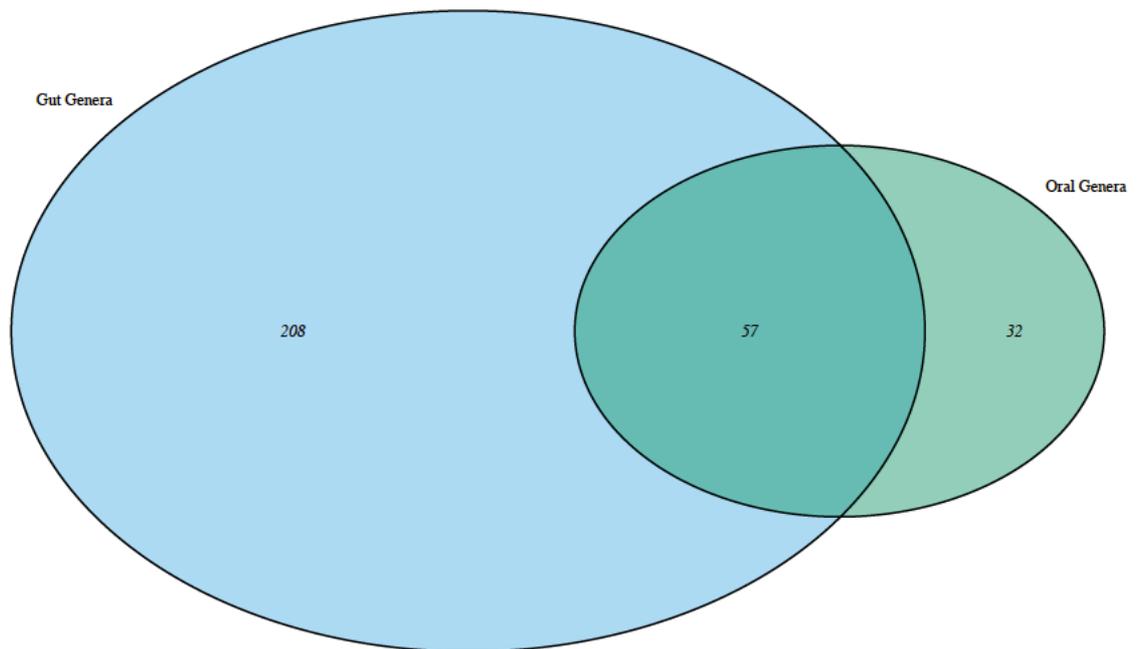

*Figure 4:6: **Venn diagram of common oral and gut genera.** There were 57 common genera between the oral and gut microbiota profiles (A). The common members are listed in Table 4:2 .*

Acidovorax	Desulfovibrio	Methylobacterium
Acinetobacter	Enterobacter	Mobiluncus
Actinomyces	Enterococcus	Mogibacterium
Aggregatibacter	Faecalibacterium	Mycoplasma
Agrobacterium	Filifactor	Neisseria
Bacillus	Flavobacterium	Odoribacter
Bacteroides	Fusobacterium	Paenibacillus
Bifidobacterium	Gemella	Paracoccus
Blautia	Geobacillus	Porphyromonas
Brevibacillus	Haemophilus	Prevotella
Bulleidia	Klebsiella	Pseudomonas
Butyrivibrio	Lachnospira	Ruminococcus
Campylobacter	Lactobacillus	Shewanella
Capnocytophaga	Lactococcus	Sphingomonas
Chryseobacterium	Lautropia	Staphylococcus
Clostridium	Leptotrichia	Stenotrophomonas
Coprococcus	Leuconostoc	Streptococcus
Desulfobulbus	Macrococcus	Treponema
Desulfomicrobium	Methanobrevibacter	Weissella

*Table 4:2: **Common genera between oral and gut microbiota of CALADRIO patients.***

4.2.3.1 B. fragilis may be associated with CB status

(A)

(B)
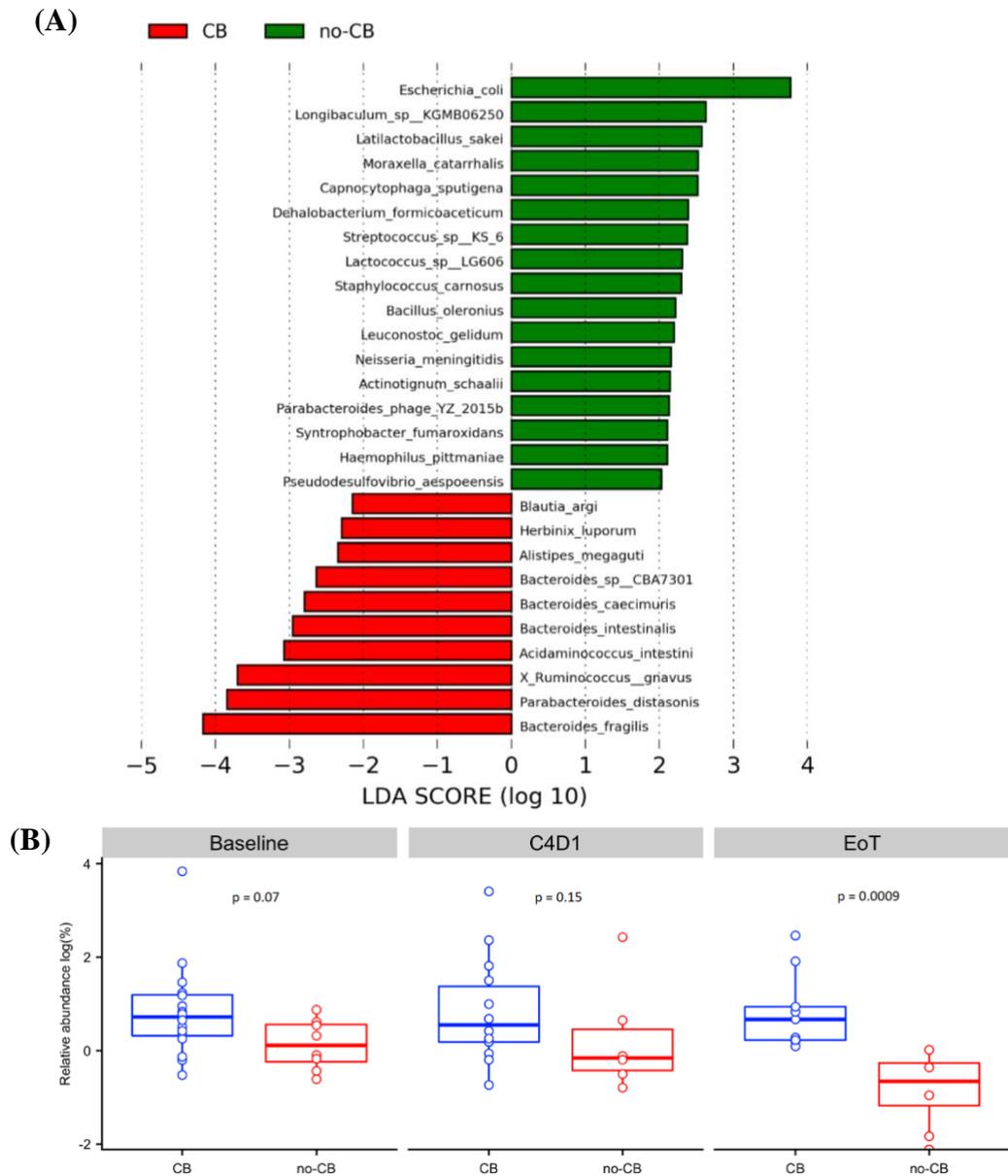
*Figure 4:7: **B. fragilis in the gut is associated with CB.** LEfSe analysis (A) showed that Bacteroides fragilis is associated with CB. Box plots (B) further showed that patients with CB consistently had a greater relative abundance of B. fragilis than no-CB. This observation was most significant at EoT (p = 0.0009, U = 0, median (CB, no-CB) = 1.95, 0.38, range (0.00-11.73)*

Although there were no overall overt differences observed in the gut microbiota profiles of patients, differential analysis identified 30 potentially significant microbiota members correlating with CB status (Figure 4:7A). The most notable features/microbes included *Bacteroides fragilis* for the CB group, and *Escherichia coli* for the no-CB group. Our investigation continued with a focus on *B. fragilis* as previous studies have shown its therapeutic potential [119, 221, 222]. Further statistical analysis indicated that the CB patients consistently had a higher relative abundance of *B. fragilis* than the no-CB group, reaching significance at EoT (p = 0.0009), shown in Figure 4:7B. There is a trend of

decreasing abundance of *B. fragilis* in no-CB compared to the CB group where the relative abundance seems to be stable across the time points. More information is necessary to determine if this is correlated to the clinical outcome, but it is possible that CB is driven by the presence of gut-associated *B. fragilis* and loss of this species translated to a less favourable clinical outcome.

4.2.4   *In vitro* validation of *B. fragilis* NCTC 9343 with MCF-7 (HR+/HER2-) BrCa cells

Based on LEfSe analysis, *B. fragilis* was a potential feature to explain CB, and given *B. fragilis* was consistently higher in CB patients than no-CB patients I believed this was a genuine observation. Therefore, I decided to elucidate possible mechanistic pathways how *B. fragilis* may influence CB status.

4.2.4.1   *Genomic screening of B. fragilis MAGs*

Given that we had shotgun metagenomics data, Dr. Raymond Kiu kindly extracted MAGs to more carefully identify the strain of *B. fragilis*. Of the MAGs that could be aligned, only five were *B. fragilis* MAGs. These were a match to the type-strain *B. fragilis* NCTC 9343, with 4/5 MAGs aligning at >98%. Apart from MAG number Z0983077.bin.48, the others came from patients who experienced a CB.

I screened the MAGs of *B. fragilis* isolates to look at functional potential, with an initial focus on antimicrobial resistance determinants. Using Abricate v1.0.1 [223] and Resfinder [224] the *B. fragilis* MAGs were screened for presence of antibiotic resistance genes (Figure 4:8), which could explain the relative higher abundance of *B. fragilis* in CB patient. The *B. fragilis* MAGs had genes conferring potential resistance to trimethoprim, D-cycloserine, cotrimoxazole and aztreonam. Of the antibiotics only trimethoprim was administered to these patients.

Screening the MAGs though the METABOLIC pipeline, it was noted that a specific cluster of sulphur cycling enzymes i.e. methionine metabolising enzymes [225], were present in these MAGs. Knowing that *B. fragilis* has immunomodulatory capacity [222] the MAGs were screened for presence of the capsular polysaccharide (PSA) and the *B. fragilis* toxin (*Bft* gene), the Genbank accession numbers are listed in: Table 8:1. Three of the five *B. fragilis* MAGs had a match to a putative protein involved in capsular polysaccharide export (Genbank: BAD47972.1) [226]. Previous hypotheses include that PSA can influence maturity of CD4+ lymphocytes, thereby directing host-commensal symbiosis [221] and

possibly influencing immunotherapies [119]. Parida *et al*., has reported *Bft* to be associated with breast tumorigenesis [227], and others have reported *Bft* to induce gut inflammation and associated with colorectal cancer [228]. However, screening our MAGs indicated no *Bft* gene (fragilysin).
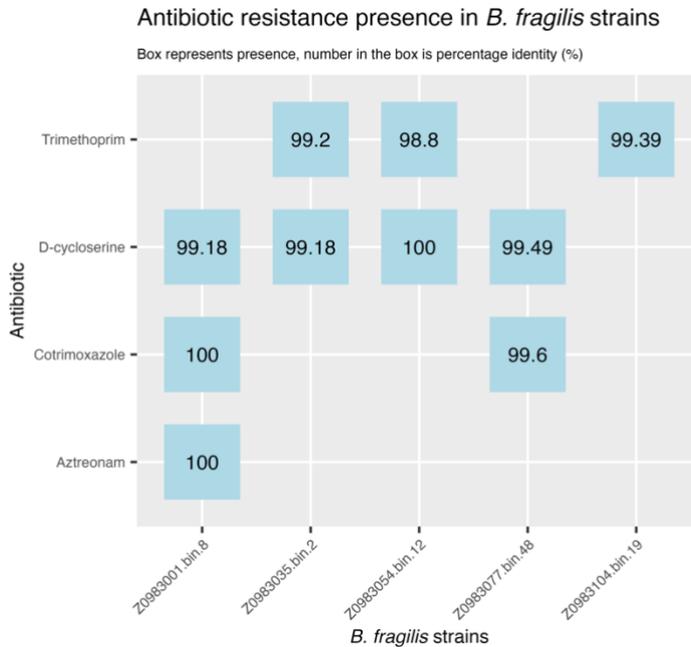

Figure 4:8: ***B. fragilis MAGs and their antibiotic resistance potential.*** *The B. fragilis MAGs were run through ResfinderFG to determine if they harbour any resistance potential which could explain their higher abundances. Only Z0983077.bin.48 came from a no-CB patient.*

4.2.4.2   *B. fragilis* NCTC 9343 growth curves and optimisation

As the MAGs were all a close match to *B. fragilis* NCTC 9343 I decided to use the type strain for downstream experimentation. I did have a BEAM study isolate, BP2A-13, which was also a match to *B. fragilis* NCTC 9343 isolated from a breast cancer patient. *B. fragilis* NCTC 9334 was kindly given to us by Dr. Regis Stentz (Quadram Institute Bioscience). Both NCTC 9343 and BP2A-13 isolates underwent whole genome sequencing to confirm identity. NCTC 9343 was a match to the type strain, BP2A-13 however was contaminated with *Eggerthella lenta*. This strain was decontaminated using Bile Esculin agar (BBE) with gentamycin (0.1mg/mL) [229], though at the time of isolation there was not enough gentamycin so it was done without (Figure 4:9C). Based on literature *E. lenta* cannot hydrolyse esculin making this an effective way of decontaminating this culture [230]. This process was ultimately successful in decontaminating the strain, and this was also confirmed with whole genome sequencing.

Next, I proceeded to determine the growth profile of *B. fragilis* by performing OD_{600} nm and CFU growth curves. OD_{600}, optical density measurement at a wavelength of 600nm, is used to estimate bacterial growth in real time as CFU growth curves are usually done

retrospectively. The technique measures the amount of light scattered through a cuvette, where a higher OD_{600} corresponds to bacteria growth resulting in media turbidity and thus greater light scattering. I sub-cultured at a 1:10 inoculation ratio to ensure similar growth kinetics. The OD600 growth curve was repeated in triplicate with differing results (Figure 4:9A and Figure 4:9B). Growth was inconsistent, with no obvious factors explaining why, and the pre-culture was performed 24h before starting the growth curve. Despite this, I was still able to estimate the length of the exponential growth phase; 16h to 18h. However, additional experiments differed from the previous collected data, where the exponential phase only lasted 12h for *B. fragilis* NCTC 9343 and 6h for *B. fragilis* BP2A-13 (Figure 4:9D and Figure 4:9E). Based on the combined CFU and OD_{600} growth curve data it was estimated that the cultures were at the following stages based on the OD_{600}: exponential (OD_{600} 1.01), late exponential (OD_{600} 1.13), stationary (OD_{600} 1.11), and death phase (OD_{600} 1.04). I set up another culture using the same ratio and used the estimated OD_{600}s to harvest the supernatant at the growth stages, this ended up being at 14h, 18h, 26h and 36h for exponential, late exponential, stationery and death phase respectively. The supernatant was filter sterilised (0.2µm) and frozen in 1mL aliquots for use in the co-culture.

**(A)** Growth curve of *B. fragilis* isolates (1st September 2022)
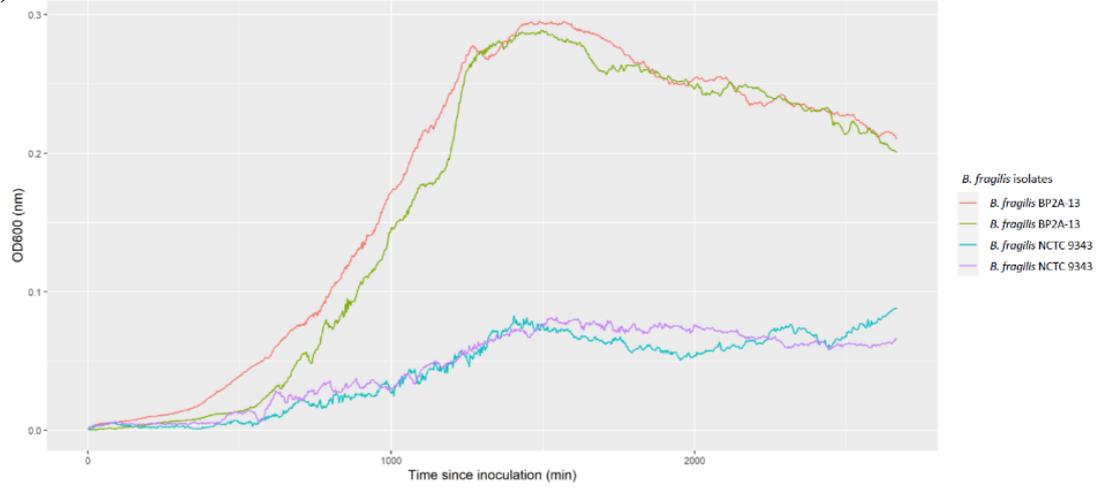
**(B)** Growth curve of *B. fragilis* isolates (8th September 2022)

(C)

(D)

(E)
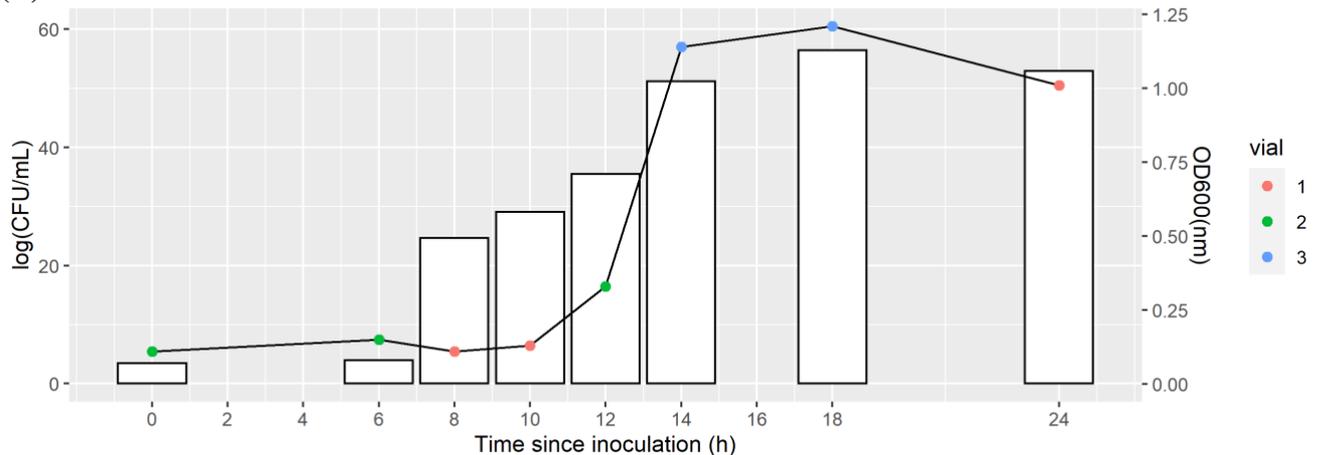
*Figure 4:9: **Growth curves of both B. fragilis strains.** OD$_{600}$ nm growth curves of B. fragilis BP2A-13 and NCTC 9343 were performed (A) and (B) with a week differing. The results were not identical however the pattern was consistent with exponential phase lasting 16-18h. I used BBE agar to isolate B. fragilis BP2A-13 from E. lenta as the former could hydrolyse esculin which turns the agar black (C). CFU and OD$_{600}$ nm growth curved was performed for B. fragilis NCTC 9343 (D) and B. fragilis BP2A-13 (E). Three vials were used to try and capture all the important growth phases. Line denotes OD$_{600}$nm measurements while bars denote log(CFU/mL). CFU measurements were done in triplicate. (A) shows four strains but two species, as they were measured at two different areas in the plate reader to determine if the sensors were malfunctioning, while (B) shows only the two species (B. fragilis NCTC and BP2A-13). All growth curves shown in (A) and (B) were done in triplicate.*

4.2.4.3 *B. fragilis NCTC 9343 cell free supernatant resulted in greater LDH release than media alone, but did not induce cell death in MCF-7 BrCa cells*

To investigate possible mechanisms of how *B. fragilis* might influence CB status of the HER2-/HR+ mBC patients, a co-culture with the cell-free supernatant (CFS) was set up. The identity of MCF-7 cells using the cell line authentication service from Eurofins was confirmed prior to starting. For the optimisation the following questions had to be first answered:

- In what bacterial media can MCF-7 cells grow without significant decrease in viability?

- Can *B. fragilis* grow well in that bacterial media?

- What density should we seed the cells?

I initially started with exposing the MCF-7 cells to neat bacterial media alone to determine what media would minimally influence the cells. The selected bacterial media were LB, nutritional broth 3, LEMP, RCM and BHI. LEMP was a homemade recipe which was a less carnivorous meat broth. This included 1% of Lab Lemco broth supplemented with 0.5% bacteriological peptone. MCF-7 cells were seeded at 1000 cells per one well in a 96-well microplate. The cells were exposed to the neat bacterial media for 24h. The growth curve was repeated twice as previous experience indicated that it is not reproducible. MCF-7 tolerated nutrient-broth 3 the best, even at a concentration of 10%, this was followed by LEMP. RCM was toxic at 10%, but only resulted in 20% decrease in relative viability. There were inconsistent results for BHI, as one would expect a decrease in viability as the concentration of BHI increases. However, this is not the case, as 2% BHI seemed to result in 20% increased relative viability i.e., growth or that 10% BHI was equal to the negative control. This was most likely not a true observation and may be explained by pipetting errors. As such, I conservatively assumed that a concentration of 5% BHI results in 80% relative viability to untreated cells, as this is consistent with the previous observations for other medias (Figure 4:10A). The *B. fragilis* strains were inoculated in the medias, both strains only grew BHI and RCM, and based on growth curves there were no significant improvement in growth in either medium. Thus, BHI was selected to continue downstream experiments as predecessors in the lab have used BHI for co-cultures with mammalian cells.

(A)

Relative viability of MCF-7 cells 24h incubation

(B)

Cell cytotoxicity 60h post exposure

(C)

Cell viability 60h post exposure
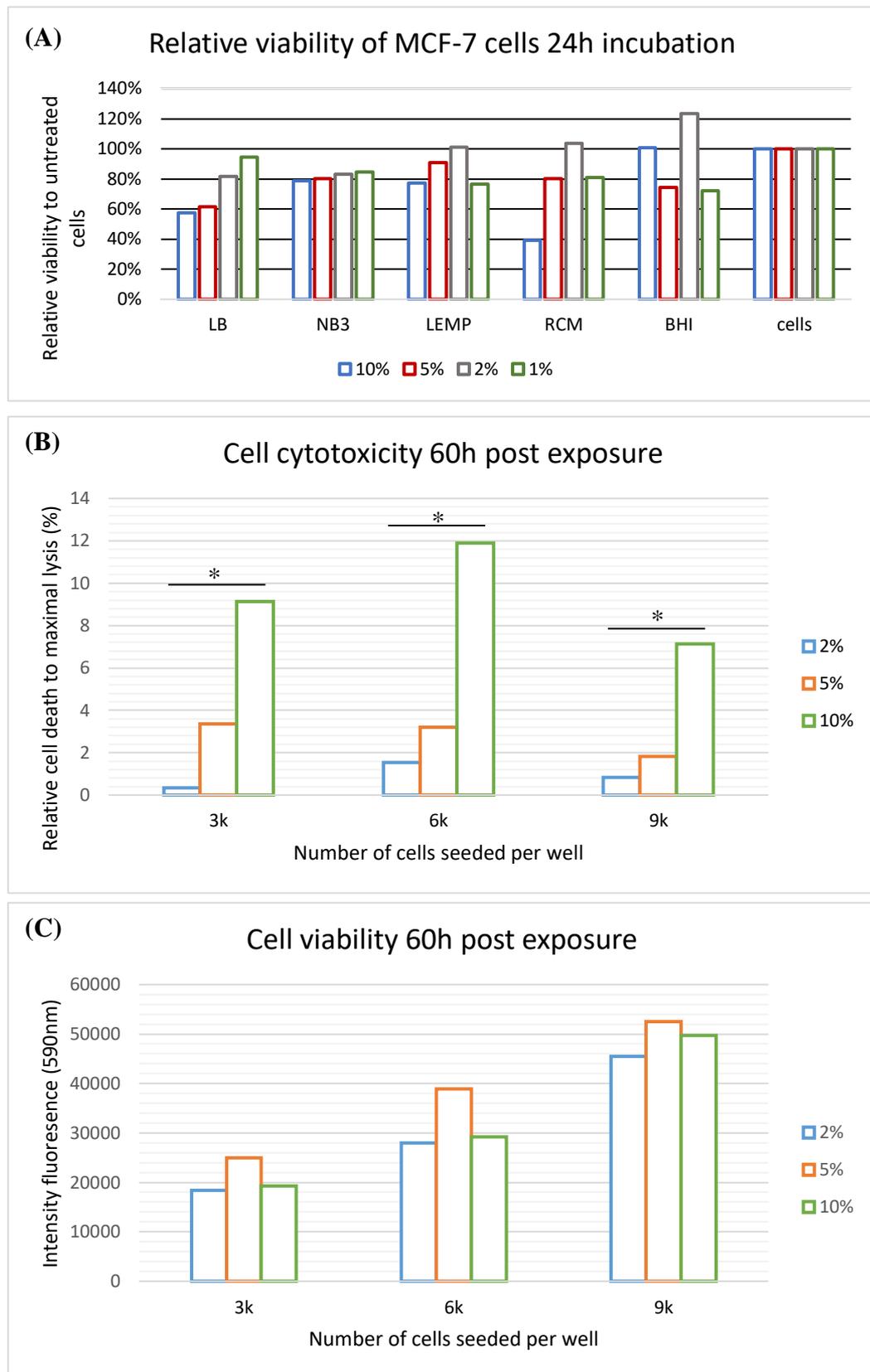
*Figure 4:10: **MCF-7 tolerance to bacterial media.** MCF-7 cells were exposed to different bacterial medias at four concentrations (1%, 2%, 5% and 10%) for 24h before relative viability to untreated cells ("cells") was measured (A). Cells were exposed to varying concentrations of FBS for 36h. 10% FBS was toxic to MCF-7 cells, while 2% barely saw any cell death. This was confirmed by Kruskal-Wallis and post-hoc Dunn's test, where * = p < 0.05 (B). Meanwhile, 5% FBS resulted in the greatest number of viable cells while 2% FBS was comparable to 10% FBS for cell viability (C). Cell viability was measured using the AlamarBlue assay kit.*

I focussed on determining the seeding density of the MCF-7 cells in a 96-well microplate. I used a range from 3000 to 9000 cells per well (which was the maximum due to potential overcrowding). Different concentrations of FBS were also included in this setup to determine what the 'minimal' media would be. After a 36h exposure time to the FBS the supernatant was removed. The cells were refreshed with fresh media with the respective concentrations of FBS and left in the incubator for another 24h before measuring viability using AlamarBlue and cell cytotoxicity using the Promega non-radioactive cytotoxicity assay. Surprisingly there was more cell death in 10% FBS than 5% FBS (Figure 4:10B), whilst the cell viability of 2% FBS was comparable to 10% FBS (Figure 4:10C). I performed a Kruskal-Wallis to determine if the FBS concentrations or number of cells seeded would have an influence on the relative viability. Kruskal-Wallis suggested that the median relative viabilities were different due to FBS concentration, $p = 0.02$. Post-hoc Dunn's test was performed, and it showed that only the medians of 10% FBS and 2% FBS was significantly different, $p = 0.02$. This suggests that 10% FBS is toxic to the cells despite growing, and instead 5% FBS is more beneficial to their growth. Two percent FBS resulted in minimal cell death even after 60h. A seeding density of 3000 cells per well was selected as the relative cell death between 3000 and 6000 was nearly double for 2% FBS, which is the minimal media. This would suggest that cells are stressed and releasing LDH into the media, possibly due to overcrowding as they are growing or depletion of nutrients. The Kruskal-Wallis test indicated that the number of cells seeded did not influence medians of either relative cytotoxicity or intensity fluorescence, $p = 0.8$ and 0.3 respectively.

The supernatant of *B. fragilis* NCTC 9343 and BP2A-13 was collected and filter-sterilised after 36h of growth, as previously mentioned. The cells were exposed at varying concentrations of CFS for 36h before the supernatant was replaced with minimal media. This experimental design was chosen to determine if the CFS had any influence on consequent growth since the doubling time is roughly 30h [231]. It was noted that the CFS resulted in high LDH release, but this was not reflected in a decreased relative viability. There was also an increased cell cytotoxicity in the *B. fragilis* BP2A-13 strain compared to *B. fragilis* NCTC 9343. Further investigations concluded that this was because the culture of BP2A-13 had double the number of bacterial cells compared to NCTC 9343, and diluting the CFS by half normalised the observation (Figure 4:11A and Figure 4:11B). This, however, did suggest a CFS 'product' being proportional to the number of bacterial cells. The change in pH of the CFS, which may induce LDH release in MCF-7 cells but not cell death was also considered as an influencing factor. After supernatants were harvested and mixed with the cell media,

their pH was checked, and they were buffered to a pH of 7.5. Satisfied that the pH is not a contributing factor I considered the phase of bacterial growth that was being used. Upon reflection, the supernatant being used was from bacteria in their death phase. Previous growth curves showed that the death phase was established at 24-30h. Lysed bacterial cells could release toxins into the CFS which could result in the phenotype observed. As such supernatants at different growth stages was harvested. An attempt to collect fresh supernatant of both *B. fragilis* strains i.e., NCTC 9343 and BP2A-13 was performed. However, due to equipment issues this could not be achieved in the designated time frame and the harvested supernatant obtained from the CFU growth curve had to be used. At this point they were a week old. After normalising the CFS based on the OD_{600} when it was harvested, the CFS was left on the MCF-7 cells for 24h before measuring viability and cytotoxicity. It should be noted that our viability assay was changed from AlamarBlue to MTS due to stock issues, however they both measure metabolism by the reduction of tetrazolium salt resulting in a formazan dye that can be detected using the spectrophotometer. This experiment was only done with *B. fragilis* NCTC 9343 as *B. fragilis* BP2A-13 could not be isolated in time. Consistent with previous observations, MCF-7 exposed to *B. fragilis* supernatant (at 14h-36h growth) released statistically more LDH than BHI media alone (Figure 4:11C). Although, this was not reflected as cell death as cells were able to reduce tetrazolium salt via NAD(P)H-dependent dehydrogenase enzymes, which is characteristic of metabolically active cells (Figure 4:11D). There was a significant difference in viability (p-value = 0.019), suggesting compromised metabolic activity in MCF-7 cells exposed to *B. fragilis* supernatant. Finally, there was a direct correlation between growth of *B. fragilis* and cytotoxicity; a linear regression confirmed a significant and positive (slope = 0.0015, p-value = $4.52x10^{-6}$ %), dose-dependent relationship (Figure 4:11E). Overall, these results suggested that products released into the supernatant can induce intracellular stress *in vitro*, albeit not severely inhibiting active metabolism in MCF-7 cells.
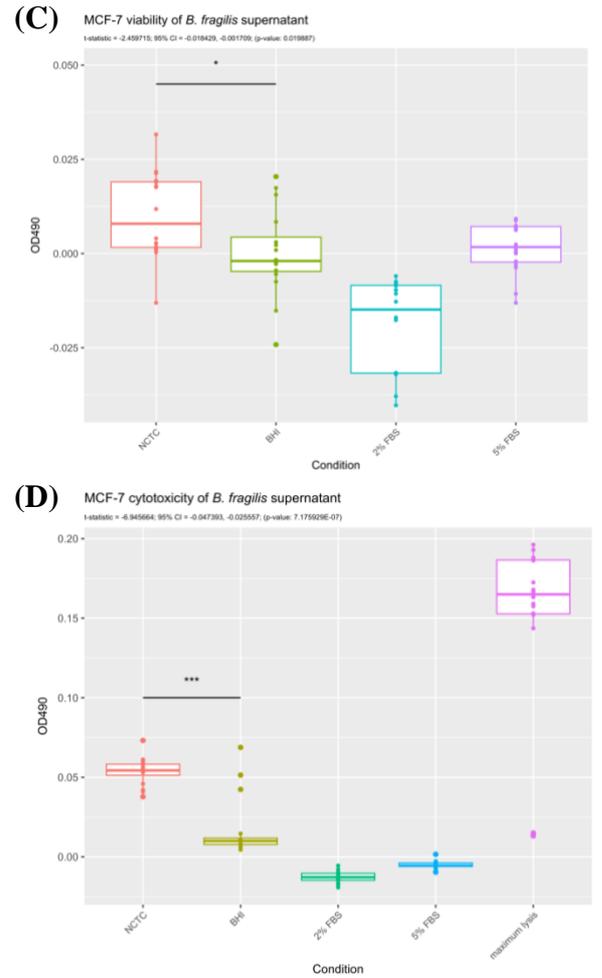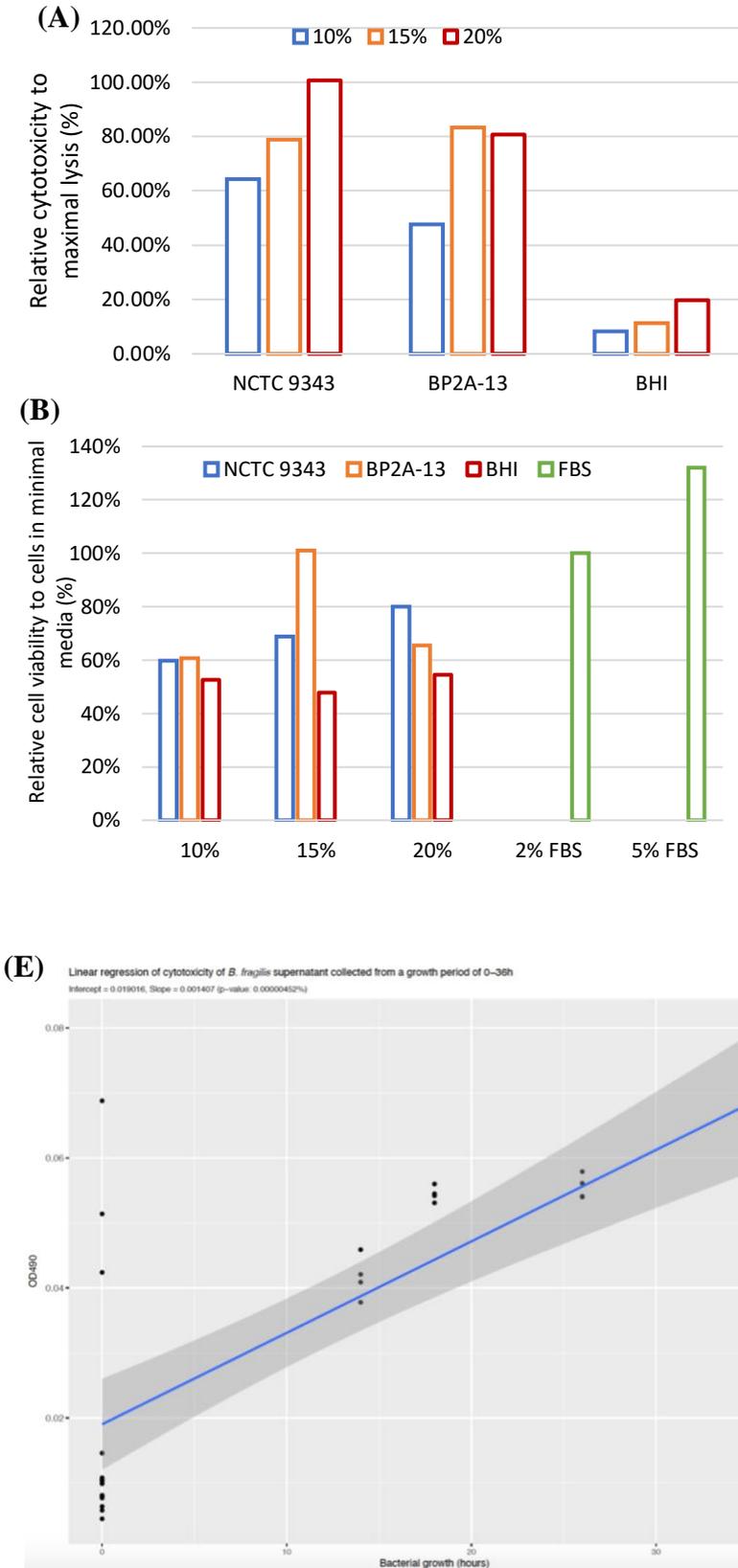
*Figure 4:11: **MCF-7 cells were exposed to B. fragilis NCTC 9343 and BP2A-13 CFS.** The CFS was a 48h culture supernatant, and cells were exposed for 60h before cytotoxicity was measured by LDH release (A) and viability was measured by an MTS assay (B). Consequently, supernatant from B. fragilis NCTC 9343 was used at different growth stages (14-36h after inoculation) and assess viability (C) and cytotoxicity (D) using the same kits as aforementioned. Linear regression showed that the relative cytotoxicity was a significant dose-dependent relationship i.e., the bacterial product was correlated to the number of bacterial cells (E). For (C) and (D) a Welch's t-test was performed. For (E) a linear regression was performed, and showed to be statistically significant, p < 0.0001. *** p < 0.0001 and * p <0.05.*

4.3 Discussion

The gut and oral microbiota are known to play key roles in host health, and more recent research has indicated these microbial communities also direct anti-cancer responses. Most studies to date exploring the relationship between microbiota and cancers have focused on melanoma or NSCLC. The CALADRIO study is one of the first that explored the relationship of the gut and oral microbiota in BrCa patients undergoing novel combination therapy against mBC. Due to its novelty, the team and I had the unique opportunity to study all aspects of how the novel combination therapy influences the oral and gut microbiota of mBC patients.

Here, I reported that pembolizumab and eribulin is not toxic to the gut or oral microbiota in treatment of mBC. *B. fragilis* appeared to be associated with CB, and further *in vitro* investigations suggested that metabolites produced by *B. fragilis* stimulate more LDH release from MCF-7 breast cancer cells than media alone. However, this did not correlate with cell death. No power calculations were undertaken prior to the analysis of these samples as it was an exploratory study that was undertaken as a sub-study of the KELLY study (phase 2 clinical trial). As such it is likely that the results reported are underpowered which could explain why, by classical statistical definition, results were not significant. Though general trends were observed e.g., gut-associated *B. fragilis* and possibly oral-associated *Streptoccocus, Capnocytophaga* and *Atopobium* being associated in CB patients.

A previous study profiled Spanish cohorts to define what is a 'normal' Spanish gut microbiota. Here the authors reported common members including *Bacteroides*, *Faecalibacterium*, *Prevotella*, *Alistipes* and *Oscillospiraceae* [232], and I also observed these genera in the CALADRIO cohort (alongside 265 genera across all samples), suggesting these as a core microbiota common to Spanish individuals, irrespective of health status. In this study I also profiled the oral microbiota, and I noted a lower diversity of genera (89) in comparison to the higher diversity of the gut samples. Between these two sites there were 57 common genera, with *Prevotella* notable between niches (although this appeared to be driven by a small number of individuals), and I also detected those normally exclusively found in the gut or oral cavity e.g., *Capnocytophaga* is common oral microbe but not gut. Other common gut and oral microbes are listed in Figure 4:6B. Of possible interest is *Fusobacterium* due to its reported associations as an opportunistic oral commensal [233], and its strong links to colorectal cancer [62, 234] and with BrCa in murine models [235] but this was not significant in our LEfSe analysis.

I further observed several known oral microbes associated with the no-CB group in the gut e.g., *Streptococcus*_sp_KS_6, *Neisseria meningitidis* and *Capnocytophaga sputigena*. *Capnocytophaga* was considered significant for both oral and gut microbiota in relation to no-CB suggesting it is common for both sites. Though when investigating further, it was determined that only two gut samples had relative high abundances of *Capnocytophaga sputigena* which would likely explain the significant observation. The significant discriminative features determined by LEfSe does open the possibility of translocation. The concept of oral to gut translocation is intriguing, and there is some evidence that perturbations of the gut by 'influx' of oral microbes is associated with the development of diseases e.g., Crohn's or CRC [213]. However, I cannot comment on direct translocation, as strain level tracking is necessary which was not possible in this study, but this could be compared in next stage clinical studies.

Other oral genera that were of interest included *Atopobium* and *Capnocytophaga* which significantly linked to no-CB and CB respectively. As aforementioned, *Atopobium* is usually associated with bacterial vaginosis. One study however did observe higher levels of *Atopobium* in cancer cases, while *Capnocytophaga* was observed more often in plaque than cancer cases. The study investigated various oral diseases and conditions which makes it difficult to draw parallels to my reported observations, and I cannot conclude if this was related to oral health or a disease/treatment association.

Streptococcus is a common member of the oral microbiota and is not usually associated with disease [236]. However, there have been studies reporting links with certain cancers associated with the respiratory tract i.e. oral *Streptococcus* with oesophageal [237] and tongue cancer [238], or faecal *Streptococcus* with gastric cancer [239]. Although these associations have been reported, the mechanisms remain unknown. Contrary to these reports i.e., *Streptococcus* being associated with cancer, I reported an association between oral *Streptococcus* and CB, though this did not reach statistical significance. I showed that at baseline, CB patients had a higher abundance of *Streptococcus* compared to no-CB patients. This signal was lost at C4D1 and EoT timepoints, however this may be partly driven by the administration of antibiotics; as all patients by C4D1 had been given two different classes of antibiotics. There are reports of streptococci being anti-inflammatory, as is the case for *Streptococcus thermophilus* [240]. More recently, a study reported anti-inflammatory properties of *Streptococcus salivarius*, a commensal oral microbe [241]. Unfortunately, due

to the nature of 16S rRNA gene amplicon sequencing, I cannot determine the species of streptococci with accuracy. In addition, 16S rRNA gene amplicon sequencing limits the ability to investigate antimicrobial resistance which would provide insight into the loss of the trend observed at baseline. Nevertheless, literature suggested that *Streptococcus* species can be anti-inflammatory, which could promote anti-tumorigenic microenvironments.

Across the gut samples, I observed no overall microbiota changes across patients and treatment that linked with different clinical outcomes. However, more specific analysis indicated a strong EoT association with *B. fragilis* and CB with the novel combination therapy. An important consideration in exploratory microbiota studies is the post-hoc analysis of sequencing data. Nearing *et al.* demonstrated elegantly the limitations of using multi-variate analysis tools used commonly in exploratory microbiota studies [104]. Conducting post-hoc analysis on in-sample data can lead to false-positives (further discussion on this topic is also outlined below). As such, *in vitro* validation should be considered to more fully explore the validity of observations drawn. In this study case I showed, *in vitro*, that there is a product produced by *B. fragilis* into the supernatant which stimulated LDH release from MCF-7 cells, but this was not reflected as cell death. Vétizou *et al.*, showed the favourable outgrowth of *B. fragilis* in CTLA-4 anti-cancer blockade in Ret melanoma, a murine model of spontaneous melanoma, and MC38 colon cancer models. In their study, they demonstrated that this may be achieved through Th1-mediated immune responses [119], possibly due to polysaccharide components of *B. fragilis*. Another study demonstrated PSA, present on *B. fragilis* NCTC 9343, can promote mucosal immunity, which aligns with the suggested mechanism of Vétizou *et al.* [242]. Components of PSA could be released into the cell-free supernatant due to natural death during the growth period [243]. Other microbial products present in the supernatant could include extracellular vesicles (EVs) or SCFAs. A study reported that EVs by *Bacteroides thetaiotaomicron* could influence host immune pathways in inflammatory bowel disease [244]. SCFAs have long been reported to exert beneficial effects on the host. Butyrate has shown to be anti-inflammatory in colitis [101], while cadaverine and lithocholic acid have been reported to reduce BrCa proliferation [124, 125]. Bacteroidota and Bacillota are the main butyrate producing members of the gut [245] and *B. fragilis* does encode the key enzyme to produce lithocholic acid [125], however it is unknown if it can produce cadaverine. It is likely that *B. fragilis* secretes immunomodulatory products into the supernatant, and further metabolomics studies could be performed to define what is present in the bacterial supernatants. The results I reported in this project in combination with published literature

suggest that these secreted products could provide an unfavourable tumour microenvironment by metabolites [116] or stimulate anti-cancer immune responses by cells [117, 119], which is in turn reflected as a CB. The metabolites themselves are most likely not causing a direct cytotoxic response to tumours but most likely are altering the environment to induce cellular stress onto the cancer cells.

Studies probing microbiota and responses to anti-PD-1 therapy have shown evidence that certain members can influence anti-tumour responses. Gopalakrishnan *et al*. assessed the oral and gut microbiota of melanoma patients receiving anti-PD1 immunotherapy and reported *Faecalibacterium* as being important in a prolonged PFS, as well as being significantly more abundant in the gut of responders. They also reported an enrichment of Bacteroidales in the faecal microbiota. However, for OTU enrichment analysis, Bacteroidales was enriched in non-responders and Clostridiales in responders [209]. Another study by Routy *et al*. reported an association between *Akkermansia muciniphila* and clinical responses by an IL-12 dependent manner [211]. Lastly, Tanoue *et al*. reported an eleven-strain consortium being capable of priming CD8 T-cells for an anti-tumour response. This consortium included seven Bacteroidales strains and four non-Bacteroidales strains [207]. No study to date has assessed the impact of eribulin alone on the gut microbiota, so I cannot comment further on this. Moreover, despite Bacteroidales not being significant in OTU enrichment analysis of one study, members of this order have been reported to be immunomodulatory and improvements in anti-tumour responses have been indicated. Despite the reported results being inconsistent across all studies, including ours, these types of studies support the hypothesis that gut microbes can influence the efficacy of anti-PD-1 therapy. The inconsistency could be due to the demographic of patients, where the studies were conducted, as well as the type of cancer i.e., BrCa vs. NSCLC and melanoma. Nevertheless, these reports support proof-of-concept exploratory studies that should be undertaken to explore adjuvant options for immunotherapies.

Whilst showing promising insights into microbiota associations with clinical outcomes my investigation does has several limitations, including the overall patient and sample numbers. However, this was an exploratory study and further carefully powered (multi-centre) studies comparing the gut microbiota with this novel treatment therapy in a larger number of BrCa patient is required. Moreover, even when a microbiota association is found to be statistically significant, depending on the sample size, it is unclear if this is a genuine or a chance observation [246]. Due to the small sample size utilised in this study even if no significance

is found it could be due to it being underpowered. Therefore, trends should be taken into considerations as this could suggest a potentially significant observation should the sample size be increased. On the other hand, despite it being an underpowered study should a significant association be found, it is most likely a genuine finding. An underpowered study may not be considered ideal, but results should not be immediately discounted. Nevertheless, undertaking *in vitro* and/or *in vivo* provides additional evidence to probe into potential mechanism of associations. Our *in vitro* work suggested that *B. fragilis* produces a metabolite/compound that stimulates MCF-7 cells to release LDH, but this is not sufficient to be cytotoxic. However, we only explored this using the CFS milieu, and further studies are needed to further define what the actual potential active components are. Finally, the oral microbiota was only profiled using limited 16S rRNA sequencing, thus further studies could apply more shotgun metagenomics profiling which would also allow strain tracking and potential translocation.

Due to the nature of stool collection, sterility cannot be guaranteed. As such, the presence of oral microbes in the stool sample may be explained by cross-contamination. I cannot determine if the significant relative abundances or oral microbiota members in some samples are due to cross-contamination or due to a genuine presence attributable to translocation. I cannot comment further on the possibility of translocation as only a small number of samples had a high relative abundance of these oral microbes, which may have led to the significant association. However, there were no oral microbes of significance associated with CB found in the gut LEfSe analysis.

4.3.1   Pitfalls of a post-hoc analysis

Whilst showing promising insights into microbiota associations with clinical outcomes my investigation is limited due to the nature of exploratory studies. Microbiota studies often involve large datasets and disentangling what is significant and what is noise can be difficult. There are methodologies to account for this which involve ML algorithms, and advancements in technology may allow researchers to make inferences of observational microbiota datasets [246]. However, we did not implement such techniques in our study and opted instead to do *in vitro* validation. Doing mechanistic work *in vitro* provides more evidence to support a microbiota association. Another limitation to our study is the presence of oral microbes in the gut sequencing data. LEfSe has indicated that some oral microbes present in the gut sequencing data may be associated with no-CB. Further investigations

showed that this was driven by some patients having a relative high abundance of these members compared to others.

4.4   Conclusion

Here, I explored the association between the gut microbiota of mBC patients undergoing a novel combination therapy, which provide novel insights into this complex relationship. I have demonstrated that the combination therapy of pembrolizumab and eribulin is likely not to confer microbiota toxicity in Spanish patients that have been diagnosed with HER2-/HR+ mBC. The novelty of the combination therapy, pembrolizumab and eribulin, makes this study one of the first to explore how these therapeutics impact the oral and gut microbiota of mBC patients. These investigations suggest that baseline oral *Streptococcus* could be a potential biomarker, however this was not significant, but larger studies could explore this in more detail. Furthermore, gut-derived *B. fragilis* could be associated with CB (at EoT), where patients with a CB had a higher relative abundance of *B. fragilis* compared no-CB. *In vitro* investigations suggest that *B. fragilis* may produce a product which can stimulate LDH release from MCF-7 (HER2-/HR+) cells, but that is not strongly cytotoxic. This initial finding shows promise for *B. fragilis* as a biomarker for this cohort of patients, with further mechanistic studies needed to elucidate potential pathways. A holistic understanding of this pathway could provide novel therapeutic avenues for anti-cancer therapy adjuvants.

4.5   Future works and impact

Alongside already established literature, this study supports the hypothesis that bacterial products present in the CFS can influence distal sites. Unfortunately, in this case I cannot conclude that the product present in the CFS by *B. fragilis* could influence the mammary gland *in vivo*. Furthermore, it would be interesting to determine what is present in the CFS which may be inducing greater LDH release in the MCF-7 cells. This could be investigated by NMR of the CFS, and should a candidate be found, isolation by HPLC to study its biological function on mammalian cells. I was mildly surprised to observe that MCF-7 cells were viable despite the relatively large amount of LDH release into the supernatant. To investigate this more, I had plans to extract RNA from these cells to determine what pathways were up- or down regulated. However, it was not possible to obtain enough RNA from the cells already harvested in the time remaining of my PhD project to address this research question.

Exploratory microbiota studies are an important tool to investigate potential biomarkers and novel therapeutic avenues. Little is known about the relationship between the gut microbiota and breast cancer, or breast health which is the knowledge gap that this PhD aims to address. HER2-/HR+ BrCa, when first-line therapies have not been successful, require novel therapeutics to improve clinical outcomes. Unfortunately, such therapies can be quite aggressive with a poor quality of life to the patient. For any cancer treatment there needs to be a balance between therapeutic success and quality of life. This is where the power of microbiota profiling can be used. As previous studies have already shown, certain profiles have been associated with better outcomes of anti-cancer therapies. This could act as a predictive marker for patients, or in the future a therapeutic adjuvant to improve clinical efficacy. In addition, understanding how the gut microbiota can influence the outcome of ICI further enhances our understanding of the gut-immune axis. In the context of BrCa this is important to consider as it is not known to be an immunogenic cancer. If the gut can influence immune cells in the context of BrCa TME it could provide novel therapeutic avenues to explore.

5   *Bifidobacterium longum* project

This chapter will describe the *in vivo* experiment I have performed during my PhD. The project arose from working collaboratively with the Robinson group, where they had planned a large animal study using chemotherapy. I had observed the relative high abundance of *Bifidobacteria* in both CALADRIO and BEAM patients. It is known that the chemotherapeutic, cyclophosphamide, is given to patients part of the BEAM study so we decided to collaborate to address if *Bifidobacterium longum* subsp. *longum* YP5B-G had an effect on cyclophosphamide efficacy which is described in this study. The handling of the mice was done by members of the Robinson group specifically, Dr. Wesley Fowler, Dr. Sally Dreger, Dr. Stephen Robinson, Mr. Christopher Price and Ms. Alicia Nicklin. Additionally, Alicia Nicklin helped me with setting up the qPCR reaction to quantify *Bifidobacterium* in the mouse faeces. Dr. Wesley Fowler kindly helped with the oral gavage preparation. Herein I describe how live oral supplementation with *B. longum* subsp. *longum,* isolated from a BrCa BEAM patient, interfered with the effectiveness of cyclophosphamide in luminal B breast cancer mouse models. I helped with the processing of animals after sacrifice this included, collecting tumour measurements and harvesting tissues. I prepared the cultures for gavage and undertook the qPCR reaction under guidance of Alicia Nicklin and lastly, I processed and analysed the data presented in this chapter.

5.1   Background

5.1.1   Bifidobacteria in anti-cancer therapies

As already discussed, studies have already established an association between *Bifidobacterium* species with enhancing anti-cancer therapies [117, 120, 247, 248]. *Bifidobacterium* is a common member of the gut microbiota and certain species have been recognised to be important in early life due to their digestion of human milk oligosaccharides, including *B. breve*, *B. longum* subsp. *infantis* and *B. bifidum* [249, 250]. The species present in an adult gut change to predominately *B. longum* subsp. *longum* and *B. adolescentis* [250], and it has been implicated as an important microbe for cross-feeding [251-253]. Studies have associated *Bifidobacterium* with the stimulation of tumoricidal activity in natural killer cells, alluding to the immunomodulatory influence of this genus [254].

In general, studies assessing the association between bifidobacteria and anti-cancer therapies are not consistent with regards to possible mechanisms. However, the presence of elevated

serum IFN-ɣ has been widely reported [117, 120, 126]. Animal studies and *in vitro* studies have set the foundation to suggest that *Bifidobacterium* species can act as adjuvants in anti-cancer therapies. Most recently a proof-of-concept study or phase 1a clinical trial (NCT03829111) was established to assess the effects of a *Bifidobacterium* live biotherapeutic product in the treatment of metastatic renal cell carcinoma patients receiving ipilimumab and nivolumab. The study suggested a trend for patients receiving the live biotherapeutic to have an improved progression free survival and overall survival, but the study sample size was too low and underpowered to detect such a significance [255]. In another study, preliminary results from NCT03775850 showed that in a cohort of triple negative BrCa patients (n=12) receiving pembrolizumab and a dose of *Bifidobacterium animalis*, 18% experienced the objective response i.e., stable disease or partial response. Recruitment and analysis for this study is still on-going but no severe adverse events have been reported in patients, which shows promise [256].

5.1.2 Bifidobacteria from the BEAM study

From the previous culturing efforts and sequencing analysis of the BEAM study I noted that *Bifidobacterium* species were in the top ten most abundant genera. Additionally, *B. longum* subsp. *longum* was cultured out of all but two BEAM patients. I was interested to see if this isolate could confer synergistic effects with a chemotherapy drug, cyclophosphamide, in murine BrCa models. Cyclophosphamide was chosen to use as the chemotherapeutic as the NNUH use this drug in clinic and we wanted to remain as accurate to clinical treatment as possible.

As part of the faecal culturing process, I confirmed each *B. longum* isolate to the type strain *Bifidobacterium longum* subsp. *longum* DSM20219. I selected the isolate YP5B-G for this experiment which had an ANI of 98% to the type strain. Once its purity was re-confirmed using whole genome sequencing, I grew up cultures to use for live gavage to luminal B breast cancer carrying mice receiving chemotherapy. Below I describe the experimental design and the results of this project, this includes the mouse experiment and its results, retrospectively quantifying the amount of *B. longum* subsp. *longum* that was present in the intestine of mice and lastly, I screened the genome to determine if this might have influenced the *in vivo* phenotype.

5.1.3   Aims and hypothesis

The hypothesis driving this experiment was: oral supplementation of *B. longum* subsp. *longum* with cyclophosphamide will increase anti-cancer therapy efficiency. Specifically, my aims were:

- To assess if oral supplementation with *B. longum* subsp. *longum* results in a greater tumour volume reduction compared to chemotherapy alone.

- To quantify the amount of *B. longum* subsp. *longum* present in the gut of the mice after treatment.

5.2   Experimental design
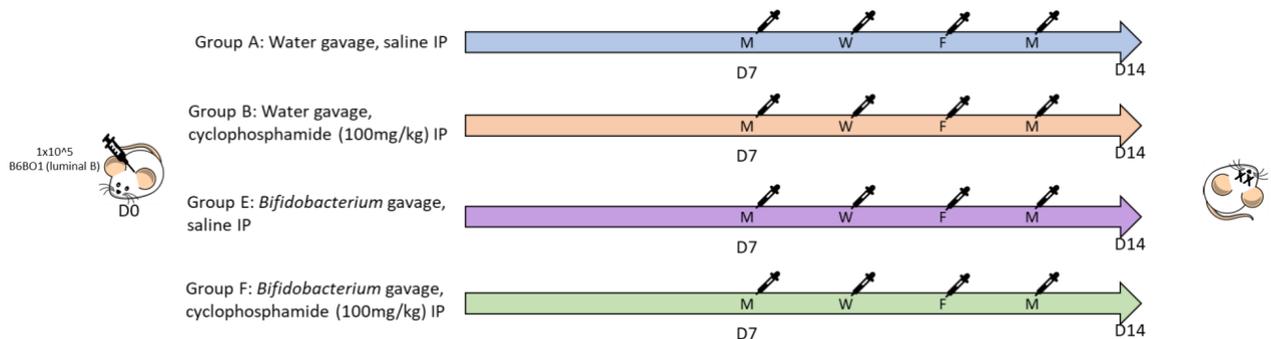
*Figure 5:1: **Experimental design of the mouse study.** Mice were injected with 10^5 B6BO1 (luminal B subtype) breast cancer cells into their left abdominal mammary fat pad. Once palpable (D7 post injection), mice were subjected to either gavage with water or Bifidobacterium and/or intraperitoneal injection (IP) of saline or cyclophosphamide. This continued every other day until D14 after which, the mice were sacrificed, and tissue was harvested.*

The number of mice was determined using power calculations based off previous work by the Robinson group. It was calculated that at least ten animals per group spread over three experimental repeats would be sufficient to detect a minimum significant difference of 20% between two different genotypes at a significance level of 5%, using a two-sided t-test, and with a power of 80%. The control groups were A (n = 9) and B (n = 10) where the former received oral gavages of water every other day with an intraperitoneal (IP) injection of saline and, the latter received water gavage and cyclophosphamide IP injection at 100mg/kg. Group E (n = 9) received oral gavage of *B. longum* subsp. *longum* YP5B-G with no cyclophosphamide while group F (n = 7) received oral gavage of *B. longum* subsp. *longum* YP5B-G with IP injection of cyclophosphamide at 100mg/kg.

5.3   Results

5.3.1   Retrospective faecal culturing could not quantify the CFU of *Bifidobacterium*

The experimental design is described in Figure 5:1. I prepared YCFA plates and media broth to retrospectively determine how many viable CFUs were gavaged to the mice each day.

The inoculum was pre-cultured 24h prior to the preparation of gavage. Four oral gavages were administered to the mice. CFU of the oral gavages were: 22 x 10^{17}, 2.4 x 10^{11}, 1 x 10^9 and 1.5 x 10^{21} CFU/mL, giving an average dose of 3.75 x 10^{20} CFU/mL a day.

I tried to quantify the number of *B. longum* subsp. *longum* viable colonies by plating out faecal slurry pellets (taken directly from the intestine of the sacrificed mice) on bifidobacteria specific media (Man Ragosa Sharpe medium supplemented with 50µg/mL Cysteine-HCl and 0.5mg/mL mupirocin). However, within 24h of anaerobic incubation the presence of colonies that did not look like bifidobacteria were noted (Figure 5:2). In spite of this, I did note that colonies only grew in group E and not group F, however I only decided to do one mouse per group as a trial. I chose the first mouse of groups E and F to do this trial on, and the order of the mice in each group was done at random i.e., whichever mouse was processed first. I decided not to culture the colonies on these plates to determine their identity as they did not look like standard *Bifidobacterium* isolates and previous attempts from other lab members under similar circumstances resulted in many other non-*Bifidobacterium* species. It is likely that there are *Bifidobacterium* isolates present on the plate however as previous reports from colleagues have shown there are many other non-*Bifidobacterium* species that grow as well. To culture, isolate and identify each colony and determine if its *Bifidobacterium* is laborious compared to an alternative method: qPCR.
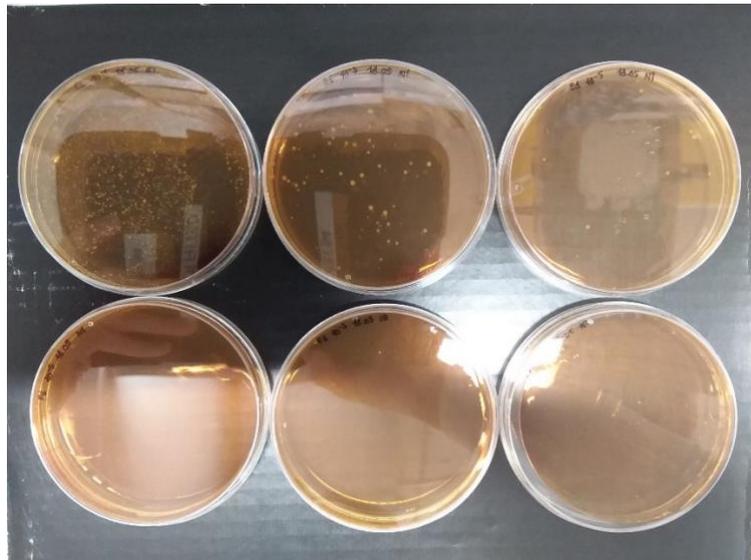

Figure 5:2: **Faecal slurry of one mouse from group E and F plated on MRS agar supplemented with 50mg/L of mupirocin and 0.5g/L cysteine.** *Plates were left to incubate anaerobically for 48h before observing colonies.*

5.3.2 qPCR of *Bifidobacterium longum* subsp. *longum* could quantify the number of bacteria present in the gut

As retrospective culturing was unsuccessful, another attempt was made to quantify the number of *Bifidobacterium* present in the gut of these mice. I extracted the genomic DNA from faecal pellets and with the help of Alicia Nicklin set up a qPCR reaction based on the methods described in [168]. I used the standard curve to set up the qPCR to quantify *groEL* gene in the faecal pellets (Figure 5:3A). There was roughly triple the number of *Bifidobacterium* in group F than group E (3.5 x 10^8 vs. 1.3 x 10^8 respectively) as seen in Figure 5:3B. This could be due to dosing inaccuracies as group F would be the gavaged after group E. The bacterial suspension may not have been resuspended properly prior to the administration resulting in a higher concentrated mixture towards the end of the treatment compared to at the start. Surprisingly, there was non-specific amplification to *B. bifidum*.

(A) Standard Curve qPCR targeting *B. longum* subsp. *longum*

(B)
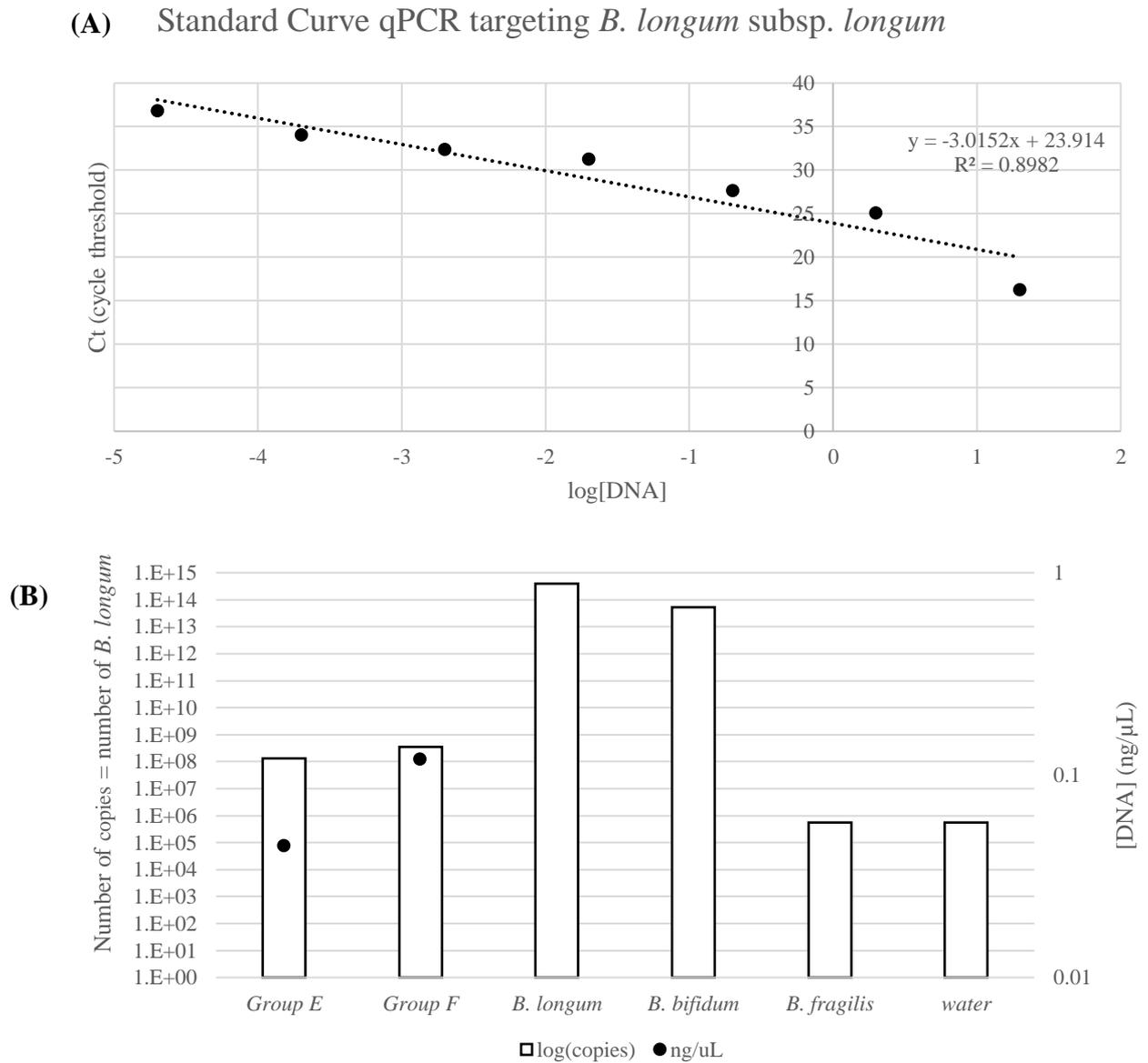
*Figure 5:3: **Quantification of B. longum in faecal pellets by qPCR targeting the groEL gene.** A standard curve was set up (A) to determine the optimum DNA concentration to quantify the number of Bifidobacterium in the mice (B). B. longum is genomic DNA extracted from B. longum subsp. longum YP5B-G, B. bifidum is B. bifidum subsp. bifidum which acted as a species-specific control for the primers as they should only target B. longum subsp. longum and B. fragilis is Bacteroides fragilis which acted as a negative control as it lacks the groEL gene so therefore should not have any amplification. Water was added as another negative control. Group E and F represent the number of B. longum subsp. longum copies present in the faecal pellets, and by extension the gut microbiota, of these mice. Primers targeting the groEL of B. longum located 250 to 840 of the corresponding nucleotide sequence were: forward: 5'-CTGAGGCTCTGGACAAGGTCG-3' and reverse: 5'-GGTGCCACGGATGTTGTTCAGG-3'. Bars show the number of copies (log(copies)) calculated based off of qPCR reactions while the dots show the DNA concentration (ng/µL) that was obtained from DNA extraction of the mice in these groups.*

5.3.3   *B. longum* subsp. *longum* is likely not to improve therapeutic efficacy

On the day of the harvest the Robinson group kindly harvested the tumours, blood, caecum and colon for me. Tumour volumes were calculated using the following formula: (length x width^2) x 0.52 [257]. Kruskal Wallis test was performed to determine if the medians of the groups were different, as the data was not normally distributed or had equal variance. The p value was 4.5 x 10^{-5}, allowing me to reject the null-hypothesis thereby providing evidence that the groups had different medians from each other. To determine which groups differed from each other I performed a post-hoc Dunn's test with Holm's correction. All comparisons showed that the median of tumour volumes were different from each other, where groups A-B and B-E had a $p < 0.001$, while groups A-F, B-F and E-F had a $p < 0.05$, as shown in Figure 5:4. Thereby providing evidence that the treatment resulted in different tumour volume. The addition of *B. longum* subsp. *longum* did not seem to reduce median tumour volume where group A, untreated, had a median of 589mm^3 compared to group E, *B. longum* subsp *longum* gavage, which had a median of 601.8mm^3. Surprisingly, the group treated *B. longum* subsp. *longum* oral gavage and cyclophosphamide (group F, median 240.5mm^3) had larger tumour volumes than the group treated with cyclophosphamide alone (group B, median 99.55mm^3). This suggested that oral gavage of *B. longum* subsp. *longum* interfered with the effectiveness of chemotherapy.
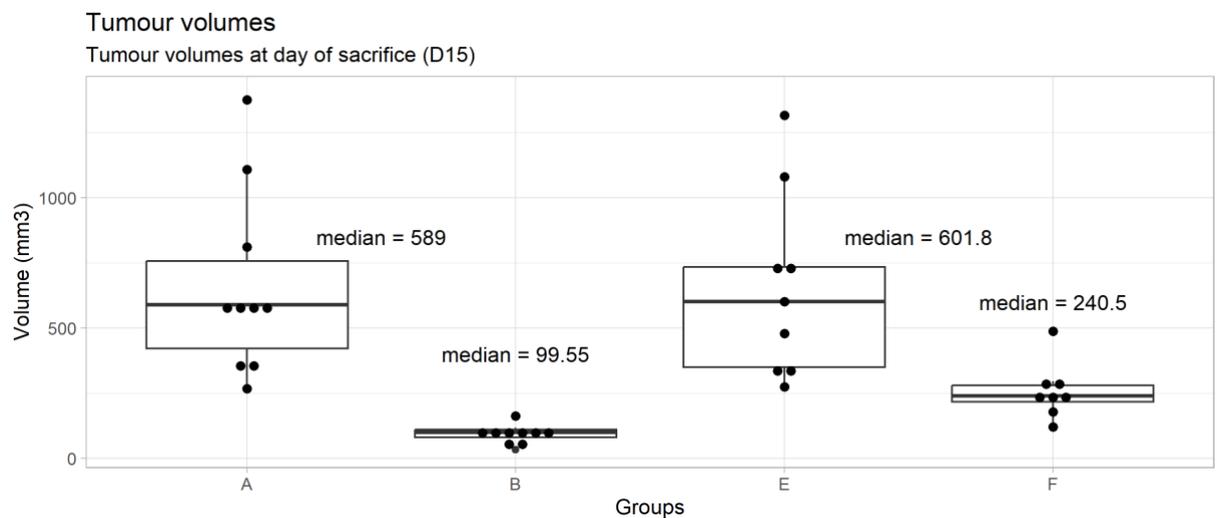

Figure 5:4: ***Tumour volumes of mice at day of sacrifice, D15.*** *Group A had ten mice treated with water gavage and saline IP injection. Group B had ten mice treated with water gavage and cyclophosphamide IP injection. Group E had eight mice treated with B. longum* subsp. *longum gavage with saline IP injection. Group F had eight mice treated with B. longum subsp. longum gavage and cyclophosphamide IP injection. Error bars denote standard deviation, dots represent a tumour volume from one animal, boxplot denotes the quartiles of the groups. Kruskal-Wallis with post-hoc Dunn's test with Holm's correction was done to determine which groups were significant. *** = p < 0.001, p values were 0.0002 and 0.0004 for A-B and B-E respectively. For groups A-F, B-F and E-F p values were 0.005, 0.001 and 0.01 respectively.*

5.3.4 *Bifidobacterium* had the potential to inactivate cyclophosphamide

A possible explanation as to why oral administration of *B. longum* subsp. *longum* interfered with the effectiveness of cyclophosphamide could be that this strain may metabolise the active component of the drug. Cyclophosphamide has to undergo several metabolic steps before reaching the active ingredient phosphoramide mustard i.e., aldophosphamide [258]. Aldehyde dehydrogenase (ALD) oxidises aldophosphamide to carboxycyclophosphamide which is an inactive form of the drug. Aldophosphamide is the component which decomposes to phosphoramide mustard and acrolein which alkylates DNA strands ultimately resulting in cell apoptosis [259].

Since ALD can inactivate phosphamide mustard it seems plausible that high levels of ALD could confer chemoresistance [260]. The mice between groups B and F only differed by their intervention which was oral gavage of *B. longum* subsp. *longum*. Accordingly I was interested to see if the isolate genome encoded ALD, which could oxidise aldophosphamide to its inactive form. I downloaded the FASTA sequences of ALD specific to *Bifidobacterium*, downloading a total of eight sequences. Using ABRicate v1.0.1[223] I screened the genome of *B. longum* subsp. *longum* YP5B-G and the type strain *B. longum* subsp. *longum* DSM 20219. The type strain had one hit (BLLJ_RS02495 aldehyde dehydrogenase family protein [*B. longum* subsp. *longum* JCM 1217]) while *B. longum* subsp. *longum* YP5B-G had two hits (BLLJ_RS02495 aldehyde dehydrogenase family protein [*B. longum* subsp. *longum* JCM 1217] and adhE bifunctional acetaldehyde-CoA/alcohol dehydrogenase [*B. pullorum*]). In the context of this specific experiment it is possible that this particular strain and its ability to inactivate cyclophosphamide resulted in the observations given. It should be noted that not all *Bifidobacterium longum* subsp. *longum* will harbour this enzyme and therefore it is not representative of the entire species.

5.4 Discussion

Studies have highlighted the beneficial potential of *Bifidobacterium* in anti-cancer therapies [117, 120, 247, 248]. However contrary to the reported literature, this experiment indicates that *B. longum* subsp. *longum* YP5B-G, an isolate cultured out from a BrCa 58-year-old patient, decreases the efficacy of cyclophosphamide. This isolate was cultured out of a post-surgery sample; however, this patient did not receive chemotherapy. The number of *Bifidobacterium* present in the faecal pellets of the mice was roughly triple in group F (*B. longum* subsp. *longum* gavage and cyclophosphamide IP) than group E (*B. longum* subsp. *longum* gavage and saline IP) which was confirmed by qPCR targeting the *Bifidobacterium*

groEL gene. Based on CFU/mL calculations, there was an average dose of 3.75×10^{20} CFU/mL a day, however it had been observed that *Bifidobacterium* did not colonise the intestinal tracts of these mice but remain in the guts of these mice for 24h before being expelled (data not shown[10]). Despite not colonising the Robinson group has shown evidence that the transient effect of *Bifidobacteria* is sufficient enough to reduce tumour volumes in animals. Work done by others have shown that Bifidobacterium species can have transient effects on the gut epithelium and throughout the body [261, 262]. Group F had 3.5×10^8 number of *Bifidobacterium* while group E had 1.3×10^8 number of *Bifidobacterium*, this is significantly less than the average CFU/mL. It is difficult to compare CFU/mL and number of copies as is the case for culturing methods vs. qPCR respectively. Live cells may have died or become stressed during transport resulting in less than 10^{20} CFU/mL being gavaged into the mice. This observation brings into question the use of retrospective CFUs for gavage experiments. Molecular quantitative methods are a more advantageous technique to quantify number of bacteria gavaged, provided a protocol exists for the species. Despite being a faster method compared to culturing, qPCR reactions require specific equipment and reagents that culturing does not necessarily need. Lastly, CFU methods do not consider the potential loss that can occur during transport, while qPCR accurately reflects the presence of bacteria in the gut of the subjects provided the reactions were successful and specific.

Primers were designed using GenBank accession numbers specific for *B. longum* the species and not strain specific. In the table there were four accession numbers given for *B. longum* however these were all partial coding sequences. The final primers designed for this experiment as reported in the paper were forward: 5'- CTGAGGCTCTGGACAAGGTCG - 3' and reverse: 5'- GGTGCCACGGATGTTGTTCAGG -3'. In the qPCR reaction I noted that there was non-specific amplification to *B. bifidum* which was unexpected as in the paper it was supposed to be species-specific. I later noted that there were two different primer sequences, one to generate qPCR standards by amplifying the *groEL* gene and another which was manually designed based off partial/complete *groEL* sequences to be used for the qPCR reaction with the sample. The primers used for this experiment were primers to be used to amplify the *groEL* gene to obtain a PCR product that can be used for qPCR standards. In principle, the primer should still work to amplify the *groEL* gene in the sample, however this was not the case in this study and the different primer could have resulted in the off-target amplification I observed with *B. bifidum* subsp. *bifidum*. Although previous results

---

[10] Discussed in Robinson lab meetings.

from the Robinson group showed that *Bifidobacterium* is not a normal gut microbiota of these mouse colonies, off-target amplification of other *Bifidobacterium* species in the gut may also confound results, and if metagenomics was performed this could have been further investigated.

Nevertheless, the gavage of *B. longum* subsp. *longum* did result in significantly different tumour volumes between groups B and F, where the former only received cyclophosphamide and the latter received cyclophosphamide and *B. longum* subsp. *longum*. Despite the number of *B. longum* being nearly triple in group F than group E, as the median tumour volumes between the groups who did not receive cyclophosphamide were the same (i.e., groups A (water gavage and saline IP) and E (*B. longum* subsp. *longum* gavage and saline IP) where the average tumour volumes were 589mm^3 and 601mm^3 respectively), I concluded that *B. longum* subsp. *longum* does not influence primary tumour growth. This was also confirmed to be not significant using Kruskal-Wallis with post-hoc Dunn's test with Holm's correction, $p = 0.88$, eliminating this as a potential influencing factor to tumour growth. When screening the genome of *B. longum* subsp. *longum* YP5B-G I observed that it possessed two genomic sequences for the ALD enzyme which have been well documented to inhibit the active metabolite of cyclophosphamide: aldophosphamide. This suggests that this isolate, has the potential function of deactivating and reducing the effectiveness of cyclophosphamide in these mice, however as I did not further pursue metabolite analysis, I cannot be certain that this is what happened.

Antibiotic administration with cyclophosphamide has previously been shown to ablate the cytotoxic activity of the drug [208]. Thereby suggesting that microbial metabolism of the drug could be important in exerting anti-tumour effects. There have also been reports of intra-tumoural bacteria resulting in the drug resistance as described in: [263, 264]. Possible mechanistic pathways include enzymes that can inactivate active compounds of the chemotherapeutic drug [197]. This concept of drug metabolism by the gut microbiota is new and is an area researchers believe could be an influential factor as to why some individuals demonstrate resistance to certain drugs; this has been named pharmacomicrobiomics.

Bifidobacterium is generally known as 'beneficial' and important in the first year of life [249]. It is also well characterised functionally for its ability to digest human-milk oligosaccharides and different carbohydrates [265]. However, other functional capacities can only be assessed with what is available on databases. I was fortunate to have the DNA

sequences of ALD specific to *Bifidobacterium* available on the NCBI database, but other potentially important enzymes are currently not annotated. Another limitation in this investigation was the fact that I only probed into potential function. It is well documented how cyclophosphamide is metabolised to become active and inactive, and all of this takes place primarily in the liver where aldophosphamide then diffuses to surrounding tissues and cells. It has not been documented how much of the aldophosphamide diffuses into the gut to become inactivated by ALD. Lastly, despite the *B. longum* subsp. *longum* YP5B-G isolate encoding for ALD, I cannot confirm that this was explicitly expressed in the mice as I only assessed the function based on gene presence rather than transcriptional responses.

5.5   Conclusion

Previous microbiota profiling of BEAM patients showed that *Bifidobacterium* was one of the top ten abundant genera in these patients. Additionally, I managed to culture out *B. longum* subsp. *longum* from nearly all BEAM patients. This, combined with literature demonstrating the anti-cancer effects of bifidobacteria, was the working hypothesis for this experiment. The aim of this experiment was to see if live supplementation of *B. longum* subsp. *longum* isolated from BrCa patients could improve anti-cancer therapy efficacy in mice. Surprisingly, it was found that oral administration of live *B. longum* subsp. *longum* with chemotherapy resulted in a larger tumour volume compared to the group without oral administration plus chemotherapy. This suggests that the bacterium may interfere with the effectiveness of cyclophosphamide. Screening the genome indicated it contained two different DNA sequence codings for a protein: aldehyde dehydrogenase, which has been documented to oxidise aldophosphamide into its inactive form carboxycyclophosphamide. This could be an explanation as to why the tumour volumes of mice receiving cyclophosphamide and *B. longum* subsp. *longum* were slightly larger than the mice receiving chemotherapy alone.

5.6   Future works and direction

Building on the results of this project I would dive more into the concept of pharmacomicrobiomics. It would be of particular interest to assess the ability of this *B. longum* subsp. *longum* isolate to inactive cyclophosphamide *in vitro* by culturing methods described by the Turnbaugh group [266]. In addition, metabolite analysis of urine or blood could also provide more insight if aldophosphamide was oxidised to its inactive form which could be due to the administration of *B. longum* subsp. *longum*. As part of the experiment, blood serum, caecum, colon and tumour tissue were harvested. Alongside metabolite

analysis of serum it would also be beneficial to do histology of the tumour and staining for e.g., CD8[+] and CD4[+] T-cells [267] to assess tumour infiltrating lymphocytes per treatment. These results could be a novel investigation demonstrating the drug-metabolising potential of *Bifidobacterium* in cancer therapies. This experiment does also demonstrate the need to perform pre-clinical experiments. *Bifidobacterium* is generally seen as 'good' and immunomodulatory, but here I observed that this was not the case for the administration of cyclophosphamide.

Pharmacomicrobiomics is a relatively novel concept and gaining appreciation as clinicians are moving towards more personalised medicine strategies in cancer therapies. Gaining a greater understanding and appreciation of how the microbiota can influence drug metabolism (either by inactivation or activation of drugs) could predict efficacy outcomes. Should a microbiota have the ability to inactivate drug metabolites it is likely that the individual would not have a strong response to chemotherapy and therefore would not make a good candidate. Some studies have already attempted this [268, 269].

6 Final discussion and future perspectives

Observational studies have suggested a potential association between the gut microbiota and breast cancer. One such study was done by Jackson *et al.,* [111] who showed that BrCa patients in their cohort had a perturbed microbiota profile compared to those who were not diagnosed with BrCa. Mouse studies have demonstrated the potential of gut microbes to influence anti-cancer therapies by modulating the systemic immune system [117, 119, 132, 133, 207-211, 255]. Furthermore, a perturbed microbiota due to antibiotic administration has shown that it could lead to either accelerated tumour growth or increased metastasis [127, 128]. These studies provided the background and working hypothesis for this thesis which aimed to profile the gut microbiota profile of BrCa patients and to determine if these signatures are associated with clinical outcomes, and also to validate observations through *in vitro* or *in vivo* models.

6.1 Summary of findings

To address the research question at hand I first needed to establish a pathway in partnership with the Breast Care clinic at the NNUH and the NRP Biorepository. This was successful and ended up being called the Breast hEalth And Microbiota (BEAM) study. However, the SARS-CoV-2 pandemic understandably impacted overall patient numbers recruited.

All samples, from May 2023, have been sequenced using 16S rRNA gene amplicon sequencing, with a subset also subjected to shotgun metagenomic sequencing. Both 16S rRNA gene amplicon and shotgun metagenomics were adequate in describing overall taxonomic changes. Metagenomic shotgun has the advantage that it can undertake strain level analysis and discriminate to a species-level, in addition to probing into potential functionality. Unfortunately, information about antibiotic administration is limited therefore I cannot assess what impact, if any, these drugs had on the gut microbiota profile. Culturing the faecal samples with YCFA and BHI led to the recovery of 298 isolates, of which there were 137 confirmed pure strains. During the culturing process two novel isolates were found and investigations done to validly publish the isolates. At least one species of the top ten genera reported in shotgun metagenomics was cultured, bar *Faecalibacterium, Coprococcus* and *Roseburia*.

BEAM focussed primarily on the overall taxonomic changes of the gut microbiota. Due to limited clinical information, I cannot assess how changes may be associated with outcome.

The CALADRIO study on the other hand aimed to specifically address whether the oral and/or gut microbiota is associated with clinical outcome. Metastatic BrCa patients received a novel combination therapy of pembrolizumab (anti PD-1 immunotherapeutic) and eribulin (mitotic inhibitor chemotherapeutic). Clinical studies with a focus on the microbiota typically focus on the gut and profiling the oral microbiota provides novelty to this study. I observed a potential biomarker in patients who experienced a clinical benefit, namely that these patients trended to having higher abundances of gut *B. fragilis*. Screening the metagenome assembled genomes of *B. fragilis* showed that these genomes were negative for the *Bft* gene, but some had the protein involved in exporting the immunogenic capsular polysaccharide protein. *In vitro* co-cultures of *B. fragilis* cell-free supernatant suggested that there was a product present in the supernatant that stimulates LDH release but is not cytotoxic. This initial finding holds promise for *B. fragilis* as a biomarker however further mechanistic studies needs to take place in order to elucidate the downstream pathways.

Finally, members of *Bifidobacterium* were the second most frequented cultured genera in our BEAM cohort. In addition, *Bifidobacterium* was consistently in the top ten genera of BEAM and CALADRIO patients. I determined that the most common *Bifidobacterium* species isolated was *B. longum* subsp. *longum*, and with the help of the Robinson group, we performed an *in vivo* study to determine if live oral supplementation with this strain could increase the anti-cancer effects of cyclophosphamide. This is based on literature suggesting that *Bifidobacterium* can have a synergistic effect of anti-cancer therapies. In this experiment we showed that *B. longum* subsp. *longum* interfered with the effectiveness of cyclophosphamide. Further screening of the isolate's genome showed that it encoded the gene for aldehyde dehydrogenase which has been reported to deactivate cyclophosphamide. Although I cannot confirm this is what occurred in the mice, it does demonstrate the ability of the gut microbiota to influence pharmacokinetics of patients.

6.2   Limitations and future works

This project attempted to investigate if first-time diagnosed BrCa patients had a gut microbiota profile associated with the disease and elucidate how the gut microbiota may influence clinical outcomes. Currently early screening is recommended for a better prognosis for BrCa, however the pandemic has unfortunately caused an unprecedented strain on health care services. Methods of detecting at risk patients would not only reduce the strain on health care services, but also provide a less invasive method of screening. The BEAM study aimed to address if BrCa patients did have an alternative gut microbiota profile, unfortunately due

to the pandemic recruitment numbers could not be reached. One limitation to this research project is having no age-matched controls to assess if BrCa patients had an alternative baseline gut profile compared to non-BrCa patients, so future works should include age-matched controls. Another limitation to be considered is that one of our initial questions was to assess the impact of prophylactic antibiotics on the gut microbiota. Collecting antibiotic data ended up being more difficult than previously anticipated so we could not obtain this information to associate antibiotics with the gut microbiota or clinical outcomes. I submitted a subset of BEAM samples for metagenomic sequencing, this method allows researchers to probe into potential function of the microbiota. This was an objective when analysing BEAM samples unfortunately I experienced software dependency issues on the HPC environment. Unfortunately, due to time constraints this issue could not be resolved, resulting in potential function not being investigated in this thesis. Nevertheless, the information and initial set up, and proven success, of the pathway can allow others to continue sample collection.

By following patients as they underwent their treatment, I had hoped to observe how their microbiota changed according to treatment and investigate if this was associated with a clinical outcome, yet this could not be achieved due to the impact of the SARS-CoV-2 pandemic. In contrast, the CALADRIO study went further and assessed if gut or oral microbiota profile could be associated with an outcome. These studies are important in understanding how the microbiota can influence anti-cancer therapies. Not only would it benefit patients, to stratify according to potential response, but it would also provide a greater understanding of the wider influence gut microbes can have on a patient's immune system. Literature has already suggested that certain microbes can influence the outcome of certain anti-cancer therapies, but studies are limited in the potential mechanistic pathways. Elucidating the mechanism of how gut bacteria can influence systemic immunity could provide novel therapeutic avenues to improve clinical outcomes. Limitations of our investigations in the CALADRIO study, like the BEAM study, includes not having a control cohort of 'healthy' age-matched females, and overall patient number limiting in-depth statistical analysis. Chemotherapy also has an influence on the gut microbiota, and I cannot determine if changes in the microbiota were due to prior therapies that patients experienced. Should research continue, it would be worth investigating what component of the cell-free supernatant from *B. fragilis* NCTC 9343 elicits the response that was observed. It would also be worthwhile to perform RNA sequencing of the MCF-7 cells to determine what pathways, e.g., immunological or death, is elicited in response to the cell-free supernatant. Previous literature has reported *B. fragilis* to be capable of influencing anti-cancer responses therefore

investigating this further would provide greater insight into possible mechanisms of how gut *B. fragilis* influences distal sites in the human body.

It should be noted that the method of DNA extraction differed in both BEAM and CALADRIO study, this could bias the reported abundances of the gut microbiota profiles. To have addressed the question whether these two methods of DNA extraction would result in different microbiota abundances I should have made a mock community and processed this using the two methods. As I did not complete this in my time undertaking this PhD I cannot conclude with certainty if the methods resulted in different abundances.

Furthermore, in my thesis the Robinson group and I, demonstrated that oral live supplementation of *B. longum* subsp. *longum* seemed to interfere with the effectiveness of cyclophosphamide in luminal B BrCa mice potentially due to *Bifidobacterium*-associated enzymatic activity. Although these are only initial studies, it does raise future considerations of how the microbiota can influence drug metabolism especially in cancer patients known as pharmamicrobiomics. To investigate this further, an *in vitro* model with the isolate and chemotherapeutic could be set up to determine if the isolate can indeed deactivate cyclophosphamide.

Overall, this project has provided key insights into the influence that the gut (and oral) microbiota has on breast cancer. To further build on this work the BEAM study should be continue for researchers to establish a database of faecal samples of first-time diagnosed BrCa patients. The flexibility in sending samples via the post allows the study to be expanded to outside of Norfolk, opening the possibility to profile individuals across the nation. As aforementioned in the

General Introduction, only one observational trial exists looking at the association of BrCa which is situated in Spain. Continuing the BEAM study allows a unique demographic to be profiled and contribute to our understanding how gut microbes can influence health and disease in the specific context of BrCa.

7 Works Cited

1. Pérez-García, J.M., et al., *Pembrolizumab plus eribulin in hormone-receptor-positive, HER2-negative, locally recurrent or metastatic breast cancer (KELLY): An open-label, multicentre, single-arm, phase II trial.* Eur J Cancer, 2021. **148**: p. 382-394.

2. Teng, N.M.Y., et al., *Exploring the impact of gut microbiota and diet on breast cancer risk and progression.* International Journal of Cancer, 2021. **149**(3): p. 494-504.

3. McPherson, K., C.M. Steel, and J.M. Dixon, *ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics.* BMJ (Clinical research ed.), 2000. **321**(7261): p. 624-628.

4. Cancer Research UK. *Breast Cancer Statistics*. 2018 6 August 2018 [cited 2019 21 May]; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer.

5. Foreman, K.J., et al., *Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories.* The Lancet, 2018. **392**(10159): p. 2052-2090.

6. Parkin, D.M., F.I. Bray, and S.S. Devesa, *Cancer burden in the year 2000. The global picture.* European Journal of Cancer, 2001. **37**: p. 4-66.

7. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: A Cancer Journal for Clinicians, 2018. **68**(6): p. 394-424.

8. Bannister, J.B.a.N., *Cancer survival by stage at diagnosis for England (experimental statistics): Adults diagnosed 2012, 2013 and 2014 and followed up to 2015*. 2016, Office for National Statistics.

9. Sun, L., et al., *Global treatment costs of breast cancer by stage: A systematic review.* PLoS One, 2018. **13**(11): p. e0207993.

10. Álvaro-Meca, A., A. Debón, R. Gil Prieto, and Á. Gil de Miguel, *Breast cancer mortality in Spain: Has it really declined for all age groups?* Public Health, 2012. **126**(10): p. 891-895.

11. Byrd, D.A., et al., *Associations of fecal microbial profiles with breast cancer and nonmalignant breast disease in the Ghana Breast Health Study.* International Journal of Cancer. **n/a**(n/a).

12. Forjaz de Lacerda, G., et al., *Breast cancer in Portugal: Temporal trends and age-specific incidence by geographic regions.* Cancer Epidemiology, 2018. **54**: p. 12-18.

13. VANNI, G., et al., *Lockdown of Breast Cancer Screening for COVID-19: Possible Scenario.* In Vivo, 2020. **34**(5): p. 3047-3053.

14. Alagoz, O., et al., *Impact of the COVID-19 Pandemic on Breast Cancer Mortality in the US: Estimates From Collaborative Simulation Modeling.* J Natl Cancer Inst, 2021. **113**(11): p. 1484-1494.

15. Maringe, C., et al., *The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study.* Lancet Oncol, 2020. **21**(8): p. 1023-1034.

16. Boyle, P. and A. Howell, *The globalisation of breast cancer.* Breast Cancer Research, 2010. **12**(4): p. S7.

17. Berger, E., et al., *The impact of lifecourse socio-economic position and individual social mobility on breast cancer risk.* BMC cancer, 2020. **20**(1): p. 1138-1138.

18. dos Santos Silva, I. and V. Beral, *Socioeconomic differences in reproductive behaviour.* IARC Sci Publ, 1997(138): p. 285-308.

19.   Ziegler, R.G., et al., *Migration Patterns and Breast Cancer Risk in Asian-American Women.* JNCI: Journal of the National Cancer Institute, 1993. **85**(22): p. 1819-1827.

20.   Hsieh, C.C., D. Trichopoulos, K. Katsouyanni, and S. Yuasa, *Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study.* Int J Cancer, 1990. **46**(5): p. 796-800.

21.   Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

22.   Cooper, G., *The Development and Causes of Cancer*, in *The Cell: A Molecular Approach*

2000, Sinauer Associates: Sunderland (MA).

23.   Soysal, S.D., A. Tzankov, and S.E. Muenst, *Role of the Tumor Microenvironment in Breast Cancer.* Pathobiology, 2015. **82**(3-4): p. 142-152.

24.   Lu, P., V.M. Weaver, and Z. Werb, *The extracellular matrix: a dynamic niche in cancer progression.* The Journal of cell biology, 2012. **196**(4): p. 395-406.

25.   Danenberg, E., et al., *Breast tumor microenvironment structures are associated with genomic features and clinical outcome.* Nature Genetics, 2022. **54**(5): p. 660-669.

26.   Acerbi, I., et al., *Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration.* Integrative Biology, 2015. **7**(10): p. 1120-1134.

27.   Gudjonsson, T., et al., *Myoepithelial cells: their origin and function in breast morphogenesis and neoplasia.* Journal of mammary gland biology and neoplasia, 2005. **10**(3): p. 261-272.

28.   Gudjonsson, T., et al., *Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition.* J Cell Sci, 2002. **115**(Pt 1): p. 39-50.

29.   Hu, M. and K. Polyak, *Microenvironmental regulation of cancer development.* Current opinion in genetics & development, 2008. **18**(1): p. 27-34.

30.   Sinn, H.P. and H. Kreipe, *A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition.* Breast Care (Basel), 2013. **8**(2): p. 149-54.

31.   Barroso-Sousa, R. and O. Metzger-Filho, *Differences between invasive lobular and invasive ductal carcinoma of the breast: results and therapeutic implications.* Ther Adv Med Oncol, 2016. **8**(4): p. 261-6.

32.   Barroso-Sousa, R. and O. Metzger-Filho, *Differences between invasive lobular and invasive ductal carcinoma of the breast: results and therapeutic implications.* Therapeutic advances in medical oncology, 2016. **8**(4): p. 261-266.

33.   Dai, X., et al., *Breast cancer intrinsic subtype classification, clinical use and future trends.* Am J Cancer Res, 2015. **5**(10): p. 2929-43.

34.   Filippini, S.E. and A. Vega, *Breast cancer genes: beyond BRCA1 and BRCA2.* Front Biosci (Landmark Ed), 2013. **18**(4): p. 1358-72.

35.   Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.

36.   Excellence, N.I.f.H.a.C., *Early and locally advanced breast cancer*
overview in *NICE guideline*. 2021.

37.   Excellence, N.I.f.H.a.C., *Suspected cancer: recognition and referral*, in *NICE Guidline*. 2015, NICE.

38.   Koh, J. and M.J. Kim, *Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage.* Korean J Radiol, 2019. **20**(1): p. 69-82.

39.   Wishart, G.C., et al., *PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer.* Breast Cancer Res, 2010. **12**(1): p. R1.

40.   Excellence, N.I.f.H.a.C., *Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer*, in *Diagnostics Guidance*. 2018.

41.   Wishart, G.C., et al., *PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2.* British journal of cancer, 2012. **107**(5): p. 800-807.

42.   Wishart, G.C., et al., *Inclusion of KI67 significantly improves performance of the PREDICT prognostication and prediction model for early breast cancer.* BMC Cancer, 2014. **14**(1): p. 908.

43.   Down, S.K., O. Lucas, J.R. Benson, and G.C. Wishart, *Effect of PREDICT on chemotherapy/trastuzumab recommendations in HER2-positive patients with early-stage breast cancer.* Oncol Lett, 2014. **8**(6): p. 2757-2761.

44.   McDonald, E.S., et al., *Clinical Diagnosis and Management of Breast Cancer.* Journal of Nuclear Medicine, 2016. **57**(Supplement 1): p. 9S-16S.

45.   Gilman, A. and F.S. Philips, *The Biological Actions and Therapeutic Applications of the B-Chloroethyl Amines and Sulfides.* Science, 1946. **103**(2675): p. 409-36.

46.   Bonadonna, G., et al., *Combination Chemotherapy as an Adjuvant Treatment in Operable Breast Cancer.* New England Journal of Medicine, 1976. **294**(8): p. 405-410.

47.   Verrill, M., *Chemotherapy for early-stage breast cancer: a brief history.* British journal of cancer, 2009. **101 Suppl 1**(Suppl 1): p. S2-S5.

48.   Gross, R. and G. Wulf, *Klinische und experimentelle Erfahrungen mit zyk lischen und nichtzyklischen Phosphamidestern des N-Losl in der Chemotherapie von Tumoren*, in *Strahlentherapie*. 1959.

49.   Arnold, H., F. Bourseaux, and N. Brock, *Chemotherapeutic Action of a Cyclic Nitrogen Mustard Phosphamide Ester (B 518-ASTA) in Experimental Tumours of the Rat.* Nature, 1958. **181**(4613): p. 931-931.

50.   Heidelberger, C., K.C. Leibman, E. Harbers, and P.M. Bhargava, *The comparative utilization of uracil-2-C14 by liver, intestinal mucosa, and Flexner-Jobling carcinoma in the rat.* Cancer Res, 1957. **17**(5): p. 399-404.

51.   Blum, J.L., et al., *Multicenter phase II study of capecitabine in paclitaxel-refractory metastatic breast cancer.* J Clin Oncol, 1999. **17**(2): p. 485-93.

52.   Lomovskaya, N., et al., *Doxorubicin Overproduction in <i>Streptomyces peucetius</i>: Cloning and Characterization of the <i>dnrU</i> Ketoreductase and<i>dnrV</i> Genes and the <i>doxA</i> Cytochrome P-450 Hydroxylase Gene.* Journal of Bacteriology, 1999. **181**(1): p. 305-318.

53.   Arcamone, F., et al., *Adriamycin, 14-hydroxydaimomycin, a new antitumor antibiotic from S. Peucetius var. caesius.* Biotechnology and Bioengineering, 1969. **11**(6): p. 1101-1110.

54.   Clarke, S.J. and L.P. Rivory, *Clinical Pharmacokinetics of Docetaxel.* Clinical Pharmacokinetics, 1999. **36**(2): p. 99-114.

55.   Bonfante, V., G. Bonadonna, F. Villani, and A. Martini, *Preliminary Clinical Experience with 4'-Epidoxorubicin in Advanced Human Neoplasia*, in *Cancer Chemo- and Immunopharmacology: 1. Chemopharmacology*, G. Mathé and F.M. Muggia, Editors. 1980, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 192-199.

56.   Drebin, J.A., V.C. Link, R.A. Weinberg, and M.I. Greene, *Inhibition of tumor growth by a monoclonal antibody reactive with an oncogene-encoded tumor antigen.* Proceedings of the National Academy of Sciences, 1986. **83**(23): p. 9129-9133.

57.   Anampa, J., D. Makower, and J.A. Sparano, *Progress in adjuvant chemotherapy for breast cancer: an overview.* BMC Medicine, 2015. **13**(1): p. 195.

58.   Pagani, O., et al., *Adjuvant exemestane with ovarian suppression in premenopausal breast cancer.* N Engl J Med, 2014. **371**(2): p. 107-18.

59.   (EBCTCG), E.B.C.T.C.G., *Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials.* The Lancet, 2005. **366**(9503): p. 2087-2106.

60.   Komorowski, A.S. and R.C. Pezo, *Untapped "-omics": the microbial metagenome, estrobolome, and their influence on the development of breast cancer and response to treatment.* Breast Cancer Res Treat, 2020. **179**(2): p. 287-300.

61.   Kaminska, M., et al., *Breast cancer risk factors.* Prz Menopauzalny, 2015. **14**(3): p. 196-202.

62.   Castellarin, M., et al., *Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma.* Genome Res, 2012. **22**(2): p. 299-306.

63.   Plummer, M., et al., *Helicobacter pylori cytotoxin-associated genotype and gastric precancerous lesions.* J Natl Cancer Inst, 2007. **99**(17): p. 1328-34.

64.   Ziegler, R.G., et al., *Migration patterns and breast cancer risk in Asian-American women.* J Natl Cancer Inst, 1993. **85**(22): p. 1819-27.

65.   Fernandez, M.F., et al., *Breast Cancer and Its Relationship with the Microbiota.* Int J Environ Res Public Health, 2018. **15**(8).

66.   Saxe, G.A., C.L. Rock, M.S. Wicha, and D. Schottenfeld, *Diet and risk for breast cancer recurrence and survival.* Breast Cancer Res Treat, 1999. **53**(3): p. 241-53.

67.   Xiao, Y., et al., *Associations between dietary patterns and the risk of breast cancer: a systematic review and meta-analysis of observational studies.* Breast Cancer Res, 2019. **21**(1): p. 16.

68.   Picon-Ruiz, M., et al., *Obesity and adverse breast cancer risk and outcome: Mechanistic insights and strategies for intervention.* CA Cancer J Clin, 2017. **67**(5): p. 378-397.

69.   Liedtke, S., et al., *Postmenopausal sex hormones in relation to body fat distribution.* Obesity (Silver Spring), 2012. **20**(5): p. 1088-95.

70.   Seitz, H.K., C. Pelucchi, V. Bagnardi, and C. La Vecchia, *Epidemiology and pathophysiology of alcohol and breast cancer: Update 2012.* Alcohol Alcohol, 2012. **47**(3): p. 204-12.

71.   Liu, Y., N. Nguyen, and G.A. Colditz, *Links between alcohol consumption and breast cancer: a look at the evidence.* Womens Health (Lond), 2015. **11**(1): p. 65-77.

72.   Segovia-Mendoza, M. and J. Morales-Montor, *Immune Tumor Microenvironment in Breast Cancer and the Participation of Estrogen and Its Receptors in Cancer Physiopathology.* Front Immunol, 2019. **10**: p. 348.

73.   Rezaei, N.J., A.M. Bazzazi, and S.A. Naseri Alavi, *Neurotoxicity of the antibiotics: A comprehensive study.* Neurol India, 2018. **66**(6): p. 1732-1740.

74.   Kallala, R., et al., *In vitro and in vivo effects of antibiotics on bone cell metabolism and fracture healing.* Expert Opin Drug Saf, 2012. **11**(1): p. 15-32.

75.   Buchta Rosean, C., et al., *Preexisting Commensal Dysbiosis Is a Host-Intrinsic Regulator of Tissue Inflammation and Tumor Cell Dissemination in Hormone Receptor-Positive Breast Cancer.* Cancer Res, 2019. **79**(14): p. 3662-3675.

76.   Velicer, C.M., et al., *Antibiotic use in relation to the risk of breast cancer.* JAMA, 2004. **291**(7): p. 827-35.

77.   Velicer, C.M., et al., *Association between antibiotic use prior to breast cancer diagnosis and breast tumour characteristics (United States).* Cancer Causes Control, 2006. **17**(3): p. 307-13.

78.   Danneskiold-Samsoe, N.B., et al., *Interplay between food and gut microbiota in health and disease.* Food Res Int, 2019. **115**: p. 23-31.

79.   Fraga, C.G., K.D. Croft, D.O. Kennedy, and F.A. Tomas-Barberan, *The effects of polyphenols and other bioactives on human health.* Food Funct, 2019. **10**(2): p. 514-528.

80.   Sun, H., et al., *The modulatory effect of polyphenols from green tea, oolong tea and black tea on human intestinal microbiota in vitro.* J Food Sci Technol, 2018. **55**(1): p. 399-407.

81.   Ley, R.E., P.J. Turnbaugh, S. Klein, and J.I. Gordon, *Microbial ecology: human gut microbes associated with obesity.* Nature, 2006. **444**(7122): p. 1022-3.

82.   Turnbaugh, P.J., et al., *An obesity-associated gut microbiome with increased capacity for energy harvest.* Nature, 2006. **444**(7122): p. 1027-31.

83.   Mutlu, E.A., et al., *Colonic microbiome is altered in alcoholism.* Am J Physiol Gastrointest Liver Physiol, 2012. **302**(9): p. G966-78.

84.   Bischoff, S.C., et al., *Intestinal permeability--a new target for disease prevention and therapy.* BMC Gastroenterol, 2014. **14**: p. 189.

85.   Shimizu, K., et al., *Normalization of reproductive function in germfree mice following bacterial contamination.* Exp Anim, 1998. **47**(3): p. 151-8.

86.   Kwa, M., C.S. Plottel, M.J. Blaser, and S. Adams, *The Intestinal Microbiome and Estrogen Receptor-Positive Female Breast Cancer.* J Natl Cancer Inst, 2016. **108**(8).

87.   Panda, S., et al., *Short-term effect of antibiotics on human gut microbiota.* PLoS One, 2014. **9**(4): p. e95476.

88.   Thursby, E. and N. Juge, *Introduction to the human gut microbiota.* The Biochemical journal, 2017. **474**(11): p. 1823-1836.

89.   Alexander, J.L., et al., *Gut microbiota modulation of chemotherapy efficacy and toxicity.* Nat Rev Gastroenterol Hepatol, 2017. **14**(6): p. 356-365.

90.   Raza, M.H., et al., *Microbiota in cancer development and treatment.* J Cancer Res Clin Oncol, 2019. **145**(1): p. 49-63.

91.   Roy, S. and G. Trinchieri, *Microbiota: a key orchestrator of cancer therapy.* Nat Rev Cancer, 2017. **17**(5): p. 271-285.

92.   Erdman, S.E. and T. Poutahidis, *Gut microbiota modulate host immune cells in cancer development and growth.* Free Radic Biol Med, 2017. **105**: p. 28-34.

93.   Wen, L. and A. Duffy, *Factors Influencing the Gut Microbiota, Inflammation, and Type 2 Diabetes.* The Journal of Nutrition, 2017. **147**(7): p. 1468S-1475S.

94.   Becattini, S., Y. Taur, and E.G. Pamer, *Antibiotic-Induced Changes in the Intestinal Microbiota and Disease.* Trends in Molecular Medicine, 2016. **22**(6): p. 458-478.

95.   Aminov, R.I., *A brief history of the antibiotic era: lessons learned and challenges for the future.* Frontiers in microbiology, 2010. **1**: p. 134-134.

96.   Ashraf, R. and N.P. Shah, *Immune System Stimulation by Probiotic Microorganisms.* Critical Reviews in Food Science and Nutrition, 2014. **54**(7): p. 938-956.

97.   Taft, D.H., et al., *Bifidobacterial Dominance of the Gut in Early Life and Acquisition of Antimicrobial Resistance.* mSphere, 2018. **3**(5): p. e00441-18.

98.   Derrien, M., F. Turroni, M. Ventura, and D. van Sinderen, *Insights into endogenous Bifidobacterium species in the human gut microbiota during adulthood.* Trends in Microbiology, 2022.

99. Heeney, D.D., M.G. Gareau, and M.L. Marco, *Intestinal Lactobacillus in health and disease, a driver or just along for the ride?* Curr Opin Biotechnol, 2018. **49**: p. 140-147.

100. Bishehsari, F., et al., *Dietary Fiber Treatment Corrects the Composition of Gut Microbiota, Promotes SCFA Production, and Suppresses Colon Carcinogenesis.* Genes, 2018. **9**(2): p. 102.

101. Zhang, M., et al., *Butyrate inhibits interleukin-17 and generates Tregs to ameliorate colorectal colitis in rats.* BMC Gastroenterology, 2016. **16**(1): p. 84.

102. Claesson, M.J., A.G. Clooney, and P.W. O'Toole, *A clinician's guide to microbiome analysis.* Nat Rev Gastroenterol Hepatol, 2017. **14**(10): p. 585-595.

103. Carlos, N., Y.W. Tang, and Z. Pei, *Pearls and pitfalls of genomics-based microbiome analysis.* Emerg Microbes Infect, 2012. **1**(12): p. e45.

104. Nearing, J.T., et al., *Microbiome differential abundance methods produce different results across 38 datasets.* Nature Communications, 2022. **13**(1): p. 342.

105. McMurdie, P.J., *Normalization of Microbiome Profiling Data*, in *Microbiome Analysis: Methods and Protocols*, R.G. Beiko, W. Hsiao, and J. Parkinson, Editors. 2018, Springer New York: New York, NY. p. 143-168.

106. McMurdie, P.J. and S. Holmes, *Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible.* PLOS Computational Biology, 2014. **10**(4): p. e1003531.

107. Thomas, A.M., et al., *Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation.* Nature Medicine, 2019. **25**(4): p. 667-678.

108. Poore, G.D., et al., *Microbiome analyses of blood and tissues suggest cancer diagnostic approach.* Nature, 2020. **579**(7800): p. 567-574.

109. Knight, R., et al., *Best practices for analysing microbiomes.* Nature Reviews Microbiology, 2018. **16**(7): p. 410-422.

110. Whalen, S., J. Schreiber, W.S. Noble, and K.S. Pollard, *Navigating the pitfalls of applying machine learning in genomics.* Nature Reviews Genetics, 2022. **23**(3): p. 169-181.

111. Jackson, M.A., et al., *Gut microbiota associations with common diseases and prescription medications in a population-based cohort.* Nature Communications, 2018. **9**(1): p. 2655.

112. Goedert, J.J., et al., *Investigation of the Association Between the Fecal Microbiota and Breast Cancer in Postmenopausal Women: a Population-Based Case-Control Pilot Study.* JNCI: Journal of the National Cancer Institute, 2015. **107**(8).

113. Goedert, J.J., et al., *Postmenopausal breast cancer and oestrogen associations with the IgA-coated and IgA-noncoated faecal microbiota.* Br J Cancer, 2018. **118**(4): p. 471-479.

114. Zhu, J., et al., *Breast cancer in postmenopausal women is associated with an altered gut metagenome.* Microbiome, 2018. **6**(1): p. 136.

115. Luu, T.H., et al., *Intestinal Proportion of Blautia sp. is Associated with Clinical Stage and Histoprognostic Grade in Patients with Early-Stage Breast Cancer.* Nutrition and Cancer, 2017. **69**(2): p. 267-275.

116. Iida, N., et al., *Commensal Bacteria Control Cancer Response to Therapy by Modulating the Tumor Microenvironment.* Science, 2013. **342**(6161): p. 967-970.

117. Sivan, A., et al., *Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–PD-L1 efficacy.* Science, 2015. **350**(6264): p. 1084-1089.

118. Abul K. Abbas, Andrew H. Lichtman, and S. Pillai, *Cellular and Molecular Immunology*. 9th Edition ed. 1991, Philadelphia: Elsevier - Health Sciences Division.

119. Vétizou, M., et al., *Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota.* Science, 2015. **350**(6264): p. 1079-1084.

120. Lee, S.-H., et al., *Bifidobacterium bifidum strains synergize with immune checkpoint inhibitors to reduce tumour burden in mice.* Nature Microbiology, 2021. **6**(3): p. 277-288.

121. Routy, B., et al., *The gut microbiota influences anticancer immunosurveillance and general health.* Nature Reviews Clinical Oncology, 2018. **15**(6): p. 382-396.

122. Liu, X., et al., *Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome.* Nature Genetics, 2022. **54**(1): p. 52-61.

123. Hezaveh, K., et al., *Tryptophan-derived microbial metabolites activate the aryl hydrocarbon receptor in tumor-associated macrophages to suppress anti-tumor immunity.* Immunity, 2022. **55**(2): p. 324-340.e8.

124. Kovács, T., et al., *Cadaverine, a metabolite of the microbiome, reduces breast cancer aggressiveness through trace amino acid receptors.* Sci Rep, 2019. **9**(1): p. 1300.

125. Miko, E., et al., *Lithocholic acid, a bacterial metabolite reduces breast cancer cell proliferation and aggressiveness.* Biochim Biophys Acta Bioenerg, 2018. **1859**(9): p. 958-974.

126. Mager, L.F., et al., *Microbiome-derived inosine modulates response to checkpoint inhibitor immunotherapy.* Science, 2020. **369**(6510): p. 1481-1489.

127. McKee, A.M., et al., *Antibiotic-induced disturbances of the gut microbiota result in accelerated breast tumor growth.* iScience, 2021. **24**(9).

128. Buchta Rosean, C., et al., *Preexisting Commensal Dysbiosis Is a Host-Intrinsic Regulator of Tissue Inflammation and Tumor Cell Dissemination in Hormone Receptor–Positive Breast Cancer.* Cancer Research, 2019. **79**(14): p. 3662-3675.

129. Kirkup, B., et al., *Perturbation of the gut microbiota by antibiotics results in accelerated breast tumour growth and metabolic dysregulation.* bioRxiv, 2019: p. 553602.

130. Buchta Rosean, C., et al., *Pre-existing commensal dysbiosis is a host-intrinsic regulator of tissue inflammation and tumor cell dissemination in hormone receptor-positive breast cancer.* Cancer Research, 2019: p. canres.3464.2018.

131. Kirkup, B.M., et al., *Antibiotic-induced disturbances of the gut microbiota result in accelerated breast tumour growth via a mast cell-dependent pathway.* bioRxiv, 2020: p. 2020.03.07.982108.

132. Davar, D., et al., *Fecal microbiota transplant overcomes resistance to anti–PD-1 therapy in melanoma patients.* Science, 2021. **371**(6529): p. 595-602.

133. Baruch, E.N., et al., *Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients.* Science, 2021. **371**(6529): p. 602-609.

134. Plaza-Díaz, J., et al., *Association of breast and gut microbiota dysbiosis and the risk of breast cancer: a case-control clinical study.* BMC cancer, 2019. **19**(1): p. 495-495.

135. Bryukhanov, A.L. and A.I. Netrusov, *Long-term storage of obligate anaerobic microorganisms in glycerol.* Applied Biochemistry and Microbiology, 2006. **42**(2): p. 177-180.

136. Prakash, O., Y. Nimonkar, and Y.S. Shouche, *Practice and prospects of microbial preservation.* FEMS Microbiology Letters, 2013. **339**(1): p. 1-9.

137. Alcon-Giner, C., et al., *Optimisation of 16S rRNA gut microbiota profiling of extremely low birth weight infants.* BMC genomics, 2017. **18**(1): p. 841-841.

138.   Weisburg, W.G., S.M. Barns, D.A. Pelletier, and D.J. Lane, *16S ribosomal DNA amplification for phylogenetic study.* Journal of bacteriology, 1991. **173**(2): p. 697-703.

139.   Weisburg, W.G., S.M. Barns, D.A. Pelletier, and D.J. Lane, *16S ribosomal DNA amplification for phylogenetic study.* J Bacteriol, 1991. **173**(2): p. 697-703.

140.   Baker, D.J., et al., *CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes.* Genome Medicine, 2021. **13**(1): p. 21.

141.   Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis, *PEAR: a fast and accurate Illumina Paired-End reAd mergeR.* Bioinformatics, 2014. **30**(5): p. 614-20.

142.   Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-2461.

143.   Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic acids research, 2013. **41**(Database issue): p. D590-D596.

144.   Huson, D.H., et al., *MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data.* PLoS Comput Biol, 2016. **12**(6): p. e1004957.

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

146.   Chen, S., Y. Zhou, Y. Chen, and J. Gu, *fastp: an ultra-fast all-in-one FASTQ preprocessor.* Bioinformatics, 2018. **34**(17): p. i884-i890.

147.   Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* Journal of computational biology : a journal of computational molecular cell biology, 2012. **19**(5): p. 455-477.

148.   Kiu, R., *BACTspeciesID: identify microbial species and genome contamination using 16S rRNA gene approach*. 2020, GitHub

149.   Jain, C., et al., *High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.* Nature Communications, 2018. **9**(1): p. 5114.

150.   Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.

151.   Ryan Wick and P. Menzel, *Filtlong: a tool for filtering long reads by quality*. 2021, GitHub

152.   Kolmogorov, M., J. Yuan, Y. Lin, and P.A. Pevzner, *Assembly of long, error-prone reads using repeat graphs.* Nature Biotechnology, 2019. **37**(5): p. 540-546.

153.   Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature Methods, 2012. **9**(4): p. 357-359.

154.   Beghini, F., et al., *Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3.* Elife, 2021. **10**.

155.   Li, D., et al., *MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.* Bioinformatics, 2015. **31**(10): p. 1674-6.

156.   Uritskiy, G.V., J. DiRuggiero, and J. Taylor, *MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis.* Microbiome, 2018. **6**(1): p. 158.

157.   Wood, D.E., J. Lu, and B. Langmead, *Improved metagenomic analysis with Kraken 2.* Genome Biology, 2019. **20**(1): p. 257.

158.   Lu, J., F.P. Breitwieser, P. Thielen, and S.L. Salzberg, *Bracken: estimating species abundance in metagenomics data.* PeerJ Computer Science, 2017. **3**: p. e104.

159.   Parks, D.H., et al., *CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.* Genome Research, 2015. **25**(7): p. 1043-1055.

160.   Finn, R.D., J. Clements, and S.R. Eddy, *HMMER web server: interactive sequence similarity searching.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W29-37.

161.   Hyatt, D., P.F. LoCascio, L.J. Hauser, and E.C. Uberbacher, *Gene and translation initiation site prediction in metagenomic sequences.* Bioinformatics, 2012. **28**(17): p. 2223-2230.

162.   Wu, Y.-W., B.A. Simmons, and S.W. Singer, *MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets.* Bioinformatics, 2015. **32**(4): p. 605-607.

163.   Kang, D.D., et al., *MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies.* PeerJ, 2019. **7**: p. e7359.

164.   Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nature Methods, 2017. **14**(4): p. 417-419.

165.   Chaumeil, P.-A., A.J. Mussig, P. Hugenholtz, and D.H. Parks, *GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.* Bioinformatics, 2019. **36**(6): p. 1925-1927.

166.   Hunt, M., et al., *Circlator: automated circularization of genome assemblies using long sequencing reads.* Genome Biol, 2015. **16**: p. 294.

167.   Browne, H.P., et al., *Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation.* Nature, 2016. **533**(7604): p. 543-546.

168.   Junick, J. and M. Blaut, *Quantification of human fecal bifidobacterium species by use of quantitative real-time PCR analysis targeting the groEL gene.* Appl Environ Microbiol, 2012. **78**(8): p. 2613-22.

169.   Parida, S. and D. Sharma, *The power of small changes: Comprehensive analyses of microbial dysbiosis in breast cancer.* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 2019. **1871**(2): p. 392-405.

170.   Wang, W.-L., et al., *Application of metagenomics in the human gut microbiome.* World journal of gastroenterology, 2015. **21**(3): p. 803-814.

171.   Aguirre, M., et al., *Evaluation of an optimal preparation of human standardized fecal inocula for in vitro fermentation studies.* Journal of Microbiological Methods, 2015. **117**: p. 78-84.

172.   Prakash, O., Y. Nimonkar, and D. Desai, *A Recent Overview of Microbes and Microbiome Preservation.* Indian Journal of Microbiology, 2020. **60**(3): p. 297-309.

173.   Adlung, L., E. Elinav, T.F. Greten, and F. Korangy, *Microbiome genomics for cancer prediction.* Nature Cancer, 2020. **1**(4): p. 379-381.

174.   Kroemer, G. and L. Zitvogel, *The breakthrough of the microbiota.* Nature Reviews Immunology, 2018. **18**: p. 87.

175.   Kamińska, M., et al., *Breast cancer risk factors.* Przeglad menopauzalny = Menopause review, 2015. **14**(3): p. 196-202.

176.   Zaura, E., et al., *Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces.* mBio, 2015. **6**(6): p. e01693-15.

177.   Mas-Lloret, J., et al., *Gut microbiome diversity detected by high-coverage 16S and shotgun sequencing of paired stool and colon sample.* Scientific Data, 2020. **7**(1): p. 92.

178.   Martinez Arbizu, P., *pairwiseAdonis: Pairwise multilevel comparison using adonis.* 2020: R package version 0.4.

179. Forster, S.C., et al., *A human gut bacterial genome and culture collection for improved metagenomic analyses.* Nature Biotechnology, 2019. **37**(2): p. 186-192.

180. Hitch, T.C.A., et al., *Automated analysis of genomic sequences facilitates high-throughput and comprehensive description of bacteria.* ISME Communications, 2021. **1**(1): p. 16.

181. Parte, A.C., et al., *List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ.* Int J Syst Evol Microbiol, 2020. **70**(11): p. 5607-5612.

182. Edgar, R.C., *MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping.* bioRxiv, 2021: p. 2021.06.20.449169.

183. Minh, B.Q., et al., *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.* Molecular Biology and Evolution, 2020. **37**(5): p. 1530-1534.

184. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v4: recent updates and new developments.* Nucleic Acids Res, 2019. **47**(W1): p. W256-w259.

185. Ndongo, S., S. Khelaifia, P.E. Fournier, and D. Raoult, *"Massiliomicrobiota timonensis," a new bacterial species isolated from the human gut.* New Microbes and New Infections, 2016. **13**: p. 25-26.

186. Meier-Kolthoff, J.P., J.S. Carbasse, R.L. Peinado-Olarte, and M. Göker, *TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes.* Nucleic Acids Research, 2021. **50**(D1): p. D801-D807.

187. Kim, D., S. Park, and J. Chun, *Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity.* Journal of Microbiology, 2021. **59**(5): p. 476-480.

188. Miller, L.T., *Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids.* J Clin Microbiol, 1982. **16**(3): p. 584-6.

189. Yoon, S.H., et al., *A large-scale evaluation of algorithms to calculate average nucleotide identity.* Antonie Van Leeuwenhoek, 2017. **110**(10): p. 1281-1286.

190. Shkoporov, A.N., et al., *Coprobacter fastidiosus gen. nov., sp. nov., a novel member of the family Porphyromonadaceae isolated from infant faeces.* Int J Syst Evol Microbiol, 2013. **63**(Pt 11): p. 4181-4188.

191. Shkoporov, A.N., et al., *Alistipes inops sp. nov. and Coprobacter secundus sp. nov., isolated from human faeces.* Int J Syst Evol Microbiol, 2015. **65**(12): p. 4580-4588.

192. Association, B.M., *The impact of the pandemic on population health and health inequalities*, in *BMA COVID Review* 2022, BMA: UK.

193. Cheng, X., et al., *Meta-analysis of 16S rRNA microbial data identified alterations of the gut microbiota in COVID-19 patients during the acute and recovery phases.* BMC Microbiol, 2022. **22**(1): p. 274.

194. Wang, M., et al., *The relationship between gut microbiota and COVID-19 progression: new insights into immunopathogenesis and treatment.* Front Immunol, 2023. **14**: p. 1180336.

195. Durazzi, F., et al., *Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota.* Scientific Reports, 2021. **11**(1): p. 3030.

196. Gingell, R., J.W. Bridges, and R.T. Williams, *Gut flora and the metabolism of prontosils in the rat.* Biochem J, 1969. **114**(1): p. 5p-6p.

197. Chrysostomou, D., L.A. Roberts, J.R. Marchesi, and J.M. Kinross, *Gut Microbiota Modulation of Efficacy and Toxicity of Cancer Chemotherapy and Immunotherapy.* Gastroenterology, 2023. **164**(2): p. 198-213.

198.   Martinez-Porchas, M., E. Villalpando-Canchola, L.E. Ortiz Suarez, and F. Vargas-Albores, *How conserved are the conserved 16S-rRNA regions?* PeerJ, 2017. **5**: p. e3036.

199.   Duncan, S.H., et al., *Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov.* Int J Syst Evol Microbiol, 2002. **52**(Pt 6): p. 2141-2146.

200.   Derrien, M., E.E. Vaughan, C.M. Plugge, and W.M. de Vos, *Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium.* Int J Syst Evol Microbiol, 2004. **54**(Pt 5): p. 1469-1476.

201.   Rainey, F.A., *Ruminococcaceae fam. nov*, in *Bergey's Manual of Systematics of Archaea and Bacteria*. p. 1-2.

202.   Feehily, C., et al., *Detailed mapping of Bifidobacterium strain transmission from mother to infant via a dual culture-based and metagenomic approach.* Nature Communications, 2023. **14**(1): p. 3015.

203.   Hilton, S.K., et al., *Metataxonomic and Metagenomic Approaches vs. Culture-Based Techniques for Clinical Pathology.* Front Microbiol, 2016. **7**: p. 484.

204.   Turner, N.C., P. Neven, S. Loibl, and F. Andre, *Advances in the treatment of advanced oestrogen-receptor-positive breast cancer.* The Lancet, 2017. **389**(10087): p. 2403-2414.

205.   Rugo, H.S., et al., *Safety and Antitumor Activity of Pembrolizumab in Patients with Estrogen Receptor–Positive/Human Epidermal Growth Factor Receptor 2–Negative Advanced Breast Cancer.* Clinical Cancer Research, 2018. **24**(12): p. 2804-2811.

206.   Cortes, J., et al., *Eribulin monotherapy versus treatment of physician's choice in patients with metastatic breast cancer (EMBRACE): a phase 3 open-label randomised study.* The Lancet, 2011. **377**(9769): p. 914-923.

207.   Tanoue, T., et al., *A defined commensal consortium elicits CD8 T cells and anti-cancer immunity.* Nature, 2019. **565**(7741): p. 600-605.

208.   Viaud, S., et al., *The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide.* Science, 2013. **342**(6161): p. 971-6.

209.   Gopalakrishnan, V., et al., *Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients.* Science, 2018. **359**(6371): p. 97-103.

210.   Iida, N., et al., *Gut commensal bacteria promote anti-tumor innate immune responses in distant tumors after immunotherapy and chemotherapy (162.5).* The Journal of Immunology, 2012. **188**(1 Supplement): p. 162.5-162.5.

211.   Routy, B., et al., *Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors.* Science, 2018. **359**(6371): p. 91-97.

212.   Burugu, S., K. Asleh-Aburaya, and T.O. Nielsen, *Immune infiltrates in the breast cancer microenvironment: detection, characterization and clinical implication.* Breast Cancer, 2017. **24**(1): p. 3-15.

213.   Kitamoto, S., et al., *The Bacterial Connection between the Oral Cavity and the Gut Diseases.* J Dent Res, 2020. **99**(9): p. 1021-1029.

214.   Seymour, L., et al., *iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics.* The Lancet Oncology, 2017. **18**(3): p. e143-e152.

215.   Libby, E.K., et al., *Atopobium vaginae triggers an innate immune response in an in vitro model of bacterial vaginosis.* Microbes Infect, 2008. **10**(4): p. 439-46.

216.   Li, Y., et al., *Composition and function of oral microbiota between gingival squamous cell carcinoma and periodontitis.* Oral Oncology, 2020. **107**: p. 104710.

217.   Dybdal-Hargreaves, N.F., A.L. Risinger, and S.L. Mooberry, *Eribulin mesylate: mechanism of action of a unique microtubule-targeting agent.* Clin Cancer Res, 2015. **21**(11): p. 2445-52.

218.   Swami, U., I. Chaudhary, M.H. Ghalib, and S. Goel, *Eribulin -- a review of preclinical and clinical studies.* Crit Rev Oncol Hematol, 2012. **81**(2): p. 163-84.

219.   Zhang, Z.Y., B.M. King, R.D. Pelletier, and Y.N. Wong, *Delineation of the interactions between the chemotherapeutic agent eribulin mesylate (E7389) and human CYP3A4.* Cancer Chemother Pharmacol, 2008. **62**(4): p. 707-16.

220.   Segata, N., et al., *Metagenomic biomarker discovery and explanation.* Genome Biol, 2011. **12**(6): p. R60.

221.   Mazmanian, S.K., C.H. Liu, A.O. Tzianabos, and D.L. Kasper, *An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system.* Cell, 2005. **122**(1): p. 107-18.

222.   Dasgupta, S., et al., *Plasmacytoid dendritic cells mediate anti-inflammatory responses to a gut commensal molecule via both innate and adaptive mechanisms.* Cell Host Microbe, 2014. **15**(4): p. 413-23.

223.   Seemann, T., *ABRicate*. 2020, Github: https://github.com/tseemann/abricate.

224.   Zankari, E., et al., *Identification of acquired antimicrobial resistance genes.* J Antimicrob Chemother, 2012. **67**(11): p. 2640-4.

225.   Barton, L.L., N.L. Ritz, G.D. Fauque, and H.C. Lin, *Sulfur Cycling and the Intestinal Microbiome.* Digestive Diseases and Sciences, 2017. **62**(9): p. 2241-2257.

226.   Krinos, C.M., et al., *Extensive surface diversity of a commensal microorganism by multiple DNA inversions.* Nature, 2001. **414**(6863): p. 555-558.

227.   Parida, S., et al., *A pro-carcinogenic colon microbe promotes breast tumorigenesis and metastatic progression and concomitantly activates Notch and ßcatenin axes.* Cancer Discovery, 2021: p. CD-20-0537.

228.   Zamani, S., et al., *Enterotoxigenic Bacteroides fragilis: A Possible Etiological Candidate for Bacterially-Induced Colorectal Precancerous and Cancerous Lesions.* Front Cell Infect Microbiol, 2019. **9**: p. 449.

229.   Tierney, D., S.D. Copsey, T. Morris, and J.D. Perry, *A new chromogenic medium for isolation of Bacteroides fragilis suitable for screening for strains with antimicrobial resistance.* Anaerobe, 2016. **39**: p. 168-172.

230.   MOORE, W.E.C., E.P. CATO, and L.V. HOLDEMAN, *Eubacterium lentum (Eggerth) Prévot 1938: Emendation of Description and Designation of the Neotype Strain.* International Journal of Systematic and Evolutionary Microbiology, 1971. **21**(4): p. 299-303.

231.   Soule, H.D., et al., *A human cell line from a pleural effusion derived from a breast carcinoma.* J Natl Cancer Inst, 1973. **51**(5): p. 1409-16.

232.   Latorre-Pérez, A., et al., *The Spanish gut microbiome reveals links between microorganisms and Mediterranean diet.* Scientific Reports, 2021. **11**(1): p. 21602.

233.   Han, Y.W., *Fusobacterium nucleatum: a commensal-turned pathogen.* Curr Opin Microbiol, 2015. **23**: p. 141-7.

234.   Kostic, A.D., et al., *Genomic analysis identifies association of Fusobacterium with colorectal carcinoma.* Genome Res, 2012. **22**(2): p. 292-8.

235.   Parhi, L., et al., *Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression.* Nat Commun, 2020. **11**(1): p. 3259.

236.   Sulyanto, R.M., et al., *The Predominant Oral Microbiota Is Acquired Early in an Organized Pattern.* Scientific Reports, 2019. **9**(1): p. 10550.

237. Kawasaki, M., et al., *Oral infectious bacteria in dental plaque and saliva as risk factors in patients with esophageal cancer.* Cancer, 2021. **127**(4): p. 512-519.

238. Mukherjee, P.K., et al., *Bacteriome and mycobiome associations in oral tongue cancer.* Oncotarget, 2017. **8**(57): p. 97273-97289.

239. Zhou, C.-B., et al., *Fecal Signatures of Streptococcus anginosus and Streptococcus constellatus for Noninvasive Screening and Early Warning of Gastric Cancer.* Gastroenterology, 2022. **162**(7): p. 1933-1947.e18.

240. Perdigon, G., et al., *Enhancement of Immune Response in Mice Fed with Streptococcus thermophilus and Lactobacillus acidophilus.* Journal of Dairy Science, 1987. **70**(5): p. 919-926.

241. Kaci, G., et al., *Anti-Inflammatory Properties of Streptococcus salivarius, a Commensal Bacterium of the Oral Cavity and Digestive Tract.* Applied and Environmental Microbiology, 2014. **80**(3): p. 928-934.

242. Chan, J.L., et al., *Non-toxigenic Bacteroides fragilis (NTBF) administration reduces bacteria-driven chronic colitis and tumor development independent of polysaccharide A.* Mucosal Immunol, 2019. **12**(1): p. 164-177.

243. Neff, C.P., et al., *Diverse Intestinal Bacteria Contain Putative Zwitterionic Capsular Polysaccharides with Anti-inflammatory Properties.* Cell Host Microbe, 2016. **20**(4): p. 535-547.

244. Gul, L., et al., *Extracellular vesicles produced by the human commensal gut bacterium Bacteroides thetaiotaomicron affect host immune pathways in a cell-type specific manner that are altered in inflammatory bowel disease.* Journal of Extracellular Vesicles, 2022. **11**(1): p. e12189.

245. Chen, Y., et al., *Butyrate from pectin fermentation inhibits intestinal cholesterol absorption and attenuates atherosclerosis in apolipoprotein E-deficient mice.* The Journal of Nutritional Biochemistry, 2018. **56**: p. 175-182.

246. Corander, J., W.P. Hanage, and J. Pensar, *Causal discovery for the microbiome.* The Lancet Microbe, 2022. **3**(11): p. e881-e887.

247. Alcon-Giner, C., et al., *Microbiota Supplementation with Bifidobacterium and Lactobacillus Modifies the Preterm Infant Gut Microbiota and Metabolome: An Observational Study.* Cell Rep Med, 2020. **1**(5): p. 100077.

248. He, F., et al., *Stimulation of the Secretion of Pro-Inflammatory Cytokines by Bifidobacterium Strains.* Microbiology and Immunology, 2002. **46**(11): p. 781-785.

249. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography.* Nature, 2012. **486**(7402): p. 222-7.

250. Derrien, M., F. Turroni, M. Ventura, and D. van Sinderen, *Insights into endogenous Bifidobacterium species in the human gut microbiota during adulthood.* Trends in Microbiology, 2022. **30**(10): p. 940-947.

251. O'Neill, I., Z. Schofield, and L.J. Hall, *Exploring the role of the microbiota member Bifidobacterium in modulating immune-linked diseases.* Emerg Top Life Sci, 2017. **1**(4): p. 333-349.

252. Lugli, G.A., et al., *Comparative genomic and phylogenomic analyses of the Bifidobacteriaceae family.* BMC Genomics, 2017. **18**(1): p. 568.

253. Milani, C., et al., *Genomics of the Genus Bifidobacterium Reveals Species-Specific Adaptation to the Glycan-Rich Gut Environment.* Appl Environ Microbiol, 2016. **82**(4): p. 980-991.

254. Miller, L.E., L. Lehtoranta, and M.J. Lehtinen, *The Effect of Bifidobacterium animalis ssp. lactis HN019 on Cellular Immune Function in Healthy Elderly Subjects: Systematic Review and Meta-Analysis.* Nutrients, 2017. **9**(3).

255. Dizman, N., et al., *Nivolumab plus ipilimumab with or without live bacterial supplementation in metastatic renal cell carcinoma: a randomized phase 1 trial.* Nature Medicine, 2022. **28**(4): p. 704-712.

256. Francisco-Anderson, L., et al., *Abstract PS11-27: A phase I/II clinical trial of EDP1503 with pembrolizumab for triple-negative breast cancer.* Cancer Research, 2021. **81**(4_Supplement): p. PS11-27-PS11-27.

257. Tomayko, M.M. and C.P. Reynolds, *Determination of subcutaneous tumor size in athymic (nude) mice.* Cancer Chemother Pharmacol, 1989. **24**(3): p. 148-54.

258. Cancer, I.A.f.R.o., *Cyclophosphamide*, in *IARC Working Group on the Evaluation of Carcinogenic Risks to Humans* 2012: Lyon, France.

259. Hall, A.G. and M.J. Tilby, *Mechanisms of action of, and modes of resistance to, alkylating agents used in the treatment of haematological malignancies.* Blood Reviews, 1992. **6**(3): p. 163-173.

260. Emadi, A., R.J. Jones, and R.A. Brodsky, *Cyclophosphamide and cancer: golden anniversary.* Nature Reviews Clinical Oncology, 2009. **6**(11): p. 638-647.

261. Kiu, R., et al., *Bifidobacterium breve UCC2003 Induces a Distinct Global Transcriptomic Program in Neonatal Murine Intestinal Epithelial Cells.* iScience, 2020. **23**(7): p. 101336.

262. Püngel, D., et al., *Bifidobacterium breve UCC2003 Exopolysaccharide Modulates the Early Life Microbiota by Acting as a Potential Dietary Substrate.* Nutrients, 2020. **12**(4).

263. Geller, L.T., et al., *Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine.* Science, 2017. **357**(6356): p. 1156-1160.

264. Shi, Y., et al., *Intratumoral accumulation of gut microbiota facilitates CD47-based immunotherapy via STING signaling.* J Exp Med, 2020. **217**(5).

265. Devika, N.T. and K. Raman, *Deciphering the metabolic capabilities of Bifidobacteria using genome-scale metabolic models.* Scientific Reports, 2019. **9**(1): p. 18222.

266. Spanogiannopoulos, P., et al., *Host and gut bacteria share metabolic pathways for anti-cancer drug metabolism.* Nat Microbiol, 2022. **7**(10): p. 1605-1620.

267. Sistigu, A., et al., *Immunomodulatory effects of cyclophosphamide and implementations for vaccine design.* Seminars in Immunopathology, 2011. **33**(4): p. 369-383.

268. Candeliere, F., et al., *β-Glucuronidase Pattern Predicted From Gut Metagenomes Indicates Potentially Diversified Pharmacomicrobiomics.* Front Microbiol, 2022. **13**: p. 826994.

269. Bhatt, A.P., et al., *Targeted inhibition of gut bacterial β-glucuronidase activity enhances anticancer drug efficacy.* Proc Natl Acad Sci U S A, 2020. **117**(13): p. 7374-7381.

8 Appendices

8.1 BEAM Human Tissue Protocol

Breast hEalth And gut Microbiota Study (BEAM)

Short title: The role of the gut microbiota in breast health

PROTOCOL:

Version 3

QIB Investigator (s):

[Nancy Teng]

[Ibrahim Sallam]

NHS Investigator(s):

[Dr. Simon Pain, NNUH]

[Dr. Leeper (Sandy) Alexander, JPUH]

Other Investigator(s):

[Dr. Lindsay Hall and Dr. Stephen Robinson]

List of contents [To create an automatic Table of contents];

Executive Summary

1.   Study Design

1.1.   Recruitment policy

We will be recruiting patients at the Norfolk and Norwich University Hospital (NNUH) and James Paget University Hospital (JPUH). The main investigator at NNUH is Dr. Simon Pain and for JPUH it is Dr. Leeper (Sandy) Alexander. For the pathway at NNUH, we will be aiming to recruit patients that have been invited to attend the breast care assessment clinic at the Hospital. These patients will be of various ages, but all have a similar reason why they are attending the clinic i.e. undergoing a breast health check. After discussions with the breast care team we have decided to have three pathways for three cohorts essentially. One pathway will capture those patients who have been diagnosed with no breast cancer i.e. they have a clear mammogram. Another pathway will recruit patients who have been diagnosed with breast cancer but will not undergo neoadjuvant treatment. The last pathway will recruit patients who have been diagnosed with breast cancer and will undergo neoadjuvant therapy. These pathways are explained in more detail below. We will collect faecal samples at various stages throughout their treatment at timely intervals. We will also collect blood and tissue samples however this is dependent if blood is already taken as part of their care or the size of the resected tissue.

We wish to recruit patients who have been newly diagnosed with invasive breast cancer. The consultants and nurses of the breast care team at the Hospital will identify potential patients and provide the PIS. Consequently, a member of staff from the Biorepository that can consent patients will intercept the patient at the next visit to the Hospital, as they would need to come in before their operation and consent them into the study if they are interested.

For JPUH the recruitment pathway is described as the following. Eligible patients (disease cohort: hormone receptor +/ HER2 -, control cohort: benign tumour) will be screened at the MDT meetings held at JPUH with the relevant clinical team. Eligible patients, both disease and control cohort will receive the patient information sheet (PIS), when they come into the clinic for their diagnosis. For control patients they will be informed about the study and consent will be taken if they agree. If consented the patient will receive a faecal collection kit to bring home to provide a faecal sample. For the disease cohort, the patient will be given the PIS and informed about the study. At JPUH, patients are invited to come in prior to surgery for a procedure involving MagTrace. At this appointment consent will be taken by an appointed research nurse or the appointed clinical fellow. After consent has been taken

the patient will receive a faecal kit. This will be the first baseline sample; consecutive samples will be taken at 6-months after diagnosis and 1-year after diagnosis. These time-points are pre-established and make up part of the routine clinical care at the JPUH. As such the patients will be informed about continuing participation in this study prior to be given a faecal kit to bring home. Anonymity will be done by the person who has taken consent i.e. the appointed nurse, clinical fellow or clinician.

At the JPUH site we will only be collecting faecal samples. The patient will receive faecal kit, explained below, to donate a faecal sample at home that can be posted safely to the Quadram Institute (QIB). At QIB an appointed clinical fellow or researcher will aliquot the samples and transfer them to the NRP Biorepository.

1.2.   Eligibility criteria for NNUH

    1.2.1.   Basic inclusion criteria

- We will be recruiting patients both male and older than 30yo who have been invited to come to the Hospital for a breast screening or examination.

   1.2.2.   Basic exclusion criteria

- Any patient who has had any antibiotics of any kind 6 months before being recruited into the study will be ineligible to participate.
- Any patient who has had cancer of any kind in the past will also be ineligible to participate in the study.
- Any patient below the age of 30.

1.3.   Eligibility criteria for JPUH

    1.3.1.   Basic inclusion criteria

- We will be recruiting patients both male and older than 30yo who have been invited to come to the Hospital for a breast screening or examination.
- First time diagnosis of low-grade HR+/HER2- breast cancer

    1.3.2.   Basic exclusion criteria

- Any patient who has had any antibiotics of any kind 6 months before being recruited into the study will be ineligible to participate.
- Any patient who has had cancer of any kind in the past will also be ineligible to participate in the study.

2.   Sample collection and Handling

We will be providing participants with a faecal collection kit. This is a kit that includes FeColl paper, simply put it is reinforced toilet paper, that can hold a faecal sample. The kit will also include gloves and a 20mL universal with a spoon lid. This allows the patient to collect a sample easily without potentially contaminating the sample either. Afterwards the universal is closed and wrapped with an absorbent sheet and enclosed in a biohazard bag. All these items are placed in a bio- safety box provided by the Royal Mail, which are specifically designed to transport biological samples. Included is an ice pack that fits within the bio-safety box which ensures that the sample stays cold for ~6hours. The participant will also have to fill in a sample record sheet. These bio-safety boxes have been prepaid with the QIB address already attached. We would like to request a minimum of 4 faecal samples as they allow us to gain an overview of how the microbiota profile changes according to their clinical care. However, we would like to include up to 2 optional samples after their final 4th sample. This is to gain a microbiota profile once their microbiota has stabilised again.

Once posted, the safety box will arrive at QIB reception. As this is still contained (in the bio-safety box and secondary contained in a sealed plastic bag with an absorbent sheet), this follows health and safety regulations. The faecal sample is then processed in the labs in the microbiology room and any leftover sample is disposed of accordingly.

JPUH will only be used for the recruitment of patients to donate faecal samples. The pathway is described previously and illustrated briefly below.
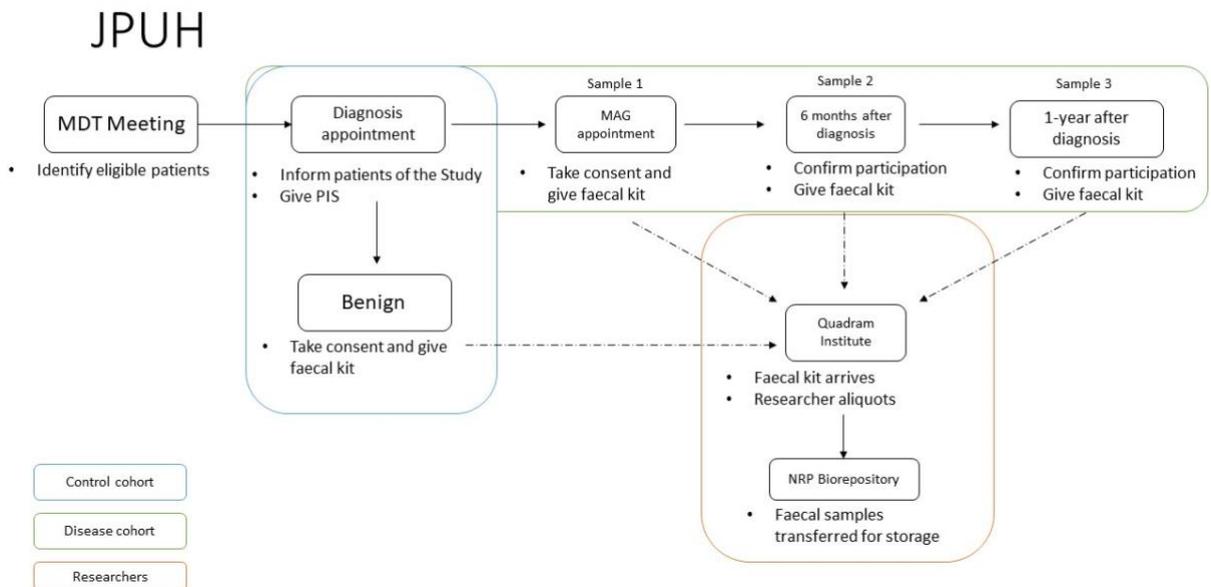
NNUH will be a site to recruit for bloods and tissue. For bloods, the patient may need to have blood drawn before their procedure. This may be done at the breast care clinic or at the day procedure unit (DPU) at NNUH. The staff members of these units have been made aware of the study and have agreed that should a patient have blood drawn as part of their care, they would include an extra vial specific for the BEAM study. The lead contact for the breast care unit is Breast Care nurse: Lynne Priestley. The lead contact for the DPU is Deputy nurse: Tracy Parker. If the patient will not have blood taken but they would like to donate blood to the study a trained and certified phlebotomist will take the sample, this may be a member of staff at the Hospital or someone from the Biorepository.

For tissue, Dr Simon Pain (consultant of the breast care team at NNUH) has said that the consultants and surgeons would be willing to donate resected tissue to the study should the tumour be larger than 15mm in size. This is to preserve the integrity

patient care. If fresh tissue will be collected the Biorepository will be on stand-by to collect and process accordingly. Fixed tissue we will obtain by requesting the block sample from the Cotman centre.

For patient recruitment we will be using one of the three pathways below dependent on the patient's care they will receive.
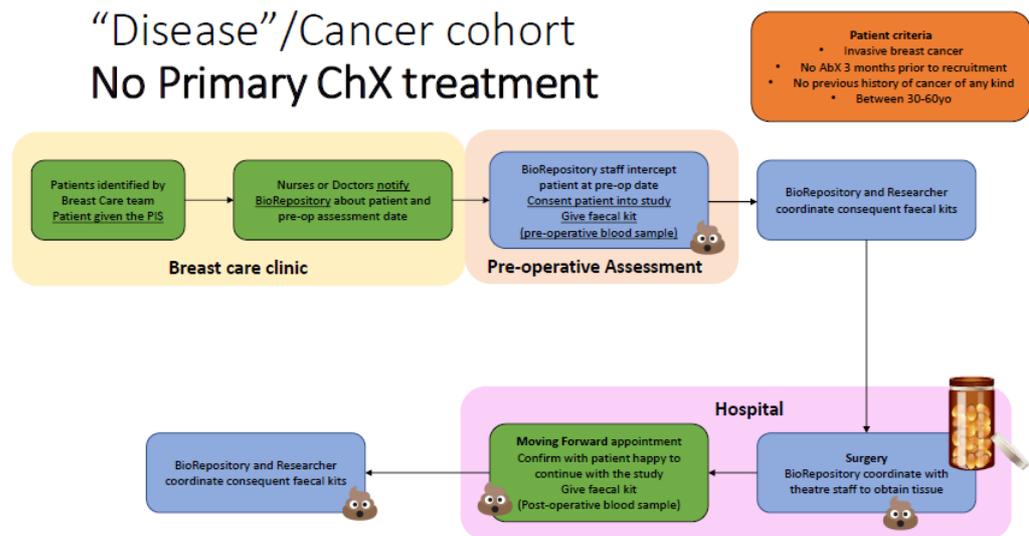

Disease cohort: No primary chemotherapy. Some patients who have been diagnosed with breast cancer will not undergo primary chemotherapy prior to surgery. To consent these patients, the breast care team at the NNUH will identify potential patients and provide them with a PIS (annex 2) if interested. If they are interested BioRepository will be notified of the patient and their pre-operative assessment date. Consequently, a member of staff of the BioRepository will intercept and consent the patient into the study at their pre-operative assessment. Once consented they will be given their first faecal sample kit to collect the sample at home and send it back to the QIB. In addition to this we will request a vial of blood to be donated to the study if the patient will have blood drawn for their appointment or a voluntary donation by the patient. The following faecal kits will be sent to patient's home addresses for them to collect their sample and send it back to the QIB. On the day of surgery, the BioRepository will coordinate with the surgical team to receive resected tissue. This donation is dependent on the size of the tumour i.e. larger than 15mm, at which point donating some tissue will not compromise the care of the patient. 6 months after initial diagnosis the patient will have a "moving forward" appointment at the breast clinic. The BioRepository will intercept the patient to confirm they are happy to

continue giving samples. If so, they will provide another faecal kit and also request a vial of blood. Their last faecal sample will be sent to their home address again.

Disease cohort: Primary chemotherapy. Patients will be identified by the breast care team. If interested they will be given the PIS. These patients will undergo neoadjuvant treatment i.e. primary chemotherapy. Their next appointment at the Hospital will be with an oncologist. The breast care team will notify the BioRepository who will intercept the patient at their oncology appointment. Here the patients will be consented into the study as previously. Once consented they will be provided with a faecal kit. The rest of the pathway follows the same way as previously described with "Disease cohort: no primary chemotherapy".

Control cohort. For our study to be valid we require a control cohort. It has been decided to do this the following way. Patients at the one-stop assessment clinic usually get a mammogram done. However due to limited resources, not everyone gets screened immediately. Some of these patients will be asked to come back later. These patients, if they fit the criteria, will receive the PIS. The breast care team will notify the BioRepository. The BioRepository will intercept the patient at their next mammogram screening and consent them into the study. Alternatively, we may consent patients once they've been given their diagnosis by the radiologist of having no breast cancer. If they provide consent, they will be provided with a faecal kit to collect their sample and if they would like to also a blood sample. Though, this is a completely voluntary sample. Afterwards the researcher and BioRepository will coordinate with each other to send the next faecal kit.

3. **Experimental methods**

We will use the faecal samples to profile the microbiota profiles of women who would have developed breast cancer. We will do this by doing 16S rRNA sequencing and whole genome sequencing. 16S rRNA sequencing allows us to resolve the microbiota down to a genus level, while whole genome sequencing allows us to resolve the microbiota down to a species level. We will utilise scripts already established in the Hall lab to resolve a microbiota profile. We would also like to probe into the metabolite profile of the microbiota by analysing faecal samples using NMR or HPLC. In doing so, we can determine to an extent, what metabolites the gut microbiota produces.

We will use blood samples to probe into the immune and metabolite profiles of these patients. We will be using the MSD QuickPlex SQ 120TM to probe into the circulatory immune cytokines. We would also like to use NMR as aforementioned to investigate the circulating metabolites. We wish to also do flow cytometry using the blood samples to probe into the peripheral blood mononuclear cells. This would allow us to see if there are changes in the circulating immune cells.

We will use fresh tissue samples to do flow cytometry or RNA extraction. Flow cytometry allows us to probe into the cellular population of the tumour using certain cell markers. RNA analysis tells us which genes are being transcribed and would

provide further insight what the gut microbiota may be influencing in the breast tumour itself.

If we have found results, we would like to validate these by undertaking a bacterial co-culture with *in-vitro* cell line models.

4.   **Statistical Analysis**

Power calculations using microbiota diversity, at 5% significance level, indicate that a single group of 72 subjects, each sampled at least three times over 12months, has a 95% power to detect a true time-averaged change (if it exists) of at least 12% variance (PASS software v8 with 100simulations on the database of Goedert et al, 2015, using unweighted UniFrac Distance metric).

We will perform multivariate statistical framework analysis using various R packages to determine whether (i) bacterial diversity and richness is correlated with analysed immune parameters and (ii) clinical metadata such as type of antibiotic and BC progression markers is also correlated to microbial diversity (using variations of Bray–Curtis distance and association analysis assessed by Spearman correlation), (iii) determine contribution of each genus and cytokine to microbiome variation using canonical correspondence analysis (CCA), and (iv) detect relationships between different members of the gut microbiota and their potential combined effect on immune output in BC patients by constructing a network of co-occurrence genus/species IDs and interrogate the network for modules using weighted gene co-expression network analysis.

5.   **Data protection and Donor confidentiality**

The researcher will have no contact with the patient. Members of staff of the Biorepository will remain the link between the researcher and the patients. For the JPUH the appointed research nurse or clinical fellow will anonymise the data and act as the link between the researcher and the patients. Extensive talks have been undertaken to determine confidentiality is kept as best as possible.

For one, the faecal collection kits will have everything prelabelled with their designated Biorepository number. This minimises the chance that the patient may write their personal information on these packs. If a faecal pack must be sent to the

patient's home address the researcher will prepare the pack with the Biorepository number and then drop it off at the Biorepository office. A member of staff will then write or label the pack and drop it off at the QIB reception where it will be sent to the patient's home address.

In the case of the JPUH site, the package will be prelabelled with the assigned number and given to the patient.

For blood and tissue the similar approach will be done. Vials will be given prelabelled to limit chances of patient identifiable information being given to the researcher. Tissue will be picked up by the Biorepository staff member who will ensure its anonymity prior to being given to the researcher. This will not apply to the JPUH site.

To gain clinical information of the patient e.g. what type of cancer they have or what the S-score was of their mammogram the following was discussed. The researcher and the Biorepository have agreed to have an 'interim' spreadsheet with the requested information on it. The patient's identity will be the assigned biorepository number. The spreadsheet will have to be sent via email to the biorepository staff member, however it will be encrypted using a password only the biorepository and researcher will know. The information will then be copied over to the master spreadsheet that is based on the researcher's computer so that the spreadsheet that is sent across via email will be blank again. Once on the researcher's computer the patient will be assigned a BEAM study ID.

If participants have any questions or concerns regarding the study, they will have been given the biorepository's contact information who will then contact the researcher for further clarification.

6.   **Ethical and regulatory considerations**
An ethical consideration that was discussed was the timing of asking consent. We have decided to provide the PIS to all patients who are eligible and interested. After discussions with the consultant we have also decided to only provide the PIS once a diagnosis has been provided. This provides patients enough time to read through it

and think about it. Then, we have decided to consent patients who will have had a biopsy as this provides us a control cohort as well as a disease cohort.

We also have considered if sending the collection kits to their home addresses was ethical to do. After speaking to potential patients and clinicians they have expressed that the compliance may be higher if it was sent to their homes. This eliminates the need to travel to a place to drop-off/pick-up items for a study.

7.   Reference

Goedert, J. J., et al. (2015). "Investigation of the Association Between the Fecal Microbiota and Breast Cancer in Postmenopausal Women: a Population-Based Case-Control Pilot Study." JNCI: Journal of the National Cancer Institute **107**(8).

UNIVERSITY OF EAST ANGLIA

FACULTY OF MEDICINE AND HEALTH SCIENCES

RESEARCH ETHICS COMMITTEE

Application Form for Ethical Approval of a Research Project

> **Please refer to the guidelines when completing this form.**
>
> **This document should help members of the FMH Ethics Committee understand the objectives of your project/research and the procedures to be conducted.**
>
> **It is ESSENTIAL that you use non-technical language that can easily be understood by non-specialists and lay members of the Committee, and all applications need to include all relevant documents.**
>
> **It is not acceptable to refer the committee to a protocol, and the information on the application, together with the attachments, should be sufficient to allow the Committee to form an opinion.**
>
> **Forms may be reviewed by the Chair and will be returned to you if you do not meet these requirements. This will delay approval of your application as applications cannot be accepted after the deadline.**

*Does the project involve the use of **drugs, or testing of new equipment**, or research on **NHS patients**?*

NO

*(If YES, it **MUST** be referred to an NHS Research Ethics Committee for approval)*

Does the project involve the use of Human Tissue?   *YES*

(If YES, it must be referred to the Faculty of Medicine and Health Sciences Research Ethics Committee)

Is the project a Service Evaluation?   *NO*

(If YES, it must be referred to the Faculty of Medicine and Health Sciences Research Ethics Committee with evidence of acceptance by the relevant NHS Trust)

Is the project an Audit?   *NO*

(If YES, it must be referred to the Faculty of Medicine and Health Sciences Research Ethics Committee with evidence of acceptance by the relevant NHS Trust)

1.   Name of applicant: ………MEI YU NANCY **TENG**……………...

(Block letters)

2. Academic address for correspondence: ……………………………………………

………QUADRAM INSTITUTE OF THE BIOSCIENCES…………………………..

………NORWICH RESEARCH PARK……………..Post code: …NR4 7UA…..

3. Tel No: ……………………………………….. Fax No: …………………………..

4. E-mail address: ………Nancy.teng@quadram.ac.uk……………

5. School (AHP, MED, NSC): ……………MED…….

6. Status of applicant (Staff, UG or PG student - and year of course): .PG-4…..

7. If Student:
   Is this study being carried out to fulfil a required part of your course? Yes

   If No:
   Please confirm contact details of supervisor

   ……………………………………………………………………………………

   Name of supervisor: ……………………………………………… … … …

8. Has this application gone to an Ethics Committee elsewhere?   NO

   If YES, please indicate where and include copies of correspondence:

   ……………………………………………………………………………………

Please send 16 copies of the application form, proposal and any other documents (please ensure all documents are fixed together in the top left-hand corner) to: *FMH Research Ethics, Research & Enterprise Office, SCI Building, Room 0.03,* University of East Anglia, Norwich NR4 7TJ; **plus an e-mail copy to fmh.ethics@uea.ac.uk on or before the deadline shown on the following intranet page.** (https://intranet.uea.ac.uk/foh/intranet/ethics-committee).

For any queries telephone: 01603 591720

Project details (please could sections 9, 10 and 11 be limited to a maximum of 3000 words.

9.	Full title: ……The role of the gut microbiota in breast health ……….

……………………………………………………………………………………………..

……………………………………………………………………………………………..

……………………………………………………………………………………………..

10.	Purpose of project:
The cascade of steps leading to breast cancer development depends on the interaction between cancer cells and their local environment. This includes the immune system. The resident microbiota is a key component linking these together. The microbiota has shown to influence the immune response and can have an impact on the surviving breast cancer post-operation. The purpose of this project is to probe the microbiota profiles of women to assess how the microbiota affects breast health. Unfortunately, some of the participants will be diagnosed with the disease in the course of this study. We wish to focus on the microbiota profile to determine factors that influence progression during early disease stages.

11.	Methodology, Procedure and Analysis:
Principal Investigator: Dr. Stephen Robinson (QIB)

NHS Investigator: Dr. Simon Pain (NNUH), Dr. Leeper (Sandy) Alexander (JPUH)

Supervisors: Dr. Lindsay Hall (TUM/QIB)

Perform a longitudinal study profiling the gut microbiota of breast cancer patients with or without previous antibiotic treatments: Statistical power calculations using microbiota diversity at 5% significance level, indicate a single group of 72 subjects (36 subjects receiving antibiotics and 36 subjects without), each over multiple time points over a period of 1 year, can provide enough data to confirm a statistical significance, if any.

We will recruit patients who will attend the breast care clinic at the Norfolk and Norwich University Hospital (NNUH) and James Paget University Hospital (JPUH). Our main aim is to collect faecal samples as this will provide a snapshot of the patient's microbiota. The patient's criteria is limited to, older than 30yo, have not had antibiotics 6 months prior to consent and first time diagnosis of invasive breast disease.

We have added JPUH as a source of recruitment for specifically low grade hormone receptor (HR) +/ HER2 – patients to boost numbers in order to answer the research question of how

the gut microbiota and breast health may be associated. The recruitment pathway for JPUH is further explained in the Study protocol. When referring to the "Biorepository" we also include authorised individuals who have received the necessary training in order to consent patients into the study as dictated by the NRP Biorepository.

After discussions with the clinicians we have decided to have three pathways to recruit patients for our final pathway. In all cases the faecal kit will be sent to the participant's home address. We have decided to do so as these patients will undergo a stressful time and did not want to contribute undue stress by requesting a faecal sample at the NNUH/JPUH or come back to do so. The pathways will be as follows will be as follows for the different patients according to treatment:
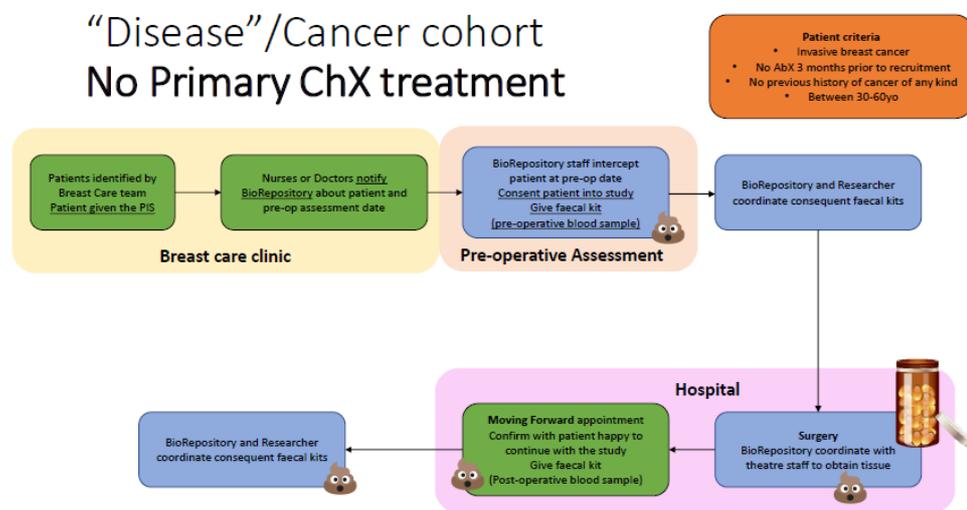
Disease cohort: No primary chemotherapy. Some patients who have been diagnosed with breast cancer will not undergo primary chemotherapy prior to surgery. To consent these patients, the breast care team at the NNUH will identify potential patients and provide them with a PIS (annex 3) if interested. If they are interested BioRepository will be notified of the patient and their pre-operative assessment date. Consequently, a member of staff of the BioRepository, will intercept and consent the patient into the study at their pre-operative assessment screening. Once consented they will be given their first faecal sample kit to collect the sample at home and send it back to the QIB. In addition to this we will request a vial of blood to be donated to the study if the patient will have blood drawn for their appointment or a voluntary donation by the patient. The following faecal kits will be sent to patient's home addresses for them to collect their sample and send it back to the QIB. On the day of surgery, the BioRepository will coordinate with the surgical team to receive resected tissue. This donation is dependent on the size of the tumour i.e. larger than 15mm, at which point donating some tissue will not compromise the care of the patient. 6 months after initial diagnosis the patient will have a "moving forward" appointment at the breast clinic. The BioRepository will intercept the patient to confirm they are happy to continue

giving samples. If so, they will provide another faecal kit and also request a vial of blood. Their last faecal sample will be sent to their home address again.

Disease cohort: Primary chemotherapy. Patients will be identified by the breast care team. If interested they will be given the PIS. These patients will undergo neoadjuvant treatment i.e. primary chemotherapy. Their next appointment at the NNUH/JPUH will be with an oncologist. The breast care team will notify the BioRepository who will intercept the patient at their oncology appointment. Here the patients will be consented into the study as previously. Once consented they will be provided with a faecal kit. The rest of the pathway follows the same way as previously described with "Disease cohort: no primary chemotherapy".

Control cohort. For our study to be valid we require a control cohort. It has been decided to do this the following way. Patients at the one-stop assessment clinic usually get a mammogram done. However due to limited resources, not everyone gets screened

immediately. Some of these patients will be asked to come back later. These patients, if they fit the criteria, will receive the PIS. The breast care team will notify the BioRepository. The BioRepository will intercept the patient at their next mammogram screening and consent them into the study. An alternative method is to consent all patients who have already received their diagnosis, of not having breast cancer, by their radiologist. If they consent, they will be provided with a faecal kit to collect their sample and if they would like to also a blood sample. Though, this is a completely voluntary sample. Afterwards the researcher and BioRepository will coordinate with each other to send the next faecal kit.

This will be an original novel study providing critical samples for and data for our research programme and wider community. The samples collected will include stool samples, and subjected to availability and consent, may include blood and tumour tissue (fresh and fixed). Samples will be collected fresh and frozen. Analysis of faecal samples include 16S rRNA and shotgun metagenomic sequencing. The former can discriminate the microbiota community down to the genus level while the latter can segregate down to a species level. We wish to use the blood and tissue samples to probe into the immune and metabolite profiles of these patients. Fixed tissue may also be used to look at the histology of the resected tissue alongside the immune profiling.

Correlate impact of microbiota alterations with mucosal and systemic immune readouts and link to clinical meta-data: Samples collected will be processed for immune and metabolite profiling. The profiling will focus on immune-mediated inflammatory profiles and several microbial-derived short-chain fatty acids. As we aim to collect samples via the NNUH or JPUH, we can also read clinical notes relevant to our study, providing information regarding antibiotic exposure, treatment regimes and clinical tumour information. We will also use an amended lifestyle and diet questionnaire based on one previously done by the BioRepository for patients recruited from the NNUH. This questionnaire has been amended to include questions in relation to breast health. The answers of the questionnaire can then be used to perform in-depth statistical analysis to determine if there are any correlations.

Correlate specific microbes with alterations in immune and metabolite responses using multivariate analysis and validate using bacterial co-culture with in vitro cell line models: We will use various statistical packages to identify if any significant changes observed in the gut microbiota of breast cancer patients can be correlated to the immune and/or metabolite responses observed. Consequently, to validate the data we aim to identify these strains and co-culture them with *in vitro* cell line models.

12.   Resources required:

We will require the cooperation and resources of the tissue bank BioRepository and clinical staff at the NNUH and JPUH.

13.   Source of Funding

Funded by the Big C, and by the QI CSF fund (JPUH cohort)

14.   Has this project been peer reviewed? Please could you include details of who the project has been peer reviewed by.

As part of the grant application, this project has been reviewed by 3 international experts.

15.   Ethical issues (Please also complete research safety checklist even if no risks are identified)

Minimal ethics issues. Collection of faecal samples are non-invasive. Blood and tumour samples will be part of routine collection as part of the patient's treatment regime. If the blood samples are not part of routine collection i.e. a voluntary donation, it will be done by a trained certified phlebotomist employed by the NNUH.

16.   Proposed start and finish dates:

Start date: …1 January 2019…… Finish date: ……30 June 2023…….

17.   Where will the research be carried out?
Quadram Institute of Biosciences

18.   Do you need to survey UEA students or staff outside the Faculty of Medicine and Health Sciences? If so, you need to get approval in principle from the Dean of Students prior to applying to the FMH Ethics Committee. Please attach a copy of approval in principle to this application form.

https://www.uea.ac.uk/polopoly_fs/1.151266!survey_form.pdf

19.   Information sheets and consent forms must be appended (c.f. NRES site for models, www.nres.npsa.nhs.uk Please ensure that participants are requested to initial the boxes on the consent forms.

      We will be utilising the approved forms from the BioRepository, attached consent form as annex 3.

20.   **Checklist (double click on each box and select 'checked' once done)**

      Have you completed all sections of the application in language
      which will be understood by lay people?   ☐

      Has your supervisor signed the form?   ☐

      Have you included your academic address (not your home
      address)?   ☐

      Have you numbered all the pages in your protocol/attachments?   ☐
      (If the pages are not numbered the Committee may return
      your application)

      Have you included the following documents, if applicable?

      Protocol   ☐
      Gatekeeper consent   ☐
      Consent forms   ☐
      Participant information sheets (using NRES format)   ☐
      Letters to participants   ☐
      Copies of questionnaires   ☐
      Copies of correspondence from other ethic committees   ☐
      Copies of all recruitment letters, emails, posters and adverts   ☐
      Research Safety Checklist   ☐
      Dean of Student Office approval in principle for survey   ☐
      Have you proof-read your application to check for   ☐
      typographical and grammatical errors?

      Have you included a header and footer on each page with your   ☐
      name, date of submission and page number?

Have you included 16 photocopies? ☐

Have you e-mailed a copy to the Research & Enterprise Office? ☐

Supervisory arrangements for **STUDENT PROJECTS ONLY**

Degree/Course …........PhD ....

School ………QIB/Medicine

Academic Supervisor ……………Dr. Lindsay Hall………….

I have read this application and can confirm that I am taking supervisory responsibility for this project.

In the case of a student research outside the normal course requirements I confirm that I am happy to take responsibility for the quality of protocol design, the provision of necessary resources, statistical support and usual supervision and governance of the student.

Project Supervisor's signature   Date
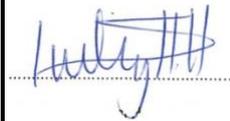
.. 27/06/2022

Post Held   Group Leader

8.3 Supplementary data for Chapter 2

8.3.1 Accession numbers for bacterial cytochrome P450

>NC_000962.3:2547749-2548939 Mycobacterium tuberculosis H37Rv, complete genome

>NC_000964.3:229525-230778 Bacillus subtilis subsp. subtilis str. 168 complete genome

>NC_000964.3:c3090413-3089226 Bacillus subtilis subsp. subtilis str. 168 complete genome

>NC_002163.1:c1343911-1342550 Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819 chromosome, complete genome

>NC_002516.2:c2794133-2792799 Pseudomonas aeruginosa PAO1, complete genome

>NC_003318.1:c919412-918249 Brucella melitensis bv. 1 str. 16M chromosome II, complete sequence

>NC_004311.2:219098-221302 Brucella suis 1330 chromosome II, complete sequence

>NC_004556.1:1960979-1962187 Xylella fastidiosa Temecula1, complete sequence

>NC_005042.1:1050143-1051237 Prochlorococcus marinus subsp. marinus str. CCMP1375, complete sequence

>NC_005126.1:1409456-1410691 Photorhabdus laumondii subsp. laumondii TTO1, complete sequence

>NC_005966.1:c791001-788356 Acinetobacter baylyi ADP1, complete sequence

>NC_006361.1:509582-510850 Nocardia farcinica IFM 10152, complete sequence

>NC_007509.1:c336937-333746 Burkholderia lata chromosome 3, complete sequence

>NC_007530.2:c2968869-2965672 Bacillus anthracis str. 'Ames Ancestor', complete sequence

>NC_007624.1:c820365-819202 Brucella abortus 2308 chromosome II, complete sequence

>NC_008095.1:c781094-779742 Myxococcus xanthus DK 1622, complete sequence

>NC_009495.1:1138095-1139354 Clostridium botulinum A str. ATCC 3502, complete sequence

>NC_010104.1:219210-221414 Brucella canis ATCC 23365 chromosome II, complete sequence

>NC_010530.1:c1238430-1236091 Cupriavidus taiwanensis LMG 19424 chromosome 2, complete sequence

>NC_010572.1:c3994707-3993502 Streptomyces griseus subsp. griseus NBRC 13350, complete sequence

>NC_011916.1:61885-63153 Caulobacter crescentus NA1000, complete genome

>NC_011963.1:c256520-255339 Cereibacter sphaeroides KD131 chromosome 1, complete sequence

>NC_012039.1:296222-297595 Campylobacter lari RM2100, complete sequence

>NC_012988.1:810458-811858 Methylorubrum extorquens DM4, complete sequence

>NC_013929.1:c665647-662435 Streptomyces scabiei 87.22, complete sequence

>NC_014034.1:c2615148-2613940 Rhodobacter capsulatus SB 1003, complete sequence

>NC_014259.1:c689149-686567 Acinetobacter oleivorans DR1, complete sequence

>NC_014334.2:983024-983215 Lacticaseibacillus paracasei, complete sequence

>NC_014541.1:c3599092-3597296 Ferrimonas balearica DSM 9799, complete sequence

>NC_014643.1:1128616-1129812 Rothia dentocariosa ATCC 17931, complete sequence

>NC_015067.1:c483668-483243 Bifidobacterium longum subsp. longum JCM 1217, complete sequence

>NC_015138.1:5471460-5474180 Acidovorax avenae subsp. avenae ATCC 19860, complete sequence

>NC_015567.1:c585825-584602 Serratia plymuthica AS9, complete sequence

>NC_015687.1:3503599-3504855 Clostridium acetobutylicum DSM 1731, complete sequence

>NC_015848.1:c4007702-4006506 Mycobacterium canettii CIPT 140010059, complete sequence

>NC_015848.1:2603573-2604763 Mycobacterium canettii CIPT 140010059, complete sequence

>NC_016887.1:c250985-249780 Nocardia cyriacigeorgica GUH-2, complete sequence

>NC_016946.1:2758486-2759772 Mycobacterium intracellulare ATCC 13950, complete sequence

>NC_016946.1:13300-14514 Mycobacterium intracellulare ATCC 13950, complete sequence

>NC_019847.1:22743-23705 Sinorhizobium meliloti GR4 plasmid pRmeGR4b, complete sequence

>NC_020272.1:1377253-1380414 Bacillus amyloliquefaciens IT-45, complete sequence

>NC_020528.1:65218-66618 Sinorhizobium meliloti 2011, complete sequence

>NC_020990.1:c240260-239061 Streptomyces albidoflavus, complete sequence

>NC_021064.1:698665-699810 Cutibacterium avidum 44067, complete sequence

>NC_021352.1:167121-168374 Corynebacterium glutamicum SCgG2, complete sequence

>NC_022663.1:27581-28813 Mycobacterium kansasii ATCC 12478, complete sequence

>NC_023018.2:3367852-3370197 Pandoraea pnomenusa, complete sequence

>NC_023150.1:650816-652189 Rhodococcus pyridinivorans SB3094, complete sequence

>NW_003206190.1:c72662-67167 Perkinsus marinus ATCC 50983 genomic scaffold scf_1104296960350, whole genome shotgun sequence

>NZ_AAUW01000001.1:c403736-402501 Roseibium aggregatum IAM 12614 1101096003783, whole genome shotgun sequence

>NZ_ACVO01000021.1:3966-5177 Rothia mucilaginosa ATCC 25296 contig00003, whole genome shotgun sequence

>NZ_AFVW02000001.1:86930-88159 Mycobacterium colombiense CECT 3035 contig00002A, whole genome shotgun sequence

>NZ_AGFX01000009.1:79759-81018 Paenibacillus peoriae KCTC 3763 contig18, whole genome shotgun sequence

>NZ_AGIZ01000001.1:c595659-594280 Fischerella thermalis JSC-11 ctg118, whole genome shotgun sequence

>NZ_AHCD03000028.1:c43660-42530 Pseudoalteromonas rubra strain DSM 6842 contig028, whole genome shotgun sequence

>NZ_AP013103.1:c18326-17139 Bradyrhizobium elkanii USDA 61 chromosome, complete genome

>NZ_AP014630.1:c2583481-2581991 Acinetobacter guillouiae strain NBRC 110550 chromosome, complete genome

>NZ_AP017600.1:c1062247-1061981 Rickettsia japonica strain M99123 chromosome

>NZ_AP017900.1:6975615-6976934 Nocardia seriolae strain UTF1 chromosome, complete genome

>NZ_AP017900.1:386806-388035 Nocardia seriolae strain UTF1 chromosome, complete genome

>NZ_AP018357.1:915305-916480 Burkholderia contaminans strain CH-1

>NZ_AP019312.1:126628-127734 Chromobacterium haemolyticum strain CH06-BL chromosome, complete genome

>NZ_AP021898.1:798228-799358 Akkermansia muciniphila strain JCM 30893 chromosome, complete genome

>NZ_AP022642.1:1490133-1491407 Pseudomonas otitidis strain MrB4 chromosome, complete genome

>NZ_AP023172.1:830939-832144 Rhodococcus qingshengii strain CS98 chromosome, complete genome

>NZ_AP024251.1:13620-14834 Mycobacterium paraintracellulare strain M011 chromosome, complete genome

>NZ_AP025193.1:c4262610-4261327 Comamonas thiooxydans strain NR4028 chromosome, complete genome

>NZ_AP025344.1:c2637814-2636597 Paenibacillus dendritiformis strain J27TS7 chromosome, complete genome

>NZ_APBN01000015.1:23348-24703 Brevibacillus borstelensis AK1 seq_num_015, whole genome shotgun sequence

>NZ_AQUJ01000001.1:c832012-830783 Sphingomonas melonis C3 SphmeDRAFT_contig1.1_C, whole genome shotgun sequence

>NZ_AUZQ01000050.1:c205051-203735 Mycobacterium avium subsp. hominissuis A5 contig050, whole genome shotgun sequence

>NZ_BBQG01000010.1:32373-33584 Streptomyces albus strain NBRC 13014, whole genome shotgun sequence

>NZ_BCUT01000001.1:c2996527-2995349 Variovorax paradoxus NBRC 15149, whole genome shotgun sequence

>NZ_BIMF01000002.1:7209-8402 Paenibacillus lautus NBRC 15380 PLA01S_CON0002, whole genome shotgun sequence

>NZ_BJOG01000001.1:252844-254172 Microbacterium oxydans strain NBRC 15586 sequence01, whole genome shotgun sequence

>NZ_BLJF01000001.1:3750913-3751146 Pseudomonas asiatica strain RYU5 RYU5_unitig_0, whole genome shotgun sequence

>NZ_CAACYD010000003.1:190281-191591 Gordonia paraffinivorans strain 3012STDY6756503, whole genome shotgun sequence

>NZ_CABEIC010000002.1:c71828-70617 Gordonia terrae strain NCTC10669, whole genome shotgun sequence

>NZ_CABKQE010000002.1:1080651-1083890 Ralstonia pickettii isolate MGYG-HGUT-01384, whole genome shotgun sequence

>NZ_CABVLY010000015.1:181704-182885 Burkholderia anthina isolate LMG 20980, whole genome shotgun sequence

>NZ_CADEVI010000002.1:c192479-191313 Burkholderia diffusa strain LMG 29043, whole genome shotgun sequence

>NZ_CADFDQ010000010.1:218294-222307 Burkholderia pseudomultivorans strain BCC1191, whole genome shotgun sequence

>NZ_CADIKT010000013.1:147161-148531 Achromobacter insuavis strain LMG 26846 isolate LMG 26846, whole genome shotgun sequence

>NZ_CAJMWN010000001.1:151830-153155 Mycobacterium riyadhense isolate MR-226, whole genome shotgun sequence

>NZ_CDBW01000035.1:c40329-38539 Aeromonas sobria strain CECT 4245, whole genome shotgun sequence

>NZ_CM000745.1:2222724-2223959 Bacillus pseudomycoides DSM 12442 chromosome, whole genome shotgun sequence

>NZ_CM000753.1:2832247-2835444 Bacillus thuringiensis serovar berliner ATCC 10792 chromosome, whole genome shotgun sequence

>NZ_CM000753.1:6235621-6236856 Bacillus thuringiensis serovar berliner ATCC 10792 chromosome, whole genome shotgun sequence

>NZ_CM001513.1:2250239-2251387 Pseudomonas lactis strain SS101 chromosome, whole genome shotgun sequence

>NZ_CP004373.1:2377895-2379307 Gluconobacter oxydans DSM 3504 chromosome, complete genome

>NZ_CP007220.1:100626-101861 Mycobacteroides chelonae CCUG 47445 chromosome, complete genome

>NZ_CP007255.1:1740107-1741474 Rhodococcus erythropolis R138 chromosome, complete genome

>NZ_CP007597.1:3418446-3420053 Stenotrophomonas rhizophila strain DSM 14405 chromosome

>NZ_CP007810.1:c28518-27175 Xanthomonas oryzae pv. oryzicola strain YM15 chromosome, complete genome

>NZ_CP008747.1:c1711953-1710679 Staphylococcus hyicus strain ATCC 11249 chromosome, complete genome

>NZ_CP008782.1:c685114-682775 Burkholderia pseudomallei strain Mahidol-1106a chromosome 2, complete sequence

>NZ_CP008782.1:c2612348-2610942 Burkholderia pseudomallei strain Mahidol-1106a chromosome 2, complete sequence

>NZ_CP008889.1:c190613-189306 Dermacoccus nishinomiyaensis strain M25 chromosome, complete genome

>NZ_CP008998.1:3429086-3430828 Xanthomonas citri pv. citri strain MN12 chromosome, complete genome

>NZ_CP009108.1:922333-923538 Bacillus altitudinis strain GR-8 chromosome, complete genome

>NZ_CP009125.1:2754136-2755323 Pectobacterium atrosepticum strain 21A chromosome, complete genome

>NZ_CP009428.1:c4464189-4462333 Paenibacillus odorifer strain DSM 15391 chromosome, complete genome

>NZ_CP009555.1:1161741-1162976 Burkholderia oklahomensis C6786 chromosome I, complete sequence

>NZ_CP009679.1:701635-704817 Bacillus velezensis strain JS25R chromosome, complete genome

>NZ_CP009709.1:1461905-1462342 Weizmannia coagulans DSM 1 = ATCC 7050 chromosome, complete genome

>NZ_CP009728.1:c1573842-1571488 Burkholderia mallei strain Turkey2 chromosome 2, complete sequence

>NZ_CP009728.1:26241-27647 Burkholderia mallei strain Turkey2 chromosome 2, complete sequence

>NZ_CP009793.1:482773-483954 Burkholderia dolosa AU0158 chromosome 2, complete sequence

>NZ_CP010025.1:c294228-293026 Paraburkholderia fungorum strain ATCC BAA-463 chromosome 3, complete sequence

>NZ_CP010650.1:2080466-2081650 Phaeobacter inhibens strain P54 chromosome, complete genome

>NZ_CP010820.1:1084867-1086648 Lysinibacillus fusiformis strain RB-21 chromosome, complete genome

>NZ_CP010896.1:c796186-793967 Pseudomonas simiae strain PCL1751 chromosome, complete genome

>NZ_CP011254.1:3792711-3793688 Serratia fonticola strain DSM 4576 chromosome, complete genome

>NZ_CP011269.1:c1607452-1606247 Mycolicibacterium fortuitum strain CT6 chromosome, complete genome

>NZ_CP011269.1:641752-643068 Mycolicibacterium fortuitum strain CT6 chromosome, complete genome

>NZ_CP011504.1:1904117-1905292 Burkholderia pyrrocinia strain DSM 10685 chromosome 2, complete sequence

>NZ_CP011530.1:27504-28892 Mycobacteroides immunogenum strain CCUG 47286 chromosome, complete genome

>NZ_CP011835.1:1092088-1094451 Azotobacter chroococcum strain B3 chromosome, complete genome

>NZ_CP012047.1:c1117900-1116632 Tetragenococcus halophilus strain MJ4 chromosome, complete genome

>NZ_CP012400.2:c1512805-1511645 Pseudomonas yamanorum strain LBUM636 chromosome, complete genome

>NZ_CP012507.1:1200673-1201947 Kocuria palustris strain MU14/1 chromosome, complete genome

>NZ_CP012746.1:2485211-2486455 Paraburkholderia caribensis MBA4 chromosome 1, complete sequence

>NZ_CP012746.1:c3136642-3135392 Paraburkholderia caribensis MBA4 chromosome 1, complete sequence

>NZ_CP012748.1:1425081-1427870 Paraburkholderia caribensis MBA4 plasmid unnamed, complete sequence

>NZ_CP012938.1:3778377-3779954 Bacteroides ovatus strain ATCC 8483 chromosome, complete genome

>NZ_CP013119.1:1316079-1317275 Alcaligenes faecalis strain ZD02 chromosome, complete genome

>NZ_CP013365.1:1097725-1099164 Burkholderia territorii strain RF8-non-BP5 chromosome 2, complete sequence
>NZ_CP013399.1:c144847-140819 Burkholderia seminalis strain FL-5-4-10-S1-D7 chromosome 3, complete sequence
>NZ_CP013403.1:c2191530-2190352 Burkholderia metallica strain FL-6-5-30-S1-D7 chromosome 2, complete sequence
>NZ_CP013413.1:28320-29726 Burkholderia thailandensis strain 2002721643 chromosome 2, complete sequence
>NZ_CP013452.1:c2260889-2259729 Burkholderia cenocepacia strain MSMB384WGS chromosome 2, complete sequence
>NZ_CP013460.1:c877811-876483 Burkholderia stagnalis strain MSMB735WGS chromosome 3, complete sequence
>NZ_CP013481.2:c3148097-3145509 Pandoraea apista strain DSM 16535 chromosome, complete genome
>NZ_CP013532.1:53028-54440 Rhizobium phaseoli strain R650 chromosome, complete genome
>NZ_CP014022.1:c1672382-1671204 Staphylococcus lugdunensis strain FDAARGOS_141 chromosome, complete genome
>NZ_CP014158.1:c4721570-4720296 Pseudomonas citronellolis strain P3B5 chromosome, complete genome
>NZ_CP014262.1:1216509-1217642 Pseudomonas corrugata strain RM1-1-4 chromosome, complete genome
>NZ_CP014347.1:c1884238-1882970 Xanthomonas phaseoli pv. dieffenbachiae LMG 695 chromosome
>NZ_CP014842.1:c898307-897087 Bacillus licheniformis strain SCDB 14 chromosome, complete genome
>NZ_CP015230.1:c1059469-1058282 Tritonibacter mobilis F1926 chromosome, complete genome
>NZ_CP015235.1:c1402064-1400862 Rhodococcus fascians D188 chromosome, complete genome
>NZ_CP015421.1:c1056410-1054221 Rhodovulum sulfidophilum strain SNK001 chromosome, complete genome
>NZ_CP015578.1:138470-139828 Campylobacter lanienae NCTC 13004 chromosome, complete genome
>NZ_CP015880.1:c984816-983488 Ensifer adhaerens strain Casida A chromosome, complete genome
>NZ_CP015941.1:c2730933-2730394 Legionella pneumophila strain C9_S chromosome, complete genome
>NZ_CP016022.1:2563859-2565145 Ralstonia insidiosa strain ATCC 49129 chromosome 1, complete sequence
>NZ_CP016809.1:505640-506902 Paenibacillus ihbetae strain IHBB 9852 chromosome, complete genome
>NZ_CP016878.1:4966834-4968027 Xanthomonas hortorum strain B07-007 chromosome, complete genome
>NZ_CP017040.1:2250802-2252097 Cutibacterium modestum strain F0672 chromosome, complete genome
>NZ_CP017060.1:2562716-2563930 Bacillus cereus strain FORC_047 chromosome, complete genome
>NZ_CP017060.1:c3109039-3105842 Bacillus cereus strain FORC_047 chromosome, complete genome
>NZ_CP017454.1:3725694-3727100 Dickeya solani strain PPO 9019 chromosome, complete genome
>NZ_CP017466.1:1653467-1654054 Staphylococcus nepalensis strain JS11 chromosome, complete genome
>NZ_CP017482.1:c2831943-2830555 Pectobacterium polaris strain NIBIO1392 chromosome, complete genome
>NZ_CP017704.1:580008-581393 Peribacillus simplex NBRC 15720 = DSM 1321 chromosome, complete genome
>NZ_CP017705.1:c2823469-2822249 Brevibacillus laterosporus DSM 25 chromosome, complete genome
>NZ_CP017707.1:c846239-844827 Chromobacterium vaccinii strain 21-1 chromosome, complete genome
>NZ_CP017886.1:c5489606-5488473 Pseudomonas frederiksbergensis strain ERDD5:01 chromosome, complete genome
>NZ_CP017962.1:275664-276920 Virgibacillus halodenitrificans strain PDB-F2 chromosome, complete genome
>NZ_CP018061.1:c2678085-2676826 Enterococcus mundtii strain DSM 4838 chromosome, complete genome
>NZ_CP018074.1:c305420-304200 Streptomyces venezuelae strain NRRL B-65442 chromosome
>NZ_CP018420.1:c11890-10688 Pseudomonas veronii strain R02 chromosome, complete genome
>NZ_CP018620.1:3109815-3111017 Paenibacillus xylanexedens strain PAMC 22703 chromosome, complete genome
>NZ_CP018725.1:1276037-1277650 Xanthomonas vesicatoria ATCC 35937 strain LMG911 chromosome, complete genome
>NZ_CP018820.1:c13107-11890 Sphingomonas koreensis strain ABOJV chromosome, complete genome
>NZ_CP019659.1:c1640682-1639465 Paenibacillus larvae subsp. larvae strain Eric_IV chromosome, complete genome
>NZ_CP019958.1:c1131623-1130487 Candidatus Liberibacter asiaticus strain JXGC chromosome, complete genome
>NZ_CP020000.1:3352075-3354657 Acinetobacter calcoaceticus strain CA16 chromosome, complete genome
>NZ_CP020398.1:961033-962211 Burkholderia multivorans strain FDAARGOS_246 chromosome 2, complete sequence
>NZ_CP020478.1:141331-142695 Campylobacter helveticus strain ATCC 51209 chromosome, complete genome
>NZ_CP020738.1:1202487-1203668 Burkholderia ubonensis subsp. mesacidophila strain ATCC 31433 chromosome 2, complete sequence
>NZ_CP020906.1:51444-52856 Rhizobium etli strain NXC12 chromosome, complete genome
>NZ_CP021047.1:c1289548-1288364 Phaeobacter gallaeciensis strain P128 chromosome, complete genome
>NZ_CP021395.1:307777-309108 Bordetella hinzii strain SV2 chromosome, complete genome
>NZ_CP021763.1:898387-899583 Ralstonia pseudosolanacearum strain RS 476 plasmid unnamed, complete sequence
>NZ_CP021894.1:868857-870245 Pectobacterium versatile strain SCC1 chromosome
>NZ_CP022046.2:c1486584-1485295 Mammaliicoccus sciuri strain FDAARGOS_285 chromosome, complete genome
>NZ_CP023011.2:1087072-1088331 Enterococcus hirae strain FDAARGOS_234 chromosome, complete genome
>NZ_CP023665.1:779223-780446 Bacillus paralicheniformis strain Bac84 chromosome, complete genome
>NZ_CP023741.1:1260571-1261812 Sphingobium yanoikuyae strain S72 chromosome, complete genome
>NZ_CP024035.1:c1790347-1789115 Priestia aryabhattai strain K13 chromosome, complete genome
>NZ_CP024109.1:1942025-1943236 Bacillus cytotoxicus strain CH_13 chromosome, complete genome
>NZ_CP024633.1:83462-84697 Mycobacteroides salmoniphilum strain DSM 43276 chromosome
>NZ_CP025003.1:c2989187-2987907 Dickeya fangzhongdai strain DSM 101947 chromosome, complete genome
>NZ_CP025070.1:c1949169-1947790 Bordetella parapertussis strain A005 chromosome, complete genome
>NZ_CP025333.1:3244171-3245463 Brevibacterium aurantiacum strain SMQ-1419 chromosome, complete genome
>NZ_CP025371.1:c1285295-1283916 Bordetella pertussis strain H640 chromosome, complete genome
>NZ_CP025799.1:1094323-1095462 Dickeya zeae strain MS2 chromosome, complete genome
>NZ_CP026105.1:c2005300-2004065 Paraburkholderia hospita strain DSM 17164 chromosome 1, complete sequence
>NZ_CP027260.1:c4043005-4041617 Pectobacterium parmentieri strain IFB5427 chromosome, complete genome
>NZ_CP027723.1:2084906-2086060 Pseudomonas orientalis strain 8B chromosome, complete genome

>NZ_CP027756.1:c2565757-2564579 Pseudomonas synxantha strain R6-28-08 chromosome, complete genome

>NZ_CP027786.1:c687261-685996 Tetragenococcus koreensis strain KCTC 3924 chromosome, complete genome

>NZ_CP027793.1:783198-784394 Rhodococcus equi strain DSSKP-R-001 chromosome, complete genome

>NZ_CP028252.1:24342-25439 Leuconostoc mesenteroides strain SRCM102733 plasmid unnamed1, complete sequence

>NZ_CP029373.1:2040581-2043301 Acidovorax citrulli strain M6 chromosome, complete genome

>NZ_CP029451.1:472676-474664 Sinorhizobium fredii CCBAU 25509 chromosome, complete genome

>NZ_CP030880.1:c691420-688892 Acinetobacter haemolyticus strain HW-2A chromosome, complete genome

>NZ_CP031253.1:1652421-1652690 Neisseria lactamica strain M17106 chromosome, complete genome

>NZ_CP031560.1:c4014673-4013033 Dickeya dianthicola strain ME23 chromosome, complete genome

>NZ_CP031611.1:c967625-966255 Campylobacter hepaticus strain HV10 chromosome, complete genome

>NZ_CP032221.1:79762-81117 Rhodococcus rhodochrous strain EP4 chromosome

>NZ_CP032365.1:c3115598-3112401 Bacillus wiedmannii strain SR52 chromosome, complete genome

>NZ_CP032617.1:c528524-527157 Bradyrhizobium diazoefficiens strain 110spc4 chromosome, complete genome

>NZ_CP032746.1:c27284-26025 Lactiplantibacillus paraplantarum strain DSM 10667 plasmid unnamed2, complete sequence

>NZ_CP033022.1:c948669-947422 Agrobacterium fabrum strain 1D132 chromosome circular, complete sequence

>NZ_CP033031.1:c707883-707707 Agrobacterium tumefaciens strain 12D1 chromosome circular, complete sequence

>NZ_CP033724.1:137188-138303 Clavibacter michiganensis subsp. michiganensis strain UF1 chromosome, complete genome

>NZ_CP034181.1:447950-449248 Mycobacteroides abscessus strain GZ002 chromosome, complete genome

>NZ_CP034655.1:2275532-2276785 Xanthomonas campestris pv. musacearum NCPPB 4379 chromosome, complete genome

>NZ_CP034725.1:4031626-4032762 Pseudomonas brassicacearum strain 3Re2-7 chromosome, complete genome

>NZ_CP034943.1:231367-232620 Bacillus subtilis subsp. spizizenii ATCC 6633 = JCM 2499 strain ATCC 6633 chromosome, complete genome

>NZ_CP035504.1:2636113-2637399 Kocuria indica strain CE7 chromosome, complete genome

>NZ_CP035901.1:c1001936-1000725 Burkholderia glumae strain 257sh-1 chromosome 2, complete sequence

>NZ_CP035997.1:152311-153546 Bacillus mycoides strain BPN36/3 chromosome, complete genome

>NZ_CP035997.1:5050498-5053695 Bacillus mycoides strain BPN36/3 chromosome, complete genome

>NZ_CP038034.1:c2444376-2443006 Achromobacter insolitus strain LCu2 chromosome, complete genome

>NZ_CP038034.1:c702621-701317 Achromobacter insolitus strain LCu2 chromosome, complete genome

>NZ_CP038034.1:c3249940-3247586 Achromobacter insolitus strain LCu2 chromosome, complete genome

>NZ_CP038663.1:c1748754-1747417 Deinococcus radiodurans ATCC 13939 chromosome I, complete sequence

>NZ_CP038855.1:c6636-5446 Pantoea vagans strain LMG 24199 plasmid pVag2, complete sequence

>NZ_CP038996.1:c2403462-2402203 Enterococcus faecium strain SRR24 chromosome, complete genome

>NZ_CP040105.1:c2715004-2712458 Acinetobacter nosocomialis M2 chromosome, complete genome

>NZ_CP040829.1:c3702865-3699689 Paenibacillus polymyxa strain ZF129 chromosome, complete genome

>NZ_CP043146.1:2422335-2423666 Bordetella holmesii strain H401 chromosome, complete genome

>NZ_CP043317.1:c17612-16428 Streptomyces olivaceus strain SCSIO T05 chromosome, complete genome

>NZ_CP043404.1:c1603849-1602644 Bacillus safensis strain PgKB20 chromosome, complete genome

>NZ_CP043428.1:287946-289319 Campylobacter volucris strain LMG 24380 chromosome, complete genome

>NZ_CP043953.1:c667087-664541 Acinetobacter baumannii strain K09-14 chromosome, complete genome

>NZ_CP044211.1:20472-21740 Rhodococcus ruber strain C1 chromosome, complete genome

>NZ_CP044483.1:c774127-771491 Acinetobacter schindleri strain HZE30-1 chromosome, complete genome

>NZ_CP045198.1:1614112-1616712 Acinetobacter indicus strain TQ23 chromosome, complete genome

>NZ_CP045927.1:847990-849267 Staphylococcus agnetis strain 1379 chromosome, complete genome

>NZ_CP046317.1:112201-113571 Campylobacter coli strain FDAARGOS_735 chromosome, complete genome

>NZ_CP046570.1:c2125104-2123845 Xanthomonas albilineans strain Xa-FJ1 chromosome, complete genome

>NZ_CP046590.1:c268505-267228 Macrococcus canis strain LI021 chromosome, complete genome

>NZ_CP047242.1:c3739250-3737844 Trichormus variabilis 0441 chromosome, complete genome

>NZ_CP047242.1:2253045-2254367 Trichormus variabilis 0441 chromosome, complete genome

>NZ_CP047495.1:2508611-2509999 Pectobacterium brasiliense strain 1692 chromosome, complete genome

>NZ_CP048261.1:275975-277207 Streptomyces rimosus subsp. rimosus ATCC 10970 chromosome, complete genome

>NZ_CP048832.1:c999110-997668 Janthinobacterium lividum strain EIF1 chromosome, complete genome

>NZ_CP049019.1:c2462625-2459428 Bacillus tropicus strain AOA-CPS1 chromosome

>NZ_CP049134.1:1989509-1990969 Paraburkholderia tropica strain IAC135 chromosome A, complete sequence

>NZ_CP049357.1:c2572432-2571287 Deinococcus wulumuqiensis R12 chromosome, complete genome

>NZ_CP049603.1:c3548504-3547224 Rouxiella badensis strain SER3 chromosome

>NZ_CP051487.1:c4204536-4203259 Pseudomonas umsongensis strain CY-1 chromosome, complete genome

>NZ_CP051652.1:834094-835482 Pectobacterium carotovorum strain WPP14 chromosome, complete genome

>NZ_CP051772.1:c331461-330094 Mesorhizobium japonicum R7A chromosome, complete genome

>NZ_CP053391.1:c678015-675433 Acinetobacter lactucae isolate QL-1 chromosome, complete genome

>NZ_CP053649.1:2522782-2523366 Xanthomonas axonopodis pv. vasculorum strain NCPPB 796 chromosome, complete genome

>NZ_CP053825.1:277339-278706 Campylobacter armoricus strain CCUG 73571 chromosome, complete genome

>NZ_CP053828.1:c1652156-1650798 Campylobacter hyointestinalis subsp. lawsonii strain CHY5 chromosome, complete genome

>NZ_CP053849.1:1502186-1503547 Campylobacter upsaliensis RM3940 chromosome, complete genome

>NZ_CP053856.1:c884179-882932 Rhizobium pusense strain 76 chromosome R76C1, complete sequence

>NZ_CP053986.1:1548699-1551053 Achromobacter denitrificans strain FDAARGOS_788 chromosome, complete genome

>NZ_CP053989.1:c3513061-3512630 Niallia circulans strain FDAARGOS_783 chromosome, complete genome

>NZ_CP054599.1:c737146-735770 Pseudosulfitobacter pseudonitzschiae strain H46 chromosome, complete genome
>NZ_CP054624.1:78590-79741 Cupriavidus gilardii strain FDAARGOS_639 chromosome 1, complete sequence
>NZ_CP054795.1:c114811-113372 Mycolicibacterium smegmatis strain FDAARGOS_679 chromosome
>NZ_CP054795.1:c4891392-4890187 Mycolicibacterium smegmatis strain FDAARGOS_679 chromosome
>NZ_CP056080.1:c1679595-1678306 Rothia nasimurium strain E1706032 chromosome, complete genome
>NZ_CP058243.1:1337053-1338246 Xanthomonas campestris pv. raphani strain MAFF106181 chromosome, complete genome
>NZ_CP058277.1:546910-548184 Mycobacterium marinum strain MMA1 chromosome, complete genome
>NZ_CP058354.1:334847-336214 Bradyrhizobium japonicum strain 5038 chromosome, complete genome
>NZ_CP059082.1:1460234-1462585 Halomonas titanicae strain SOB56 chromosome, complete genome
>NZ_CP060273.1:c37043-35775 Priestia flexa strain SSAI1 plasmid unnamed1, complete sequence
>NZ_CP061079.1:c2599920-2598778 Pseudomonas chlororaphis strain qlu-1 chromosome, complete genome
>NZ_CP062148.1:6192-7358 Novacetimonas hansenii strain C110 plasmid pKHC110_2, complete sequence
>NZ_CP062158.2:2106392-2107549 Pseudomonas lundensis strain 2T.2.5.2 chromosome, complete genome
>NZ_CP063993.1:3529110-3530150 Xanthomonas translucens pv. undulosa strain XtLr8 chromosome, complete genome
>NZ_CP064063.1:164910-166304 Brucella anthropi strain PBO chromosome 2, complete sequence
>NZ_CP064875.1:c1820618-1819413 Bacillus toyonensis strain P18 chromosome, complete genome
>NZ_CP065044.1:c4601163-4597984 Pectobacterium aroidearum strain L6 chromosome, complete genome
>NZ_CP065253.1:48153-49514 Streptomyces clavuligerus strain F1D7 chromosome, complete genome
>NZ_CP065534.1:754650-755789 Lonsdalea populi strain N-5-1 chromosome, complete genome
>NZ_CP065640.1:3477693-3478919 Serratia rubidaea strain FDAARGOS_926 chromosome, complete genome
>NZ_CP065682.1:c1425925-1424720 Brevibacterium casei strain FDAARGOS_902 chromosome, complete genome
>NZ_CP065729.1:c735388-734111 Macrococcus caseolyticus strain FDAARGOS_868 chromosome, complete genome
>NZ_CP065820.1:c719446-716897 Acinetobacter seifertii strain S21 chromosome, complete genome
>NZ_CP065921.1:859578-860777 Staphylococcus pseudintermedius strain SP_11304-3A chromosome, complete genome
>NZ_CP066038.1:441818-442999 Burkholderia ambifaria strain FDAARGOS_1027 chromosome 2, complete sequence
>NZ_CP066119.1:3922861-3925503 Acinetobacter bereziniae strain GD03185 chromosome, complete genome
>NZ_CP066699.1:c463381-461996 Rhodopseudomonas palustris strain RCB100 chromosome, complete genome
>NZ_CP067086.1:c1763234-1760883 Comamonas testosteroni strain G1 chromosome
>NZ_CP068049.1:c2072586-2071390 Burkholderia gladioli strain BBB-01 chromosome 1, complete sequence
>NZ_CP068061.1:666724-668013 Mammaliicoccus vitulinus strain FDAARGOS_1153 chromosome, complete genome
>NZ_CP068161.1:c1746934-1745744 Corynebacterium propinquum strain FDAARGOS_1112 chromosome, complete genome
>NZ_CP068168.1:c2104667-2103840 Corynebacterium amycolatum strain FDAARGOS_1108 chromosome, complete genome
>NZ_CP068998.1:c679937-678594 Sulfitobacter mediterraneus strain SC7-37 chromosome, complete genome
>NZ_CP069587.1:c1261547-1260453 Chromobacterium violaceum strain FDAARGOS_1273 chromosome, complete genome
>NZ_CP070242.1:c348407-347187 Streptomyces californicus strain FDAARGOS_1209 chromosome
>NZ_CP071454.1:1869485-1870897 Rhizobium lentis strain BLR27 chromosome, complete genome
>NZ_CP071883.1:1692209-1693453 Curtobacterium flaccumfaciens pv. flaccumfaciens strain BRIP:70614 chromosome, complete genome
>NZ_CP072268.1:2493113-2494321 Xanthomonas euvesicatoria pv. alfalfae strain CFBP3836 chromosome, complete genome
>NZ_CP072387.1:c639509-638157 Cellulosimicrobium cellulans strain ORNL-0100 chromosome, complete genome
>NZ_CP072549.1:c2975462-2972832 Acinetobacter lwoffii strain H7 chromosome, complete genome
>NZ_CP072888.1:1224411-1225670 Enterococcus raffinosus strain F162_2 chromosome, complete genome
>NZ_CP073011.1:392326-393585 Virgibacillus pantothenticus strain DSM 26 chromosome, complete genome
>NZ_CP073241.1:398383-399603 Acinetobacter soli strain M3-1-68 chromosome, complete genome
>NZ_CP073250.1:c2525943-2524591 Cylindrospermopsis raciborskii N8 chromosome, complete genome
>NZ_CP074376.1:c59443-58130 Mycolicibacterium neoaurum strain MN2019 chromosome, complete genome
>NZ_CP077367.1:24647-25837 Pantoea agglomerans strain FDAARGOS 1447 plasmid unnamed1, complete sequence
>NZ_CP079880.1:c2255736-2254477 Enterococcus lactis strain CX 2-6_2 chromosome, complete genome
>NZ_CP080954.1:163407-164636 Rhodococcus opacus PD630 chromosome, complete genome
>NZ_CP081178.1:3983476-3984600 Pseudomonas mandelii strain KGI_MA19 chromosome, complete genome
>NZ_CP082886.1:4023501-4024037 Bacteroides nordii strain FDAARGOS_1461 chromosome, complete genome
>NZ_CP084662.1:c3155426-3154167 UNVERIFIED_ORG: Clostridium sporogenes strain FDAARGOS_1532 chromosome
>NZ_CP085200.1:177695-178936 Mycobacterium ulcerans strain JKD8049 chromosome, complete genome
>NZ_CP085785.1:985900-987111 Synechococcus elongatus PCC 6301 chromosome, complete genome
>NZ_CP086102.1:c83656-82406 Streptomyces anulatus strain YINM00001 chromosome, complete genome
>NZ_CP086328.1:c2633252-2632038 Bacillus pacificus strain anQ-h4 chromosome, complete genome
>NZ_CP086328.1:2110832-2114029 Bacillus pacificus strain anQ-h4 chromosome, complete genome
>NZ_CP087678.1:789775-791004 Xylella taiwanensis strain PLS206 chromosome, complete genome
>NZ_CP091882.1:c343508-342123 Peribacillus frigoritolerans strain JHS1 chromosome, complete genome
>NZ_CYPU01000039.1:249154-250404 Ruegeria atlantica strain CECT 4292, whole genome shotgun sequence
>NZ_CYYA01000001.1:c18452-17133 Eubacterium ramulus strain 2789STDY5608891, whole genome shotgun sequence
>NZ_FNJH01000006.1:291914-294436 Pseudomonas congelans strain DSM 14939, whole genome shotgun sequence
>NZ_FNKR01000003.1:c879355-877130 Pseudomonas gessardii strain LMG 21604, whole genome shotgun sequence
>NZ_FPLG01000008.1:30182-31978 Moritella viscosa isolate LFI 5006, whole genome shotgun sequence
>NZ_FWZG01000048.1:c3777-580 Bacillus mobilis strain 16-00177, whole genome shotgun sequence
>NZ_FWZG01000102.1:c16071-14857 Bacillus mobilis strain 16-00177, whole genome shotgun sequence

>NZ_FYAX01000013.1:477192-478553 Cupriavidus metallidurans isolate NDB4MOL1, whole genome shotgun sequence

>NZ_FZPB01000001.1:1777153-1779678 Stenotrophomonas lactitubi strain YR347, whole genome shotgun sequence

>NZ_FZPB01000002.1:c773480-771876 Stenotrophomonas lactitubi strain YR347, whole genome shotgun sequence

>NZ_HE997646.1:c29013-27700 Yersinia massiliensis CCUG 53443, whole genome shotgun sequence

>NZ_HG938353.1:c1125164-1123896 Neorhizobium galegae bv. orientalis str. HAMBI 540 chromosome I, complete sequence

>NZ_HG964936.1:c466850-465615 Mycobacterium asiaticum DSM 44297, whole genome shotgun sequence

>NZ_HG999362.1:c3617904-3616711 Xanthomonas arboricola pv. juglandis isolate Xanthomonas arboricola pv. juglandis CPBF 1494 isolated from C. illinoinensis chromosome 1, complete sequence

>NZ_JAAGEP010000088.1:14615-15871 Paraburkholderia aspalathi strain R-69781 contig00088, whole genome shotgun sequence

>NZ_JAASRC010000001.1:799458-800651 Xanthomonas arboricola strain CFBP 6825 Ga0372521_01, whole genome shotgun sequence

>NZ_JAAXQT010000003.1:c570144-568900 Rhizobium laguerreae strain TLR6 Scaffold_3, whole genome shotgun sequence

>NZ_JABSUS010000002.1:c340633-339248 Streptomyces lunaelactis strain MMun143 NODE_2_length_690312_cov_17.17, whole genome shotgun sequence

>NZ_JABXWH010000001.1:780696-782177 Streptomyces caniscabiei strain ND05-3B NODE_1_length_932813_cov_28.566274, whole genome shotgun sequence

>NZ_JACVKS010000081.1:4911-6287 Francisella noatunensis strain FSC1145 FSC1145_80_len6451, whole genome shotgun sequence

>NZ_JADNLF010000004.1:c272510-271203 Bacteroides clarus strain 1001095IJ_161003_B3 NODE_4_length_295142_cov_25.5295, whole genome shotgun sequence

>NZ_JAEIKQ010000025.1:c118008-116830 Pseudomonas carnis strain S1_2A YA0719_25, whole genome shotgun sequence

>NZ_JAEUBJ010000003.1:932898-934178 Mammaliicoccus lentus strain GFQ9D124P GFQ9D124P_contig_3, whole genome shotgun sequence

>NZ_JAFBRL010000001.1:112311-113501 Sulfitobacter geojensis strain TR60-85 GNM50492__scaffold1, whole genome shotgun sequence

>NZ_JAFCJS010000002.1:363342-364757 Bradyrhizobium liaoningense strain SZCCT0400 NODE_2, whole genome shotgun sequence

>NZ_JAFIMT010000001.1:679963-681162 Corallococcus exiguus strain NCCRE002 000000F_arrow, whole genome shotgun sequence

>NZ_JAGJ01000004.1:199902-200273 Cellulosimicrobium cellulans J36 CelceDRAFT_scaffold_3.4_C, whole genome shotgun sequence

>NZ_JAGJBY010000001.1:c146479-145277 Streptomyces acidiscabies strain NRRL B-16521 1, whole genome shotgun sequence

>NZ_JAHUVX010000001.1:206296-207579 Nocardia nova strain LGO-A14 A14_AS2_SC01, whole genome shotgun sequence

>NZ_JAINWB010000038.1:c19488-18664 Thalassospira xiamenensis strain IOP_1 contig38, whole genome shotgun sequence

>NZ_JAIQWY010000001.1:743133-746327 Cupriavidus pauculus strain MF1 contig001, whole genome shotgun sequence

>NZ_JAIVLB010000027.1:18449-19645 Bacillus stratosphericus strain KcN0-8 KcN0-8_ctg027, whole genome shotgun sequence

>NZ_JAIZPY010000003.1:830355-831794 Burkholderia cepacia strain AU41368 NODE_3_length_964771_cov_43.904678, whole genome shotgun sequence

>NZ_JAJIWN010000013.1:c19269-18088 Xanthomonas perforans strain DC05T6 Contig_13_111.816, whole genome shotgun sequence

>NZ_JAKJKO010000001.1:c1243839-1240606 Corynebacterium parakroppenstedtii strain MC-29 NODE_1_length_1486533_cov_243.274920, whole genome shotgun sequence

>NZ_JFKD01000005.1:c117546-116191 Marivita cryptomonadis strain CL-SK44 contig5, whole genome shotgun sequence

>NZ_JH994110.1:c194706-193543 Brachyspira hampsonii 30446 isolate IDAC 161111-01 scaffold00003, whole genome shotgun sequence

>NZ_JIAH01000008.1:207431-208621 Corynebacterium pseudodiphtheriticum DSM 44287 O146DRAFT_scaffold00004.4_C, whole genome shotgun sequence

>NZ_JPJK01000044.1:c29123-28059 Gallibacterium anatis strain 12158-5 contig000038, whole genome shotgun sequence

>NZ_JXSK01000001.1:c569192-567867 Photorhabdus luminescens subsp. luminescens strain DSM 3368 Contig001, whole genome shotgun sequence

>NZ_JXYX01000001.1:1405369-1406673 Microcystis aeruginosa NIES-88 scaffold1, whole genome shotgun sequence

>NZ_JYLD01000002.1:276917-279454 Pseudomonas helleri strain DSM 29165 2_617688_39.188, whole genome shotgun sequence

>NZ_KB849557.1:145883-148426 Acinetobacter venetianus RAG-1 = CIP 110063 acLse-supercont1.2, whole genome shotgun sequence

>NZ_KB849705.1:855940-858603 Acinetobacter johnsonii ANC 3681 acLro-supercont1.4, whole genome shotgun sequence

>NZ_KB850260.1:c101663-100176 Acinetobacter ursingii NIPH 706 acLZz-supercont1.15, whole genome shotgun sequence

>NZ_KB912923.1:c162073-160709 Duganella zoogloeoides ATCC 25935 F460DRAFT_scaffold00005.5, whole genome shotgun sequence

>NZ_KB913036.1:c217073-215904 Salinispora arenicola DSM 45545 strain CNS-991 scaffold1, whole genome shotgun sequence

>NZ_KE340376.1:165186-167723 Acinetobacter colistiniresistens strain NIPH 2036 acVBK-supercont1.11, whole genome shotgun sequence

>NZ_KI973153.1:c3110597-3109377 Enterobacter kobei strain UCI 24 adjce-supercont-complete, whole genome shotgun sequence

>NZ_KN046795.1:6640868-6643486 Delftia acidovorans strain 2167 scaffold1, whole genome shotgun sequence

>NZ_KV757140.1:c171341-169908 Streptomyces hygroscopicus subsp. hygroscopicus strain OsiSh-2 Scaffold2, whole genome shotgun sequence

>NZ_KZ846567.1:c1183511-1182228 Enterococcus gallinarum strain 298EA1 aekeE-supercont-complete, whole genome shotgun sequence

>NZ_LAHB01000001.1:23459-24850 Nostoc linckia z6 scaffold1.1, whole genome shotgun sequence

>NZ_LAWW01000004.1:c272479-269930 Pseudomonas fulva strain YAB-1 contig4, whole genome shotgun sequence

>NZ_LCZH01000011.1:730567-731952 Delftia tsuruhatensis strain MTQ3 scaffold2_split2, whole genome shotgun sequence

>NZ_LCZH01000019.1:2155-4773 Delftia tsuruhatensis strain MTQ3 scaffold3_split7, whole genome shotgun sequence

>NZ_LCZH01000034.1:c28633-26237 Delftia tsuruhatensis strain MTQ3 scaffold8, whole genome shotgun sequence

>NZ_LDIP01000011.1:169890-171050 Bacillus sonorensis strain G25-136 G25-136_contig000011, whole genome shotgun sequence

>NZ_LDWY01000085.1:c12564-11206 Campylobacter vulpis strain 73/13 seq_10, whole genome shotgun sequence

>NZ_LGDU01000036.1:44457-44831 Streptomyces sp. AS58 P432contig13.1, whole genome shotgun sequence

>NZ_LGUE01000004.1:144101-147259 Rossellomorea marisflavi strain JCM 11544 scaffold2, whole genome shotgun sequence

>NZ_LMVL02000003.1:c281897-280653 Agrobacterium vitis strain NCPPB 3554 contig3, whole genome shotgun sequence

>NZ_LOWL01000032.1:c27444-26257 Burkholderia vietnamiensis strain FL-2-3-10-S3-D0 FL-2-3-10-S3-D0_39, whole genome shotgun sequence

>NZ_LQOP01000004.1:c89795-88566 Mycolicibacterium conceptionense strain CCUG 50187 contig_12, whole genome shotgun sequence

>NZ_LR025743.1:543127-544434 Burkholderia stabilis isolate E chromosome II, complete sequence

>NZ_LR134326.1:c1373890-1372733 Bordetella bronchiseptica strain NCTC10543 chromosome 1, complete sequence

>NZ_LR134338.1:c1625263-1624040 Brevibacillus brevis strain NCTC2611 chromosome 1, complete sequence

>NZ_LR134406.1:c2213779-2212682 Arachnia propionica strain NCTC12967 chromosome 1, complete sequence

>NZ_LS483377.1:c1103123-1101519 Stenotrophomonas maltophilia strain NCTC10258 chromosome 1, complete sequence

>NZ_LS999205.1:c2826590-2825451 Pseudomonas protegens CHA0 chromosome 1, complete sequence

>NZ_LT593929.1:c2448320-2447106 Propionibacterium freudenreichii isolate PFRJS14 chromosome I, complete sequence

>NZ_LT629687.1:172030-173133 Pseudomonas koreensis strain LMG 21318 chromosome I

>NZ_LT629702.1:c249575-247350 Pseudomonas azotoformans strain LMG 21611 chromosome I

>NZ_LT629706.1:2659850-2662375 Pseudomonas poae strain LMG 21465 chromosome I

>NZ_LT629788.1:c1125675-1124566 Pseudomonas moraviensis strain LMG 24280 chromosome I

>NZ_LT853882.1:224759-227278 Xanthomonas fragariae strain PD885 chromosome 1, complete sequence

>NZ_LT906453.1:c91101-89842 Dermatophilus congolensis strain NCTC13039 chromosome 1, complete sequence

>NZ_LT907842.1:c5438711-5436531 Pseudomonas fluorescens strain ATCC 13525 chromosome I

>NZ_LWAG01000001.1:189396-192554 Priestia endophytica strain 3617_2C 3617_2C_contig_1, whole genome shotgun sequence

>NZ_LWAG01000005.1:c17285-16053 Priestia endophytica strain 3617_2C 3617_2C_contig_102, whole genome shotgun sequence

>NZ_LWAG01000027.1:c16411-15158 Priestia endophytica strain 3617_2C 3617_2C_contig_30, whole genome shotgun sequence

>NZ_LWAG01000057.1:c14185-11003 Priestia endophytica strain 3617_2C 3617_2C_contig_59, whole genome shotgun sequence

>NZ_LWAG01000075.1:10828-12000 Priestia endophytica strain 3617_2C 3617_2C_contig_76, whole genome shotgun sequence

>NZ_LYMI01000004.1:c16589-15549 Xanthomonas nasturtii strain WHRI 8853 scf_22201_12, whole genome shotgun sequence

>NZ_MLIK01000003.1:86201-87583 Mycobacteroides franklinii strain 1559 NODE_6_length_91717_cov_35.6471, whole genome shotgun sequence

>NZ_MPRU01000019.1:c27157-25946 Solemya velum gill symbiont isolate NC-DML14 DML14_sym_scf18, whole genome shotgun sequence

>NZ_MRTI01000001.1:c472008-470806 Paenibacillus amylolyticus strain FSL H7-0692 NODE_1_length_1366088_cov_1.56855_ID_2737, whole genome shotgun sequence

>NZ_MSCQ01000001.1:c888920-888669 Photobacterium phosphoreum strain JCM 21184 scaffold00001, whole genome shotgun sequence

>NZ_MTLN01000009.1:c51503-50241 Pseudomonas psychrotolerans strain SDS18 SO_6301contig_9, whole genome shotgun sequence

>NZ_MWQA01000001.1:c13059-11869 Mycobacterium persicum strain 12MK PseudoContig_CP009483.1, whole genome shotgun sequence

>NZ_MWUT01000001.1:c142781-141513 Staphylococcus delphini strain 215100905101-2 NODE_1_length_422887_cov_81.2253_ID_1, whole genome shotgun sequence

>NZ_NBSL01000002.1:c1529481-1528222 Enterococcus avium strain FDAARGOS_182 scf7180000000006_trim_quiver_pilon, whole genome shotgun sequence

>NZ_NIWA01000001.1:c369965-369456 Mameliella alba strain JL351 contig1, whole genome shotgun sequence

>NZ_NJQO01000006.1:c489786-486589 Bacillus paranthracis strain 14-9 14-9_R1_(paired)_contig_6, whole genome shotgun sequence

>NZ_NJQO01000013.1:c59464-58238 Bacillus paranthracis strain 14-9 14-9_R1_(paired)_contig_13, whole genome shotgun sequence

>NZ_NKAQ01000003.1:1003221-1004435 Priestia megaterium strain 22-2 22-2_R1_(paired)_contig_3, whole genome shotgun sequence

>NZ_NOIV01000004.1:243969-245165 Rothia kristinae strain ATCC 27570 NODE_4_length_259270_cov_327.474, whole genome shotgun sequence

>NZ_NPBN01000048.1:c34219-33011 Alkalihalobacillus clausii strain 7520-2 contig00048, whole genome shotgun sequence

>NZ_PEMR01000096.1:341-1504 Thermus scotoductus strain 6_S6 NODE_187_length_2988_cov_55.0696_ID_373, whole genome shotgun sequence

>NZ_PQNX01000007.1:5539-8757 Corynebacterium bovis strain 4826 7, whole genome shotgun sequence

>NZ_PTXV01000001.1:c842319-841123 Bacillus pumilus strain Ha06YP001 Contig_01, whole genome shotgun sequence

>NZ_PYLU01000001.1:167167-167418 Photobacterium iliopiscarium strain NCIMB 13355 CFSAN065522_1, whole genome shotgun sequence

>NZ_PZBL01000001.1:607408-608685 Staphylococcus chromogenes strain SNUC 1341 contig001, whole genome shotgun sequence

>NZ_QCXN01000010.1:80737-81951 Legionella taurinensis strain Genessee04 NODE_10_length_112531_cov_20.1295_ID_1363, whole genome shotgun sequence

>NZ_QGGH01000002.1:4837-9069 Mesorhizobium loti strain DSM 2626 Ga0215673_102, whole genome shotgun sequence

>NZ_QGHF01000001.1:c415007-414672 Pantoea allii strain PNA 200-10 Ga0215836_101, whole genome shotgun sequence

>NZ_RBIZ01000004.1:c107099-105882 Yokenella regensburgei strain DSM 5079 Ga0215677_102, whole genome shotgun sequence

>NZ_SDFS01000003.1:c299533-296336 Bacillus albus strain PG 26 NODE_3_length_681754_cov_41.0844, whole genome shotgun sequence

>NZ_SHMF01000001.1:c521437-518933 Pseudoxanthomonas winnipegensis strain NML 140781 NML140781_1, whole genome shotgun sequence

>NZ_SILH01000001.1:c1395479-1394235 Rhizobium leguminosarum strain SM52 chrom_SM52, whole genome shotgun sequence

>NZ_SMZT01000001.1:482832-484115 Kocuria rosea strain S-A3 NODE_1_length_935051_cov_277.960614, whole genome shotgun sequence

>NZ_SMZT01000002.1:416431-417690 Kocuria rosea strain S-A3 NODE_2_length_641051_cov_300.244472, whole genome shotgun sequence

>NZ_SPNK01000006.1:c101862-100576 Kocuria rhizophila strain 4R-31 NODE_6_length_164217_cov_317.048156, whole genome shotgun sequence

>NZ_UFTC01000001.1:c2818381-2816573 Cytobacillus firmus strain NCTC10335, whole genome shotgun sequence

>NZ_UFUJ01000001.1:c1051254-1049968 Bordetella trematum strain NCTC12995, whole genome shotgun sequence

>NZ_UGJF01000001.1:136916-138286 Helicobacter pullorum strain NCTC13156, whole genome shotgun sequence

>NZ_UGSH01000003.1:155391-156782 Brucella intermedia strain NCTC12171, whole genome shotgun sequence

>NZ_VARQ01000077.1:2767-4140 Vibrio tasmaniensis strain LMG 20012 645600077, whole genome shotgun sequence

>NZ_VDNQ01000008.1:c65721-64459 Pseudomonas oryzihabitans strain DE0585 NODE_8_length_154164_cov_28.407019, whole genome shotgun sequence

>NZ_VEKZ01000011.1:c76647-75391 Siminovitchia fortis strain DE0258 NODE_11_length_80562_cov_37.572740, whole genome shotgun sequence

>NZ_VIGJ01000001.1:1020134-1021609 Pseudoalteromonas luteoviolacea strain H2 1, whole genome shotgun sequence

>NZ_VISK01000002.1:c561117-559954 Azospirillum brasilense strain Sp 7 Ga0060187_unitig_3_quiver.2, whole genome shotgun sequence

>NZ_VITA01000004.1:409854-411254 Sinorhizobium medicae strain USDA1037 Ga0310605_104, whole genome shotgun sequence

>NZ_VLKI01000001.1:c546642-545452 Cytobacillus oceanisediminis strain CGMCC 1.10115 IQ19DRAFT_scaffold00001.1, whole genome shotgun sequence

>NZ_VSST01000004.1:37058-38278 Bradyrhizobium canariense strain BTA-1 BTA-1_S1_R1_contig_4, whole genome shotgun sequence

>NZ_VTFG01000001.1:c867737-866520 Pseudomonas marginalis strain PgKB35 contig1, whole genome shotgun sequence

>NZ_WBKI01000002.1:c218508-217444 Serratia proteamaculans strain CCUG 14510 contig_0000002, whole genome shotgun sequence

>NZ_WIFZ01000002.1:1011221-1011871 Pseudomonas balearica strain KOL14.W.20.10 Scaffold_2_length_3079602_pilon, whole genome shotgun sequence

>NZ_WMDN01000005.1:c164201-163950 Photobacterium carnosum strain TMW2.2147 5, whole genome shotgun sequence

>NZ_WQDA01000003.1:313999-315177 Ruegeria arenilitoris strain HKCCA0515 NODE_3, whole genome shotgun sequence

8.3.2   Accession numbers for *Bacteroides fragilis* fragilysin

>AB026626.1:345-1538 Bacteroides fragilis bft-2 gene for metalloprotease, complete cds

>CP098482.1:c5312584-5311391 Bacteroides fragilis strain 86-5443-2-2 chromosome, complete genome

>JAHYPF010000020.1:55345-56538 Bacteroides fragilis strain BJH_183 20, whole genome shotgun sequence

>JALFMY010000010.1:81936-83129 MAG: Bacteroides fragilis isolate SUG827 k141_67695, whole genome shotgun sequence

>JANUSS010000001.1:3806129-3807322 Bacteroides fragilis strain BFG-391 contig_6_segment0_pilon, whole genome shotgun sequence

>JANUTD010000001.1:2513275-2514468 Bacteroides fragilis strain BFG-482 contig_1_segment0, whole genome shotgun sequence

>JANUTH010000001.1:2513065-2514258 Bacteroides fragilis strain BFG-479 contig_1_segment0_pilon, whole genome shotgun sequence

>JANUTZ010000001.1:347095-348288 Bacteroides fragilis strain BFG-4 contig_1_segment0_pilon, whole genome shotgun sequence

>JANUTZ010000002.1:1981850-1983043 Bacteroides fragilis strain BFG-4 contig_2_segment0_pilon, whole genome shotgun sequence

>JANUUF010000001.1:347004-348197 Bacteroides fragilis strain BFG-439 contig_1_segment0_pilon, whole genome shotgun sequence

>JANUUF010000001.1:2527119-2528312 Bacteroides fragilis strain BFG-439 contig_1_segment0_pilon, whole genome shotgun sequence

>NZ_CP011073.1:4564579-4565772 Bacteroides fragilis strain BOB25 chromosome, complete genome

>NZ_CP098482.1:c5312584-5311391 Bacteroides fragilis strain 86-5443-2-2 chromosome, complete genome

>NZ_JAPUAE010000018.1:56787-57980 Bacteroides fragilis strain BF_BC_ODE_DK_2016 contig00018, whole genome shotgun sequence

>NZ_JGEF01000023.1:64376-65569 Bacteroides fragilis str. 20793-3 gbf207933.contig.22, whole genome shotgun sequence

>NZ_JH724206.1:c3476585-3475392 Bacteroides fragilis CL07T00C01 supercont1.1, whole genome shotgun sequence

>NZ_JH724218.1:c72677-71484 Bacteroides fragilis CL07T12C05 supercont1.4, whole genome shotgun sequence

>NZ_LIDS01000027.1:55915-57108 Bacteroides fragilis strain 86-5443-2-2 contig00027, whole genome shotgun sequence

>NZ_LIDT01000031.1:c8881-7688 Bacteroides fragilis strain 20793-3 contig00031, whole genome shotgun sequence

>NZ_LIDV01000087.1:c315654-314461 Bacteroides fragilis strain 2-078382-3 contig00089, whole genome shotgun sequence

>NZ_PDCT01000007.1:c60762-59569 Bacteroides fragilis strain CM13 contig0007, whole genome shotgun sequence

>CP098482.1:2670205-2671398 Bacteroides fragilis strain 86-5443-2-2 chromosome, complete genome

8.4 Supplementary data for chapter 4

Accession number (Genbank)	Genome location (bp-bp), (c, chromosome)
AB026626.1	345-1538
CP098482.1	2670205-2671398
CP098482.1	c5312584-5311391
JAHYPF010000020.1	55345-56538
JALFMY010000010.1	81936-83129
JANUSS010000001.1	3806129-3807322
JANUTD010000001.1	2513275-2514468
JANUTH010000001.1	2513065-2514258
JANUTZ010000001.1	347095-348288
JANUTZ010000002.1	1981850-1983043
JANUUF010000001.1	2527119-2528312
JANUUF010000001.1	347004-348197
NZ_CP011073.1	4564579-4565772
NZ_CP098482.1	c5312584-5311391
NZ_JAPUAE010000018.1	56787-57980
NZ_JGEF01000023.1	64376-65569
NZ_JH724206.1	c3476585-3475392
NZ_JH724218.1	c72677-71484
NZ_LIDS01000027.1	55915-57108
NZ_LIDT01000031.1	c8881-7688
NZ_LIDV01000087.1	c315654-314461
NZ_PDCT01000007.1	c60762-59569

Table 8:1:Genbank accession numbers for the Bft, fragilysin, gene in Bacteroides fragilis.

8.5 Peer reviewed articles

REVIEW

IJC
International Journal of Cancer

Exploring the impact of gut microbiota and diet on breast cancer risk and progression

Nancy M. Y. Teng[1] | Christopher A. Price[1] | Alastair M. McKee[1] |
Lindsay J. Hall[1,2,3] | Stephen D. Robinson[1,4]

[1]Gut Microbes & Health, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK

[2]Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, UK

[3]Chair of Intestinal Microbiome, School of Life Sciences, ZIEL-Institute for Food & Health, Technical University of Munich, Freising, Germany

[4]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, UK

Correspondence
Stephen D. Robinson, Quadram Institute Bioscience & University of East Anglia, Norwich, UK.
Email: stephen.robinson@quadram.ac.uk

Funding information
BigC Cancer Charity, Grant/Award Numbers: 17-16R, 18-15R; Biotechnology and Biological Sciences Research Council, Grant/Award Numbers: BB/R012490/1, BBS/E/F/000PR10353, BBS/E/F/000PR10356; Breast Cancer Now, Grant/Award Number: 2017NovPhD973; Wellcome Trust, Grant/Award Number: 100/974/C/13/Z

Abstract

There is emerging evidence that resident microbiota communities, that is, the microbiota, play a key role in cancer outcomes and anticancer responses. Although this has been relatively well studied in colorectal cancer and melanoma, other cancers, such as breast cancer (BrCa), have been largely overlooked to date. Importantly, many of the environmental factors associated with BrCa incidence and progression are also known to impact the microbiota, for example, diet and antibiotics. Here, we explore BrCa risk factors from large epidemiology studies and microbiota associations, and more recent studies that have directly profiled BrCa patients' gut microbiotas. We also discuss how in vivo studies have begun to unravel the immune mechanisms whereby the microbiota may influence BrCa responses, and finally we examine how diet and specific nutrients are also linked to BrCa outcomes. We also consider future research avenues and important considerations with respect to study design and implementation, and we highlight some of the important unresolved questions, which currently limit our overall understanding of the mechanisms underpinning microbiota-BrCa responses.

KEYWORDS
antibiotics, breast cancer, diet, immune, microbiota

1 | INTRODUCTION

Annually, breast cancer (BrCa) is predicted to affect over 2 million new patients, with more than 600 000 BrCa-related deaths worldwide, second only to lung cancer in incidence and mortality.[1] The financial implications of the disease on patients and health services are equally staggering with average treatment costs ranging from

~£22 000 to £115 000 for a single patient depending on disease stage.[2] Moreover, while patients diagnosed in early stages usually have good prognostic outcomes, those diagnosed in late stages of the disease have very poor 5-year survival rates, less than 30%.[2] Thus, understanding the factors that drive BrCa development and progression is important not only for patient outcomes but also for alleviating financial burdens on healthcare systems.

BrCa is an extremely heterogeneous disease and thus traditional therapeutic approaches are dependent on disease classification. There are six molecular subtypes of the disease, luminal A, luminal B (HER2+ of HER2−), HER2-enriched, normal-like and basal-like or triple-negative breast cancer (TNBC) (Table 1),

which are classified according to the expression of several proteins.[3-6]

Crucially, expression of these proteins determines which therapies clinicians employ to treat the disease such as neoadjuvant hormone therapies, chemotherapies and/or radiotherapies.[7,8] Successful treatment outcomes rely on efficient activation of anticancer responses, therefore delineating factors that beneficially or negatively impact associated immune responses are key. One such emerging modulator of BrCa aetiology is the gut microbiota, which may represent a viable therapeutic target for altering the course of the disease.

2 | BrCa RISK FACTORS: IS THE MICROBIOTA A MISSING LINK?

Alongside known genetic factors, all cancers are considered to have an "environmental" element associated with increased risk and disease progression. These associations are often uncovered in large human epidemiological studies that correlate lifestyle factors (eg, smoking, drug exposure and diet) with cancer onset and clinical outcomes.[9-11] As many of these factors are also known to alter the gut microbiota, population-based association studies at least intimate that the gut microbiota dictates cancer outcomes.[9,12,13] A stable and diverse microbial ecosystem (in adults) is considered optimal for health, although the exact taxa that confer beneficial effects is an ever changing conundrum, and is dependent on age, diet, medications, host genetics and numerous other external factors. In many cases, a reduction in alpha and beta diversity is associated with increased disease risk for many conditions, but correlations with specific microbial taxa may change between patient cohorts. In terms of "beneficial" bacterial members, those best studied include some of the following: *Bifidobacterium*, *Lactobacillus*, *Akkermanisa*, *Ruminococcus*, *Bacteroides* and *Faecalibacterium*; however, species and strain-level differences are key considerations. More recent studies indicate that overall function of the total microbial community may allow more "healthy" signatures to be found, as recently described for colorectal cancer (CRC) and choline degradation.[14] For reviews on this subject, see References 15 and 16. The gut-tumour connection includes sites known to have direct cross talk between the host and the gut microbiota (eg, CRC), but also in sites distal to

the gut (eg, the skin, liver and breast). It is likely that, what applies in one cancer setting is by no means universal, and BrCa, by its extreme heterogeneity and relative low incidence of genetic predisposition, is particularly unique. Consequently, there are many large studies focused on understanding how different environmental factors influence BrCa, and each of these factors influence (and are influenced by) the microbiota. These include the following:

1. *Diet*: In the 1990s, several groups investigated the association between diet and BrCa risk. For example, a low-fat diet elicited a lower risk of relapse after tumour resection.[15] Recent meta-analyses of cohort studies continue to correlate dietary patterns with BrCa risk.[16] This topic is covered in more detail in Section 3 of this review.

2. *Obesity*: Complementary to a low-fat diet, obesity is associated with increased risk of developing postmenopausal BrCa with a worse clinical outcome. Meta-analysis of nine studies showed increased BrCa risk with increased body mass index (BMI).[17] Associations between obesity and postmenopausal BrCa may be due to adipose tissue catalysing the formation of oestrogen after menopause, thereby increasing circulating oestrogen levels[17,18]; see Point (4).

3. *Alcohol consumption*: Excessive alcohol intake is also recognised as a risk factor for BrCa.[19] Although the specific molecular mechanisms driving this correlation remain unknown, ethanol may (a) induce molecular damage in mammary cells; (b) inhibit oestrogen-metabolising enzymes in the liver; (c) increase aromatase activity in the liver, which has been reported to facilitate the conversion of testosterone to oestrogen[20]; see Point (4).

4. *Changes in circulating hormonal levels*: Alongside uterine, ovarian and prostate cancers, some forms of BrCa are oestrogen driven. Both a late menarche and an early menopause decrease the risk of developing BrCa.[11] For a recent review on this subject, see Reference 21.

5. *Antibiotic exposure*: Use of antibiotics is becoming increasingly controversial, with unexpected adverse effects being reported in several disease contexts.[22-24] In 2004, Velicer et al concluded that cumulative days of antibiotic exposure were associated with increased risk of BrCa.[25] A follow-up study also showed that antibiotic use may be associated with less favourable tumour features.[26]

TABLE 1 BrCa subtypes defined by receptor status and proliferative potential as based on Ki67 expression

BrCa Subtype	Receptor status			Ki67 expression	Clinical prognosis
	ER	PR	HER2		
Luminal A	+	+	−	Low	Good
Normal-like	+/−	+/−	−	Low	Intermediate
Luminal B (HER2−)	+	+/−	−	Any	Intermediate
Luminal B (HER2+)	+	+/−	+	High	Intermediate
HER2-enriched	−	−	+	Any	Poor
Basal-like/TNBC	−	−	−	Separate basal markers used, for example, claudin	Poor

TABLE 2 BrCa risk factor and microbiota associations

Factor	Influence on the gut microbiota

Diet (covered in more detail in Section 3)	• Members of the microbiota can digest otherwise indigestible components of our diet (eg, dietary fibre)
	• Dietary fibre constituents can (a) boost nutritional intake, (b) act as a substrate for other microbiota members to colonise and (c) act as a metabolite[27]
	  ○ Short-chain fatty acids (SCFA), a constituent of metabolised dietary fibre, can module host immune responses
	  ○ Bioactive compounds, a constituent of metabolised polyphenols, encourage growth of beneficial bacteria for example, *Bifidobacterium* and *Lactobacillus* and production of SCFA[28,29]

Obesity	• Gut microbiota profiles differ among obese and lean patients and between those with metabolic syndrome[30] • In mice, studies showed that an obese microbiota profile had a greater nutritional intake capacity[31] ○ Members of an obese microbiota profile encoded enzymes that could more efficiently degrade polysaccharides

Alcohol	• Perturbations of the gut microbiota profile was observed in alcoholics vs nonalcoholics[32]
	  ○ This resulted in lower abundance of Bacteroidetes and higher abundance of Proteobacteria
	  ○ Alcoholics also had higher levels of serum endotoxin
	• Alcoholics tended to have greater gut permeability, which could lead to a local inflammatory state and disease, for example, alcohol-related liver disease[33]
	• It can be hypothesised that changes in microbiota members due to alcoholism alter the metabolites available by the host to use for other physiological processes including gut barrier function

Hormones	• In 1998, a group observed that germ-free mice, which do not have a gut microbiota, regained normal oestrous levels upon accidental bacterial contamination
	  ○ This suggested a link between gut bacteria and reproductive capacity[34]
	• Microbiota members possess ß-glucoronidase, which can deconjugate already metabolised oestrogen
	  ○ Thereby increasing levels of systemic oestrogen, increasing the risk of ER+ breast cancer[35]
	• A population-based study demonstrated an association between oestrogen metabolism and phylogenetic diversity of the gut microbiota, suggesting a link between the gut bacteria and circulating reproductive hormones[34]

Antibiotics	• Antibiotics severely impact the gut microbiota, most notably they reduce microbial diversity
	  ○ After depletion due to antibiotics it became easier for pathogenic bacteria, for example, *Salmonella* to colonise due to lack of competitive exclusion[36]
	  ○ The change in microbiota members consequently influenced the availability of metabolites used by the host, which could influence, for example, host immune responses[36]

One commonality between each of these risk factors is that they significantly alter the profile of the gut microbiota (see Table 2), suggesting a strong link between microbiota make-up and BrCa development. This suggests that a perturbed microbiota and therefore altered microbial-associated functions may impact BrCa risk. Generally, in terms of microbiome changes and duration of these alterations in response to different factors, certain "resilient" individuals may have a rapid but only transient change in microbiome profiles, while others may have large-scale changes that persist over many years (often compounded during the early life window), after introduction and removal of, for example, antibiotics and dietary interventions.[37] However, this is dependent on the baseline microbiota of individuals, and in many cases we do not understand the impact of these factors on specific microbial communities, species and strains, which may play a key and oversized beneficial role. Further studies are needed in this area, including within BrCa patient cohorts.

3 | GROWING EVIDENCE LINKING THE GUT MICROBIOTA AND BrCa

Changes in the abundance (ie, levels) of particular microbes have been associated with several cancers. A well-known example is *Helicobacter*

pylori, which initiates gastric inflammation and the formation of pre-cancerous lesions.[38] Likewise, increases in *Fusobacterium* spp. are associated with increased risk of developing CRC.[39]

Functional pathways linking specific microbiota members and BrCa have yet to be shown, but gut microbiota profiling of BrCa patients has indicated that there may be microbial signatures associated with disease stage and outcomes. Microbiota profiling methods and analysis tools are varied and there are important points to be considered when designing or interpreting these types of studies. Microbiota profiling of the Twins UK cohort, which comprises a large number of older women, indicated that BrCa incidence correlated with "disrupted" or a non-healthy microbiota signature,[40] suggesting that the gut microbiota may represent a useful biomarker and/or treatment focus for BrCa patients. Studies looking at postmenopausal BrCa patients have shown that these women had an enrichment of 38 species compared to postmenopausal controls, these included species positively (albeit) weakly associated with oestrogen metabolism (*Shewanella putrefaciens* and *Erwinia amylovora*), short-chain fatty acid butyrate producing species (*Roseburia inulinivorans*) and a species that was negatively but weakly associated with tumour-infiltrating immune cells (*Actinomyces* sp. HPA0247).[41] In a different study by Goedert et al, an altered microbiota signature in the same subtype of BrCa patients, that is, postmenopausal BrCa was observed; lower alpha diversity and increased relative abundance in

classes Clostridiaceae and Ruminococcaceae and in *Faecalibacterium*, and a relative decrease in abundance of *Dorea* and *Lachnospiraceae*.[42] In short, these studies suggest that postmenopausal BrCa patients do have an altered microbiota signature, which was reported in the Twin UK cohort study.[40] Guan et al assessed microbiota profiles of HER2 metastatic BrCa patients undergoing metronomic Capecitabine. They observed different microbiota profiles and reduced diversity in Capecitabine vs conventional patients. Further microbiota probing suggested that *Blautia obeum* was significantly associated with progression-free survival in Capecitabine, while *Slackia* was negatively associated with progression-free survival.[43] Wu et al recruited 37 BrCa patients and assessed the microbiota profile based on their tumour characteristics. The group observed that patients who were positive for HER2 had a significantly lower alpha diversity. In addition, they found that patients with a higher tumour grade were associated with a higher abundance of *Clostridium* and *Veillonella*, and a lower abundance of Erysipelotrichaceae.[45] Both *Veillonella* and Erysipelotrichaeceae have previously been reported to correlate with inflammatory conditions.[44] In the same study, the group assessed microbiota changes to recognised BrCa risk factors, for example, BMI and menarche. They observed that an early age of menarche was associated with a lower microbiota diversity—which may suggest a link with circulating oestrogen.[44] Furthermore, larger studies (considering other microbiome confounders) are required to explore patterns/relationships, which may be at the functional level rather than through shared taxa—as recently indicated in CRC (and choline degradation).[14]

Although longitudinal human studies facilitate important insights into the complex relationship between the gut microbiota and BrCa, there are numerous ethical and logistical issues that preclude their use for understanding detailed mechanisms. Thus, in vivo models are crucial to better define the underlying mechanisms driving specific observations under robust controlled conditions.

4 | ANIMAL STUDIES OF THE LINKS BETWEEN THE GUT MICROBIOTA AND BrCa

To date, there is still a relatively limited body of research exploring the role of the microbiota and different in vivo BrCa models, although interest is growing, and lessons learned from other cancers may be applicable in this underresearched area. Crucially, studies to date have indicated that microbiota modulation of the immune system may represent a key cross-roads determining disease and treatment outcomes. A distinct advantage of using preclinical models is the ability to explore the impact of microbiota on different BrCa subtypes, which is key given the heterogeneous nature of this cancer type (see Table 1).

4.1 | Evidence of microbiota involvement in non-BrCa disease

Understanding mechanistic links between the gut microbiota and BrCa is in its infancy. Thus, to better gauge the potential for the microbiota

to influence BrCa occurrence and progression, it is important to consider the larger body of literature confirming such links in non-BrCa disease. Disruption of gut homeostasis and effects on local inflammatory diseases have been well researched in animal models. Thus, it is unsurprising that CRCs were some of the first to be linked to changes in microbial communities, for example, increased *Fusobacterium nucleatum* abundance has been heavily associated with colorectal carcinogenesis (as mentioned earlier). However, without the use of animal models, it is difficult to conclude whether such changes are causative of disease or simply a product of it. Recently, animal studies have been able to explore these relationships in more detail, for example, Yu et al identified that subcutaneous xenograft tumours, derived from SW480 colon adenocarcinoma cells, intratumorally injected with *F. nucleatum* were resistant to oxaliplatin chemotherapy through an autophagy-dependent pathway,[46] suggesting a protumorigenic influence of the bacteria.

Microbes have also been observed to play antitumorigenic roles, which in many cases appears to be via education of the host immune system. One of the seminal studies in the microbiome cancer field identified an association between *Bifidobacterium* abundance and reduced melanoma tumorigenesis in a subcutaneous allograft B16. F10.SIY model in C57/BL6 mice.[47] The same study observed that oral administration of a cocktail of *Bifidobacterium* species combined with an immune checkpoint inhibitor immunotherapy targeting the PD1—PD-L1 signalling pathway using an anti-PD-1 monoclonal antibody promoted activation of CD8+ cytotoxic T cells and significantly reduced tumour outgrowth.[47] Tanoue et al presented similar findings when administering mice with a consortium of 11 bacterial strains isolated from healthy human faeces, including those of the genus *Bifidobacterium* and *Lactobacillus*, describing improved CD8 T-cell activation and improved efficacy of immune checkpoint inhibitor therapies.[48] Furthermore, Gopalakrishnan et al went on to confirm a similar outcome in human patients, whereby those with an increased microbial diversity responded more favourably to anti-PD-1 immunotherapies than patients with lower diversity.[49]

4.2 | BrCa literature

With the awareness that non-BrCa disease is influenced by changes in the microbiota, we can now ask if similar affects are observed in BrCa models. Unlike in melanoma, the lower expression of immune-modulating proteins, such as PD-1 and CTLA-4, in BrCa means it is less amenable to checkpoint inhibition therapies.[50,51] However, the immune system still plays a key role in both pro- and anticancer responses at different stages of BrCa progression. In one of the earlier microbiota—BrCa studies, Rao et al observed that *Helicobacter hepaticus* infection in C57/BL6 Apc$^{Min/+}$ mice (spontaneous CRC model) resulted in the formation of tumours in the breast as well as in the colon.[52] In a similar model deficient in the *Rag2* gene, thus lacking mature lymphocytes, tumours were more frequent with increased infiltration of F4/80+ myeloid cells, suggesting a protective role of lymphocytes in reducing innate immune inflammation associated with tumourigenesis. Indeed, dosage of *Rag2*-deficient animals with

CD4$^+$CD45^{low}CD25$^+$ T-regulatory cells from wild-type animals significantly reduced breast tumourigenesis, including in *H hepaticus* infected animals.[52] This supports the consensus that microbes have an integral role in priming the immune system, particularly the adaptive components, to act against cancers distal to the GI tract.

More recent studies have focused on the impact of loss of known mutualistic bacterial genera and species, and the impact of microbial metabolites on BrCa progression (also discussed later). One of the most influential factors contributing to disruption of the gut microbiota is antibiotic use, which has been linked to increased risk of several cancers including BrCa.[53,54] In a comprehensive study, Buchta Rosean et al administered C57/BL6 mice with a robust antibiotic cocktail, comprising vancomycin, neomycin, metronidazole, gentamycin and ampicillin, for 2 weeks to ablate the gut microbiota.[24] Following a recolonisation period of 4 days, animals were orthotopically induced with either a poorly metastatic, hormone receptor positive model or a more aggressive PyMT-derived model. A pre-perturbed microbiota significantly increased metastasis to the lungs in both models, without influencing primary tumour growth kinetics.[24] Subsequent analysis of immune cell infiltration and cytokine analysis of mammary tissue prior to tumour induction identified increased abundance of myeloid cells and elevated myeloid recruitment components including CXCL10 and CCL2, suggesting that metastatic potential was promoted through antibiotic-induced inflammatory pathways independent of tumour status. McKee et al used a similar antibiotic cocktail (swapping gentamycin for amphotericin), but continued treatment throughout the experimental period.[55] In contrast to Buchta Rosean et al, they did observe a significant increase in primary tumour kinetics when using orthotopic syngeneic models for both luminal (PyMT-BO1) and basal-like (EO771) BrCa. Downstream scRNA-seq revealed an increased stromal signature (PyMT-BO1 only) and histological analysis revealed increased abundance of mast cells tumour stroma in PyMT-BO1 and EO771 tumours from antibiotic-treated animals. Notably, inhibition of mast-cell activation confirmed this immune population to be driving enhanced primary tumour growth after antibiotic-induced microbiota disturbances. Importantly, the same study also demonstrated that using a clinically relevant cephalosporin antibiotic promoted the same increase in primary tumour growth and was associated with similar increases in tumour mast cells.[55]

Based on the similarity of the orthotopic models used in these two studies, it is likely that the differences in primary tumour growth between them were due to the differences in treatment regimen. Buchta-Rosean et al allowed for a bacterial recolonisation period of 4 days, which likely aided in slowing primary tumour growth, possibly through an immunological re-priming.[24] However, McKee et al undertook an uninterrupted antibiotic treatment, which may have prevented mutualistic bacteria from recolonising, leaving the primary tumour to grow "unchecked".[55] Nonetheless, both studies suggest a healthy microbiota positively regulates antitumour immune pathways. Thus, it is surprising that to date there does not appear to be any mechanistic in vivo studies into whether the administration of beneficial or "probiotic" genera such as *Bifidobacterium* or *Lactobacillus* may influence BrCa progression or support therapeutic intervention against it.

The metabolome of the microbiota is also known to play a key role in host health, influencing an array of biological pathways including cellular proliferation, metabolism and immunity. Although studies focusing on microbial metabolites in BrCa models are very limited, a previous study determined that cadaverine (produced during microbial breakdown of animal tissue) supplementation reduced both primary and metastatic burden in an orthotopic 4T1 triple-negative-like BrCa model in BALB/c mice.[56] As highlighted previously, short-chain fatty acid (SCFAs, particularly butyrate) also have links to cancer and are known to promote an invasive/aggressive phenotype in BrCa in vitro.[57] Crucially, these microbial metabolites are derived after fermentation/metabolism of dietary components and therefore diet may act as a key overriding factor that impacts the microbiota, their metabolites and subsequent host interactions, leading to differential BrCa outcomes.

5 | DIET, THE GUT MICROBIOTA AND BrCa

As already mentioned earlier (and in Table 2), extensive epidemiological studies have laid the groundwork for understanding that diet has a major role on cancer risk and progression.[58-60] One of the key roles played by the microbiota is breakdown of complex dietary substrates into their constituent bioactive compounds; therefore, there is growing interest in understanding functional outcomes and the underlying mechanisms governing diet-microbe interactions with respect to cancer.

5.1 | Diet and BrCa

Although the correlations between BrCa risk and dietary intake have been intensively studied, the underlying associations or effector mechanisms remain poorly understood. Historically, increased risk of BrCa has been tied to high intake of red meat and animal fat,[61,62] with decreased risk being concurrently linked to fruit and vegetables consumption.[63] The overall field remains conflicted, as epidemiological links between individual foods and BrCa appear difficult to rationalise within the confines of even large-scale observational studies. This point is emphasised by the 2017 third expert report on "diet, nutrition and physical activity in BrCa" by the World Cancer Research Fund/American Institute for Cancer Research, which stated that although body fatness (as adjudged by BMI and waist-to-hip ratio) was a probable risk factor for BrCa in premenopausal and postmenopausal women, there is only limited or suggestive evidence for contribution of any single food group.[64] It is possible that these findings are a result of the requirement for a change in total dietary pattern to have significant effect on the gut microbiota and disease risk, or because of the difficulty in conducting longitudinal human studies which focus on a single food group. More recently, research has moved toward assessment of dietary patterns (rather than specific foodstuffs), which indicate that "Western" diets (ie, those that are high in processed

meat, sugar and fat) increase BrCa risk, while a more "healthy" diet (ie, high fresh fruit, vegetables and fish) decreases BrCa risk.[16] Importantly, and specific to BrCa, menopausal status and BrCa subtype (ie, receptor status) are important confounders when assessing dietary links. Western diet effects, for example, are only significant in postmenopausal patients with hormone receptor-positive tumours, while "healthy" diet effects are only significant in premenopausal women, but across receptor-positive and receptor-negative tumours.[16] Alcohol intake is also a significant risk factor, with high consumption linked with disease recurrence and reduced survival.[65] Another dietary pattern linked to BrCa is the Mediterranean Diet, with recent studies showing an inverse relationship, particularly in the context of triple-negative disease.[66,67]

5.2 | Diet and the gut microbiota

There is a strong evolutionary relationship between the gut microbiota and diet. Certain members genomically encode enzymes such as glycoside hydrolases, which allow poly- and/or oligosaccharide carbohydrates to be metabolised.[68] Some "generalist" microbes, such as members of the Bacteroides phyla (eg, *Bacteroides thetaiotaomicron*), degrade a wide array of carbohydrates, while other "specialist" gut microbes (eg, *Roseburia intestinalis*) degrade specific oligosaccharides.[69] Ingestion of dietary fibre elicits a dynamic response from communities of these metabolising microbes through extensive primary and secondary degradation. Here, primary degraders (eg, *B. thetaiotaomicron*) convert polysaccharides into oligosaccharides and secondary metabolites (eg, SCFAs), which can then be utilised by secondary degraders (eg, *Eubacterium rectale*) to further enhance nutrient bioavailability (eg, breakdown to monosaccharides) and support community colonisation.[70] This so-called "cross-feeding" is a determinant of gut population dynamics, as increased metabolism often affords selective advantages, which increase microbe abundance and subsequent digestion efficiency.[69]

Previous work in CRC has indicated that microbial-derived SCFAs, particularly butyrate, have anticancer effects (demonstrated in cancer cell cultures[71,72] and animal models[73]); however, clinical supplementation studies have proved difficult due to issues with bioavailability and toxicity.[74] Therefore, a more nuanced microbiota and defined diet approach may represent a more realistic avenue to improve cancer outcomes. This highlights the interlinking relationship between microbiota composition and metabolite production, with both factors requiring consideration if we are to realistically improve cancer outcomes.

More recently, it is now appreciated that specific components from fruit and vegetables (eg, polyphenols) may also be processed and influence the gut microbiota by acting as prebiotics (defined as dietary substrates, which can be utilised by host microorganisms to confer a health benefit).[75] Dietary polyphenols have well-documented effects on the host, which has been reviewed elsewhere.[28] Studies have shown that increased polyphenol intake is associated with higher levels of beneficial bacteria (such as *Bifidobacterium* and *Lactobacillus*) and SCFAs in humans, while also decreasing levels of bacteria that have been associated with disease, so-called pathobionts.[29] Polyphenols are known to undergo extensive metabolism via the gut microbiota during conversion to bioavailable metabolites. The magnitude and complexity of these interactions have made detailed studies difficult, but some examples of known biotransformations include *Flavonifractor plautii* conversion of catechin and epicatechin into valerolactones and valeric acids, and soy isoflavones into equol and/or O-desmethylangolensin by *Slackia isoflavoniconvertens* and *Slackia equolifaciens*. Biochemical groups of microbiota-modulating polyphenols include flavanols (eg, catechin),[76] resveratrol[77] and anthocyanins,[78] which are highly concentrated in foodstuffs such as green tea, berries and red-wine. While human mechanistic studies involving polyphenols and disease are lacking, the aforementioned polyphenols are able to rescue high-fat diet (HFD)-induced microbiota perturbations and elevate murine type II diabetes symptoms through increases in gut *Akkermansia* species.[79-81] Given the emerging importance of the gut microbiota in tumour progression and cancer therapy, dietary manipulation of the system (via dietary fibre and polyphenols) is an important theme to be explored in humans and via mechanistic in vivo studies.

5.3 | Mechanistic links between diet and BrCa

The majority of in vivo mechanistic knowledge is confined to tumours "local" to the digestive tract (such as gastric, colon and CRC, as discussed earlier). This is perhaps expected, given gastrointestinal tissues come into direct contact with bioactive compounds resulting from digestion. Studies probing the mechanistic impact of diet on BrCa are limited, and most studies to date have explored how HFDs are linked to primary and metastatic tumours. One particular study indicated that an HFD promoted formation of pre-metastatic niches and lung metastases in mice through activation of myeloid-derived suppressor cells (MDSC),[82] but this HFD-induced metastasis could be significantly ameliorated through treatment with a saponin called glycyrrhizic acid, a plant-derived phytochemical. Glycyrrhizic acid significantly altered microbiota composition in this context, and concurrently reduced colonic lipopolysaccharide (LPS), NF-kB and macrophage activity—which correlated with decreases in MDSC infiltration to pre-metastatic niches. Interestingly, the effect of glycyrrhizic acid was ablated through microbiota depletion with antibiotics, causing re-introduction of HFD levels of colonic LPS and implicating the gut microbiota as the key effector in this model system. Mechanistic links between diet and BrCa metastasis have also been made elsewhere, as it has been shown the calorie restriction during radiotherapy (for TNBC) causes decreased metastatic burden in mice compared to animals on an ad libitum diet.[83] This decrease is caused by downregulation of the oncogenic insulin-like growth factor-1/Akt pathway—a known regulator of tumour metastasis.[84]

As previously mentioned, dietary polyphenol intake has been linked with improved BrCa outcomes. To date, many of the studies examining the anticancer properties of polyphenols have been performed in vitro, and have often also used unmetabolised purified compounds, therefore reflecting a direct anti-tumour therapeutic

approach.[85-87] Contrastingly, studies exploring dietary polyphenol intake and their downstream effects on BrCa, within the context of "normal digestion", are thin on the ground. Of the limited mouse studies completed so far, dietary delivery of quercetin was shown to be effective in reducing tumour number and volume in the C3/SV40 Tag model of BrCa.[88] Elsewhere, oral delivery of grape polyphenols (resveratrol, catechins and quercetin) decreased primary tumour growth and metastases in a mouse xenograft model via downregulation of NFkB,[89] and oral piceatannol administration decreased 4T1 breast tumour metastasis through inhibited macrophage infiltration and angiogenesis.[90]

Although it is clear that diet, and specific components like fibre and polyphenols, may play an important role in BrCa, comprehensive studies directly linking the gut microbiota, dietary components and BrCa progression are required to fully understand how we can manipulate the system.

6 | FUTURE PERSPECTIVES

The field of microbiome and cancer is rapidly expanding, yet studies focused on BrCa are currently limited. There is now a strong body of evidence indicating the gut microbiota and associated factors such as diet and antibiotics may play a key role in BrCa risk and outcomes, therefore detailed mechanistic studies in preclinical models that help underpin next-stage translational projects are needed. However, there are some important considerations specifically in relation to BrCa that need to be evaluated.

One such consideration relates to the differential risks between BrCa subtypes (see Table 1), including treatment-resistant (eg, TNBC) tumours. To date this fundamental BrCa factor has been somewhat ignored, therefore future microbiota profiling studies of BrCa should consider clear stratification of patients by histological and molecular subtypes, which may allow comparison between groups and associated pathological outcomes. Age will also need to be carefully considered in these groups as the microbiota diversity does decrease in elderly populations.[91] Previous studies have indicated that healthy individuals living at home have a more "robust" microbiota compared to those living in long-term residential care (which may be linked to diet). There are also alterations in microbiome composition associated with frailty indices and inflammatory status.[92] This is obviously important for BrCa as, the older a woman is, the more likely she is to get BrCa; rates are highest in women over 70.[10] Furthermore, to date, nearly all BrCa microbiome studies have focused on high-income country patient populations. As BrCa rates are increasing in low-to-middle income countries, which also have differing patient population characteristics (age and BrCa subtype),[93] further studies should establish signatures in diverse BrCa patient cohorts as this is crucial for next-stage clinical trials and optimal patient outcomes. An important consideration for microbiome and BrCa studies also relates to standard clinical care pathways. Typically, BrCa patients will undergo one or more of the following: surgery, chemotherapy and/or radiotherapy. Several studies are currently ongoing (either still recruiting or still to

publish results) to determine the impact of chemotherapy and the gut microbiota on BrCa patients, for example, NCT03702868, NCT04138979 or NCT03222856; however, findings are yet to be published. Conversely, there appears to be no studies that are solely focused on the relationship between radiotherapy in BrCa patients and its impact on the gut microbiota and clinical outcomes, although there appears to be studies with a radiotherapy aspect (in combination with other therapies, that is, immune-checkpoint) but not as a primary outcome (NCT04435964). Interestingly, in vivo studies from the 1960s have suggested that radiotherapy on germ-free mice was less effective[94]; however, further studies—mechanistic and clinical—are required to understand the influence of these differing factors on the gut microbiota and BrCa outcomes.

Although the limited studies to date have indicated an altered microbiota signature in BrCa patients, further research into the strain-level variation and functional capacity of the microbiota, including a focus on nonbacterial members, that is, viruses, fungi and archaea, and also impact of external factors is required. This may allow more specific biomarker profiling in relation to BrCa risk in relation to the many microbiota-modulating factors earlier in life may predispose to onset of cancer, including BrCa. Thus, longer-term longitudinal in-depth profiling studies are key to understanding the impact of lifestyle factors and routine medications (eg, antibiotics) on the microbiota prior to detection of primary breast tumours. Enhanced microbiota profiling studies may also allow patient stratification for treatment trials (identifying potentially nonresponsive patients) in tandem with current or new therapies (as described earlier). Moreover, more robust microbiome-profiling approaches and tools are expected to facilitate development and interventions/treatment with microbiota therapies such as probiotics and/or live biotherapeutic products. Probiotics are products that contain live microorganisms, which when administered in adequate amounts can confer health benefits to the host.[95] Certain types of these probiotic microbes have been shown to stimulate immune responses and also shown in vitro to exert antitumorigenic potential via immunoregulatory pathways.[96] However, to date, there has only been a very small number of clinical trials to evaluate probiotic supplementation in BrCa patients: NCT03760653 and NCT03358511. Unfortunately, the former was prematurely terminated, while the latter has recently completed with only 7 participants enrolled. It is clear that larger-scale randomised placebo controlled trials are needed, but ideally these should be based on evaluation of probiotics or live biotherapeutic products in vitro and in vivo to pinpoint potential multifactorial mechanisms underlying beneficial effects, and thus optimal strain(s) combined to take forward into BrCa patients. For a recent review on this topic, see Reference 96.

As diet is a significant confounder for all microbiota studies, and the apparent close links between diet and BrCa, food diaries (using new apps) may provide important insight into the patient's dietary habits. However, current dietary patterns may not reflect previous eating habits, which may have contributed to disease development or susceptibility. Future observational or clinical trials should seek to control, or capture associated meta-data, which will be crucial for teasing out causative or associative factors with respect to BrCa risk.

Moreover, studies focused on diet/metabolites may also need to consider if dietary interventions should be personalised if they are to be successfully utilised for future therapy. Indeed, recent studies have highlighted the personalised response to individuals (and the microbiota) to the same diet,[97] which highlights the limitations and challenges for next-stage studies of this kind.

It is clear that more comprehensive preclinical data are required in the field, and that choice of in vivo models is key. Consideration of the models used, their clinical relevance, microbiota modulatory factors (eg, diet and antibiotics) and the approaches used to integrate microbial and host components must be carefully reviewed. Although subcutaneous tumour studies demonstrate to some extent that microbes can influence carcinogenesis, it is beneficial to test such hypotheses in more physiologically relevant orthotopic or spontaneous models, which better underpin the physiological and biomolecular mechanisms involved in such phenotypes. Indeed, spontaneous in vivo BrCa models may allow further insights into microbiome and microbiota-modulating factors and BrCa risk, while orthotopic models may allow testing of acute immune mechanisms and testing of new promising therapies. The greater translational impact of these models should provide greater clarity for moving preclinical research to human studies. Moreover, studies focusing on dietary or antibiotic interventions in mouse models should include downstream analyses of the gut microbiota, mucosal and systemic immunity, as well as microbial and host metabolism, so to develop a multisystem picture of the effects of microbiota manipulations on BrCa risk and progression. An important caveat to any resulting findings, however, is that mice of the same genetic background housed at different animal facilities will likely have different microbiota profiles, which may influence experimental cancer outcomes.[98] Overall, more comprehensive preclinical findings will better inform and facilitate design of translational human studies looking at the impact of defined dietary components on the gut microbiota and BrCa outcomes, as well as profiling responders vs nonresponders to treatments in the context of diet.

7 | CONCLUSIONS

The microbiome is emerging as a central player in cancer risk and anticancer responses, and recent studies also suggest this may be the case for BrCa. However, as many intrinsic and extrinsic factors are known to influence the microbiota, and indeed the cancer itself may also impact community composition, robust experimental design from sequencing the microbiota to capturing meta-data, and the choice of in vivo models, is required. A combination of approaches is most likely to narrow down important causative or associative factors, which may allow interventions to reduce BrCa risk, and guide the development of novel therapeutic approaches to alleviate symptoms or improve prognosis in BrCa patients.

CONFLICT OF INTEREST
The authors declare no conflicts of interest.

ORCID
Stephen D. Robinson https://orcid.org/0000-0002-6606-7588

REFERENCES
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394-424.
2. Sun L, Legood R, Dos-Santos-Silva I, Gaiha SM, Sadique Z. Global treatment costs of breast cancer by stage: a systematic review. *PLoS One.* 2018;13:e0207993.
3. Scholzen T, Gerdes J. The Ki-67 protein: from the known and the unknown. *J Cell Physiol.* 2000;182:311-322.
4. Yersal O, Barutca S. Biological subtypes of breast cancer: prognostic and therapeutic implications. *World J Clin Oncol.* 2014;5:412-424.
5. Dai X, Li T, Bai Z, et al. Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res.* 2015;5:2929-2943.
6. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn HJ. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol.* 2011; 22:1736-1747.
7. Heng B, Lim CK, Lovejoy DB, Bessede A, Gluch L, Guillemin GJ. Understanding the role of the kynurenine pathway in human breast cancer immunobiology. *Oncotarget.* 2016;7:6506-6520.
8. Fallahpour S, Navaneelan T, De P, Borgo A. Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data. *CMAJ Open.* 2017;5:E734-E739.
9. Komorowski AS, Pezo RC. Untapped "-omics": the microbial metagenome, estrobolome, and their influence on the development of breast cancer and response to treatment. *Breast Cancer Res Treat.* 2020;179:287-300.
10. Kaminska M, Ciszewski T, Lopacka-Szatan K, Miotla P, Staroslawska E. Breast cancer risk factors. *Prz Menopauzalny.* 2015; 14:196-202.
11. Hsieh CC, Trichopoulos D, Katsouyanni K, Yuasa S. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study. *Int J Cancer.* 1990;46:796-800.
12. Ziegler RG, Hoover RN, Pike MC, et al. Migration patterns and breast cancer risk in Asian-American women. *J Natl Cancer Inst.* 1993;85: 1819-1827.
13. Fernandez MF, Reina-Perez I, Astorga JM, Rodriguez-Carrillo A, Plaza-Diaz J, Fontana L. Breast cancer and its relationship with the microbiota. *Int J Environ Res Public Health.* 2018;15:1747-1767.
14. Thomas AM, Manghi P, Asnicar F, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med.* 2019;25:667-678.
15. Saxe GA, Rock CL, Wicha MS, Schottenfeld D. Diet and risk for breast cancer recurrence and survival. *Breast Cancer Res Treat.* 1999;53: 241-253.

16. Xiao Y, Xia J, Li L, et al. Associations between dietary patterns and the risk of breast cancer: a systematic review and meta-analysis of observational studies. Breast Cancer Res. 2019;21:16.

17. Picon-Ruiz M, Morata-Tarifa C, Valle-Goffin JJ, Friedman ER, Slingerland JM. Obesity and adverse breast cancer risk and outcome: mechanistic insights and strategies for intervention. CA Cancer J Clin. 2017;67:378-397.

18. Liedtke S, Schmidt ME, Vrieling A, et al. Postmenopausal sex hormones in relation to body fat distribution. Obesity (Silver Spring). 2012;20:1088-1095.

19. Seitz HK, Pelucchi C, Bagnardi V, La Vecchia C. Epidemiology and pathophysiology of alcohol and breast cancer: update 2012. Alcohol Alcohol. 2012;47:204-212.

20. Liu Y, Nguyen N, Colditz GA. Links between alcohol consumption and breast cancer: a look at the evidence. Womens Health (Lond). 2015;11: 65-77.

21. Segovia-Mendoza M, Morales-Montor J. Immune tumor microenvironment in breast cancer and the participation of estrogen and its receptors in cancer physiopathology. Front Immunol. 2019;10:348.

22. Rezaei NJ, Bazzazi AM, Naseri Alavi SA. Neurotoxicity of the antibiotics: a comprehensive study. Neurol India. 2018;66:1732-1740.

23. Kallala R, Graham SM, Nikkhah D, et al. In vitro and in vivo effects of antibiotics on bone cell metabolism and fracture healing. Expert Opin Drug Saf. 2012;11:15-32.

24. Buchta Rosean C, Bostic RR, Ferey JCM, et al. Preexisting commensal dysbiosis is a host-intrinsic regulator of tissue inflammation and tumor cell dissemination in hormone receptor-positive breast cancer. Cancer Res. 2019;79:3662-3675.

25. Velicer CM, Heckbert SR, Lampe JW, Potter JD, Robertson CA, Taplin SH. Antibiotic use in relation to the risk of breast cancer. JAMA. 2004;291:827-835.

26. Velicer CM, Heckbert SR, Rutter C, Lampe JW, Malone K. Association between antibiotic use prior to breast cancer diagnosis and breast tumour characteristics (United States). Cancer Causes Control. 2006; 17:307-313.

27. Danneskiold-Samsoe NB, Dias de Freitas Queiroz Barros H, Santos R, et al. Interplay between food and gut microbiota in health and disease. Food Res Int. 2019;115:23-31.

28. Fraga CG, Croft KD, Kennedy DO, Tomas-Barberan FA. The effects of polyphenols and other bioactives on human health. Food Funct. 2019;10:514-528.

29. Sun H, Chen Y, Cheng M, Zhang X, Zheng X, Zhang Z. The modulatory effect of polyphenols from green tea, oolong tea and black tea on human intestinal microbiota in vitro. J Food Sci Technol. 2018;55: 399-407.

30. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature. 2006;444:1022-1023.

31. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444:1027-1031.

32. Mutlu EA, Gillevet PM, Rangwala H, et al. Colonic microbiome is altered in alcoholism. Am J Physiol Gastrointest Liver Physiol. 2012; 302:G966-G978.

33. Bischoff SC, Barbara G, Buurman W, et al. Intestinal permeability—a new target for disease prevention and therapy. BMC Gastroenterol. 2014;14:189.

34. Shimizu K, Muranaka Y, Fujimura R, Ishida H, Tazume S, Shimamura T. Normalization of reproductive function in germfree mice following bacterial contamination. Exp Anim. 1998;47:151-158.

35. Kwa M, Plottel CS, Blaser MJ, Adams S. The intestinal microbiome and estrogen receptor-positive female breast cancer. J Natl Cancer Inst. 2016;108:1-10.

36. Panda S, El Khader I, Casellas F, et al. Short-term effect of antibiotics on human gut microbiota. PLoS One. 2014;9:e95476.

37. Priya S, Blekhman R. Population dynamics of the human gut microbiome: change is the only constant. Genome Biol. 2019;20:150.

38. Plummer M, van Doorn LJ, Franceschi S, et al. Helicobacter pylori cytotoxin-associated genotype and gastric precancerous lesions. J Natl Cancer Inst. 2007;99:1328-1334.

39. Adlung L, Elinav E, Greten TF, Korangy F. Microbiome genomics for cancer prediction. Nat Cancer. 2020;1:379-381.

40. Jackson MA, Verdi S, Maxan ME, et al. Gut microbiota associations with common diseases and prescription medications in a population-based cohort. Nat Commun. 2018;9:2655.

41. Zhu J, Liao M, Yao Z, et al. Breast cancer in postmenopausal women is associated with an altered gut metagenome. Microbiome. 2018; 6:136.

42. Goedert JJ, Jones G, Hua X, et al. Investigation of the association between the fecal microbiota and breast cancer in postmenopausal women: a population-based case-control pilot study. J Natl Cancer Inst. 2015;107:djv147.

43. Guan X, Ma F, Sun X, et al. Gut microbiota profiling in patients with HER2-negative metastatic breast cancer receiving metronomic chemotherapy of capecitabine compared to those under conventional dosage. Front Oncol. 2020;10:902.

44. Wu AH, Tseng C, Vigen C, et al. Gut microbiome associations with breast cancer risk factors and tumor characteristics: a pilot study. Breast Cancer Res Treat. 2020;182:451-463.

45. Kaakoush NO. Insights into the role of Erysipelotrichaceae in the human host. Front Cell Infect Microbiol. 2015;5:84.

46. Yu T, Guo F, Yu Y, et al. Fusobacterium nucleatum promotes chemoresistance to colorectal cancer by modulating autophagy. Cell. 2017;170:548-63.e16.

47. Sivan A, Corrales L, Hubert N, et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science. 2015;350:1084-1089.

48. Tanoue T, Morita S, Plichta DR, et al. A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. Nature. 2019;565: 600-605.

49. Gopalakrishnan V, Spencer CN, Nezi L, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science. 2018;359:97-103.

50. Hammerl D, Smid M, Timmermans AM, Sleijfer S, Martens JWM, Debets R. Breast cancer genomics and immuno-oncological markers to guide immune therapies. Semin Cancer Biol. 2018;52: 178-188.

51. Sobral-Leite M, Van de Vijver K, Michaut M, et al. Assessment of PD-L1 expression across breast cancer molecular subtypes, in relation to mutation rate, BRCA1-like status, tumor-infiltrating immune cells and survival. Onco Targets Ther. 2018;7:e1509820.

52. Rao VP, Poutahidis T, Ge Z, et al. Innate immune inflammatory response against enteric bacteria Helicobacter hepaticus induces mammary adenocarcinoma in mice. Cancer Res. 2006;66:7395-7400.

53. Kilkkinen A, Rissanen H, Klaukka T, et al. Antibiotic use predicts an increased risk of cancer. Int J Cancer. 2008;123:2152-2155.

54. Boursi B, Mamtani R, Haynes K, Yang YX. Recurrent antibiotic exposure may promote cancer formation—another step in understanding the role of the human microbiota? Eur J Cancer. 2015;51:2655-2664.

55. McKee A, Kirkup B, Madgwick M, et al. Antibiotic-induced disturbances of the gut microbiota result in accelerated breast tumour growth via a mast cell-dependent pathway. bioRxiv. 2020. https://doi.org/10.1101/2020.03.07.982108.

56. Kovacs T, Miko E, Vida A, et al. Cadaverine, a metabolite of the microbiome, reduces breast cancer aggressiveness through trace amino acid receptors. Sci Rep. 2019;9:1300.

57. Thirunavukkarasan M, Wang C, Rao A, et al. Short-chain fatty acid receptors inhibit invasive phenotypes in breast cancer cells. PLoS One. 2017;12:e0186334.

58. Armstrong B, Doll R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *Int J Cancer*. 1975;15:617-631.

59. Block G, Patterson B, Subar A. Fruit, vegetables, and cancer prevention: a review of the epidemiological evidence. *Nutr Cancer*. 1992;18: 1-29.

60. Bradbury KE, Appleby PN, Key TJ. Fruit, vegetable, and fiber intake in relation to cancer risk: findings from the European Prospective Investigation into Cancer and Nutrition (EPIC). *Am J Clin Nutr*. 2014;100 (Suppl 1):394S-398S.

61. Levi F, La Vecchia C, Gulie C, Negri E. Dietary factors and breast cancer risk in Vaud, Switzerland. *Nutr Cancer*. 1993;19:327-335.

62. La Vecchia C, Decarli A, Franceschi S, Gentile A, Negri E, Parazzini F. Dietary factors and the risk of breast cancer. *Nutr Cancer*. 1987;10:205-214.

63. Malin AS, Qi D, Shu XO, et al. Intake of fruits, vegetables and selected micronutrients in relation to the risk of breast cancer. *Int J Cancer*. 2003;105:413-418.

64. World Cancer Research Fund International/American Institute for Cancer Research. Continuous Update Project Report: Diet, Nutrition, Physical Activity and Breast Cancer. 2017. Available at: wcrf. org/breast-cancer-2017. All CUP reports are available at wcrf. org/cupreports.

65. Schwedhelm C, Boeing H, Hoffmann G, Aleksandrova K, Schwingshackl L. Effect of diet on mortality and cancer recurrence among cancer survivors: a systematic review and meta-analysis of cohort studies. *Nutr Rev*. 2016;74:737-748.

66. van den Brandt PA, Schulpen M. Mediterranean diet adherence and risk of postmenopausal breast cancer: results of a cohort study and meta-analysis. *Int J Cancer*. 2017;140:2220-2231.

67. Shaikh AA, Braakhuis AJ, Bishop KS. The Mediterranean diet and breast cancer: a personalised approach. *Healthcare (Basel)*. 2019;7.

68. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;42:D490-D495.

69. Cockburn DW, Koropatkin NM. Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. *J Mol Biol*. 2016;428:3230-3252.

70. Elhenawy W, Debelyy MO, Feldman MF. Preferential packing of acidic glycosidases and proteases into Bacteroides outer membrane vesicles. *mBio*. 2014;5:e00909-e00914.

71. Li Q, Cao L, Tian Y, et al. Butyrate suppresses the proliferation of colorectal cancer cells via targeting pyruvate kinase M2 and metabolic reprogramming. *Mol Cell Proteomics*. 2018;17:1531-1545.

72. Li Q, Ding D, Meng T, et al. Butyrate suppresses motility of colorectal cancer cells via deactivating Akt/ERK signaling in histone deacetylase dependent manner. *J Pharmacol Sci*. 2017;135:148-155.

73. Tian Y, Xu Q, Sun L, Ye Y, Ji G. Short-chain fatty acids administration is protective in colitis-associated colorectal cancer development. *J Nutr Biochem*. 2018;57:103-109.

74. Douillard JY, Bennouna J, Vavasseur F, et al. Phase I trial of interleukin-2 and high-dose arginine butyrate in metastatic colorectal cancer. *Cancer Immunol Immunother*. 2000;49:56-61.

75. Gibson GR, Hutkins R, Sanders ME, et al. Expert consensus document: the international scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nat Rev Gastroenterol Hepatol*. 2017;14:491-502.

76. Jin JS, Touyama M, Hisada T, Benno Y. Effects of green tea consumption on human fecal microbiota with special reference to Bifidobacterium species. *Microbiol Immunol*. 2012;56:729-739.

77. Moreno-Indias I, Sanchez-Alcoholado L, Perez-Martinez P, et al. Red wine polyphenols modulate fecal microbiota and reduce markers of the metabolic syndrome in obese patients. *Food Funct*. 2016;7:1775-1787.

78. Neyrinck AM, Van Hee VF, Bindels LB, De Backer F, Cani PD, Delzenne NM. Polyphenol-rich extract of pomegranate peel alleviates tissue inflammation and hypercholesterolaemia in high-fat diet-induced obese mice: potential implication of the gut microbiota. *Br J Nutr*. 2013;109:802-809.

79. Yang Q, Liang Q, Balakrishnan B, Belobrajdic DP, Feng QJ, Zhang W. Role of dietary nutrients in the modulation of gut microbiota: a narrative review. *Nutrients*. 2020;12.

80. Roopchand DE, Carmody RN, Kuhn P, et al. Dietary polyphenols promote growth of the gut bacterium *Akkermansia muciniphila* and attenuate high-fat diet-induced metabolic syndrome. *Diabetes*. 2015;64:2847-2858.

81. Tomas-Barberan FA, Selma MV, Espin JC. Interactions of gut microbiota with dietary polyphenols and consequences to human health. *Curr Opin Clin Nutr Metab Care*. 2016;19:471-476.

82. Qiu M, Huang K, Liu Y, et al. Modulation of intestinal microbiota by glycyrrhizic acid prevents high-fat diet-enhanced pre-metastatic niche formation and metastasis. *Mucosal Immunol*. 2019;12:945-957.

83. Simone BA, Dan T, Palagani A, et al. Caloric restriction coupled with radiation decreases metastatic burden in triple negative breast cancer. *Cell Cycle*. 2016;15:2265-2274.

84. Shastri AA, Saleh A, Savage JE, DeAngelis T, Camphausen K, Simone NL. Dietary alterations modulate the microRNA 29/30 and IGF-1/AKT signaling axis in breast cancer liver metastasis. *Nutr Metab (Lond)*. 2020;17:23.

85. Scarlatti F, Maffei R, Beau I, Codogno P, Ghidoni R. Role of non-canonical Beclin 1-independent autophagy in cell death induced by resveratrol in human breast cancer cells. *Cell Death Differ*. 2008;15:1318-1329.

86. Ranganathan S, Halagowder D, Sivasithambaram ND. Quercetin suppresses twist to induce apoptosis in MCF-7 breast cancer cells. *PLoS One*. 2015;10:e0141370.

87. Huang C, Lee SY, Lin CL, et al. Co-treatment with quercetin and 1,2,3,4,6-penta-O-galloyl-beta-D-glucose causes cell cycle arrest and apoptosis in human breast cancer MDA-MB-231 and AU565 cells. *J Agric Food Chem*. 2013;61:6430-6445.

88. Steiner J, Davis J, McClellan J, et al. Dose-dependent benefits of quercetin on tumorigenesis in the C3(1)/SV40Tag transgenic mouse model of breast cancer. *Cancer Biol Ther*. 2014;15:1456-1467.

89. Castillo-Pichardo L, Martinez-Montemayor MM, Martinez JE, Wall KM, Cubano LA, Dharmawardhane S. Inhibition of mammary tumor growth and metastases to bone and liver by dietary grape polyphenols. *Clin Exp Metastasis*. 2009;26:505-516.

90. Song H, Jung JI, Cho HJ, et al. Inhibition of tumor progression by oral piceatannol in mouse 4T1 mammary cancer is associated with decreased angiogenesis and macrophage infiltration. *J Nutr Biochem*. 2015;26:1368-1378.

91. Jeffery IB, Lynch DB, O'Toole PW. Composition and temporal stability of the gut microbiota in older persons. *ISME J*. 2016;10:170-182.

92. O'Toole PW, Jeffery IB. Microbiome-health interactions in older people. *Cell Mol Life Sci*. 2018;75:119-128.

93. Anderson BO, Ilbawi AM, El Saghir NS. Breast cancer in low and middle income countries (LMICs): a shifting tide in global health. *Breast J*. 2015;21:111-118.

94. McLaughlin MM, Dacquisto MP, Jacobus DP, Horowitz RE. Effects of the germfree state on responses of mice to whole-body irradiation. *Radiat Res*. 1964;23:333-349.

95. Hill C, Guarner F, Reid G, et al. The international scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat Rev Gastroenterol Hepatol*. 2014;11:506-514.

96. Ranjbar S, Seyednejad SA, Azimi H, Rezaeizadeh H, Rahimi R. Emerging roles of probiotics in prevention and treatment of breast cancer: a

comprehensive review of their therapeutic potential. *Nutr Cancer.* 2019;71:1-12.

97. Deehan EC, Yang C, Perez-Munoz ME, et al. Precision microbiome modulation with discrete dietary fiber structures directs short-chain fatty acid production. *Cell Host Microbe.* 2020;27:389-404.e6.

98. Leystra AA, Clapper ML. Gut microbiota influences experimental outcomes in mouse models of colorectal cancer. *Genes (Basel).* 2019;10:900.

Allocoprobacillus halotolerans gen. nov., sp. nov and *Coprobacter tertius* sp. nov., isolated from human gut microbiota

Nancy M. Y. Teng[1,2], Raymond Kiu[1], Rhiannon Evans[1], David J. Baker[1], Christian Zenner[3], Stephen D. Robinson[1,4,*] and Lindsay J. Hall[1,2,3,*]

Abstract

Two novel bacterial isolates were cultured from faecal samples of patients attending the Breast Care clinic at the Norwich and Norfolk University Hospital. Strain LH1062[T] was isolated from a 58-year-old female diagnosed with invasive adenocarcinoma with ductal carcinoma *in situ*. Strain LH1063[T] was isolated from a healthy 51-year-old female. Isolate LH1062[T] was predicted to be a potential novel genus most closely related to *Coprobacillus*, whilst LH1063[T] was predicted to be a novel species belonging to *Coprobacter*. Both strains were characterized by polyphasic approaches including 16S rRNA gene analysis, core-genome analysis, average nucleotide identity (ANI) comparisons and phenotypic analysis. Initial screening of the 16S rRNA gene of LH1062[T] returned a nucleotide identity of 93.4% to *Longibaculum muris*. For LH1063[T], nucleotide identity was a 92.6% to *Coprobacter secundus*. Further investigations showed that LH1062[T] had a genome size of 2.9 Mb and G+C content of 31.3 mol%. LH1063[T] had a genome size of 3.3Mb and G+C content of 39.2 mol%. Digital DNA-DNA hybridization (dDDH) and ANI values of LH1062[T] with its closest relative, *Coprobacillus cateniformis* JCM 10604[T], were 20.9 and 79.54%, respectively. For LH1063[T], the dDDH and ANI values with its closest relative, *Coprobacter secundus* 177[T], were 19.3 and 77.81%, respectively. Phenotypic testing confirmed that LH1062[T] could not be matched to a known validly published isolate in any database; thereby indicating a novel genus for which the name *Allocoprobacillus* gen. nov. is now proposed with LH1062[T] (=DSM 114537[T]=NCTC 14686[T]) being the type strain of the proposed novel species *Allocoprobacillus halotolerans* sp. nov. Strain LH1063[T] (=DSM 114538[T]=NCTC 14698[T]) fits within the genus *Coprobacter* and, it being the third species within this genus, the name *Coprobacter tertius* sp. nov. is proposed.

INTRODUCTION

The metagenomic era has allowed researchers to delve into the microbiota diversity of the human gut and associate certain taxa with disease status. However, in order to understand underlying mechanisms and develop new microbiota-based therapies, pure and well-characterized isolates are required. Certain taxa can be problematic to culture due to their fastidious nature, including their acute sensitivity to oxygen, which has limited progress in this area. Thus, we sought to apply a culturing approach to our observational trial. In the Breast hEalth And Microbiota (BEAM) study, we isolated two bacterial strains that did not have a match to the Type Strain Genome Server (TYGS) database. Further genomic investigations suggested the novelty of these isolates. Herein we describe two novel bacterial strains: *Allocoprobacillus halotolerans* LH1062[T] and *Coprobacter tertius* LH1063[T].

RESULTS

Isolation and ecology

Faecal samples were donated by breast cancer patients as part of the BEAM study in partnership with the Norwich and Norfolk University Hospital. This study gained favourable ethical approval by the Faculty of Medicine and Health Ethics board at the University of East Anglia (201819-092HT). Patients who were aged between 30–60 years old, with first time diagnosis of invasive breast cancer and did not have any antibiotics 3 months prior to consenting were eligible to partake. LH1062^T was isolated from a 58-year-old patient diagnosed with adenocarcinoma with ductal carcinoma *in situ*. The tumour was stage 1C with no presence of invasion into lymph nodes. Isolate LH1063^T was isolated from a healthy 51-year-old female. The protocol for faecal collection was laid out by the Norwich Research Park (NRP) Biorepository (Norwich, UK), and was in accordance with the terms of the Human Tissue Act 2004 (HTA) and approved with license number 11208 by the HTA.

Approximately 1 g faecal sample was transferred into preservation medium (20% glycerol in sterile PBS) and stored at –80 °C until further use; 100 mg was taken from the glycerol frozen aliquot and homogenized in sterile reduced PBS. Consequently, a serial dilution was prepared and 200 µl spread on a 14 cm agar plate with yeast–casitone–fatty acid (YCFA) medium supplemented with carbohydrates (glucose, maltose and cellobiose) and brain heart infusion (BHI) medium [1]. The plates were incubated anaerobically at 37 °C in an atmosphere containing N_2, CO_2, H_2 (85, 5 and 10%, respectively) for 72 h before colonies were picked and purified by re-streaking at least three times with 48 h growth periods in between. A pure liquid culture was prepared for long-term storage using 20% glycerol solution.

Genomic characterization

A. halotolerans LH1062^T was grown in BHI media and *C. tertius* LH1063^T in YCFA, both for 48 h. Genomic DNA was extracted using the FastDNA Spin Kit for Soil (MP Biomedicals, 116560200), following the manufacturer's protocol with an amendment of 3 min bead-beating procedure as described in [2]. The genome for LH1062^T, *Allocoprobacillus halotolerans*, was sequenced using the Nanopore MinION sequencing platform. Sequencing for LH1062^T was performed following Oxford Nanopore's native barcoding genomic DNA protocol (SQK-LSK109). The sequence reads were initially filtered through Filtlong version 0.2.1 [3], with only the top 90% quality of reads remained for subsequent genome assembly. Consequently, the genome was assembled using Flye version 2.9 [4], resulting in one contig of 2.93 Mbp and a G+C content of 31.26 mol%. Sequencing for LH1063^T was carried out on an Illumina NextSeq500 instrument; libraries were prepared using a novel modified Illumina DNA prep tagmentation approach [5]. The genome was quality-filtered using Fastp version 0.20.0 with option -p 20 [6]. *De novo* assembly was performed using Spades version 3.11 [7]. BactspeciesID version 1.2 was used to check for contamination and provided a preliminary identity [8]. Using the output of Bactspe-ciesID, if available, the type strain was downloaded and FastANI version 1.33 used to compare the query whole genome to the type strain whole genome to confirm identity [9]. For isolates that did not provide a BactspeciesID output, the 16S rRNA gene sequence was extracted *in silico* using BactspeciesID and run through BLASTn [10] and type strain whole genomes downloaded to use for final confirmation. The two query genomes were screened using GTDB (release R207) [11], not resulting in any matches. The assembled draft genome for LH1063^T had 25 contigs, a genome size of 3.3 Mbp and a G+C content of 39.23 mol%. Using Protologger [12], we identified LH1062^T as representing a novel genus whilst LH1063^T as representing a novel species [12]. The genomes were also run on the TYGS [13] suggesting two potential novel species for both isolates.

Allocoprobacillus halotolerans LH1062^T

The full-length 16S rRNA gene sequences (1.5 Kb) of 35 species representing 35 genera within the family Erysipelotrichaceae were downloaded from the List of Prokaryotic names with Standing in Nomenclature (LSPN: June 2022) [14]. The 16S rRNA gene sequence for *Coprobacillus cateniformis* was also included after Protologger suggested it was the closest relative based on ANI results. The 16S rRNA gene sequences were aligned using MUSCLE version 3.8.31 [15] prior to the reconstruction of a maximum-likelihood phylogenetic tree using IQ-TREE version 2.0.5 [16] with the TEST model at 1000 bootstrap replications and subsequent visualization using iTOL version 6 [17]. LH1062^T was placed next to *Massiliomicrobiota timonensis* SN16 (Fig. 1a), but according to LSPN this has yet to be validated as an official new genus and species. The 16S rRNA gene sequence similarity between LH1062^T and *M. timonensis* SN16 was 96.49%. According to the Fig. 1a, *Intestinibaculum porci* KCTC 15725^T seems to be the closest relative to LH1062^T; however, the 16S rRNA percentage identity with *I. porci* KCTC 15725^T is only 89.32%, whereas that with *L. muris* DSM 29487^T is 93.04%. We also compared the 16S rRNA gene sequence of LH1062^T with that of *C. cateniformis* JCM 10604^T, as suggested by the Protologger result, which had a nucleotide identity of 90.80%. We reconstructed a phylogenomic tree using PhyloPhlan version 3.0.51 after downloading the genomes of the species used in Fig. 1a. No whole genome sequences could be found for *Breznakia pachnodae* Pei061^T and *Absiella argi* N6H1-5^T. The configuration file specified the use of DIAMOND version 0.9.19 and MAFFT version 7.515 as the aligner. Sequences were trimmed using TRIMAL version 2.4.rev15 and the tree reconstructed using IQ-TREE version 2.1.4. The tree was reconstructed with the PhyloPhlAn options –diversity medium and –accurate. Fig. 1d shows the genomic tree. As suggested by PhyloPhlAn, LH1062^T is placed amongst the family Coprobacillaceae and is closely related to *L. muris* DSM 29487^T and *C. cateniformis* JCM 10604^T. However, based on the 16S rRNA gene sequences, it has a higher identity percentage to *L. muris* DSM 29487^T than *C. cateniformis* JCM 10604^T. Further genomic investigation between LH1062^T and *L. muris* DSMZ 29487^T indicated dDDH was estimated at 21.7% (TYGS), whilst
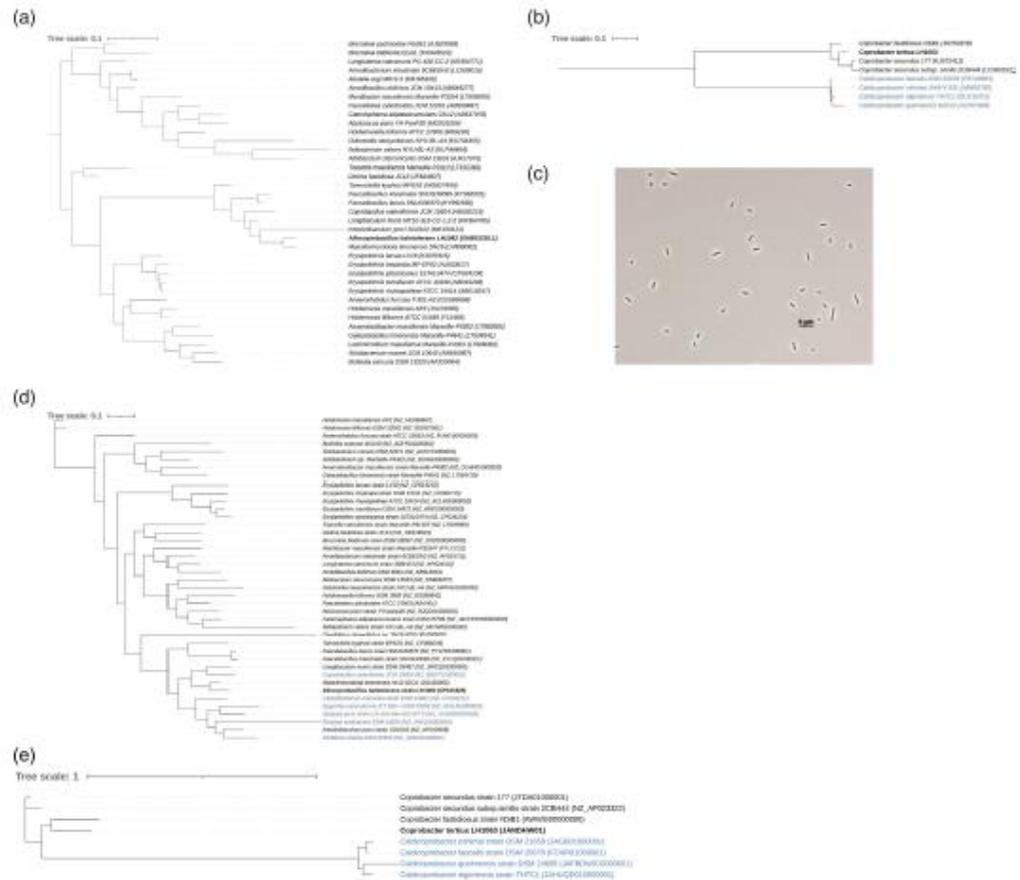
Fig. 1. (a) A mid-point rooted maximum-likelihood phylogenetic tree of *Allocoprobacillus halotolerans* LH1062^T relative to 16S rRNA gene sequences of the genera in the family Erysipelotrichaceae. (b) Mid-point rooted maximum-likelihood phylogenetic tree of Coprobacter tertius LH1063^T, relative to the 16S rRNA gene sequences of the species in the genus Coprobacter. The outgroups are members of the genus *Caldicoprobacter*, belonging to the family Caldicoprobacteraceae. (c) Phase contract microscopy image of LH1063^T. (d) Genome tree reconstructed using PhyloPhlAn for LH1062^T. In blue are outgroup genomes, which are the species in the genus *Coprobacillus*, family Coprobacillaceae. (e) Genome tree reconstructed using PhyloPhlAn for LH1063^T. In blue are outgroup genomes, which are the species belonging to the genus *Caldicoprobacter*, family Caldicoprobacteraceae.

the ANI is 79.3% (fastANI v1.3) [9, 13]. The dDDH comparison between LH1062^T and *C. cateniformis* JCM 10604^T is 21% and ANI 78.8%. The highest ANI (79.3%) and dDDH (21.7%) values were significantly below the intra-species thresholds of 95 and 70% for ANI and dDDH, respectively. We used EzAAI version 1.2.1 [18] to calculate average amino acid identity (AAI) using the genomes shown in Fig. 1a. The highest percentage was matched to *C. cateniformis* JCM 10604^T at 73.25%. This was followed by *L. muris* DSM 29487^T at 71.22% and *I. porci* SG0102^T (KCTC15725^T) at 62.29%. AAI with *M. timonensis* SN16 was only 50.65%. Using Protologger [12], the percentage of conserved proteins (PoCP) analysis assigned LH1062^T to *Clostridium* with a value of 50.08%, which is borderline to be suggestive of a novel genus. However, using BLASTN and limiting the search to the 'Bacillus/Clostridium group', the highest 16S rRNA gene sequence from the genus *Clostridium* was 89%, which is even lower than to *L. muris* DSM 29487^T and to *C. cateniformis* JCM 10604^T. Taken together and given the inconsistencies within the genus *Clostridium*, a novel genus *Allocoprobacillus* is proposed, with *Allocoprobacillus halotolerans* sp. nov. as the type species and LH1062^T representing the type strain.

Phenotypic investigations were carried out by DSMZ Services, Leibniz-Institute DSMZ–Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany. This involved: cell and colony morphology, salt and temperature tolerance, fermentation profiles of different carbohydrates, catalase activity, oxidase activity, and fatty acid analysis. Despite the strain being isolated in BHI, it was found to grow better in PY+X medium (DSMZ: 104b), which was subsequently used for the phenotypic testing. LH1062^T cells were found to grow in rods and in chains and were Gram-positive. They were negative for catalase and oxidase activity but positive for haemolytic activity. The bacterium grew in a relatively broad range of salt conditions (1–20%), with growth delayed between 7–20%. It failed to grow at temperatures below 25°C, but grew normally up to a maximum of 40°C; however, weak growth was observed at 45°C. Optimum growth was observed between 30–40°C. Biochemical characteristics were observed using API50CHB strips. Weak activity for D-arabinose, L-arabinose, D-xylose, L-xylose, fructose, mannose, sucrose, turanose and D-lyxose and positive activity for glucose, sorbose, aesculin, gentiobiose, D-tagatose and 5-ketogluconate was observed. The isolate was also incubated in the Gen III Biolog MicroPlate using Medium A. The inoculum was grown to a transmission of 93% turbidity before anaerobically incubating with the substrates for 48h at 37°C; after which, the plate was read using Biolog's Microbial Identification Systems software. The bacterium was positive for gentiobiose, D-fructose, D-fucose, L-fucose, L-rhamnose, D-serine, D-fructose-6-PO4, minocycline, L-galactonic acid lactone, D-glucuronic acid, glucuronamide and sodium butyrate (Table 1). We noted inconsistencies between the API strips and the Gen III Biolog MicroPlate results, which may be due to the fact that the API50CHB strip was incubated for 12 days at 37°C aerobically (covered in paraffin), whereas the Biolog MicroPlate assay was performed anaerobically. Comparing the reaction patterns with the not validly published isolate *M. timonensis* SN16, there is a distinct difference in the ability of LH1062^T to react with these substrates. Although the closest relative *L. muris* DSM 29487^T was negative for the acidification of carbohydrates, LH1062^T was not (Table 1). *C. cateniformis* JCM 10604^T was found to have acidification of glucose, mannose, galactose, fructose, sucrose, maltose, cellobiose, lactose and trehalose, following a similar profile to LH1062^T. Cellular fatty acids were detected after converting them into fatty acid methyl esters (FAMEs) following a modified protocol [19]. The FAME mixture was separated by gas chromatography and detected by a flame ionization detector using Sherlock Microbial Identification System (MIDI) based on the TSBA6 database. C$_{16:0}$ was the most abundant fatty acid for LH1062^T at 19.08%. This was also the major fatty acid for the not validly published isolate *M. timonensis* SN16, at 41% and for *L. muris* DSM 29487^T at 30.1%.

Based on the genomic and phenotypic results presented above, we propose LH1062^T as the type strain of a new genus *Allocoprobacillus* gen. nov., naming it like its closest genomic relative based on 16S rRNA gene sequence results, i.e. *Coprobacillus*. Strain LH1062^T is suggested reperesent a novel species named *Allocoprobacillus halotolerans* sp. nov.

Coprobacter tertius LH1063^T

For *Coprobacter tertius* LH1063^T, the 16S rRNA gene sequences representative of six *Coprobacter* species and two *Coprobacter secundus* subspecies type strains were downloaded from LSPN (LSPN: June 2022) [14]. The maximum-likelihood phylogenetic tree was generated as aforementioned for *A. halotolerans* LH1062^T. Strain LH1063^T was placed next to *Coprobacter fastidiosus* and *Coprobacter secundus* (Fig. 1b), based on the 16S rRNA gene sequences. From the phylogenomic tree (Fig. 1e), strain LH1063^T is more closely related to *C. fastidiosus* NSB1^T than *C. secundus* species. Based on 16S rRNA gene sequences comparison (Protologger), the closest relative was *C. secundus* with a nucleotide identity of 91.5%. *C. secundus* was also the closest match based on ANI at 77.81% (Table 2). We note that the ANI values for *C. secundus* 177^T reported by Protologger and OrthoANI [20] are different, being 77.81 and 72.8%, respectively. This discrepancy is explained due to Protologger using fastANI while OrthoANI uses USearch. OrthoANI was used as opposed to fastANI as fastANI would not provide an output if the ANI<80%, which it was for each species in the genus *Coprobacter*. The dDDH values between LH1063^T and *C. fastidiosus* DSMZ 26242^T, *C. secundus* 177^T and *C. secundus* subsp. *similis* 2CBH44^T were 20.1, 19.4 and 19.3% respectively. LH1603^T had a genome size of 3.3 Mbp and a G+C content of 39.23 mol%, whilst the genome size and G+C content for *C. secundus* 177^T are 4.1 Mbp and 37.8 mol%, respectively.

Phenotypic investigations were also carried out by DSMZ Services, Leibniz-Institute DSMZ–Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany. This involved: cell morphology, salt, bile and temperature tolerance, fermentation profiles of different carbohydrates and fatty acid analysis. *C. tertius* LH1063^T cells grow as rods in pairs measuring roughly 3 μm long (Fig. 1c). The strain was Gram-stain-negative and negative for catalase, oxidase and haemolytic activity. The bacterium tolerated up to 3% salinity and was shown to grow well between 30–40°C, with weak growth at 25°C. Unlike *C. fastidiosus* and *C. secundus*, LH1063^T failed to grow in any concentration of ox gall [21, 22]. Biochemical characteristics were observed using API20A strips and the inoculation was grown anaerobically at 37°C for 24h before the test was performed. The strain produced acid from glucose, lactose, maltose, mannose, raffinose and trehalose. The strain could hydrolyse gelatin and aesculin, and was weakly positive for acid production from mannitol, sucrose and melezitose. In APIrID32A assays, the strain was positive for α-galactosidase, β-galactosidase, α-glucosidase, N-acetyl-β-glucosaminidase, raffinose and glutamic acid decarboxylase fermentation, alkaline phosphatase, arginine phosphatase, leucyl-glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, tyrosine arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl-glutamic acid arylamidase, and serine arylamidase. The bacterium also had a weak reaction for mannose fermentation. Strain LH1063^T was additionally run on the Gen III Biolog MicroPlate with the same conditions described previously for strain LH1062^T. It was positive

Table 1. Comparison of analytical profile indexes of strain LH1062ᵀ, *Massiliomicobiota timonensis* SN16 and *Coprobacillus cateniformis* JCM 10604ᵀ

Two tests were used, API50CHB and Biolog GenIII Microplate, to test for a broader range of substrates for *A. halotolerans* LH1062ᵀ. Blank cells indicate that the substrate was not present in the test. The *M. timonensis* SN16 profile was determined using previously published literature [23]. +, Positive; –, negative; –/ +, borderline/weak.

	API50CHB for LH1062ᵀ	*M. timonensis* SN16	Biolog III microplate for LH1062ᵀ	*C. cateniformis* JCM 10604ᵀ
Methyl α-ᴅ-glucoside	–	–		
Methyl α-ᴅ-mannoside	–	+		
2-Ketogluconate	–	–		
5-Ketogluconate	+	+		
Adonitol	–	+		
Amygdalin	–	+		–
Arbutin	–	+		
Cellobiose	–	+		+
ᴅ-Arabinose	–/+	–		–
ᴅ-Arabitol	–	–	–	
ᴅ-Fucose	–	+	+	
ᴅ-Lyxose	–/+	+		
ᴅ-Tagatose	+	+		
Turanose	–/+	+	–/+	
ᴅ-Xylose	–/+	–		–
Dulcitol	–	+		
Erythritol	–	+		–
Aesculin	+	+		–
Fructose	–/+	+	+	+
Galactose	–	+	–/+	+
Gentibiose	+	+	+	
Gluconate	–	+		
Glucose	+	+	–/+	+
Glycerol	–	–	–	
Glycogen	–	+		–
Inositol	–	–	–	–
Inulin	–	+		
ʟ-Arabinose	–/+	–		
ʟ-Arabitol	–			
ʟ-Fucose	–	+	+	
ʟ-Xylose	–/+	–		–
Lactose	–	+	–	+
Maltose	–	+	–	+
Mannitol	–	–	–	–
Mannose	–/+	+	–	+

Continued

5

Table 1. Continued

	API50CHB for LH1062^T	*M. timonensis* SN16	Biolog III microplate for LH1062^T	C.cateniformis JCM 10604
Melibiose	–	+	–/+	
Melizitose	+	+		–
N-Acetylglucosamine	–	–	–	
Raffinose	–	+	–	–
Rhamnose	–	–	+	–
Ribose	+	–		–
Salicin	–	+	–	+
Sorbitol	–	–	–	–
Sorbose	+	–		
Methyl β-D-xyloside	–			
Starch	–	+		–
Sucrose	–/+	+	–/+	
Trehalose	–	+	–	+
Xylitol	–	+		

after 48 h incubation for gentiobiose, melibiose, α-D-glucose, D-mannose, D-fructose, D-galactose, 3-methyl glucose, D-fucose, L-fucose, L-rhamnose, D-glucose 6-phosphate, D-fructose 6-phosphate, minocycline, D-galacturonic acid and D-glucuronic acid. The acidification of the various carbohydrates including α-glucose, D-mannose and D-fructose to name a few was also confirmed using the Gen III Biolog MicroPlate. The biochemical profile of strain LH1063^T is quite similar to that of *C. fastidiosus* NSB1^T and slightly different from *C. secundus* 177^T (Table 3). Cellular fatty acids were detected using the same method for LH1062^T. The major fatty acid produced by LH1603^T was anteiso-C$_{15:0}$ at just 25%, followed closely by iso-C$_{15:0}$ at 20%. This was similar to *C. fastidiosus* NSB1^T and *C. secundus* 177^T. For *C. fastidiosus* NSB1^T it was 23–27% and 26–27% and for *C. secundus* 177^T it was 0.24–0.34% and 0.59–0.70% for anteiso-C$_{15:0}$ and iso-C$_{15:0}$, respectively.

Based on the genomic and phenotypic results presented above, we propose LH1063^T as representing a novel species within the genus *Coprobacter*. We propose the epithet *tertius*, as this is the third species of this genus.

DESCRIPTION OF *ALLOCOPROBACILLUS* GEN. NOV.

Allocoprobacillus (Al.lo.co.pro.ba.cil.lus. Gr. masc. adj. *allos*, another, other, different; Gr. fem. n. *kopros*, excrement, ordure, faeces; L. masc. n. *bacillus*, a small rod; N.L. masc. n. *Allocoprobacillus*, another small rod isolates from faeces). The genus is placed into the

Table 2. Digital DNA–DNA hybridization (dDDH) and average nucleotide identity (ANI) percentages between species in *Coprobacter* compared to *Coprobacter tertrius* LH1063^T

dDDH was determined using TYGS [13]. ANI was determined using the EzBioCloud ANI calculator [20].

Species	dDDH	ANI
Coprobacter fastidiosus DSM 26242^T	20.1	71.9
Coprobacter secundus 177^T	19.4	72.8
Coprobacter secundus subsp. *similis* 2CBH44^T	19.3	72.0
Caldicoprobacter oshimai DSM 21659^T	18.7	63.3
Caldicoprobacter guelmensis DSM 24605^T	18.6	63.3
Caldicoprobacter faecalis DSM 20678^T	18.5	61.8
Caldicoprobacter algeriensis DSM 22661^T	18.2	60.6

Table 3. Comparison of analytical profile indexes of strain LH1063ᵀ and other members of the genus *Coprobacter*

Three tests were used: API 20A, API rID 32A and Biolog Gen III MicroPlate. The profiles of *C. fastidiosus* and *C. secundus* were based on previous published literature [21, 22]. +, Positive; –, negative; –/+, borderline/weak.

API 20A	*C. tertius* LH1063ᵀ	*C. fastidiosus* NSB1ᵀ	*C. secundus* 177ᵀ	API rID 32A/Biolog GenIII MicroPlate	*C. tertius* LH1063ᵀ	*C. fastidiosus* NSB1ᵀ	*C. secundus* 177ᵀ
Acid from arabinose	–	–	–	α-Arabinosidase	–	–	–
Acid from cellobiose	–	–	+	α-Fucosidase	–	–	+
Acid from glucose	+	+	+	α-Galactosidase	+	+	+
Acid from glycerol	–	–	–	α-Glucosidase	+	+	+
Acid from lactose	+	+	+	Alanine arylamidase	+	+	+
Acid from maltose	+	+	+	Alkaline phosphatase	+	+	+
Acid from mannitol	–/+	–	–	Arginine arylamidase	+	–	–
Acid from mannose	+	+	+	Arginine dihydrolase	–	–	–
Acid from melezitose	–/+	–	–	β-Galactosidase	+	+	+
Acid from raffinose	+	+	+	β-Galactosidase 6-phosphate	–	–	–
Acid from rhamnose	–	–	+	β-Glucosidase	–	–	+
Acid from salicin	–	–	–/+	β-Glucuronidase	–	–	+
Acid from sorbitol	–	–	–	Glutamic acid decarboxylase	+	+	–
Acid from sucrose	–/+	–	+	Glutamyl-glutamic acid arylamidase	+	–	–
Acid from trehalose	+	–	+	Glycine arylamidase	+	–	–
Acid from xylose	–	–	–	Histidine arylamidase	+	–	–
Aesculin hydrolysis	+	–	–	Indole production	–	–	–
Gelatin hydrolysis	+	+	–	Leucine arylamidase	+	–	–
Indole production	–	–	–	Leucyl-glycine arylamidase	+	+	+
Urease	–	–	–	Mannose fermentation	–/+	+	–/+
				N-Acetyl-β-Glucosaminidase	+	+	+
				Nitrate reduction	–	–	–
				Phenylalanine arylamidase	+	–	–
				Proline arylamidase	–	–	–
				Pyroglutamic acid arylamidase	–	–	–
				Raffinose fermentation	+	+	+
				Serine arylamidase	+	–	–
				Tyrosine arylamidase	+	–	–
				Urease	–	–	–

family *Coprobacillaceae* (phylum Bacillota) based GTDB-Tk comparison, but placed into the family Erysipelotrichaceae (phylum Firmicutes) based on 16S rRNA gene analysis. The ANI value of the type strain LH1062ᵀ with *Coprobacillus cateniformis* JCM 10604ᵀ, which was suggested to be the closest relative based on GTDB-Tk comparison, was 78.8%. The closest relative based on 16S rRNA gene analysis is *Longibaculum muris* DSM 29487ᵀ with a nucleotide identity of 91.84%. The higher nucleotide identity to *L. muris* suggests that LH1062ᵀ belongs to the family *Erysipelotrichaceae*. PoCP was 50.08% with *Clostridium*, furthermore suggesting that this could be novel. Based on phenotypic characterisation, *L. muris* DSM29487ᵀ was negative for carbohydrate acidification while LH1062ᵀ was not. LH1062ᵀ had more in common with the metabolic profile of *C. cateniformis* JCM 10604ᵀ. A

novel genus, *Allocoprobacillus*, is proposed within the currently validly named family *Erysipelotrichaceae* to accommodate isolate LH1062[T], with the type species being *Allocoprobacillus halotolerans*.

DESCRIPTION OF *ALLOCOPROBACILLUS HALOTOLERANS* SP. NOV.

Allocoprobacillus halotolerans [ha.lo.to'le.rans. Gr. masc. n. *hals* (gen. *halos*), salt; L. pres. part. *tolerans*, tolerating, enduring; N.L. part. adj. *halotolerans*, salt-tolerating].

Description is based on a single strain. Cells are Gram-positive, facultative anaerobic, haemolytic, rod-shaped and grows in chains. The bacterium grows well in a temperature range of 30–40 °C, with weak growth at 45 °C. It tolerates a range of NaCl concentrations (1–20%), with delayed growth between 7–20% in PY-X medium. Colonies on BHI medium after 48 h are circular, smooth, shiny with entire margins roughly 0.3–0.4 mm in diameter, which are positive for ribose, glucose, sorbose, aesculin, melizitose, gentibiose, D-tagatose and 5-ketogluconate. The major fatty acid produced is $C_{16:0}$. The type strain, LH1062[T] (DSM 114537[T]=NCTC 14686[T]), was isolated from a donated faecal sample from a 58-year-old breast cancer patient. The genome size is 2.92 Mbp with a G+C content of 31.26 mol%.

DESCRIPTION OF *COPROBACTER TERTIUS* SP. NOV.

Coprobacter tertius (ter.ti.us. L. masc. adj. *tertius* third, referring to the fact that this is the third species to be described within the genus *Coprobacter*).

Description is based on a single strain. Cells are Gram-negative, facultative anaerobic, non-haemolytic and absent for catalase and oxidase. LH1063[T] cells are rod shaped and grow in pairs. The bacterium tolerates NaCl concentrations between 1–3% and a temperature range of 30–40 °C with weak growth observed at 25 °C when grown in YCFA. Colonies on YCFA medium after 48 h are circular, shiny, convex with no clear margins roughly 0.1 mm in diameter and does not tolerate any concentration of ox bile. The strain is positive for gelatine and aesculin hydrolysis, α-galactosidase, β-galactosidase, α-glucosidase, N-acetyl-β-glucosaminidase, glutamic acid decarboxylase, alkaline phosphatase, arginine arylamidase, leucyl-glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, tyrosine arylamidase, alanine arylamidase, histidine arylamidase, glutamyl-glutamic acid arylamidase and serine arylamidase. The strain also produces acid from glucose, lactose, maltose, mannose, raffinose and trehalose. The major fatty acid produced is anteiso-$C_{15:0}$. The type strain, LH1063[T] (DSM 114538[T]=NCTC 14698[T]), was isolated from a faecal sample from healthy 51-year-old female. The genome size is 3.3 Mbp with a G+C content of 39.23 mol%.

References
1. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, et al. Culturing of "Unculturable" human Microbiota reveals novel Taxa and extensive Sporulation. *Nature* 2016;533:543–546.
2. Alcon-Giner C, Dalby MJ, Caim S, Ketshemety J, Shaw A, et al. Microbiota supplementation with *Bifidobacterium* and *Lactobacillus* modifies the Preterm infant gut Microbiota and Metabolome: an observational study. *Cell Rep Med* 2020;1:100077.
3. Wick R, Menzel P. Filtlong: a tool for filtering long reads by quality. https://github.com/rrwick/Filtlong 2021.
4. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
5. Baker DJ, Aydin A, Le-Viet T, Kay GL, Rudder S, et al. CoronaHiT: high-throughput sequencing of SARS-Cov-2 Genomes. *Genome Med* 2021;13:21.

6. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.

7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.

8. Kiu R. BACTspeciesID: identify microbial species and genome contamination using 16S rRNA gene approach; 2020. https://github.com/raymondkiu/bactspeciesID

9. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.

10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.

11. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;50:D785–D794.

12. Hitch TCA, Riedel T, Oren A, Overmann J, Lawley TD, et al. Automated analysis of genomic sequences facilitates high-throughput and comprehensive description of bacteria. *ISME Commun* 2021;1:16.

13. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 2022;50:D801–D807.

14. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;70:5607–5612.

15. Edgar RC. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. *bioRxiv* 2021.

16. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020;37:2461.

17. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.

18. Kim D, Park S, Chun J. Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. *J Microbiol* 2021;59:476–480.

19. Miller LT. Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *J Clin Microbiol* 1982;16:584–586.

20. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* 2017;110:1281–1286.

21. Shkoporov AN, Khokhlova EV, Chaplin AV, Kafarskaia LI, Nikolin AA, et al. *Coprobacter fastidiosus* gen. nov., sp. nov., a novel member of the family *Porphyromonadaceae* isolated from infant faeces. *Int J Syst Evol Microbiol* 2013;63:4181–4188.

22. Shkoporov AN, Chaplin AV, Khokhlova EV, Shcherbakova VA, Motuzova OV, et al. *Alistipes inops* sp. nov. and *Coprobacter secundus* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 2015;65:4580–4588.

23. Ndongo S, Khelaifia S, Fournier P-E, Raoult D. "*Massiliomicrobiota timonensis*," a new bacterial species isolated from the human gut. *New Microbes New Infect* 2016;13:25–26.