The Gastric Cancer Microbiome

Mary Elizabeth Booth

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Medicine

October 2023

Intellectual Property Statement

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

This thesis would not have been possible without the selfless donation of tissue by patients with gastric cancer. For that, I am truly grateful.

I would like to express my sincere gratitude to the Wellcome Trust for its financial support of this work through the 4Ward North Clinical Research Training PhD Fellowship.

I would like to thank my supervisors – Professor Heike Grabsch, Dr Henry Wood, Professor Phil Quirke, and Professor Mark Travis – for the guidance and support they have provided throughout my fellowship. Through their supervision, I have been supported to follow my research interests and develop my research skillset. Despite potentially challenging logistics due to starting my PhD in 2020 in the wake of the COVID-19 pandemic, I was supported in adapting my project to allow me to learn new skills and achieve my research aims.

I am grateful for the support of the Quirke laboratory. In particular, Mr Dan Bottomley for support with sequencing training and shipping of samples to Guelph, Ms Gemma Hemmings for help with processing fixed tissue, and Alice Westwood for supporting aspects of the ethics application. I would like to thank Dr Kate Marks for the peer-mentoring and friendship we have shared throughout my fellowship. I am also grateful for the time, enthusiasm, and patient-perspective provided by Mr John Barnes through his help with project development and patient information sheet design.

I am grateful to the upper gastro-intestinal surgical team at St James' University Hospital for their support with collecting tissue. I would like to thank Mr Andy Cockbain for his support and advice in the initial planning process. I am also grateful to Ms Cath Moriarty for her guidance and support in navigating the local processes required to set up a study within Leeds Teaching Hospitals NHS Trust.

I will be forever grateful to the entire Allen-Vercoe laboratory for the kindness and knowledge shared with me during my time in Guelph. I'm extremely grateful to Professor Emma Allen-Vercoe, for welcoming me into her laboratory with

such enthusiasm and for the invaluable insight and advice offered regarding my work. In addition to specific microbiology knowledge, I have learnt from her demonstration of excellent leadership skills and mentorship. A special thanks goes to Ms Sarah Vancuren for the training support throughout my time in Guelph, and in particular the initial crash-course in bacterial culture which enabled me to maximise my time at the University of Guelph and achieve more than my initial goals. I am grateful for all the friendships gained during my time in Guelph, in particular that of Ms Gwen Freeze, with whom I hold many happy memories of, both in the laboratory and on cross-country ski tracks.

Finally, I would like to thank my friends and family for the love and support over the past three years. I am fortunate to have unconditional support from my Mum and Dad, who always believe in me more than I believe in myself. I am eternally grateful to my husband James, for living the ups and downs of my PhD journey, accepting that it has taken over our lives at times, and for reminding me of the importance of life outside of work. Lastly, I would like to thank our daughter, Skye, for being an excellent motivation to get my thesis completed and for holding on until after my viva before making an appearance into the world.

Author's declaration of work performed

All of the work presented in this thesis was conducted by Dr Mary Booth, with the following exceptions:

Dr Henry Wood obtained the sequencing data of the TCGA stomach adenocarcinoma project from the National Institute of Health NCI Genomic Data Commons Data Portal. Dr Wood also generated microbiome data from sequencing data for both the TCGA and 100,000 Genomes Project data by removing the human reads from tumour sequencing data, and realigning remaining reads to microbial genome databases. Dr Alice Westwood drafted the initial participation information sheets for the IRAS ethics application. Prof Heike Grabsch provided the Lauren classification of TCGA gastric cancer cases for TCGA, where this was missing and where slides were available. Prof Grabsch also reviewed the digital slides of fresh samples with Dr Mary Booth, to determine the presence or absence of adenocarcinoma features. Ms Gemma Hemmings performed formalin-fixation and paraffin-embedding of tissue collected in Leeds, as well as haematoxylin and eosin staining of these slides. Ms Sarah Vancuren assisted with the UKG7 bioreactor bacterial isolations and the assembly of the UKG6 defined community. Ms Vancuren processed the raw Illumina reads from the bioreactor samples and processing and profiling of the nuclear magnetic resonance spectra from bioreactor samples.

Abstract

Helicobacter Pylori infection is recognised as an established risk factor for developing gastric cancer, although only 3% of infected patients develop gastric cancer. Previous sequencing-based studies have indicated changes in the gastric mucosal microbiota along the carcinogenic cascade. Gastric cancer is characterised by significant heterogeneity with respect to anatomical, histopathological, and molecular features, in addition to epidemiological heterogeneity. The relationship between these clinical and pathological factors and the intratumoural gastric cancer microbiome has not previously been systematically examined. Furthermore, current investigations of the gastric cancer microbiome are predominantly culture-independent studies, limiting interpretations to associations alone.

The work presented within this thesis investigated the intratumoural microbiome of established gastric cancer through use of existing databases and prospectively collected gastric cancer tumour tissue.

Using two large sequencing databases, the 100,000 Genomes Project and the Cancer Genome Atlas, the gastric cancer microbiome was characterised and the relationships between the microbiome and clinicopathological variables were explored. Microsatellite instability, low pathological depth of invasion, intestinal histological phenotype, and non-Asian origin were all associated with increased microbial abundance and alpha diversity.

Prospectively collected gastric cancer samples were used to investigate methods of characterising the microbiome through culture-dependent and culture-independent methods. Bacterial isolations were performed from five gastric tumour samples, resulting in the greatest number of bacterial genera isolated from any gastric cancer study.

A bioreactor model was developed, capable of supporting gastric cancer microbiota in continuous culture. This model was used to explore the impact of pH on the intratumoural gastric cancer community, demonstrating an increase in lactic acid bacilli at lower pH. Measured metabolites produced by the microbial communities differed to what has previously been observed by the production of colonic microbiota.

The findings presented in this thesis indicate that the gastric cancer microbiota differs according to clinicopathological factors which should be accounted for in future research. Different methodological approaches were explored and refined; culture-dependent and culture-independent methods are complimentary and should be used to further progress the understanding of the role of the microbiome in gastric cancer.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACRG	Asian Cancer Research Group
AER	aerobic
AN	anaerobic
CagA	cytotoxin-associated gene A
CIBERSORT	cell-type identification by estimating relative subsets of RNA transcripts
CIN	chromosomal instability
CRF	case report form
dMMR	mismatch repair deficiency
DNA	deoxyribonucleic acid
EBV	Epstein-Barr virus
EMP	Earth Microbiome Project
FAA	Fastidious Anaerobe Agar
FFPE	formalin-fixed paraffin-embedded
FMT	faecal microbiota transplantation
FNB	Fastidious Anaerobe Agar without blood
GC	gastric cancer
GeCIP	Genomics England Clinical Interpretation Partnership
GS	genomically stable
H. pylori	*Helicobacter pylori*
H&E	haematoxylin and eosin
HRA	Health Research Authority
ID	identity
IHC	immunohistochemistry
IRAS	Integrated Research Application System
JVN	josamycin, vancomycin, norfloxacin
LTHT	Leeds Teaching Hospitals NHS Trust
MA	microaerobic
M-SHIME	mucus simulator of the human intestinal microbial ecosystem

MaAsLin2	Multivariable Association Discovery in Population-scale Meta-omics Studies
MMR	mismatch repair
MSI	microsatellite instability
MSKCC	Memorial Sloan Kettering Cancer Center
MSS	microsatellite stable
MSS/EMT	microsatellite stable with epithelial mesenchymal transition
mut/Mb	mutations per Mb
NCI	National Cancer Institute
NGRL	National Genomic Research Library
NGS	next-generation sequencing
NHS	National Health Service
NMR	nuclear magnetic resonance
OTU	operational taxonomic unit
P-ECSIM	Proximal Environmental Control System for Intestinal Microbiota
P. micra	*Parvimonas micra*
PCR	polymerase chain reaction
PD1	programmed cell death protein-1
PD-L1	programmed cell death ligand
pDCs	plasmacytoid dendritic cells
PERMANOVA	permutational multivariate analysis of variance
pN	pathological lymph node status
PPI	proton pump inhibitor
pT	pathological depth of invasion
qPCR	quantitative polymerase chain reaction
R. gnavus	*Ruminococcus gnavus*
REC	Research Ethics Committee
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
SCV	somatic coding variant

SIMGI	SIMulator Gastro-Intestinal
SJUH	St James' University Hospital
sTSB	supplemented tryptic soy broth
TCGA	The Cancer Genome Atlas
TIM	TNO Gastro-Intestinal Model
TP53	tumour protein 53
Tregs	regulatory T cells
TRM	tissue-resident memory T
UoG	University of Guelph
UoL	University of Leeds
UoM	University of Manchester
UK	United Kingdom
V	hypervariable
VacA	vacuolating cytotoxin A
WGS	whole genome sequencing

Chapter 1 Introduction

In this chapter, gastric cancer (GC) and the GC microbiome is introduced. The current knowledge and major challenges of the GC microbiome is described. The chapter concludes by outlining the performed work presented in this thesis.

1.1 Gastric cancer epidemiology and heterogeneity

GC is the fifth most common cancer globally and is the fourth most common cause of cancer-related deaths worldwide; in 2020 there were over 1,000,000 new cases and 769,000 deaths attributable to GC [1]. Prognosis is poor, with significant geographical variation; five year survival rates are 60.3% and 68.9% in Japan and South Korea, respectively, but only 20.7% and 33.1% in the United Kingdom (UK) and United States, respectively [2]. GC is twice as common in men than in women and over 70% of GC occur in Asia [3].

Gastric adenocarcinoma represents over 90% of all GC and is characterised by a heterogeneous histological phenotype. The Laurén classification divides gastric adenocarcinomas into intestinal, diffuse, mixed, and indeterminate subtypes [4], whereas the 5th edition World Health Organisation (2019) classification divides gastric adenocarcinomas into tubular, papillary, poorly cohesive, mucinous, mixed, as well as other, rarer, subtypes [5]. Diffuse-type cancers are more commonly seen in Eastern than Western populations [6]. Whilst the incidence of intestinal-type cancers is declining worldwide, the incidence of diffuse-type GC appears to be stable or increasing in some countries [7,8].

GC may be categorised according to anatomical location. Moving proximally to distally, the stomach may be anatomically divided into the cardia, fundus, body, and antrum. The stomach is bordered by the gastroesophageal junction proximally and the duodenal bulb distally. The gastric mucosal epithelium is a columnar epithelium which folds into gastric pits, within which lie various glands. The mucous glands of the cardiac mucosa produce an alkaline mucus which protects surface epithelial cells. The principle glands within the fundus and body of the stomach are responsible for producing gastric fluid. Principle glands include parietal cells, which secrete hydrochloric acid, and chief cells, which produce pepsinogen and gastric lipase, as well as neuroendocrine cells. The antral mucosa consists of mucus producing cells, cells which secrete

progastricsin, precursor of the proteolytic enzyme gastricsin, and neuroendocrine cells.

Cardia GC and non-cardia GC differ with respect to risk factors and geographical distribution. Risk factors for cardia GC include obesity [9] and gastroesophageal reflux disease [10], whereas infection with the bacterium *Helicobacter pylori* (*H. pylori*) is a risk factor for developing non-cardia GC [11,12]. The highest rates of both cardia and non-cardia GC occur in Eastern and South-eastern Asia [13]. In most countries non-cardia GC rates are higher than cardia GC. However, for men in Australia, the United States and the UK, rates of cardia and non-cardia GC are similar [13].

There are a number of molecular classifications of GC [14]. The Cancer Genome Atlas (TCGA) Research Network carried out an analysis of 295 GC samples based on six molecular platforms (array-based somatic copy number analysis, whole-exome sequencing, array-based DNA methylation profiling, messenger ribonucleic acid (RNA) sequencing, microRNA sequencing and reverse-phase protein array). Through cluster analysis of data from this multiomics approach, four distinct subtypes of GC were defined [15]; Epstein-Barr virus (EBV)-related, GC with microsatellite instability (MSI), genomically stable (GS) GC, and GC with chromosomal instability (CIN). The four TCGA subtypes are associated with different molecular, histological, and clinical features. EBV-related GC was found to be associated with male sex, predominance in the gastric fundus or gastric body, an extreme CpG island methylator phenotype, *PIK3CA* mutation, and increased levels of programmed cell death ligand (PD-L1) expression [16]. MSI GC was found to be associated with increased mutation rates, in particular hypermethylation at the *MLH1* promotor. Clinically MSI GC has increased prevalence in females and older age, and tends to respond differently to systemic anticancer treatment, as discussed further below. GS GC was found to be associated with the diffuse histological phenotype, earlier age of diagnosis, and mutations in *CDH1* and *RHOA*. CIN GC represented 50% of the analysed TCGA cases and was found to be associated with intestinal histological phenotype, cancers of the gastric cardia or gastro-oesophageal junction, tumour protein 53 (*TP53*) mutations, and amplification of receptor tyrosine kinases.

In parallel to the TCGA, the Asian Cancer Research Group (ACRG) proposed a four-tier classification based on gene expression profiling, genome-wide copy

number microarrays and targeted gene sequencing of 300 GC samples [17]. GCs were subsequently assigned to one of four subtypes: MSI, microsatellite stable (MSS) with epithelial mesenchymal transition (MSS/EMT), MSS *TP53*-active, and MSS *TP53*-inactive. Although there is overlap, particularly with the MSI groups, the TCGA and ACRG classifications are not directly comparable. Within this classification, MSI GC was associated with gastric antral location, intestinal histological phenotype, earlier stage, better prognosis, hypermutation, and mutations in *KRAS* and the PI3K-PTEN-mTOR pathway. MSS/EMT GC was associated with diagnosis at a younger age, diffuse histological subtype, later stage, worse prognosis, and decreased mutation events. MSS *TP53*-active GC was associated with EBV infection and mutations in *APC*, *ARID1A*, *KRAS*, *PIK3CA* and *SMAD4*. MSS *TP53*-inactive GC was associated with the highest rate of *TP53* mutations as well as amplifications of *ERBB2*, *EGFR*, *CCNE1*, *CCND1*, *GATA6* and *MYC*. The TCGA and ACRG GC classifications are summarised in Figure 1.

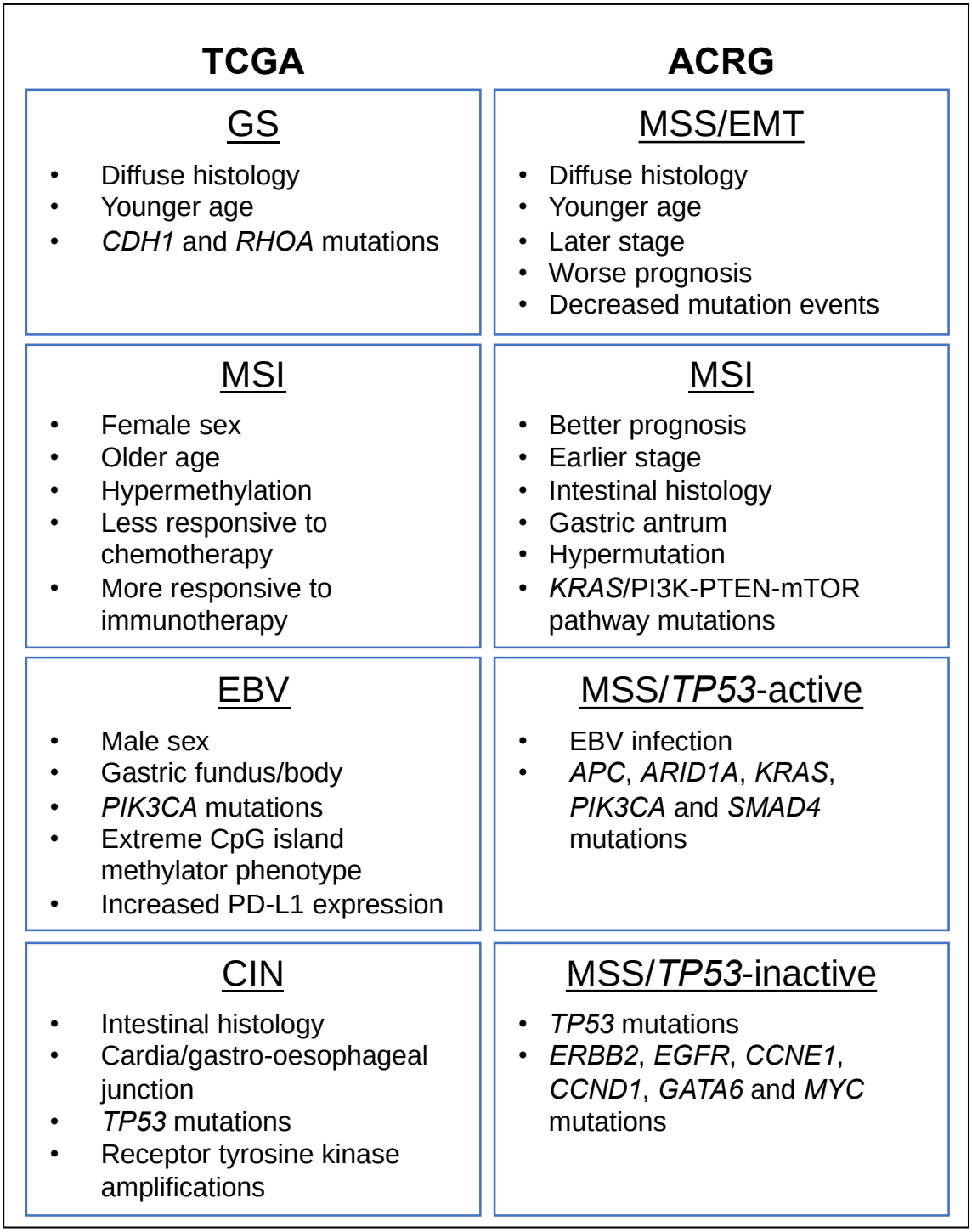


Figure 1. Overview of the TCGA and ACRG gastric cancer classifications.

ACRG, Asian Cancer Research Group; CIN, chromosomal instability; EBV, Epstein Barr virus; EMT, epithelial mesenchymal transition; GS, genomically stable; MSI, microsatellite instability; MSS, microsatellite stable; TCGA, The Cancer Genome Atlas; TP53, tumour protein 53.

MSI is of particular interest in GC as it can inform treatment decisions. Microsatellites are genome-wide short tandem repeats of one to six nucleotides [18]. MSI is the hyper-mutable phenotype which occurs at these microsatellites, resulting from deficient mismatch repair (MMR) mechanisms [19,20]. In normal

tissue, deoxyribonucleic acid (DNA) MMR proteins recognise and repair microsatellites. Defective MMR pathways can therefore result in unrepaired microsatellites, leading to MSI. MSI status may be evaluated through a number of different methodologies, either through detection of MMR deficiency (dMMR) through immunohistochemistry (IHC), or through direct MSI testing. IHC detection testing involves staining of four MMR proteins (MLH1, PMS2, MSH2 and MSH6). Absence of any of these proteins indicates dMMR and is interpreted as a surrogate for MSI. MSI testing may be through polymerase chain reaction (PCR) by fragment analysis of five satellite regions (commonly NR-21, BAT-25, MONO-27, NR-24, and BAT-26) [21,22]. Next-generation sequencing (NGS) based MSI testing is also possible, through targeting relevant microsatellite loci within an NGS panel [23].

Post-hoc analyses of adjuvant [24] and perioperative [25] chemotherapy trials has indicated that patients with MSI GC do not benefit from chemotherapy in addition to surgery, in contrast to MSS patients. However, the standard of care neoadjuvant chemotherapy for non-Asian patients is now FLOT, a regimen containing docetaxel. FLOT was not present in the trials analysed post-hoc, and which may activity in MSI cancers. Therefore, it is presently debated whether or not patients with MSI resectable GC should receive chemotherapy [26,27]. A meta-analysis of randomised controlled trials in GC suggested that GC with MSI responds well to immunotherapy [28]; this is currently being investigated in an ongoing prospective phase II study [29].

Despite multiple histological and molecular classifications being proposed over the last decade, these classifications are not currently used to guide treatment, with the exception of patients with MSI GC. Current classifications are therefore of limited benefit to GC patients.

1.2 The human microbiome

The microbiome refers to the collection of genomes of all microorganisms within an environment. The human body hosts at least as many bacterial cells as human cells [30]. The human microbiota consists of bacteria, viruses, fungi, archaea, and protozoa which inhabit the human body.

1.2.1 Microbiome study methodology

Current microbiome research encompasses a vast variety of study designs, techniques and analyses. Historically, examination of microbiota relied upon culture techniques, growing individual microbes *in vitro* and performing biochemical tests to characterise their phenotype, e.g. the Gram stain to characterise the bacterial cell wall. Today, a range of culture-independent and culture-based methodologies are available for the investigation of individual microbes and microbial communities. Community characterisation may be performed through sequencing of microbial DNA of marker genes. The most commonly employed example of this is the 16S ribosomal RNA (rRNA) gene, present in all bacteria [31] and used to infer bacterial taxonomy. Alternatively, microbial community characterisation may be performed through metagenomic sequencing, whereby genomic DNA (not limited to marker genes) from a given sample is sequenced. This may be used to infer both taxonomic composition as well as predicting gene function [32]. NGS allows high throughput characterisation of tissue through both 16S rRNA gene sequencing and metagenomic sequencing. For a more in depth description of 16S rRNA gene sequencing and metagenomic sequencing, the reader is directed to Chapter 3, section 3.1.4 and Chapter 2, section 2.1.1, respectively.

In addition to the observational characterisation of tissue, a variety of interventional methodologies exist, incorporating *in vitro* studies, *in vivo* studies, and clinical trials. Such studies allow phenotypical investigation of microbial isolates and testing of hypotheses generated though observational studies. Table 1 summarises the different methodologies available.

Table 1. Overview of microbiome study methodologies. 16S rRNA, 16S ribosomal ribonucleic acid; IHC, immunohistochemistry; qPCR, quantitative polymerase chain reaction. References of example studies included.

Design	Aims	Materials/Subjects	Methods	Notes
Observational	・Characterise microbiome ・Compare cohorts [33]	Subjects: ・Patients with cancer/ precancerous conditions [34–36] ・Healthy controls [37,38] Human tissue, e.g.: ・Gastric mucosa [39–45] ・Gastric fluid [42,46] ・Stool [47] ・Tongue coatings [48]	・16S rRNA gene sequencing ・Metagenomic sequencing ・qPCR ・IHC ・Specialist stains ・Culture of microbial isolates	Exploratory analyses may involve comparisons of microbiomes between populations, or with clinical/pathological features. Findings may be validated on a separate cohort.
In vitro interventional	・Test impact of certain microbiota on cancer cells ・Test behaviour of microbes/microbial communities under certain conditions	・Cell lines ・Individual microbes ・Microbial communities	・Microbial culture under reproducible conditions ・Co-culture of microbiota with cell lines [49]	Microbial culture may vary from simple monoculture, to microbial communities within bioreactor environments that mimic those of the gastrointestinal tract [50].

Design	Aims	Materials/Subjects	Methods	Notes
In vivo interventional	Evaluate the impact of certain microbiota to evaluate: ・Potentially causative microbiota ・Effect of different microbiota on survival or surrogate outcomes ・Effect of manipulation of microbiota	・Animal models ・Microbial communities	Controlled experiment involving manipulation of microbiome: ・Microbiome transplants [51] ・Antibiotic treatment ・Probiotic treatment	Microbiomes may be manipulated to test effects of certain microbiota. Tumours may be examined to analyse effect. Specific pathogen free mouse models reduce confounding variables, enabling mechanistic investigation of the impact of specific bacteria [52].
Human interventional	Evaluate impact of intervention on prespecified outcome measures	・Patients with cancer/other illness ・Healthy individuals	Clinical trial of: ・Manipulation of microbiome ・Microbiome specific patient pathways	Trials may assess microbiome manipulation for therapeutic benefit [53–55] or evaluation of microbiome stratified screening pathways [56].

1.2.2 The human microbiome in health and disease

Although it is sometimes assumed that the human microbiome refers only to the luminal lower gastrointestinal microbiome, almost all anatomical body sites host a microbiome. The composition of the human microbiome varies according to anatomical site, as well as between individuals at the same site [57]. The different communities present in different body sites are a reflection of the environments provided by the host, and the conditions required by the microbiota. For example, obligate anaerobes (microbes which would be killed by atmospheric oxygen levels), which may be part of the normal flora of the lower gastrointestinal tract lumen, could not tolerate the aerobic environment of the skin epidermis. Bacteria such as lactobacilli, which are highly abundant in the low-pH environment of the vagina, may be out-competed by other microbes in the oral cavity. Studies such as the Human Microbiome Project [58] have aimed to characterise the so-called “normal” microbiome of healthy individuals.

Dysbiosis refers to the imbalance of the microbiota composition and function, which may be a result of a reduction of beneficial organisms, growth of potentially harmful organisms, or loss of microbial diversity [59,60]. Changes in the intestinal microbiome have been associated with a range of adverse health conditions, including inflammatory bowel disease [61,62], obesity [63], arthritis [64], and depression [65]. Beyond the gastrointestinal system, microbiome dysbiosis is associated with disease not limited to infection. For example, reduced skin microbiome diversity has been correlated with atopic dermatitis [66] and changes in the vaginal microbiome have been associated with pre-term birth [67].

1.2.3 The microbiome and cancer

The microbiome has been investigated across a wide range of cancers. In the 2022 update of the “Hallmarks of Cancer” paper, “polymorphic microbes” was recognised as a new hallmark, demonstrating the increasing body of evidence linking the microbiome to cancer [68]. The understanding of the microbiome’s significance in cancer is rapidly expanding, with evidence of its role in cancer causation [3,69], diagnosis [70], response to treatment [71–74] and outcomes [51]. Such relationships have been observed across a range of cancers, including, but not limited to, cervical [3], colorectal [69,74], pancreatic [51,71], and melanoma [73]. The mechanisms through which the human microbiome

may be involved in cancer initiation, promotion and progression are varied, and incompletely understood at present [75].

Examples of bacteria which have been implicated in cancer development include *pks+ Escherichia coli*, Enterotoxigenic *Bacteroides fragilis*, and *Fusobacterium nucleatum* in colorectal cancer and *Salmonella typhi* in gall bladder cancer [76]. Despite the expanding understanding of the role of these bacteria in cancer development, the only bacterium established as a class one (definite) carcinogen is *H. pylori* [11]. *H. pylori* is discussed in detail in section 1.3.2. In addition, many viruses are recognised as established carcinogens, including human papillomavirus in cervical cancer, hepatitis B virus and hepatitis C virus in hepatocellular carcinoma, EBV in lymphoma, and human T-cell lymphotropic viruses in leukaemia and lymphoma [3].

An important distinction to be aware of when interpreting the literature surrounding cancer and the microbiome, is the environment or tissue to which the microbiome refers. Studies have investigated the faecal microbiome in relation to a range of cancers [69,73,77]. In addition to the faecal microbiome, microbes residing at mucosal sites may become part of the intratumoural microbiome; thus studies may investigate either the microbial content of the tumour itself, or the adjacent mucosa [78]. Importantly, the intratumoural microbiome is typically analysed by determining the microbial composition of a sample containing tumour as well as the surrounding tumour microenvironment. As such, intratumoural microbiome samples may include both intracellular and extracellular tumour cell microbiota, as well as the microbiota from surrounding vasculature, extracellular matrix, stroma, fluid, and immune cells.

1.3 Current understanding of the gastric cancer microbiome

1.3.1 The gastric microbiome in health and disease

The stomach was originally thought to be sterile organ due to the low pH and digestive enzymes within the gastric fluid; any microbes isolated from gastric contents were assumed to be transient [79]. In 1982, *H. pylori* was first cultured from the stomach of patients with gastritis and ulceration by Marshall and Warren [80,81], thus demonstrating that the stomach was capable of supporting microbes.

The first large-scale sequencing study to characterise the human stomach was performed by Bik *et al.* in 2006 [82]. This study, in which 16S rRNA gene sequencing was performed on 23 gastric endoscopic biopsy samples, reported that the majority of sequences belonged to five phyla: Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, and Fusobacteria. Further studies have supported the presence of these phyla in the human stomach [83,84]. In 2021, following the International Committee on Systematics of Prokaryotes vote to include phylum in the rules of the International Code of Nomenclature of Prokaryotes, the names of 42 phyla of bacteria and archaea were changed, with the full list of updated phyla being published by Oren *et al*. [85]. Therefore, the post-2021 names of these phyla most commonly found in the human stomach are: Pseudomonadota, Bacillota, Bacteroidota, Actinomycetota, and Fusobacteriota, respectively [85]. At genus level, sequencing studies of the gastric mucosa of healthy individuals has demonstrated the following dominant genera: *Streptococcus*, *Prevotella*, *Rothia*, *Veillonella*, *Neisseriae*, *Fusobacterium*, *Haemophilus*, *Porphyromonas*, and *Pasturellaceae* [82,86,87].

Gastric mucosal infection with *H. pylori* infection known to cause gastritis, in addition to other disease, as discussed in section 1.3.2. [88,89]. Beyond gastritis, study of the gastric microbiome has largely been limited to the context of GC.

1.3.2 *Helicobacter pylori*

H. pylori is a Gram-negative bacterium that colonises the gastric mucosa of over 50% of the world's population, with significant geographic variation [90]. Its ability to create urease, and therefore neutralise gastric acidity by converting urea to ammonia and carbon dioxide, enables the bacterium to colonise the acidic gastric environment. *H. pylori*'s mobility is facilitated by its flagella, enabling it to move from the gastric lumen through the mucus layer, to the mucosa.

H. pylori is an established risk factor for developing GC [11]. Correa [91] proposed a cascade of events leading to intestinal-type GC, whereby *H. pylori* infection initiates the sequential development of precancerous stages through non-atrophic gastritis, multifocal atrophic gastritis, intestinal metaplasia, dysplasia and ultimately cancer. A meta-analysis of randomised controlled trials of *H. pylori* eradication demonstrated that eradiation reduced the incidence of

GC [92]. However, fewer than 3% of individuals colonised with *H. pylori* will go on to develop GC [93]. This disparity has led to the study of *H. pylori* virulence factors. The two most studied virulence factors associated with increased GC risk are cytotoxin-associated gene A (CagA) and vacuolating cytotoxin (VacA) [94–96], both present only within *H. pylori*, and not found within other *Helicobacter* species [97].

The *cag* pathogenicity island, a locus containing 31 genes, is found in some *H. pylori* strains [91]. The terminal gene in the island, *cagA*, is used as a marker for the entire *cag* locus, defining strains as *cagA* positive or negative. Several *cag* island genes encode type IV secretion system proteins, enabling the protein CagA, a recognised oncoprotein [91], to be translocated into epithelial cells [98]. CagA subsequently increases cell proliferation through activation of the Wnt-β-catenin pathway[99]. *cagA*+ *H. pylori* strains are associated with higher grades of gastric inflammation [100], increased risk for precancerous gastric lesions [101] and GC [94,95,102].

All *H. pylori* strains contain a vacuolating cytotoxin (*vacA*) gene and most secrete the protein it encodes, the toxin VacA [96]. VacA, induces vacuole formation in the cytoplasm of gastric epithelial cells, as well as a variety of other effects following internalisation. Such effects include interference in mitochondrial function [103], stimulation of apoptosis [104] and inhibition of immune cells [105–107]. There is variation in VacA toxin activity with different *H. pylori* strains [102,108]. Infection with *vacA* s1m1 positive *H. pylori* is associated with increased risk of GC [102].

The role of the non-*Helicobacter* microbiome in the development of GC is less well studied, compared to *H. pylori*. Despite *H. pylori* infection being a risk factor for developing GC, *H. pylori* is typically present at lower levels in patients with GC and premalignant conditions than benign conditions [34,109]. de Assumpcao *et al*. [110] proposed that this is due to changes in the gastric environment initiated by *H. pylori*, whereby persistent inflammation leads to atrophic gastritis, and therefore fewer acid-secreting gastric parietal cells, facilitating the colonisation by other bacteria as the conditions become less acidic. However, beyond *H. pylori*, knowledge of the role of the microbiome in GC development is limited.

1.3.3 Epstein-Barr Virus

EBV, also known as *Human gammaherpesvirus 4*, infects 90% of the population worldwide, asymptomatically in most cases [111]. It has been demonstrated that EBV genomes are present within GC cells and adjacent dysplastic epithelium although absent in surrounding normal cells [112–114]. Representing approximately 9% of cases [15,115,116], EBV positive GCs are more common in males [115], in more proximal GCs [115], and are associated with improved survival relative to EBV negative GC [117]. Whilst the role of EBV in gastric carcinogenesis is not fully understood, it is thought that transcription and translation products of EBV positive cells, together with genetic abnormalities of host cells co-potentiate tumour development [118].

1.3.4 Microbiome composition in gastric cancer

As discussed in section 1.2.3, the intratumoural microbiome consists of both the intra- and extracellular tumour microbiome, as well as the microbiome of the surrounding tumour microenvironment. Within this thesis, the term ‘GC microbiome’ will refer to the microbiome of a GC tumour sample, inclusive of any extracellular and cells of the local tumour microenvironment.

Increased availability and reduced cost of microbial sequencing has progressed GC microbiome research beyond focusing on single microorganisms, towards analysis of whole microbiomes of GC patient cohorts. Such studies have analysed the differential compositions of the microbiota of patients along the GC carcinogenic cascade. Table 2 summarises findings of nine observational sequencing studies which have characterised the gastric microbiome genera of GC patients with either healthy or precancerous gastric mucosa. Since these studies focussed on relative microbial differences between mucosa of GC versus healthy or precancerous states, these studies cannot be used to differentiate whether a particular genus may be present at a lower abundance or not at all, when described as less abundant in a particular environment. Whilst there is some overlap in findings between studies, overall the results are very mixed. As such, a consensus dysbiotic microbiome associated with GC has not yet been identified.

Gastric mucosal samples have previously been compared with gastric fluid samples of patients along the gastric carcinogenic cascade (inclusive of patients with superficial gastritis [n=61], intestinal metaplasia [n=55], and GC

[n=64] [119]. In this study, alpha diversity was greater in fluid samples compared to mucosal samples, where *Helicobacter*, *Lactococcus*, and *Bacillus* appeared to dominate. At phylum level, Bacteroidota, Fusobacteriota and Pseudomonadota were enriched in the gastric fluid, whereas Bacillota was enriched in the gastric mucosa.

Previous GC microbiome descriptive studies have highlighted the fact that genera identified through GC characterisation are often oral bacteria [40,46,120]. The Human Microbiome Project analysis of the oral cavity found that *Streptococcus* was the most dominant bacterial genera, with *Haemophilus*, *Prevotella*, and *Veillonella* also occurring frequently [121]. An analysis of the intratumoural GC microbiome of 276 patients with GC found *Helicobacter*, *Streptococcus*, *Halomonas*, *Lactobacillus*, and *Veillonella* to be the most prevalent bacteria genera, with *Prevotella* also featuring among the ten most abundant genera. Whilst there is significant overlap between species reported in the oral and the GC microbiome, *Helicobacter* and *Lactobacillus* are found more commonly in the GC microbiome than the oral microbiome, and *Haemophilus* is found less commonly in the GC microbiome.

Table 2. Summary of genera found at increased (blue) or decreased (green) abundances in GC compared to controls across nine studies. CG, chronic gastritis; FD, functional dysplasia; GC, gastric cancer; HC, healthy control; NAG, non-atrophic gastritis; SG, superficial gastritis. *Nested by H. pylori positivity.

Genus	**GC vs HC** [122]	**GC vs HC** [38]	**GC vs FD*** [39]	**GC vs NAG** [123]	**GC vs SG** [119]	**GC vs SG** [40]	**GC vs CG** [124]	**GC vs CG** [34]	**GC vs CG** [46]
Achromobacter									
Actinomyces									
Alloprevotella									
Anoxybacillus									
Aquabacterium									
Arthrobacter									
Atopobium									
Bosea									
Brachybacterium									
Bradyrhizobium									
Brevibacillus									
Burholderia									
Carnobacterium									
Citrobacter									
Cloacibacterium									
Clostridium sensu stricto									
Clostridium									
Deinococcus									
Devosia									
Dietzia									
Dolosigranulum									
Escherichia-Shigella									
Eubacteriaceae									
Eubacterium									
Flavonifractor									
Fusobacterium									
Gemella									
Gordonia									
Granulicatella									
Haemophilus									
Halomonas									
Helicobacter									
Hydrogenophaga									
Lachnoanaerobaculum									
Lachnospiraceae									
Lactobacillus									
Lactococcus									
Leptotrichia									

Genus	**GC vs HC** [122]	**GC vs HC** [38]	**GC vs FD*** [39]	**GC vs NAG** [123]	**GC vs SG** [119]	**GC vs SG** [40]	**GC vs CG** [124]	**GC vs CG** [34]	**GC vs CG** [46]
Methylobacterium									
Microbacterium									
Moryella									
Neisseria									
Nitrospirae									
Novosphingobium									
Ochrobactrum									
Paenibacillus									
Pasteurellaceae									
Phyllobacterium									
Porphyromomas									
Prevotella									
Propionibacterium									
Pseudoxanthomonas									
Psudonocardia									
Ralstonia									
Rhicobium									
Rhodcoccus									
Rothia									
Salinivibrio									
Sarcina									
Selenomas									
Serratia									
Sphingomonas									
Staphylococcus									
Streptococcus									
Tepidimonas									
Tsukamurella									
Unibacterium									
Urubuella									
Veillonella									

1.3.5 Microbiome diversity

In addition to exploring the abundance of particular microbes, microbiome sequencing enables microbial diversity to be analysed. Alpha diversity represents the richness and evenness of an average sample within a habitat, whereas beta diversity represents the variability between samples. Both measures can provide useful insight into the compositions and differences between different habitats.

Many studies have found reduced mucosal microbial alpha diversity in GC, compared to precancerous conditions [34,40,123,125] and healthy controls [38,126–128]. A meta-analysis of the 825 gastric biopsy samples (including superficial gastritis, atrophic gastritis, intestinal metaplasia, and GC) also reported lower alpha diversity in GC, relative to pre-cancerous subtypes [129]. In contrast to these findings, He *et al*. reported an increased alpha diversity of GC samples relative to precancerous lesions [119]. It is unclear why this finding was in contrast to other studies, but may be reflective of differences in the patient populations studied, such as rates of *H. pylori* infection.

Comparisons of alpha diversity between matched tumour and adjacent tissue microbiota have had mixed findings. Liu *et al*. [130] found reduced alpha diversity within peritumoural tissue, and tumour tissue of late stage GC when compared to matched normal tissue. This is consistent with two other studies, which also reported increased alpha diversity in the tumour tissue, relative to macroscopically normal issue [131,132]. In contrast, Chen *et al*. [133] found increased alpha diversity in cancerous tissue compared to non-cancerous tissue. It is possible this inconsistency could be a reflection of underlying population differences confounding the results.

There is consensus that the microbiome of tumour and adjacent non-cancerous tissue differ using beta diversity measures. Studies have found significant differences in the beta diversity of microbiomes of GC compared to precancerous stages [34,40,46,48] and the gastric mucosa of healthy controls [122]. In addition, early and advanced GC have been found to differ according to beta diversity [134]. Collectively, these studies are suggestive of microbiome compositional changes along the GC carcinogenic cascade. However, as described in section 1.3.4, studies to date have not consistently revealed what such compositional changes are.

To date, there has been little investigation of the relationship between the GC microbiome diversity and clinical features. Studies have reported no difference in the alpha diversity of gastric mucosal or GC microbiome according to anatomical site [82,120,124], age, sex [134].

1.4 Challenges of gastric cancer microbiome research

Some of the challenges in GC microbiome research are common across many studies of microbiome in human health. In addition, there are some concerns which are particularly challenging in the study of the GC microbiome.

Across all microbiome studies, variation in methodological approaches can have great impact upon study findings. The many different methodologies of sample collection, processing, analysing and interpreting results may all impact findings and limit comparisons between studies [135–138]. Standardisation within certain areas of microbiome research, by creation of, and adherence to an agreed set of guidelines may decrease such variation. However, since all methodologies have specific advantages and disadvantages, there is no single superior protocol which will answer all microbiome research questions.

Contaminants must also be considered, as local water and the results of other processing steps may also influence and confound findings [139–143]. In addition to environmental contamination, contamination may occur due to sampling. Sampling via endoscopic routes exposes tissue to luminal contaminants along the entire pathway from entry of the endoscope (the oral cavity in the case of upper gastrointestinal biopsy), to the sampling site [144]. Contamination is of particular importance in tissues with lower microbial biomass, such as GC, than in studies of material with greater microbial biomass (e.g. stool samples) as they may represent a greater proportion of sequenced microbes, therefore resulting in greater distortion of results. Contamination is discussed in greater detail in Chapter 2, section 2.1.4.

There is also significant inter-individual heterogeneity in the human microbiome, meaning that large sample sizes are often required in order to have sufficient power to detect significant differences in observational studies. Collection of fresh gastric tissue is an invasive procedure, thus, tissue is inevitably more difficult to obtain than blood or stool samples. Where existing tissue sequencing databases exist, they can offer an opportunity to obtain microbial information on large samples sizes through metagenomic approaches. *In vitro* [49,145] and *in vivo* [146] models may offer opportunities for microbiome investigations, independent of host variation.

Confounding factors represent an important consideration in GC microbiome research. Patient- and tumour-specific factors responsible for GC heterogeneity, as described in section 1.1, as well as diet, and pharmaceutical therapies (including antibiotics, pH altering medications, and probiotics) may influence, or be influenced by the GC microbiota. However, the relationships between such factors and the microbiome remains largely unknown. These possible confounders should therefore be measured where possible, and adjusted for where appropriate in analyses, although to do so relies upon the adequate collection of such clinical data.

Different methodological and analysis approaches may also impact upon study findings. Sequencing studies may all introduce different biases, dependent upon the sequencing methodologies and bioinformatic pipelines employed [147,148]. Cross-study comparisons should be interpreted with caution. Whilst it is not possible to remove such confounding factors completely, efforts should be made to identify and minimise such confounders where possible.

1.5 Thesis overview

Despite the existing GC microbiome studies, significant challenges remain in the field of the GC microbiome and many questions remain unanswered. GC heterogeneity and differences in methodological approaches are potential barriers to microbiome characterisation studies and should be explored in order to inform observational studies. Across microbiome-cancer studies, the number of observational studies is substantially greater than mechanistic studies [75]. Developing models upon which to study the GC microbiome is necessary to understand the importance of any observed correlations from observational studies.

The overarching aim of this thesis was to **progress the understanding of the intratumoural GC microbiome** and **how best to study it**. This was achieved through investigation of *in silico* sequencing data and investigations upon prospectively collected GC tissue. The specific objectives addressed are described below.

The objectives addressed in Chapter 2 were to **define the intratumoural GC microbiome** and **explore how the intratumoural GC microbiome is**

associated with clinical and pathological factors. Through this work, the importance of *in silico* decontamination was emphasised, and used to inform methodology used within Chapter 3.

The objective addressed in Chapter 3 was to **evaluate different methodological approaches to defining the intratumoural GC microbiome**. This work involved use of prospectively collected tissue, which was analysed through culture-independent and culture-based methodologies.

The objective addressed in Chapter 4 was to **develop an *in vitro* model for the culture of intratumoural GC microbiota in a continuous system**. This work involved adaptation of an existing model used to investigate colorectal cancer, to reflect the intratumoural GC environment.

Chapter 2 Characterising the gastric cancer microbiome using *in silico* data

2.1 Introduction

In this chapter, the methods and resources used to perform metagenomic analysis of sequencing databases are described. Using sequencing data from the 100,000 Genomes Project and TCGA, the GC microbiome is characterised, and relationships between clinicopathological characteristics and the microbiome are explored.

2.1.1 Metagenomic sequencing

NGS is a large-scale DNA sequencing method which can be used to analyse the entire genome (whole genome sequencing; WGS) or the exons within genes (whole exome sequencing; WES) [149]. When DNA is extracted from a clinical sample, it includes not only human DNA, but also DNA of the microbiota present in the sample. In contrast to traditional sequencing analysis techniques of human DNA, where non-human DNA sequences are discarded, in metagenomic NGS human DNA sequences are removed and the remaining non-human sequences represent the DNA of the tissue microbiota. DNA sequences are subsequently aligned to a reference database to identify microbial taxa present. Although the aim of WES is to target human protein encoding exons only, this process results in sufficient off-target sequencing of microbial reads to allow the metagenomic approach to be applied to exome sequencing data. Metagenomic sequencing can sequence bacteria, viruses, fungi, and archaea, but is more expensive than 16s rRNA gene sequencing per sample. A key advantage of metagenomic approaches is that it allows post-hoc analysis of WGS or WES databases where sequencing data have been collected for other uses, increasing the research potential of such databases.

2.1.2 The 100,000 Genomes Project

The 100,000 Genomes Project is a project run by Genomics England and was set up to sequence 100,000 whole genomes from National Health Service (NHS) patients, focussing on patients with rare diseases and patients with cancer. Patients were recruited from over 90 hospitals in England via one of 13 NHS genomic medicine centres, where tissue preparation, DNA extraction and quantification was performed according to standardised protocols [150].

Genomics England Clinical Interpretation Partnership (GeCIP) is the system through which funders, researchers, and clinicians collaborate to analyse and improve the clinical interpretation of the genomic and health data held within the National Genomic Research Library (NGRL). The Genomics England Research Environment is the platform through which approved GeCIP members can have controlled secure access to genomic and clinical data from the NGRL. To date, there are no published studies of metagenomic analyses of cancers from the 100,000 Genomes Project, although a characterisation of 2023 colorectal cancer samples is available as a pre-print [151], describing the colorectal intratumoural microbiome within this cohort.

2.1.3 The Cancer Genome Atlas

The TCGA is a programme in which 33 cancer types were molecularly characterised [152]. The TCGA was run by the National Cancer Institute (NCI) and the National Human Genome Research Institute, both of which lie within the National Institutes of Health, United States Department of Health and Human Services [153]. TCGA tissue was collected from sites across the world. Overview publications are available for each of the cancer types, detailing the characterisations performed and the initial analysis. The overview publication of the first 295 gastric adenocarcinoma cases was published in 2014 and described the six molecular platforms used to characterise samples, which included WES [116]. Within this article by the TCGA Research Network, the TCGA molecular subtypes were first described: EBV, MSI, CIN, and GS.

TCGA sequencing data of fresh-frozen gastric adenocarcinoma and matched blood samples is publicly accessible via the NCI Genomic Data Commons Data Portal [154]. The University of California Xena TCGA hub (https://tcga.xenahubs.net) is an online platform containing TCGA data where genomic and clinical data may be visualised and downloaded from. Microbiome analyses have been performed using TCGA data across a number of cancers. However, TCGA microbiome analyses which have included GC samples have predominantly focussed on differences between tumour tissue and adjacent tissue [70,155,156], as described in more detail in Chapter 1.

2.1.4 Decontamination processes

One rarely considered factor in microbiome analyses, including but not limited to metagenomic analyses, is contaminating microbial DNA related to the material sampling process or the laboratory environment [139–143]. Contamination may distort results and its effect has been shown to be exaggerated in studies of low microbial biomass [157,158]. Most GC microbiome studies, so far, have not included an *in silico* decontamination processes; this may have contributed to previously reported conflicting results.

In silico decontamination methods broadly fall into two categories: 1) exclude-list processes and 2) include-list processes. Exclude-list decontamination involves the removal of sequences or taxa previously identified through prior investigations as contaminants, e.g. as performed by Poore *et al*. [70]. This decontamination method may be seen as less stringent approach to decontamination as whilst the contaminants removed will have a high likelihood to be true contaminants, it may miss contaminants not previously recognised as such, and may not identify study-specific contaminants [159].

Included-list processes are those where a list of taxa is established which contains only taxa which are unlikely to be contaminants. Such include-lists are subsequently used to filter microbiome data such that only taxa contained on the include-list are used for downstream analysis following the decontamination process. Dohlman *et al.* [158] developed an include-list approach to a metagenomic analysis of WES data across cancers from the TCGA, using a statistical approach based on the assumption that taxa detected from blood samples were likely contaminants. Specifically, the authors considered species to be “tissue-resident” if they were both prevalent in fewer than 20% of blood samples and were statistically significantly more prevalent within tissue samples than in blood. Such “tissue-resident” taxa defined the include-list used for the *in silico* decontamination process. This more sophisticated approach to decontamination offers the advantage that it is study specific, however, relies on a relatively large sample size for sufficient power to identify true tissue-resident species.

2.1.5 Gastric cancer clinicopathological characteristics and the microbiome

As discussed in Chapter 1, GC is a heterogeneous disease, with variation in incidence and patient outcomes according to geographical origin, patient sex, histological phenotype, and molecular phenotype [1–3,5,17]. However, the contribution of patient and tumour characteristics to microbiome variation has not been systematically investigated and microbiome research has largely considered GC as a single disease.

A small number of gastric microbiome studies have explored the relationship between age or sex. Of the three identified studies which investigated the GC microbiome with age and sex, none found significant associations with age or sex [38,123,134]. These studies have been relatively small (the largest n=105, consisting of 76 patients with GC and 29 healthy controls) and all included mixed cohorts of patients with and without GC. When the relationship between age and sex, and the non-cancerous gastric mucosal microbiome has been investigated, results have been mixed. The largest of such studies found no association between gastric microbial features and age or sex in a population of 80 Chinese and 80 Mexican patients [120]. However, Yang *et al*. [33] found certain bacterial taxa to be correlated with patient age within both high and low GC risk Colombian populations. Wang *et al*. [160] identified a relationship between age and the microbiome composition in patients receiving upper gastrointestinal endoscopy who did not have GC. As such, there is currently no clear evidence of a relationship between the GC microbiome and age or sex, although this has not yet been examined in a large (n >100) population of patients with GC.

Given that different anatomical sites within the stomach harbour different environments, as described in Chapter 1, it may be reasonable to question whether the microbiota differs according to anatomical site. Despite the anatomical microenvironmental differences arising from the different areas of the stomach, studies have found no difference in composition or alpha diversity of the gastric mucosal or GC microbiome according to anatomical site [82,120,124].

The relationship between the GC microbiome diversity and TCGA molecular subtypes was investigated in a recent study by Abate *et al*. [156]. Within this

study, which included 286 TCGA GC samples and 234 samples from a GC separate cohort from the Memorial Sloan Kettering Cancer Center (MSKCC), no difference in alpha (within sample) diversity was observed between MSI, GS, and CIN GC. This study did identify a number of bacterial taxa enriched in GC classified as the MSI TCGA molecular subtype relative to GC of other molecular subtypes across the two cohorts, with *Lactobacillus* being the only genus statistically enriched across both cohorts in the MSI TCGA molecular subtype.

Given the multitude of factors which could potentially affect the GC microbiome, the clinical and pathological heterogeneity of GC, and the methodological differences in microbiome research, it is not surprising that previous findings have been conflicting. In this chapter, the relationship between clinical and pathological features and the GC is explored, using two distinct sequencing databases, the 100,000 Genomes Project and TCGA, with a large combined total number of patients. Through use of these two cohorts, and applying a decontamination step, a systematic explorative investigation of associations between clinical and pathological characteristics and the microbiome was performed.

2.1.6 Inferred immune cellular fractions and immune subtypes of TCGA data

In order to understand any relationships between the microbiome and GC molecular characteristics, it may be necessary to consider the immune cells of the local tumour microenvironment. The microenvironment, encompassing tumour cells, infiltrating inflammatory cells, stromal cells, as well as secreted components and extracellular matrix, is shared by the GC and its microbiota.

Cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) is a computational approach which uses gene expression data to provide an estimation of abundances of different cell types [161]. This methodology was applied to TCGA data by Thorsson *et al*. [162] and is publicly available.

In addition to estimating fractions of selected immune cells, Thorsson *et al*. also used cluster analysis based on 160 gene expression signatures to define six immune subtypes across TCGA cancers with distinct immune expression

signatures. These six immune subtypes are: wound healing, IFN-g dominant, inflammatory, lymphocyte depleted, immunologically quiet, and TGF-b dominant, named C1-C6, respectively. TCGA samples demonstrated distinct characteristics in macrophage or lymphocyte signatures, type 1 T helper: type 2 T helper cell ratio, extent of intratumoural heterogeneity, aneuploidy, extent of neoantigen load, overall cell proliferation, expression of immunomodulatory genes, and prognosis according to immune subtype.

2.2 Aims and objectives

The primary aims of the work described in this chapter were to define the intratumoural GC microbiome and to explore the relationships between clinicopathological features and the GC microbiome. Specific objectives were to:

- Curate databases of clinical, pathological and microbiome data from GC cases from: 1) the 100,000 Genomes Project and 2) TCGA.

- Further develop the *in silico* decontamination process, based upon that described by Dohlman *et al*. [158] and apply to both databases.

- Use the curated, decontaminated databases to determine most common taxa present within both cohorts.

- Use the curated, decontaminated databases to explore the relationship between microbiome abundance, diversity and composition, and clinical and pathological characteristics.

2.3 Methods

2.3.1 100,000 Genomes Project data acquisition

An application to become a GeCIP member was submitted in November 2020, under the ‘Upper gastrointestinal cancer’ domain. This application was approved and relevant information governance training was performed, prior to accessing the online Genomics England Research Environment. All analyses of Genomics England data were performed within the Research Environment. Source sample metadata was stored within the Research Environment, in tables within the data management software, LabKey.

WGS data of fresh-frozen primary gastric adenocarcinoma and matched blood samples from the 100,000 Genomes Project were accessed via GeCIP, via the Genomics England Research Environment under Research Registry project 307 [163].

A metadata database of relevant clinical, pathological and molecular information was created using data from two separate source tables from main_programme_v11_2020-12-17 within LabKey; cancer_analysis and av_tumour. The cancer_analysis table contained data generated from Genomics England, which had passed through the Genomics England quality checks. The av_tumour data was data generated from the National Cancer Registration and Analysis Service. Every participant identifier listed under the 'UPPER_GASTROINTESTINAL' disease subtype within cancer_analysis was cross-referenced within av_tumour; in cases where the pathology was stated as adenocarcinoma and site was either stomach or gastro-oesophageal junction, relevant pathology data, patient demographics and clinical information was inputted into the metadata database from the two source tables. Only those identified as primary adenocarcinoma of stomach or gastro-oesophageal junction were included for subsequent analyses. In one case, two gastric tumours were listed: one gastric gastrointestinal stromal tumour and one gastric adenocarcinoma. This case was not included in subsequent analyses as it was impossible to determine whether the associated microbiome data corresponded to the gastrointestinal stroma tumour or the gastric adenocarcinoma. Samples from the metadata, genus, and species tables were aligned by sample identifier.

DNA ploidy was obtained from Genomics England data via the "Tumour Bam" location in the cancer_analysis table. Ploidy data were previously calculated using copy number call data. Specifically, read counts from defined areas of each chromosomes were input into the "Battenburg" algorithm [164], which produced an estimate of the average number of copy gain across the entire genome, based upon tumour cell content and ploidy mixtures.

All clinical, pathological and molecular data used for exploratory analysis were obtained from the relevant Labkey tables. Year of birth and patient sex were obtained from the cancer_analysis table. Year of surgery, tumour location, pathological depth of invasion (pT), pathological lymph node status (pN), and histological phenotype were obtained from the av_tumour table. Age was determined by subtracting year of birth from year of surgery. Somatic coding

variants (SCVs), used to infer MSI status and TCGA subtype as described in section 2.3.4 was obtained from the cancer_analysis table.

2.3.2 TCGA data acquisition

Exome sequencing data of fresh-frozen primary gastric adenocarcinoma and matched blood samples from the TCGA stomach adenocarcinoma project was obtained from the National Institute of Health NCI Genomic Data Commons Data Portal [154] by Dr Henry Wood, using a controlled data access agreement (ID 25180). Basic clinical characteristics were obtained from the University of California Xena TCGA hub, (https://tcga.xenahubs.net) and Liu *et al*. [165]. TCGA GC molecular subclass data were obtained from the TCGA Research Network [116]. For the majority of cases, Lauren histological classification was publicly available [166]. For cases where Lauren classification was not available, the classification was provided by Professor Heike Grabsch after reviewing the TCGA GC virtual slides available from the National Institute of Health NCI Genomic Data Commons Data Portal [154], where slides were available.

2.3.3 Immune cellular fraction and immune subtype data acquisition

All CIBERSORT [161] immune cellular fraction estimates and immune subtypes were obtained from Thorsson *et al*. [162]. Estimates of six pre-selected immune cell subtypes (lymphocytes, neutrophils, macrophages, dendritic cells, eosinophils, and mast cells) and immune cell subtype (C1-C6) were analysed in exploratory analyses.

2.3.4 Microsatellite instability status and TCGA molecular subtypes

For the TCGA cohort, MSI status was obtained from previously published data [165]. These data were generated from PCR analysis of mononucleotide repeat sequences. In the present study, MSI-low cases were grouped with MSS cases, since MSS and MSI-low were previously reported to be similar with respect to mutations per Mb (mut/Mb) [167]. The TCGA molecular subtype of the TCGA GC cohort was obtained from the TCGA Research Network classifications [116].

Neither MSI (via PCR or MMR testing results), nor TCGA subtype data were available for the 100,000 Genomes Project cohort. Therefore, MSI status and TCGA molecular subtype were inferred using the following data from the Genomics England Research Environment: number of SCVs per sample, EBV sequencing count (virions per human cell), and DNA ploidy. Histograms of these variables (in log scale for SCV and EBV) were plotted to determine the thresholds for classifying samples. SCV and EBV data showed a bi-modal distribution (see results section, Figure 3 and Figure 4); as a result, for both variables the trough value between the two distributions was chosen as the threshold value. Since DNA ploidy demonstrated a multimodal distribution, the trough value between the first and second peak was selected as the threshold to represent the difference between diploid and polyploid samples.

These threshold values were subsequently used to infer MSI status and TCGA molecular subtype for 100,000 Genomes Project samples. Samples from the 100,000 Genomes Project were classed as MSS if they were in the lower SCV distribution and MSI if they were in the higher SCV distribution, defined by the SCV threshold value. Figure 2 demonstrates how (a) the TCGA cohort was classified into TCGA subtypes by Bass *et al*. [15] and (b) how the 100,000 Genomes Project was classified into inferred TCGA subtypes in the present study. For 100,000 Genomes Project samples, inferred TCGA molecular subtypes (EBV, MSI, CIN, GS) were established using threshold values, determined as described above. Samples were classed as EBV subtype if the EBV sequencing count was greater than or equal to the EBV threshold value. Samples not assigned to the EBV subtype, with SCVs greater than or equal to the SCV threshold value were assigned to the MSI subtype. Remaining samples were assigned to the GS subtype if the ploidy value was below the ploidy threshold value or the CIN subtype if the DNA ploidy value was greater than or equal to the DNA ploidy threshold value.

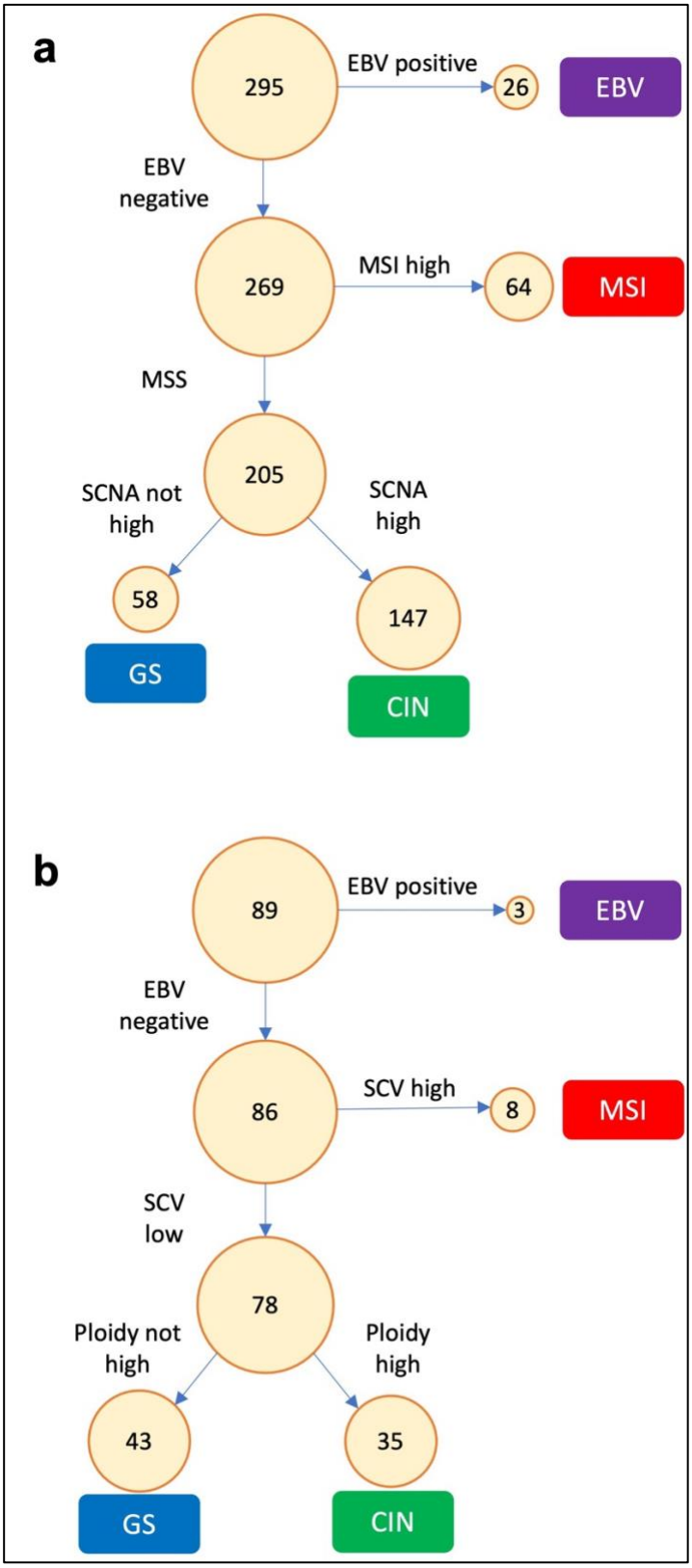


Figure 2. Classification of TCGA molecular subtypes (a) method used to classify samples within TCGA cohort, as previously described [15]**, (b) method used to categorist samples from the 100,000 Genomes Project.**

CIN, chromosomal instability; EBV, Epstein-Barr virus; GS genomically stable; MSI; MSS, microsatellite stable; microsatellite instability; SCNA, somatic copy-number alterations; SCV, somatic coding variants

2.3.5 Metagenomic profiling

Microbiome data were generated by Dr Henry Wood from sequencing data by removing the human reads from tumour sequencing data, and realigning remaining reads to microbial genome databases. This was performed using the GATK PathSeq algorithm, aligned against the default PathSeq microbial databases [168]. The PathSeq 'score' output was used for microbial sequencing reads, except for the decontamination steps where unambiguous reads were used. Within both databases, this resulted in both species and genera tables, where each row represented a clinical sample and each column represented a unique taxon. Values within the body of the table represented the microbial cells per human cell for the 100,000 Genomes Project cohort and number of sequencing reads of each taxon in the TCGA cohort per clinical sample. Within the 100,000 Genomes Project cohort, microbial cells per human cell was calculated using the following formula:

$$microbes\ per\ human\ cell = \frac{(microbial\ reads \div microbial\ genome\ size)}{(human\ reads \div human\ genome\ size)}$$

after adjustment of the human genome size for DNA ploidy and tumour cell content, as found in LabKey tables. No adjustment for human genome size was made using the TCGA data, since data were derived from exome reads and therefore human reads were not representative of the whole genome due to overrepresentation of exons (relative to intergenic regions).

2.3.6 Decontamination

A modified version of the methodology described by Dohlman *et al*. [158] was used to perform *in silico* decontamination. Prevalence was defined as at least two unambiguous reads per taxa per sample. For each species, blood prevalence was compared to tissue prevalence in both the 100,000 Genomes Project and TCGA cohorts. One sided Fisher's exact test was performed for each species-specific comparison, using a multiple comparison-corrected significance threshold of $q < 0.05$. An include-list was created from species more prevalent in tissue than in blood ($q < 0.05$) in the 100,000 Genomes Project or TCGA, where blood prevalence was <20% of samples in both cohorts. For all species with q values ≥0.05 and <0.4 from the TCGA cohort, the literature was reviewed and species identified as inhabitants of the respiratory tract or digestive system were manually added to the include-list. In addition, EBV was manually added to the include-list. A flow diagram (Figure 6) of the generation of the include-list may be found in the results section 2.4.3. Decontaminated databases for both cohorts for downstream analysis were created by filtering

the genus and species reads to include only species present on the include-list. All downstream analysis used only the decontaminated databases.

2.3.7 Statistical analyses

Analyses were conducted in R [169] within RStudio [170], using stringr [171], dplyr [172], qvalue [173], and vegan [174] packages; pheatmap [175], Microsoft Excel, Microsoft PowerPoint, and ggplot2 [176] were used to create heatmaps, MaAsLin2 plots, flow diagrams, and other figures, respectively. For both cohorts, the 50 most abundant species and 30 most abundant genera were displayed in heatmaps to balance clarity of visualisation with representation of the largest proportion of signal.

Due to differences in sequencing methodologies and availability of metadata between the two cohorts, analyses were performed in either one or both cohorts depending on data availability. Clinicopathological variables used in analyses included: age, sex, tumour location, geographical origin, pT, pN, histological phenotype, MSI status, and TCGA molecular subtype.

To evaluate microbial abundance, total microbial count was calculated for each sample. Within the 100,000 Genomes Project cohort, microbial abundance was represented by microbes per human cell. Within the TCGA cohort, microbial abundance was represented by the sum of microbial sequencing reads. Shannon index [177] was calculated for each sample as a measure of alpha diversity (i.e. intra-sample variability). Wilcoxon and Kruskal-Wallis tests were applied for comparisons of categorical variables with two, or more than two groups, respectively. Spearman’s rank correlation coefficient was calculated for correlation analyses.

Permutational multivariate analysis of variance (PERMANOVA, Adonis), using Bray-Curtis dissimilarity index [178], was used to analyse beta diversity (i.e. inter-sample variability) in species between subgroups. The PERMANOVA analysis considered clinicopathological variables with over 85% completeness. To avoid overlapping variables, TCGA molecular subtype was not included since it is a composite variable which includes MSI status, which was already included in PERMANOVA analysis. Samples with no taxa were removed prior to PERMANOVA analysis. Variables significant at PERMANOVA analysis were

included in analysis of differential abundance according to clinicopathological variables. Multivariable Association Discovery in Population-scale Meta-omics Studies (MaAsLin2) [179] was used to perform differential abundance analyses.

2.4 Results

2.4.1 Samples available for analysis

A total of 89 tumour samples from patients with gastric or gastro-oesophageal junction adenocarcinoma were identified in the 100,000 Genomes Project cohort. All 89 GC samples had WGS data for tumour and blood which were used to generate microbial sequencing data. TCGA microbial sequencing data was generated from whole exome sequencing data from 441 tumour samples and 396 matched blood samples from the TCGA GC cohort. Histological review of available slides from the TCGA GC cohort identified one sample as squamous cell carcinoma, therefore this case was removed from subsequent analysis. The distribution of clinicopathological features within both the 100,000 Genomes Project and TCGA cohorts can be found in Table 3. Only data relating to tumour samples were used for both cohorts as 1) previous studies have investigated differences between normal and adjacent mucosa and 2) the data regarding where the adjacent mucosa samples were taken from was insufficient.

Table 3. Clinicopathological characteristics of samples used within this study, from 100,000 Genomes Project (gastric and gastro-oesophageal junction adenocarcinoma) and TCGA (gastric adenocarcinoma) cohorts. CIN, chromosomal instability; EBV, Epstein-Barr virus-positive; GS, genomically stable; MSI, microsatellite instability; MSS, microsatellite stabile; n, number; *t* = t test statistic; χ^2, chi squared statistic; * statistically significant (p <0.05); †Inferred as described in methods. ††Immune subtype according to Thorsson *et al*. [162] available for TCGA cohort only.

Characteristic	**100,000 Genomes Project** total n=89 n (%)	**TCGA** total n=440 n (%)	**Test statistic**	**p**
Age, (interquartile range)	69 (62-77)	67 (58-73)	*t* = 1.48	0.139
Unknown	0 (0)	5 (1)		
Sex			χ^2=5.80	0.016*
Male	69 (78)	283 (64)		
Female	20 (23)	157 (36)		
Tumour location			χ^2=5.74	0.017*
Cardia	27 (30)	90 (20)		
Non-cardia	44 (49)	280 (64)		
Unknown	18 (20)	70 (16)		
Geographic origin			χ^2=18.78	<0.0001*
Asia	0 (0)	69 (15)		
Not Asia	89 (100)	314 (71)		
Unknown	0 (0)	57 (13)		
Pathological depth of invasion (pT)			χ^2=6.94	0.074
1	8 (9)	23 (5)		
2	11 (12)	93 (21)		
3	32 (36)	198 (45)		
4	29 (33)	117 (27)		
Unknown	9 (10)	9 (2)		
Pathological lymph node status (pN)			χ^2=3.41	0.33
0	20 (23)	131 (30)		
1	25 (28)	117 (27)		
2	22 (25)	86 (20)		
3	13 (15)	88 (20)		
Unknown	9 (10)	18 (4)		

Characteristic	**100,000 Genomes Project** total n=89 n (%)	**TCGA** total n=440 n (%)	**Test statistic**	**p**
Histological phenotype			χ^2=3.19	0.364
Diffuse	10 (11)	76 (17)		
Intestinal	23 (26)	278 (63)		
Mixed	2 (2)	21 (4)		
Mucinous	0 (0)	19 (4)		
Unknown	54 (61)	46 (10)		
MSI status			χ^2=5.59	0.018*
MSS	81 (91)†	308 (70)		
MSI	8 (9)†	75 (17)		
Unknown	0 (0)	57 (13)		
TCGA subtype			χ^2=31.20	<0.0001*
EBV	3 (3)*	26 (6)		
MSI	8 (9)*	64 (15)		
CIN	35 (39)*	145 (33)		
GS	43 (48)*	58 (13)		
Unknown	0 (0)*	147 (33)		
Immune subtype††			-	-
C1	0 (0)	129 (31)		
C2	0 (0)	209 (51)		
C3	0 (0)	35 (9)		
C4	0 (0)	9 (2)		
C5	0 (0)	0 (0)		
C6	0 (0)	7 (2)		
Unknown	89 (100)	51 (12)		

2.4.2 MSI status and TCGA molecular subtypes

MSI status was available for 383 (87%) TCGA GC from previously published data [180]. MSI data were not directly available for the 100,000 Genomes Project data and an inferred MSI status was used for analysis. Samples where SCV ≥20 mut/Mb were inferred as MSI; samples where SCV <20 mut/Mb were inferred as MSS, demonstrated in Figure 3.

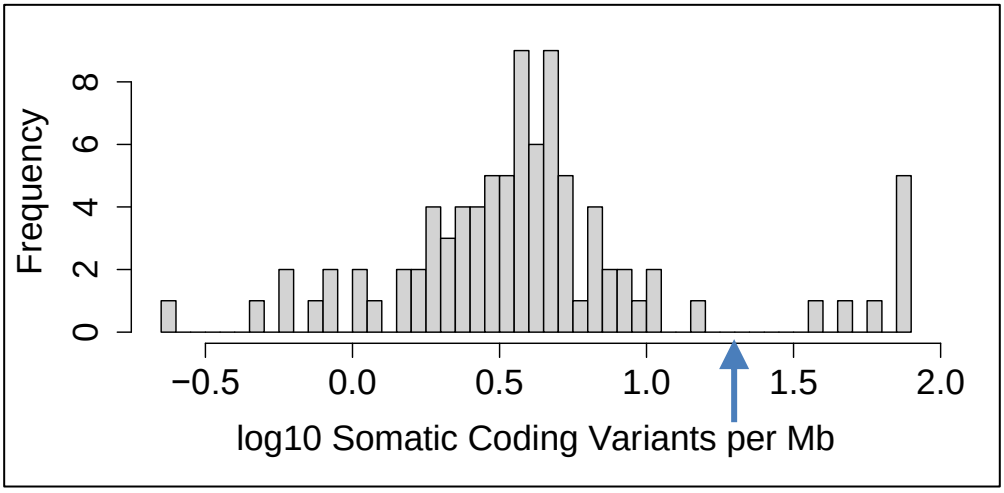


Figure 3. Histogram of somatic coding variants (log10) in 89 gastric and junctional adenocarcinomas within the 100,000 Genomes Project. Blue arrow represents SCV threshold (log10(20) ≈ 1.3) used to define MSI and TCGA subtypes as described in main text, section 2.3.4.

TCGA molecular subtype status was available for 293/440 (67%) GC from the TCGA cohort. Within the 100,000 Genomes Project, thresholds used to determine inferred TCGA molecular subtype were: EBV count $1x10^{-3}$ per human cell (Figure 4); SCV 20mut/Mb (Figure 3); and total DNA ploidy 2.5 (Figure 5). The distribution of samples according to inferred MSI and TCGA molecular subtype is shown in Table 3.

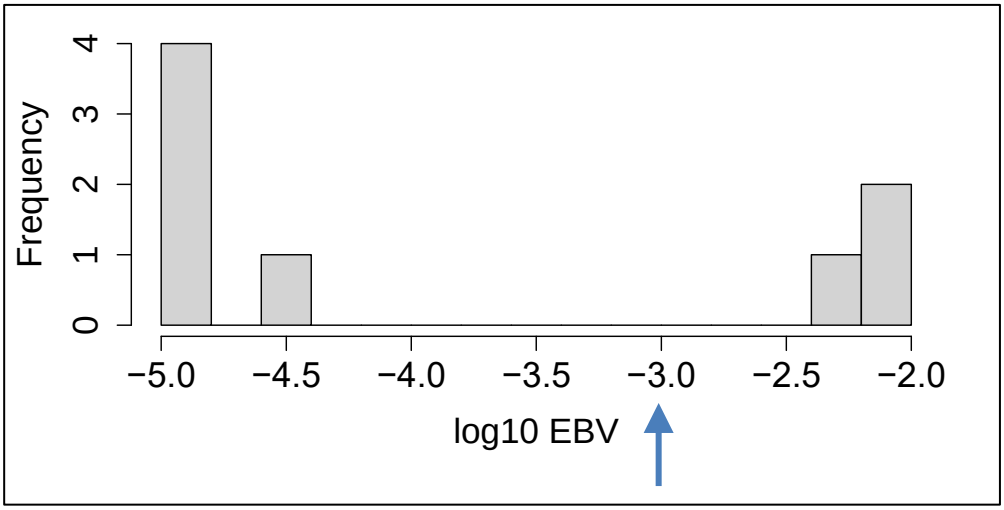


Figure 4. Histogram of non-zero EBV counts (log10) in 89 gastric and junctional adenocarcinomas within the 100,000 Genomes Project. Blue arrow represents EBV threshold (EBV count $1x10^{-3}$ per human cell) used to define TCGA subtypes as described in main text, section 2.3.4.

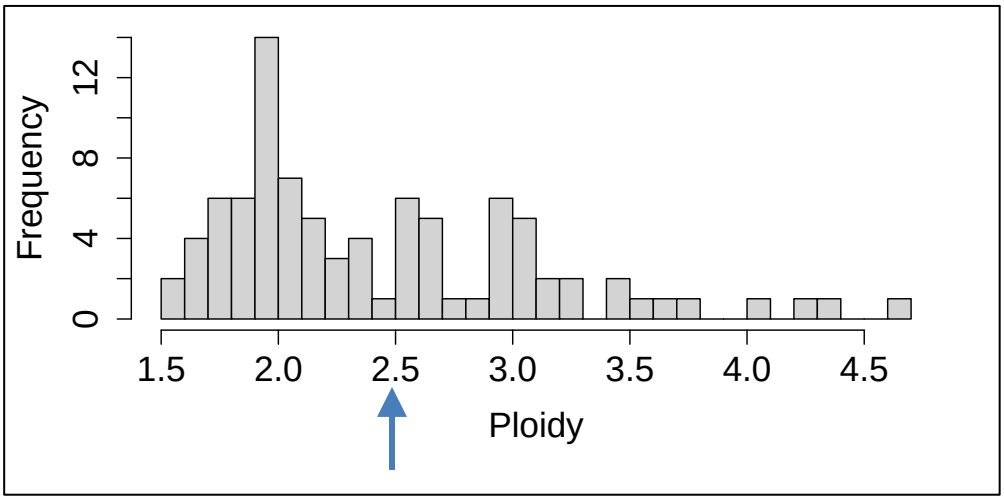


Figure 5. Histogram of total DNA ploidy in 89 gastric and junctional adenocarcinomas within the 100,000 Genomes Project. Blue arrow represents total ploidy threshold (2.5) used to define TCGA subtypes as described in main text, section 2.3.4.

2.4.3 Decontamination

Figure 6 illustrates the decontamination process performed to generate the species include-list. Prior to decontamination, 5261 species from 946 genera were present across at least one sample from samples of the 100,000 Genomes Project cohort; 1491 species from 374 genera were present across at least one sample from the TCGA cohort. The greater number of species identified from the 100,000 Genomes Project cohort is likely a result of the low-coverage WGS metagenomic approach resulting in improved microbial detection, relative the WES metagenomic approach used in TCGA. After the comparing the prevalence of taxa in blood and tissue samples, 480 species from the 100,000 Genomes Project cohort and 56 species from the TCGA cohort were identified as tissue-resident as they were more prevalent in tumour than blood samples ($q < 0.05$) and prevalent in <20% of blood samples. These species were therefore added to the include-list. A further 50 species from a total of 95 species from the TCGA cohort where $q \geq 0.05$ and < 0.4 and prevalent in <20% of blood samples were identified as inhabitants of the respiratory tract or digestive system. These species were therefore added to the include-list. None of the species identified as tissue-resident from only one cohort had a blood prevalence ≥20% in either cohort. EBV was not statistically assigned to either list due to its low prevalence in the 100,000 Genomes Project cohort (4/89 tumour samples; 0/89 blood samples) and absence within the TCGA cohort (0/441 tumour samples; 0/396 blood samples) and was manually added to the include-list. The final include-list consisted of 496 species from 105 genera. These taxa were assumed to be ‘tissue-resident’ taxa, i.e. not external contaminants. The total number of species (496) is less than the sum of the species added from each cohort as there was overlap with some species being identified as tissue-resident from both cohorts. Whilst Figure 6 depicts an include-list and exclude-list, only the include-list was used for downstream decontamination; the exclude-list is shown for clarity. Species not included on the include-list were assumed to be environmental contaminants, e.g. from laboratory reagents.

The full species include-list used to decontaminate both databases may be found in Appendix B, (Supplemental) Table 19.

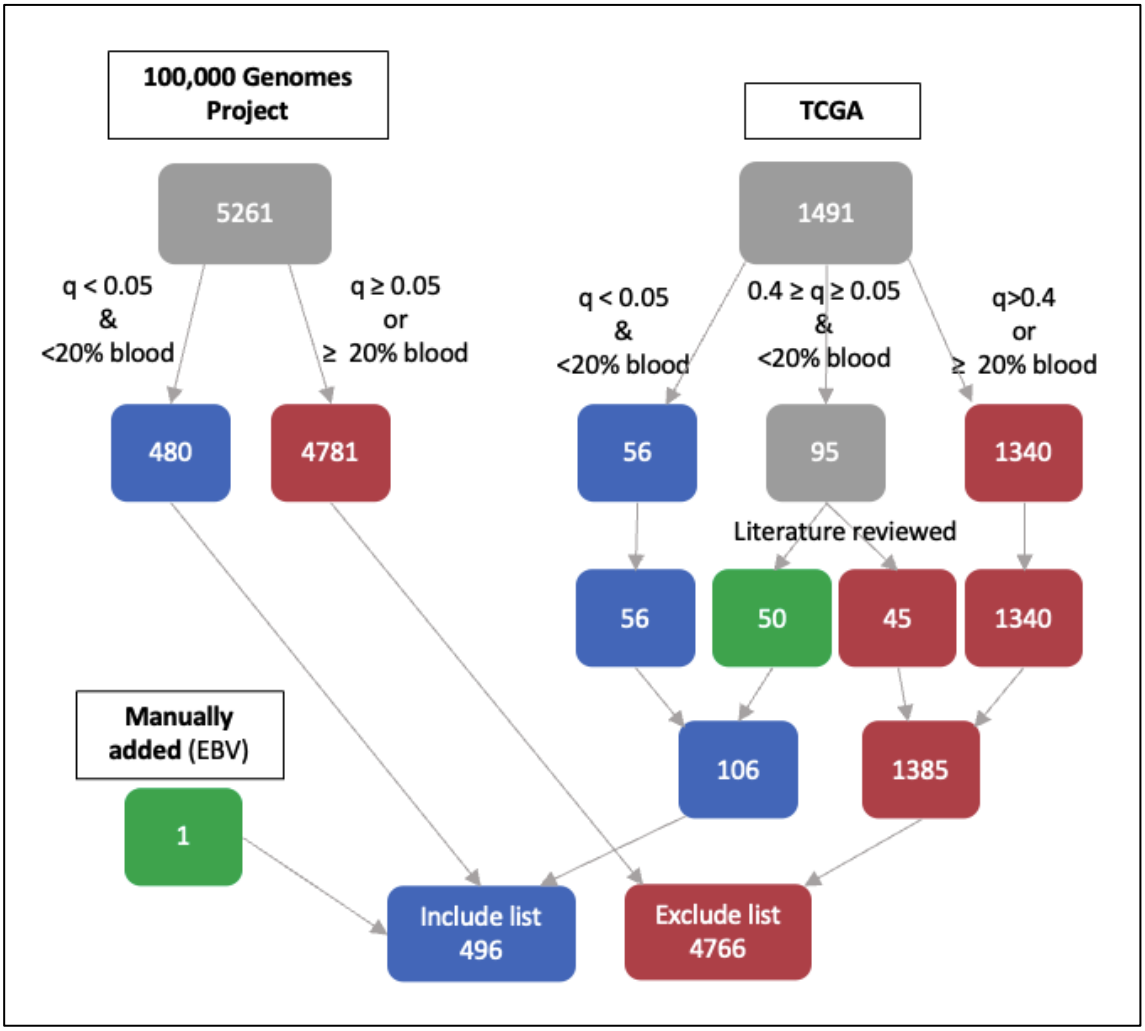


Figure 6. Flow chart demonstrating categorisation of species into the study include-list. Species-level prevalence data from both the 100,000 Genomes Project and TCGA GC cohorts were used to generate the include-list. EBV, Epstein-Barr Virus; TCGA, The Cancer Genome Atlas. Grey boxes represent species prior to sorting into include- or exclude-lists. Blue boxes represent species added to include-list; red boxes represent species added to exclude-list (i.e. not on include-list). Green boxes represent species added to include list following a manual process. Final include- and exclude-list numbers are lower than the sums of the numbers from 100,000 Genomes Project and TCGA cohort due to overlap of species detected in both cohorts.

The decontaminated databases represented 53% and 78% of the original microbial content for the 100,000 Genomes Project and TCGA, respectively. For the TCGA cohort, 251/440 (57%) of GC samples from the TCGA cohort had fewer than five total microbial reads following the decontamination process, demonstrating negligible microbial content (see Figure 8). All 100,000 Genomes Project samples had taxa remaining after decontamination (see Figure 7).

2.4.4 Taxonomic composition

Following decontamination, six genera (*Prevotella*, *Selenomonas*, *Stomatobaculum*, *Streptococcus*, *Lactobacillus*, and *Lachnospiraceae* [family, genus undefined]) were found to be in the 10 most abundant genera in both cohorts. The inclusion of the "genus" *Lachnospiraceae* [family, genus undefined] reflects the fact that in both cohorts, genera belonging to the family *Lachnospiraceae* were present, which were unable to taxonomically further classified when compared to the reference database. The ten most abundant genera for both cohorts are displayed in Table 4.

Table 4. Table showing ten most abundant genera within the 100,000 Genomes Project (n=89) and TCGA (n=440) cohorts following decontamination, ordered by decreasing abundance. The six genera found in the most abundant ten bacterial genera across both cohorts are shown in blue. TCGA, The Cancer Genome Atlas.

100,000 Genomes Project	**TCGA**
Lactobacillus	*Prevotella*
Stomabaculum	*Selenomonas*
Leptotrichia	*Capnocytophaga*
Prevotella	*Veillonella*
Lachnospiraceae [family, genus undefined]	*Stomatobaculum*
Roseolovirus	*Streptococcus*
Streptococcus	*Lactobacillus*
Helicobacter	*Lachnospiraceae [family, genus undefined]*
Clostridium	*Porphyromonas*
Selenomonas	*Campylobacter*

Two species (*Stomatobaculum longum* and *Lachnospiraceae bacterium oral taxon 082*) were found to be in the 10 most abundant species within both the 100,000 Genomes Project and TCGA cohorts. The ten most abundant species for both cohorts are displayed in Table 5. Table showing ten most abundant species within the 100,000 Genomes Project (n=89) and TCGA (n=440) cohorts following decontamination, ordered by decreasing abundance.

Table 5. Table showing ten most abundant species within the 100,000 Genomes Project (n=89) and TCGA (n=440) cohorts following decontamination, ordered by decreasing abundance. The two species found in the most abundant ten bacterial species across both cohorts are shown in blue. TCGA, The Cancer Genome Atlas.

100,000 Genomes Project	**TCGA**
Stomabaculum longum	*Capnocytophaga sp. oral taxon 329*
Leptotrichia wadei	*Prevotella multiformis*
Lactobacillus johnsonii	*Selenomonas_sputigena*
Lachnospiraceae bacterium oral taxon 082	*Selenomonas sp. CM52*
Lactobacillus gasseri	*Prevotella denticola*
Human betaherpes virus 7	*Stomatobaculum longum*
Helicobacter pylori	*Prevotella veroralis*
Clostridium perfringens	*Lachnospiraceae bacterium oral taxon 082*
Lactobacillus vaginalis	*Prevotella nigrescens*
Leptotrichia shaii	*Veillonella sp. 6_1_27*

Heatmaps for the 30 most abundant genera and 50 most abundant species from the 100,000 Genomes Project, representing 99% and 92% of the total microbial content, respectively, are shown in Figure 7. As demonstrated in Figure 7, certain genera appear to be universally or almost universally present across all samples, e.g. *Streptococcus* and *Prevotella*. Some taxa appear to be present at higher abundances when present, or completely absent, e.g. *H. pylori*.

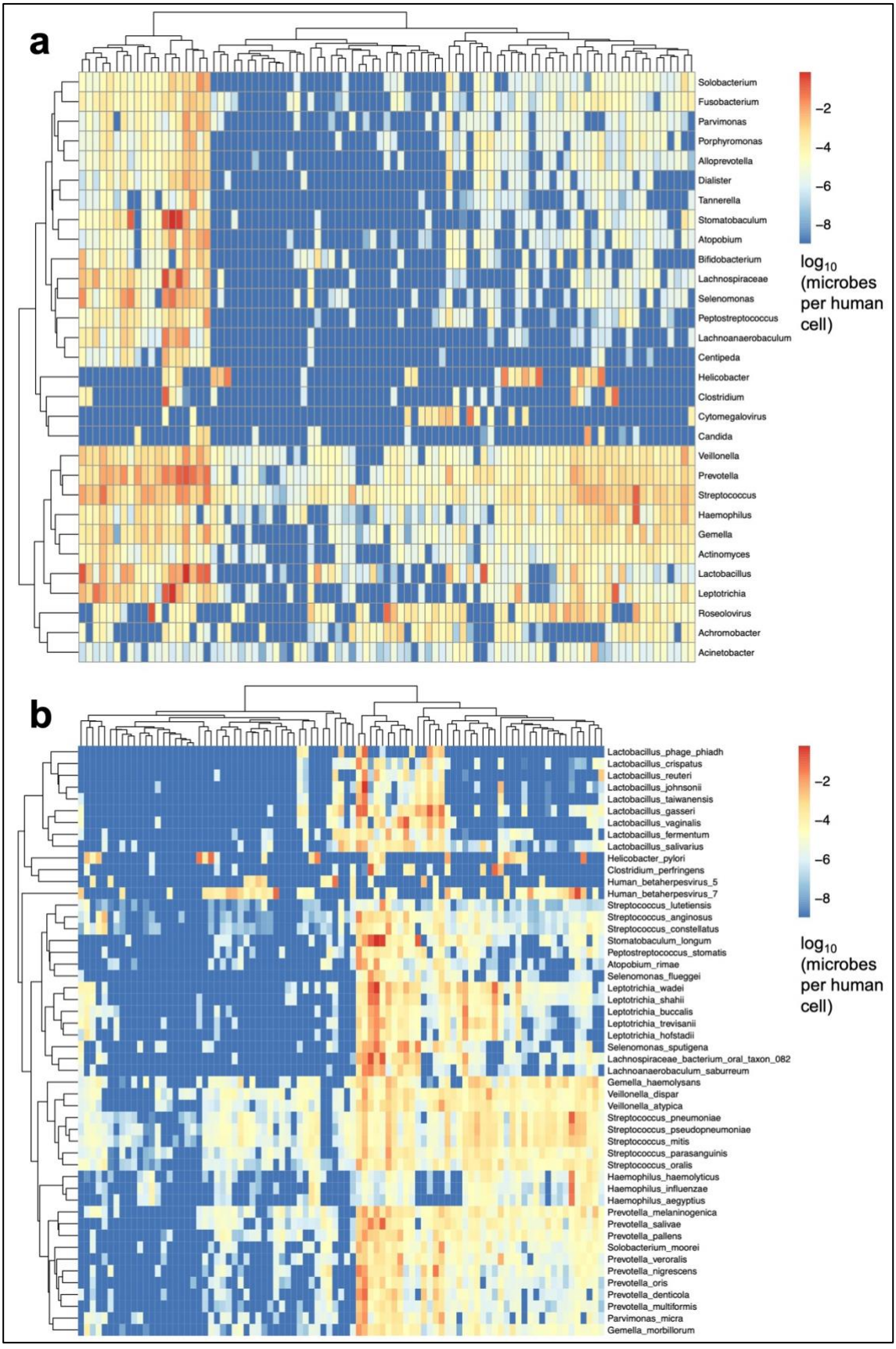

a
Solobacterium
Fusobacterium
Parvimonas
Porphyromonas
Alloprevotella
Dialister
Tannerella
Stomatobaculum
Atopobium
Bifidobacterium
Lachnospiraceae
Selenomonas
Peptostreptococcus
Lachnoanaerobaculum
Centipeda
Helicobacter
Clostridium
Cytomegalovirus
Candida
Veillonella
Prevotella
Streptococcus
Haemophilus
Gemella
Actinomyces
Lactobacillus
Leptotrichia
Roseolovirus
Achromobacter
Acinetobacter
−2
−4
−6
−8
log10 (microbes per human cell)
b
Lactobacillus_phage_phiadh
Lactobacillus_crispatus
Lactobacillus_reuteri
Lactobacillus_johnsonii
Lactobacillus_taiwanensis
Lactobacillus_gasseri
Lactobacillus_vaginalis
Lactobacillus_fermentum
Lactobacillus_salivarius
Helicobacter_pylori
Clostridium_perfringens
Human_betaherpesvirus_5
Human_betaherpesvirus_7
Streptococcus_lutetiensis
Streptococcus_anginosus
Streptococcus_constellatus
Stomatobaculum_longum
Peptostreptococcus_stomatis
Atopobium_rimae
Selenomonas_flueggei
Leptotrichia_wadei
Leptotrichia_shahii
Leptotrichia_buccalis
Leptotrichia_trevisanii
Leptotrichia_hofstadii
Selenomonas_sputigena
Lachnospiraceae_bacterium_oral_taxon_082
Lachnoanaerobaculum_saburreum
Gemella_haemolysans
Veillonella_dispar
Veillonella_atypica
Streptococcus_pneumoniae
Streptococcus_pseudopneumoniae
Streptococcus_mitis
Streptococcus_parasanguinis
Streptococcus_oralis
Haemophilus_haemolyticus
Haemophilus_influenzae
Haemophilus_aegyptius
Prevotella_melaninogenica
Prevotella_salivae
Prevotella_pallens
Solobacterium_moorei
Prevotella_veroralis
Prevotella_nigrescens
Prevotella_oris
Prevotella_denticola
Prevotella_multiformis
Parvimonas_micra
Gemella_morbillorum
−2
−4
−6
−8
log10 (microbes per human cell)

Figure 7. Heatmaps of microbes per human cell in 89 gastric cancer samples from the 100,000 Genomes Project cohort, showing (a) 30 most abundant genera and (b) 50 most abundant species. Cells are coloured according to log10(microbes per human cell) of taxa from red (highest) to blue (lowest). Samples are represented in columns. The dendrogram branching on the left illustrates clustering taxa; branching at the top represents samples with similar compositions of taxa.

Heatmaps for the 30 most abundant genera and 50 most abundant species from TCGA, representing 97% and 83% of the total microbial content, respectively, are shown in Figure 8. As described in section 2.4.3, a considerable number (251/440; 57%) of samples had no taxa following the decontamination process. This is demonstrated in Figure 8 by the large number of heatmap columns consisting of only blue cells.

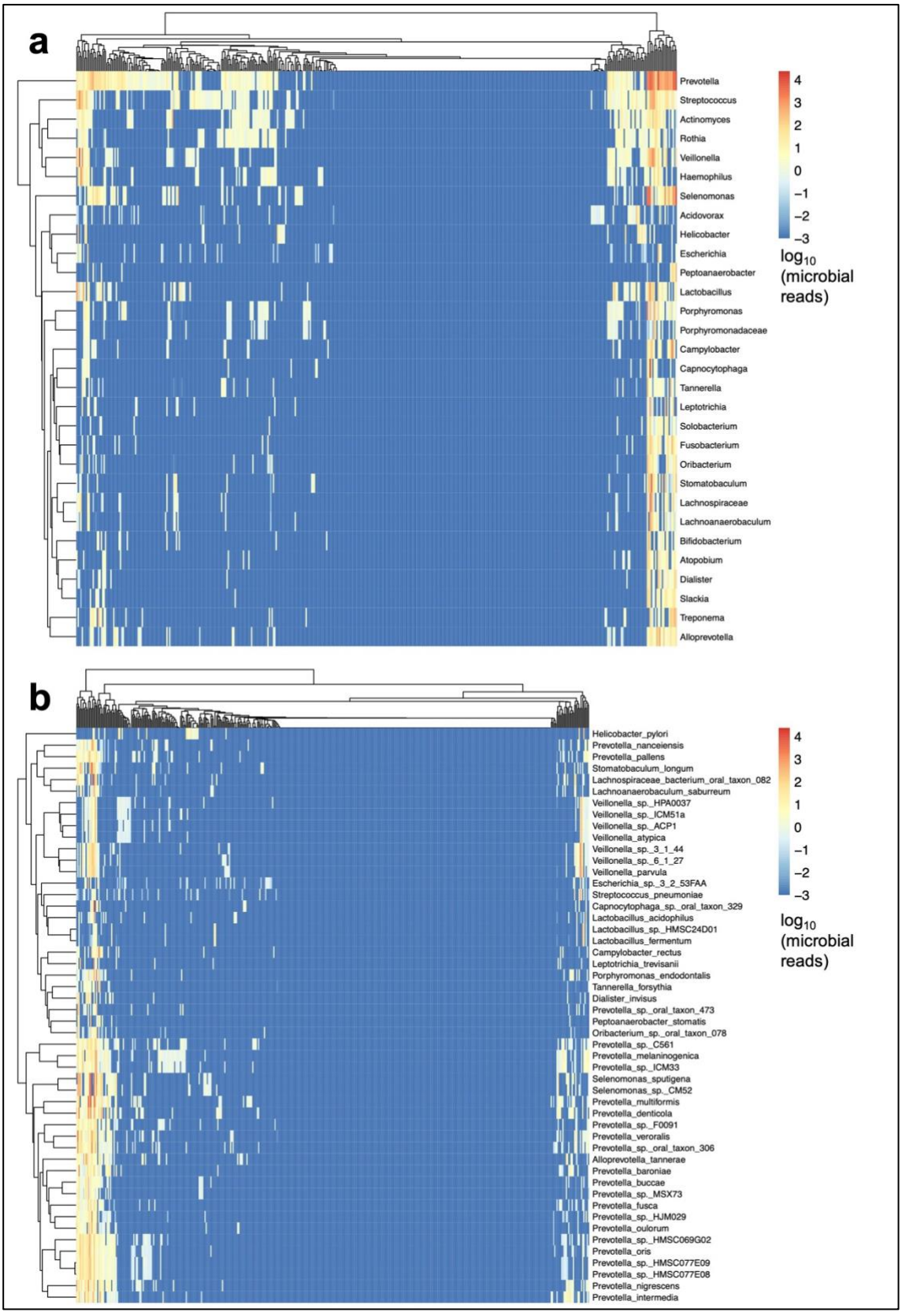

a
Prevotella
Streptococcus
Actinomyces
Rothia
Veillonella
Haemophilus
Selenomonas
Acidovorax
Helicobacter
Escherichia
Peptoanaerobacter
Lactobacillus
Porphyromonas
Porphyromonadaceae
Campylobacter
Capnocytophaga
Tannerella
Leptotrichia
Solobacterium
Fusobacterium
Oribacterium
Stomatobaculum
Lachnospiraceae
Lachnoanaerobaculum
Bifidobacterium
Atopobium
Dialister
Slackia
Treponema
Alloprevotella
4
3
2
1
0
−1
−2
−3
log10 (microbial reads)
b
Helicobacter_pylori
Prevotella_nanceiensis
Prevotella_pallens
Stomatobaculum_longum
Lachnospiraceae_bacterium_oral_taxon_082
Lachnoanaerobaculum_saburreum
Veillonella_sp._HPA0037
Veillonella_sp._ICM51a
Veillonella_sp._ACP1
Veillonella_atypica
Veillonella_sp._3_1_44
Veillonella_sp._6_1_27
Veillonella_parvula
Escherichia_sp._3_2_53FAA
Streptococcus_pneumoniae
Capnocytophaga_sp._oral_taxon_329
Lactobacillus_acidophilus
Lactobacillus_sp._HMSC24D01
Lactobacillus_fermentum
Campylobacter_rectus
Leptotrichia_trevisanii
Porphyromonas_endodontalis
Tannerella_forsythia
Dialister_invisus
Prevotella_sp._oral_taxon_473
Peptoanaerobacter_stomatis
Oribacterium_sp._oral_taxon_078
Prevotella_sp._C561
Prevotella_melaninogenica
Prevotella_sp._ICM33
Selenomonas_sputigena
Selenomonas_sp._CM52
Prevotella_multiformis
Prevotella_denticola
Prevotella_sp._F0091
Prevotella_veroralis
Prevotella_sp._oral_taxon_306
Alloprevotella_tannerae
Prevotella_baroniae
Prevotella_buccae
Prevotella_sp._MSX73
Prevotella_fusca
Prevotella_sp._HJM029
Prevotella_oulorum
Prevotella_sp._HMSC069G02
Prevotella_oris
Prevotella_sp._HMSC077E09
Prevotella_sp._HMSC077E08
Prevotella_nigrescens
Prevotella_intermedia
4
3
2
1
0
−1
−2
−3
log10 (microbial reads)

Figure 8. Heatmaps of microbial abundance in 440 gastric cancer samples from the TCGA cohort, showing (a) 30 most abundant genera and (b) 50 most abundant species. Cells are coloured according to log10(microbial reads) of taxa from red (highest) to blue (lowest). Samples are represented in columns. The dendrogram branching on the left illustrates clustering taxa; branching at the top represents samples with similar compositions of taxa. The large area of blue demonstrates the large proportion (57%) of samples with a negligible microbial read content following the decontamination process.

2.4.5 Microbial abundance and alpha diversity according to clinicopathological factors

Within the 100,000 Genomes Project cohort, no statistically significant relationships were identified between the analysed clinicopathological variables (age, sex, tumour location, pT, pN, histological phenotype, MSI status, and TCGA molecular subtype) and either microbial abundance or Shannon index (see (Supplemental) Table 20). However, when the inferred MSI status of the 100,000 Genomes Project GC cohort was examined in relation to microbial abundance, MSI samples tended to have greater microbial abundance than MSS samples, although this did not reach statistical significance (p=0.061). No relationship was identified between MSI status and Shannon index.

In the TCGA cohort, presence of MSI was associated with greater microbial abundance (p=0.0012) and Shannon index (p=0.0011) than MSS (Figure 9).

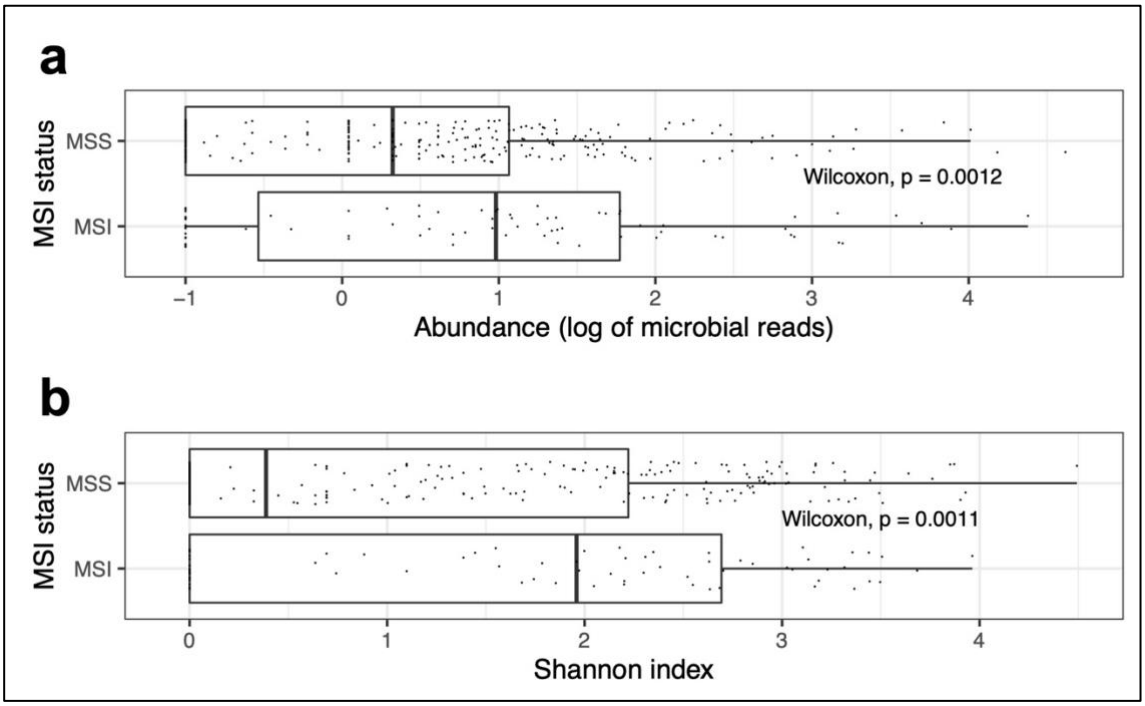


Figure 9. Boxplots of (a) microbial abundance and (b) Shannon index of 383 TCGA samples, according to MSI status. MSI GC samples were associated with greater microbial abundance (a) and Shannon index (b) than MSS GC samples. Boxes represent lower (left border) and upper (right border) interquartile ranges; vertical line within box represents median value; whiskers represent minimum and maximum values, excluding outliers (defined as 1.5 times the respective upper/lower interquartile range). MSI, microsatellite instability (n=75); MSS, microsatellite stable (n=308).

In the TCGA cohort, TCGA molecular subtype was associated with microbial abundance (p <0.001) and Shannon index (p <0.001). The boxplots (Figure 10) appear to demonstrate highest microbial abundance and Shannon index in the MSI TCGA subtype, and lowest abundance and Shannon index in the GS subtype, although pair-wise comparisons were not performed due to low statistical power and risk of multiple hypothesis testing.

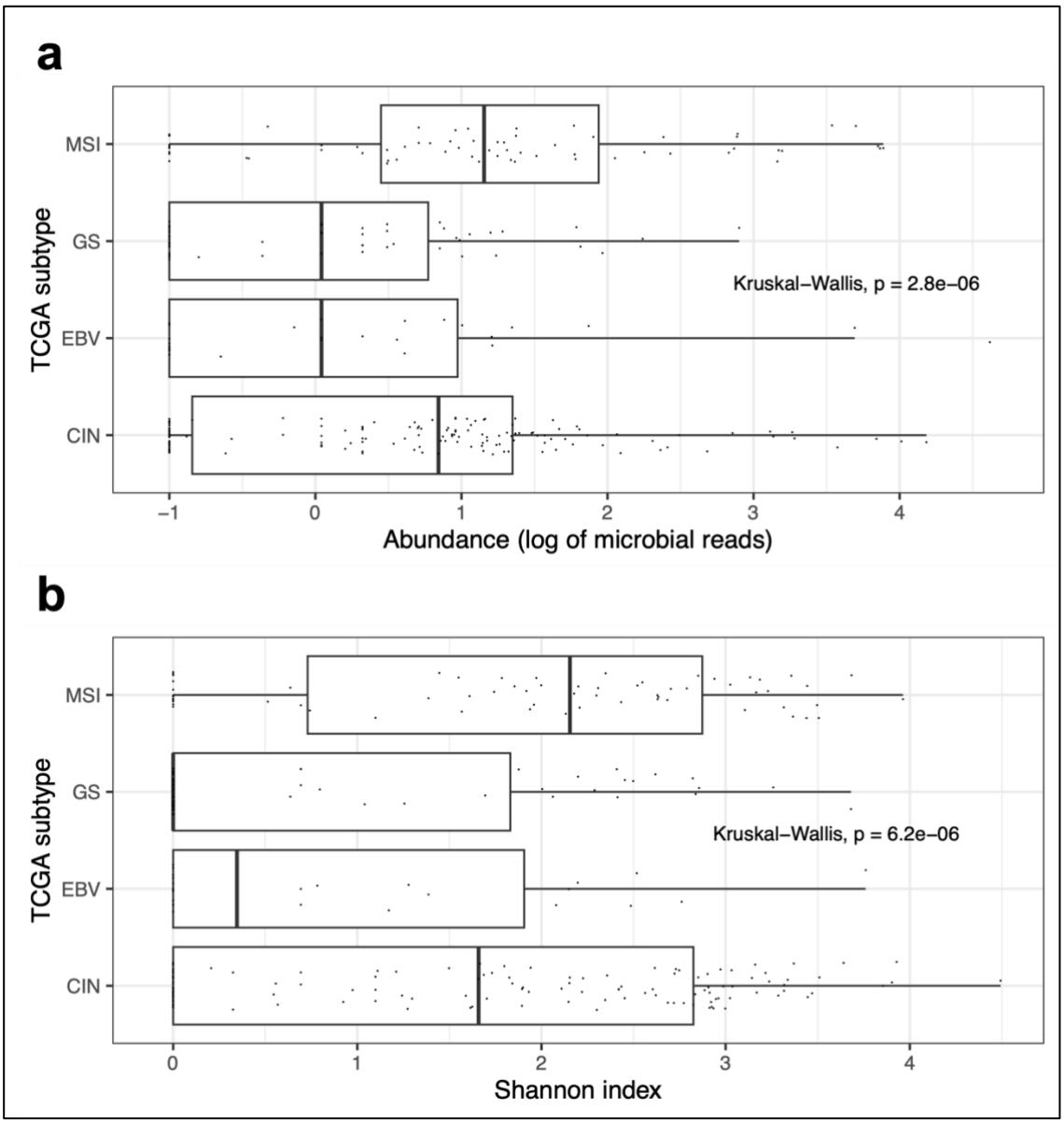


Figure 10. Boxplots of (a) microbial abundance and (b) Shannon index of 293 TCGA samples, according to TCGA subtype. TCGA subtype was associated with abundance (a) and Shannon index (b); the MSI subtype had the greatest numerical microbial abundance and greatest Shannon index. Boxes represent lower (left border) and upper (right border) interquartile ranges; vertical line within box represents median value; whiskers represent minimum and maximum values, excluding outliers (defined as 1.5 times the respective upper/lower interquartile range). CIN, chromosomal instability (n=145); EBV, Epstein-Barr virus-positive (26); GS, genomically stable (58); MSI, microsatellite instability (64).

In the TCGA cohort, lower pT category (pT1 and pT2), compared to higher pT category (pT3 and pT4) was associated with both greater microbial abundance (p=0.0040), and greater Shannon index (p=0.00017), demonstrated in Figure 11.

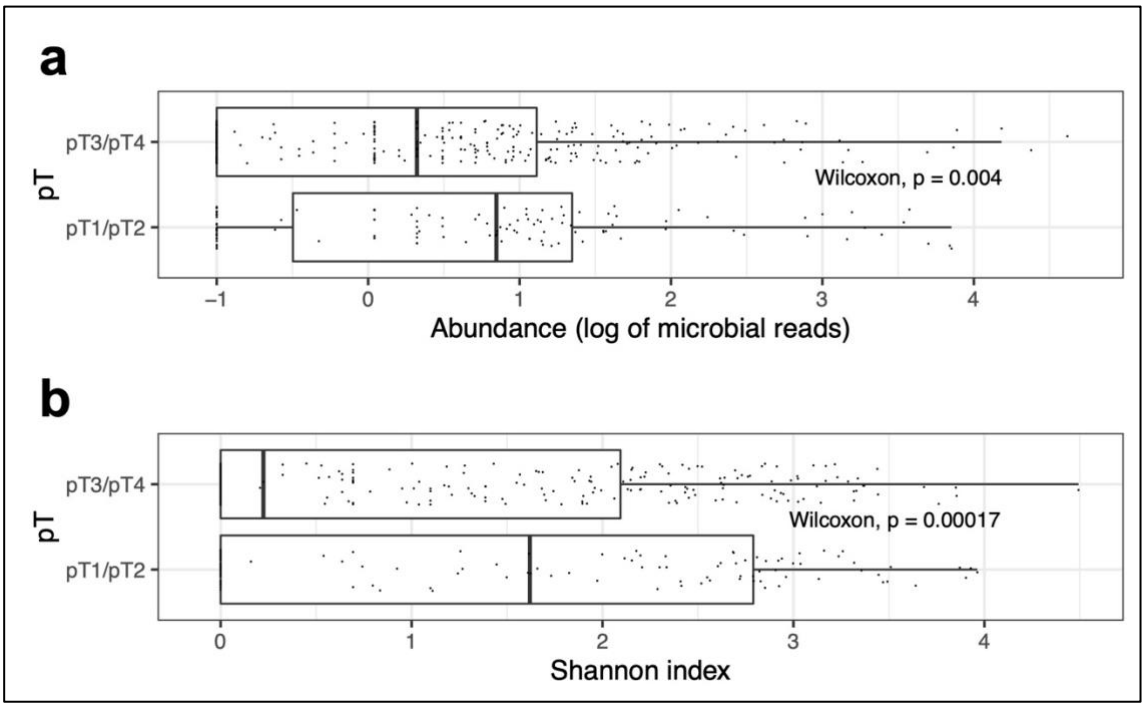


Figure 11 . Boxplots of (a) microbial abundance and (b) Shannon index of 431 TCGA samples, according to pathological depth of invasion (pT). GC samples with a lower pT category (pT1/2, n=116) had greater microbial abundance (a) and Shannon index (b) than GC samples with a higher pT category (pT3/4, n=315). Boxes represent lower (left border) and upper (right border) interquartile ranges; vertical line within box represents median value; whiskers represent minimum and maximum values, excluding outliers (defined as 1.5 times the respective upper/lower interquartile range).

In the TCGA cohort, GC from Asia had lower microbial abundance (p=0.026) and lower Shannon index (p=0.041) than samples not from Asia (Figure 12).

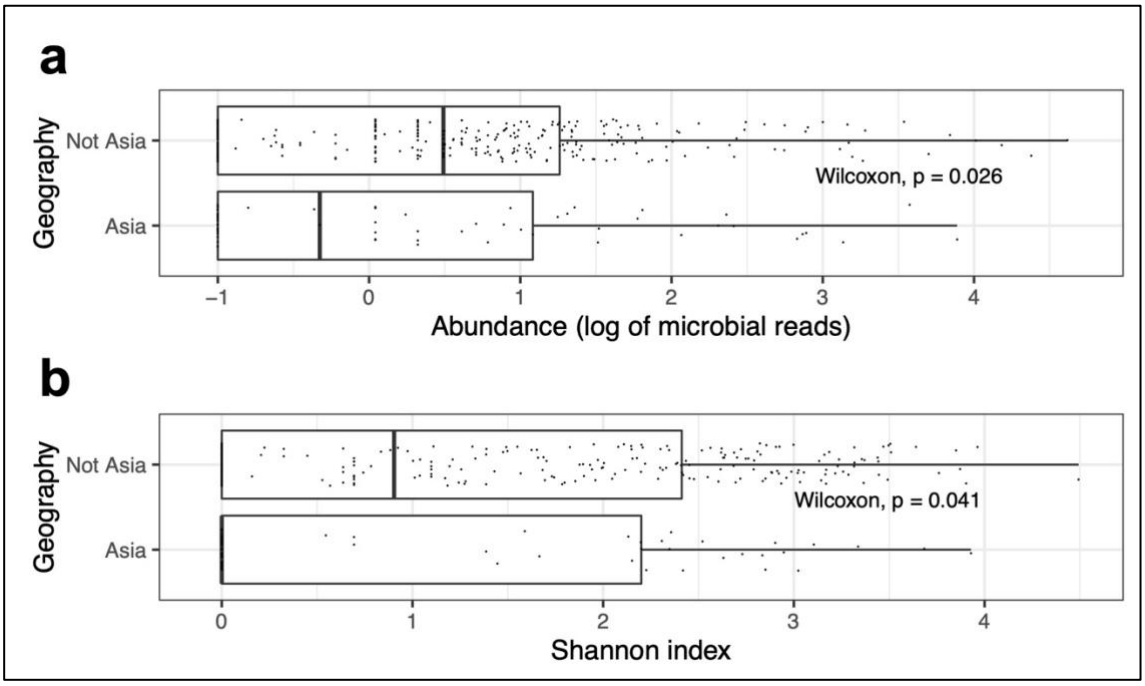


Figure 12. Boxplots of (a) microbial abundance and (b) Shannon index of 383 TCGA samples, according to geography. GC samples from Asia (n=69) had lower microbial abundance (a) and Shannon index (b) than GC samples not Asia (n=314). Boxes represent lower (left border) and upper (right border) interquartile ranges; vertical line within box represents median value; whiskers represent minimum and maximum values, excluding outliers (defined as 1.5 times the respective upper/lower interquartile range).

In the TCGA cohort, histological phenotype was associated with microbial abundance (p=0.018) and Shannon index (p=0.036). The boxplots (Figure 13) appear to demonstrate higher microbial abundance and Shannon diversity in mucinous GC and intestinal-type GC than diffuse-type GC, although pair-wise comparisons were not performed due to low statistical power and risk of multiple hypothesis testing.

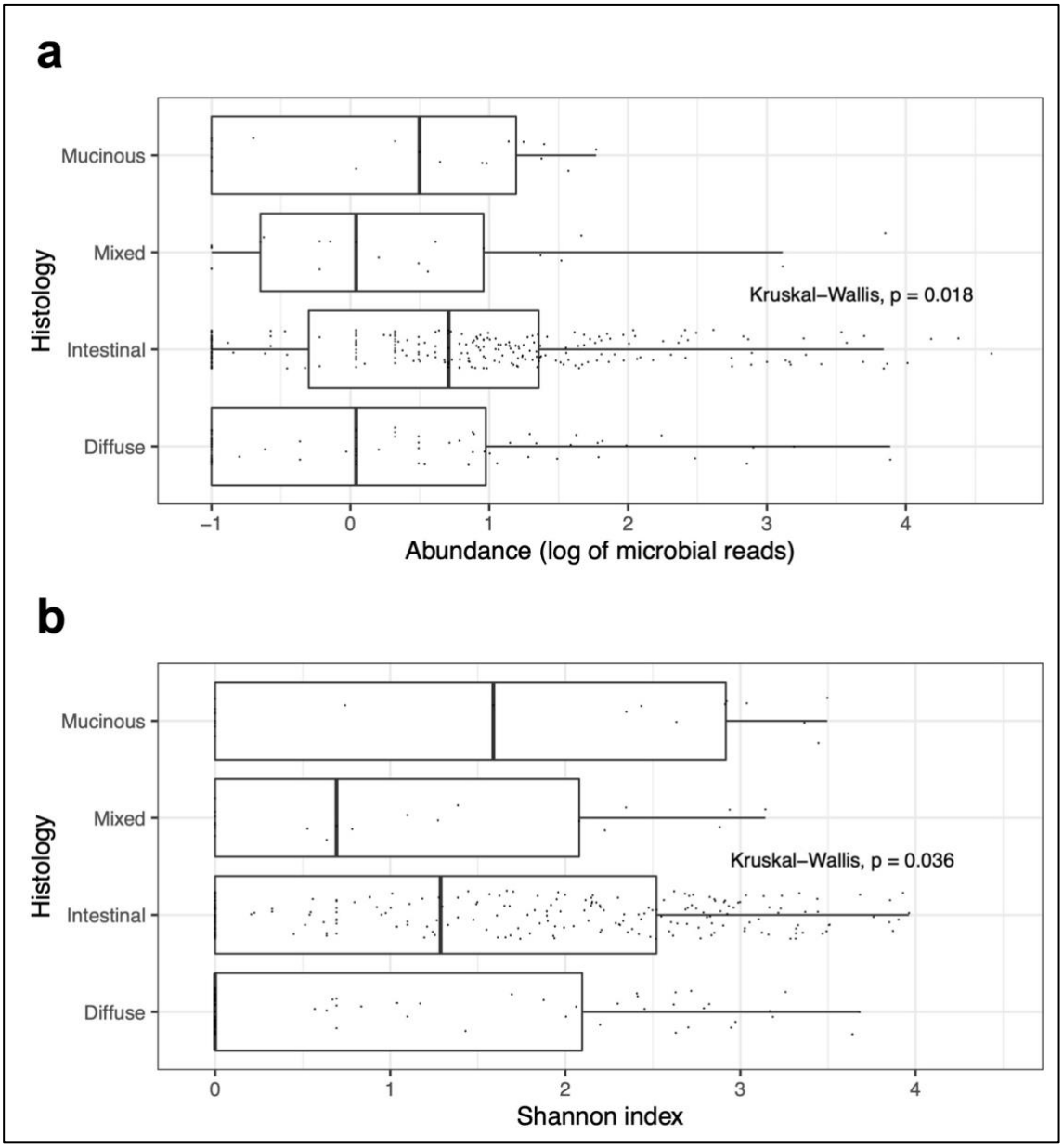


Figure 13. Boxplots of (a) microbial abundance and (b) Shannon index of 394 TCGA samples, according to histological subtype. Histological subtype was associated with abundance (a) and Shannon index (b). Intestinal-type (n=278) and mucinous (n=19) GC had greater numerical microbial abundance and Shannon diversity than mixed (n=21) and diffuse (n=76) GC. Boxes represent lower (left border) and upper (right border) interquartile ranges; vertical line within box represents median value; whiskers represent minimum and maximum values, excluding outliers (defined as 1.5 times the respective upper/lower interquartile range).

Patient sex, patient age, tumour location (cardia versus non-cardia), and pN category were not related to microbial abundance or Shannon index in either the 100,000 Genomes Project nor the TCGA cohort.

2.4.6 Beta diversity

To evaluate beta (between sample) diversity differences according to clinicopathological variables, PERMANOVA analysis was performed. The PERMANOVA analyses considered variables with over 85% completeness, in a pre-defined order of: geographical origin (TCGA only), age, sex, histological phenotype (TCGA only), MSI status, pT category, and pN category. Removal of samples with missing variables (both cohorts) and no microbial reads (TCGA only), resulted in 80 samples from the 100,000 Genomes Project and 240 samples from TCGA, available for PERMANOVA analysis.

The PERMANOVA analysis demonstrated an association between MSI status and microbial composition within both the 100,000 Genomes Project cohort (1.7% of variance; p=0.036) and the TCGA cohort (0.74% of variance; p=0.046). In addition, within the TCGA cohort, sex (0.085% of variance; p=0.024) and pT category (0.078% variance; p=0.036) were associated with differences in microbial composition (Table 6). These PERMANOVA findings demonstrate the numerical magnitude and the statistical significance of the association between these clinicopathological variables and the GC microbiome composition but cannot indicate what such differences are.

Table 6. Permutational multivariate analysis of variance (PERMANOVA) for intratumoural species within 100,000 Genomes Project and TCGA. formulations. R2, the proportion of the variance in the microbiome associated with each variable; MSI microsatellite instability-high versus non high; pT, pathological depth of invasion (pT1/pT2 versus pT3/pT4); pN, pathological lymph node status (pN0 versus >pN0). * statistically significant (p <0.05); †Inferred MSI status for 100,000 Genomes Project cohort.

	100,000 Genomes Project (n=80)		**TCGA (n=240)**	
	R2	P value	R2	P value
Geography	-	-	0.00592	0.1246
Sex	0.01415	0.2184	0.00845	0.0237*
Age	0.01383	0.2634	0.00716	0.0559
MSI†	0.01727	0.0360*	0.00740	0.0455*
Histology	-	-	0.01448	0.2042
pT	0.01391	0.386	0.00776	0.0363*
pN	0.01306	0.308	0.00337	0.6058
Residuals	0.92777		0.94546	-

In order to explore how the relative abundances of individual microbial taxa were contributing to the observed compositional differences highlighted by the PERMANOVA analysis, MaAsLin2 analysis was performed, evaluating only the clinicopathological variables identified as significant by PERMANOVA.

In the 100,000 Genomes Project cohort, MaAsLin2 analysis detected 12 species and six genera with statistically significant differential abundances between inferred MSI and MSS inferred subtypes (Figure 14 and Figure 15). All taxa identified as being differentially abundant according to inferred MSI status were taxa which were more prevalent in the MSI GC samples, relative to MSS GC samples. *Stomatobaculum longum* and *Stomatobaculum* were the species and genus, respectively, which were increased by the largest magnitude in the MSI GC, relative to MSS GC.

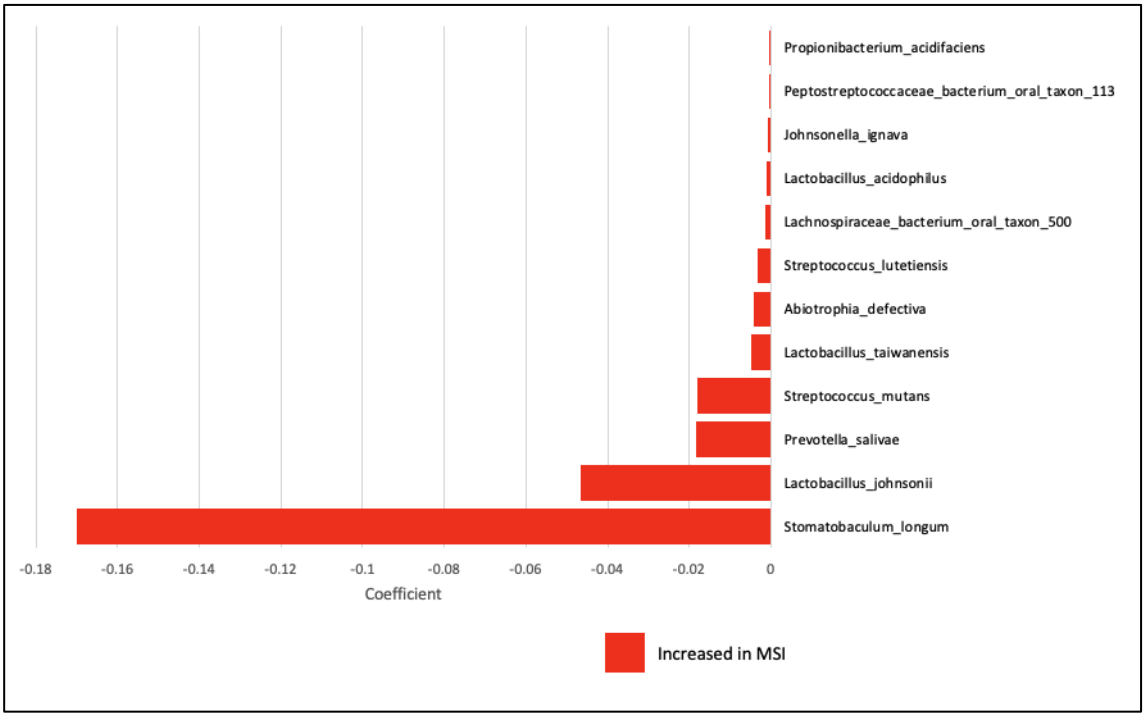


Figure 14. Association of specific species within the 100,000 Genomes Project samples, according to inferred MSI status by MaAsLin2 (n = 89). The 12 species depicted are the 12 species in which a statistically significant difference in abundance between MSI GC and MSS GC was found. The x-axis represents the MaAsLin2 coefficient, a measure of effect size. The negative coefficient of all species within the plot indicates that all species in which there was a statistically significant difference in abundance between MSI and MSS GC were found in greater abundance in the MSI group. MSI, microsatellite instability; MSS, microsatellite stable.

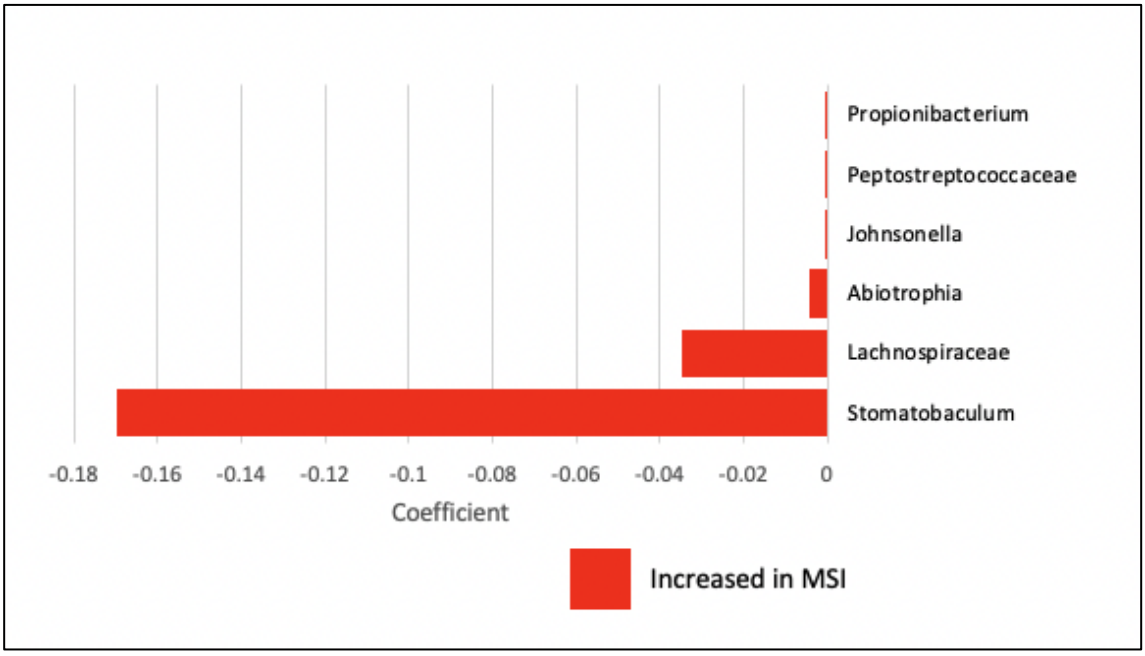
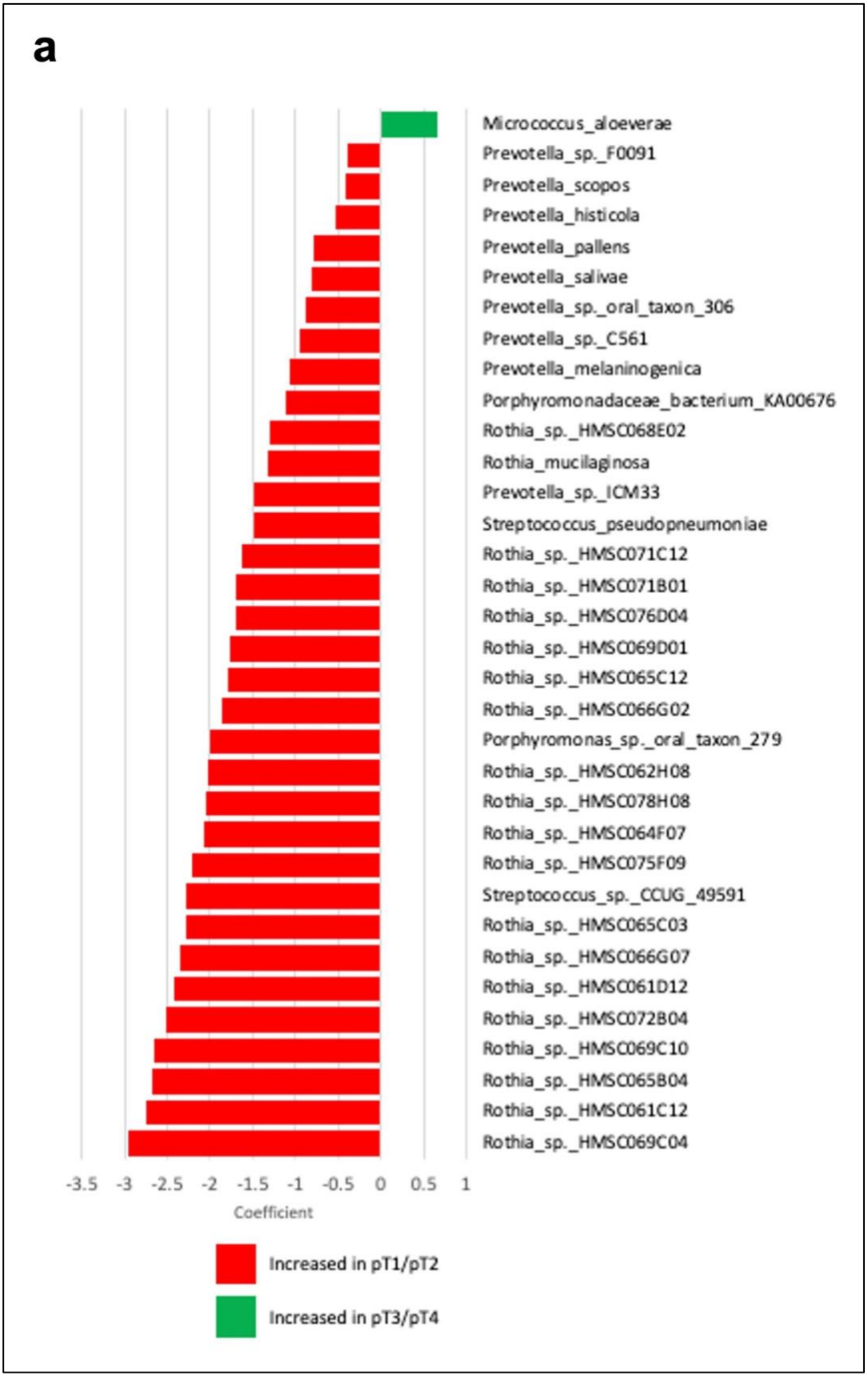


Figure 15. Association of specific genera within the 100,000 Genomes Project samples, according to inferred MSI status by MaAsLin2 (n = 89). The six genera depicted are the six genera in which a statistically significant difference in abundance between MSI GC and MSS GC was found. The x-axis represents the MaAsLin2 coefficient, a measure of effect size. The negative coefficient of all species within the plot indicates that all species in which there was a statistically significant difference in abundance between MSI and MSS GC were found in greater abundance in the MSI group. MSI, microsatellite instability; MSS, microsatellite stable.

In the TCGA cohort, MaAsLin2 analysis identified 45 species and 12 genera differentially abundant across sex, MSI status and pT category (Figure 16 and Figure 17). All differentially abundant taxa according to pT category were found more commonly in pT1/pT2 GC compared to pT3/pT4 GC, except for the *Micrococcus* genus and *Micrococcus aloeverae* species, found in greater abundance in pT3/pT4 GC. All differentially abundant taxa according to MSI were found more commonly in MSI GC compared to MSS GC, except for the *Neisseria* genus, which was found in greater abundance in MSS GC. All differentially abundant taxa according to sex were found more commonly in males than in females.

Of note, none of the species or genera identified as differentially abundant according to MSI status within the 100,000 Genomes Project were differentially abundant according to MSI status on multivariate analysis in the TCGA cohort.


a
Micrococcus_aloeverae
Prevotella_sp._F0091
Prevotella_scopos
Prevotella_histicola
Prevotella_pallens
Prevotella_salivae
Prevotella_sp._oral_taxon_306
Prevotella_sp._C561
Prevotella_melaninogenica
Porphyromonadaceae_bacterium_KA00676
Rothia_sp._HMSC068E02
Rothia_mucilaginosa
Prevotella_sp._ICM33
Streptococcus_pseudopneumoniae
Rothia_sp._HMSC071C12
Rothia_sp._HMSC071B01
Rothia_sp._HMSC076D04
Rothia_sp._HMSC069D01
Rothia_sp._HMSC065C12
Rothia_sp._HMSC066G02
Porphyromonas_sp._oral_taxon_279
Rothia_sp._HMSC062H08
Rothia_sp._HMSC078H08
Rothia_sp._HMSC064F07
Rothia_sp._HMSC075F09
Streptococcus_sp._CCUG_49591
Rothia_sp._HMSC065C03
Rothia_sp._HMSC066G07
Rothia_sp._HMSC061D12
Rothia_sp._HMSC072B04
Rothia_sp._HMSC069C10
Rothia_sp._HMSC065B04
Rothia_sp._HMSC061C12
Rothia_sp._HMSC069C04
-3.5 -3 -2.5 -2 -1.5 -1 -0.5 0 0.5 1
Coefficient
Increased in pT1/pT2
Increased in pT3/pT4

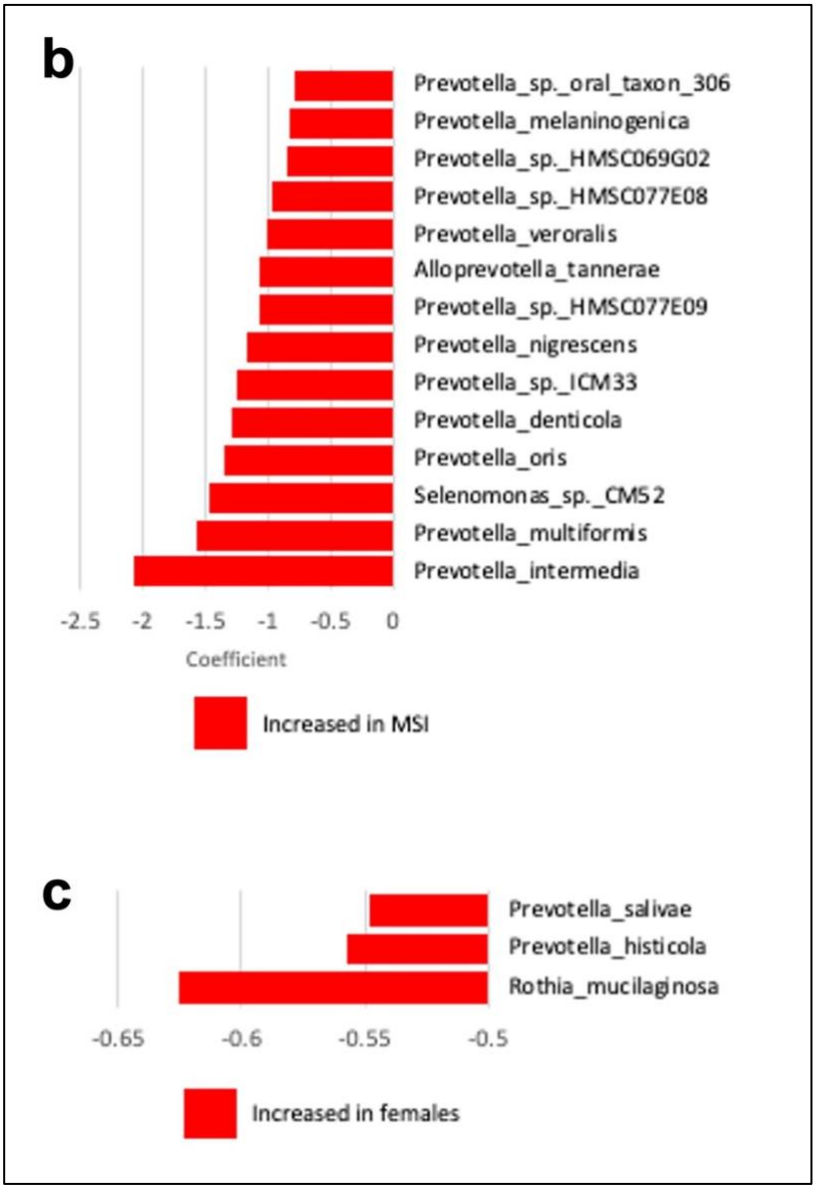
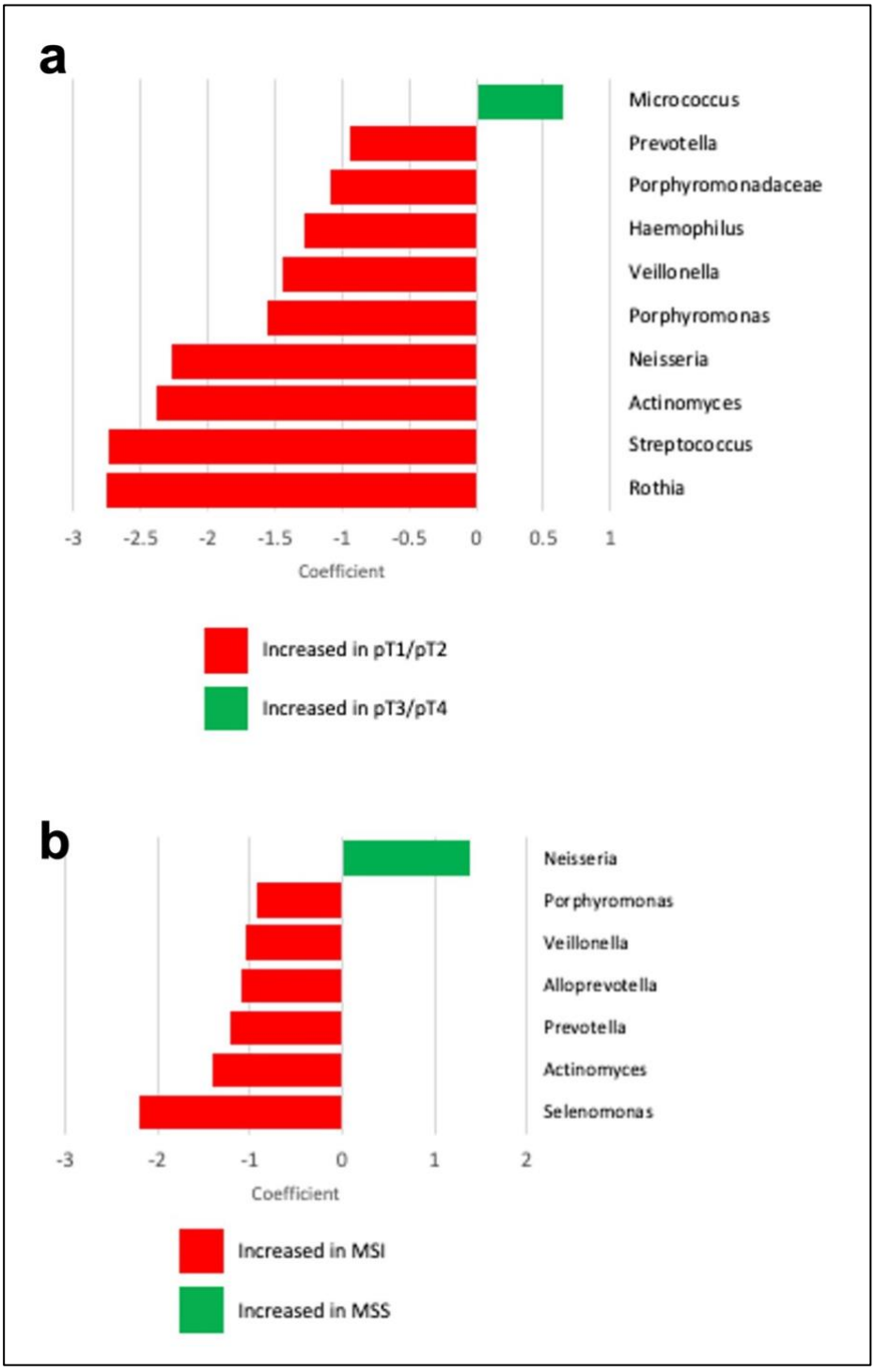


Figure 16. Association of specific species within TCGA samples according to pT category (a), MSI status (b), and sex (c) by MaAsLin2 (n=375). The species depicted are the 45 species in which a statistically significant difference in species abundance between the respective variables was found. Red bars and a negative coefficient indicate species enriched in pT1/pT2 versus pT3/pT4 (a), MSI versus MSS (b) and in females versus males (c); green bars and a positive coefficient indicate species enriched in pT3/pT4 versus pT1/pT2 (a). MSI, microsatellite instability; MSS, microsatellite stable; pT, pathological depth of invasion.


a
Micrococcus
Prevotella
Porphyromonadaceae
Haemophilus
Veillonella
Porphyromonas
Neisseria
Actinomyces
Streptococcus
Rothia
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
Coefficient
Increased in pT1/pT2
Increased in pT3/pT4
b
Neisseria
Porphyromonas
Veillonella
Alloprevotella
Prevotella
Actinomyces
Selenomonas
-3
-2
-1
0
1
2
Coefficient
Increased in MSI
Increased in MSS

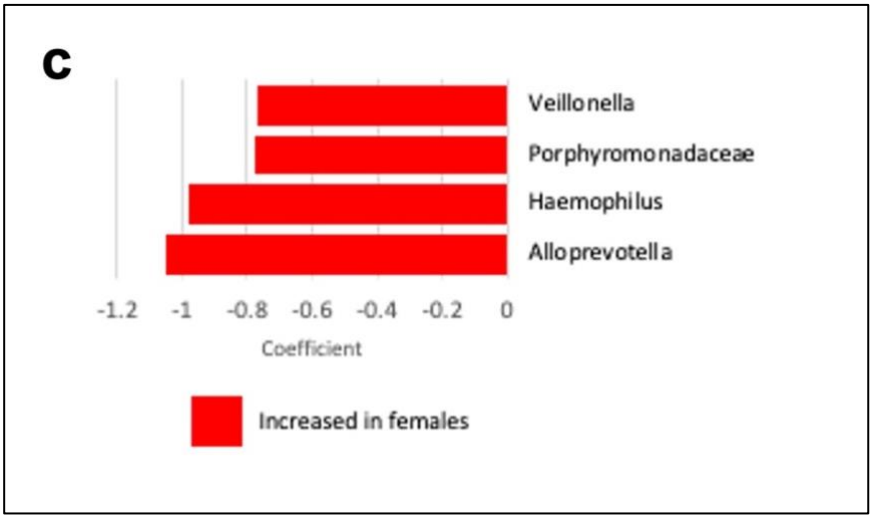


Figure 17. Association of specific genera within TCGA samples according to pT category (a), MSI status (b), and sex (c) by MaAsLin2 (n=375). The genera depicted are the 12 genera in which a statistically significant difference in genus abundance between the respective variables was found. Red bars and a negative coefficient indicate species enriched in pT1/pT2 versus pT3/pT4 (a), MSI versus MSS (b) and in females versus males (c); green bars and a positive coefficient indicate species enriched in pT3/pT4 versus pT1/pT2 (a) and MSS versus MSI (b). MSI, microsatellite instability; MSS, microsatellite stable; pT, pathological depth of invasion.

2.4.7 Immune cell estimates

Immune cellular fraction estimates and immune subtypes data generated from Thorsson *et al*. [162] were available for 389/440 (88%) TCGA GC (Table 3).

Estimates of cellular fractions of six immune cell types (lymphocytes, neutrophils, macrophages, dendritic cells, eosinophils, and mast cells) were plotted against both total microbial count and Shannon index. No notable relationships were identified between any of the six immune cell types and either abundance or Shannon index (data not shown). Immune subtype (C1 - C6), according to Thorsson *et al*), [162] was associated with a statistically significant difference in both microbial abundance ($p=0.0070$) and Shannon index ($p < 0.001$) (Figure 18). The boxplots appear to demonstrate lowest microbial abundance and Shannon index in the C3 immune subtype and highest Shannon index in the C6 subtype, although pair-wise comparisons were not performed due to low statistical power and risk of multiple hypothesis testing.

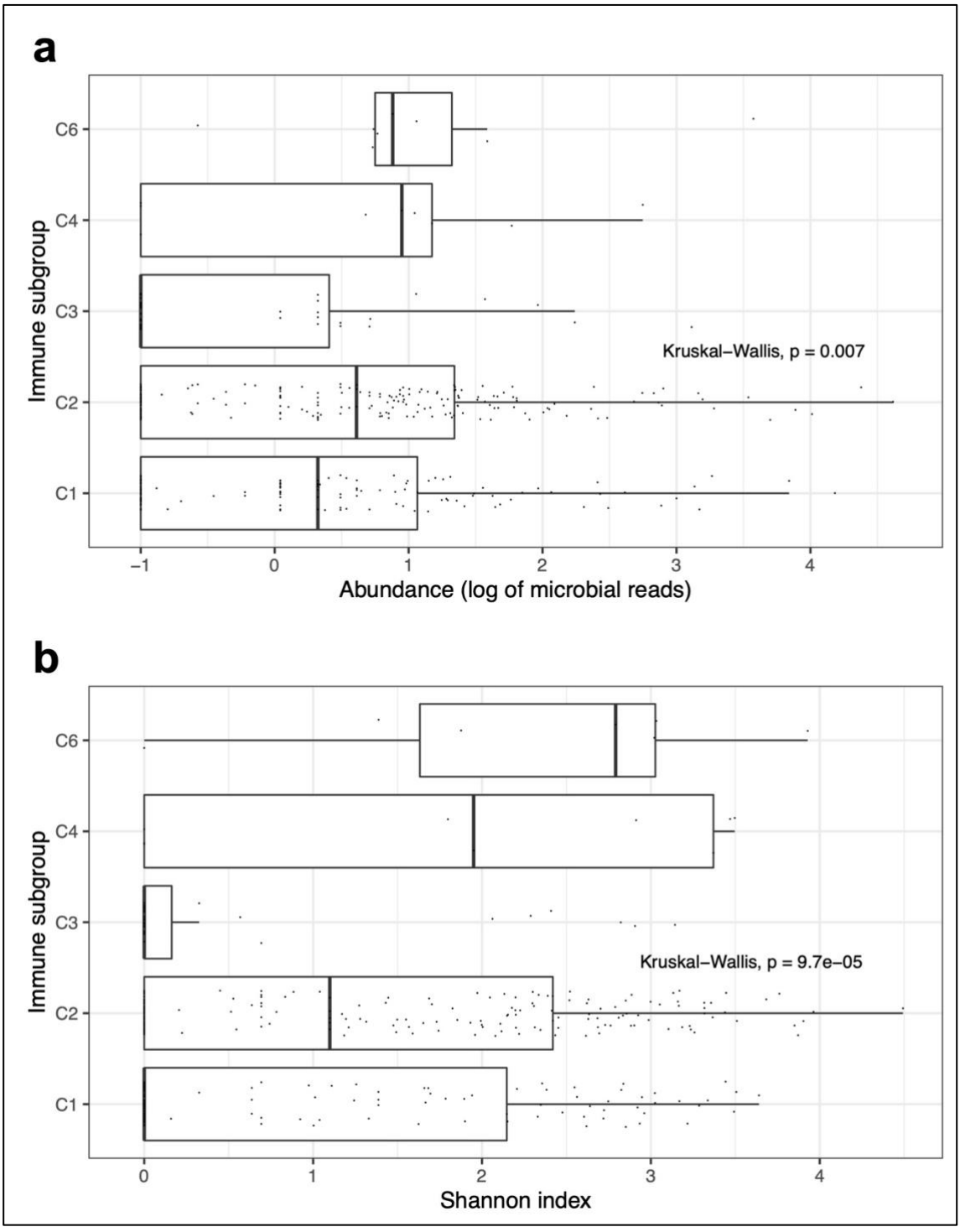


Figure 18. Boxplots of (a) microbial abundance and (b) Shannon index of 389 TCGA samples, according to immune subtype. Immune subtype was associated with abundance (a) and Shannon index (b). The C3 immune subtype had the lowest numerical microbial abundance and lowest Shannon index.

Boxes represent lower (left border) and upper (right border) interquartile ranges; vertical line within box represents median value; whiskers represent minimum and maximum values, excluding outliers (defined as 1.5 times the respective upper/lower interquartile range).

2.5 Discussion

The aims of the work described in this chapter were to characterise the GC microbiome and to explore the relationship between patient- and tumour-specific factors and the GC microbiome. In this chapter, the results from an exploratory study of the intratumoural GC microbiome in a total of 529 GC patients are presented, with analysis of WGS data from the 100,000 Genomes Project and WES data from TCGA. This is the largest study to date to use WGS and WES data to characterise the clinicopathological features associated with the GC microbiome. Through this analysis, clinicopathological features have been identified which associated with microbial abundance, alpha diversity, and beta diversity.

2.5.1 Decontamination is important in gastric cancer microbiome studies

Work presented in this chapter further developed the decontamination process as initially described by Dohlman *et al*. [158], incorporating two separate cohorts to maximise the number of genuine tumour taxa and minimise contamination. The similarities in microbial composition observed between the 100,000 Genomes Project and TCGA cohorts following decontamination suggest that GC does have a consistently detectable microbiome. However, the large proportion of the TCGA cohort which had a negligible microbiome left following the decontamination process suggests that using exome data may be near the limit of detection thresholds for GC. Decontamination is rarely performed in GC studies; however, the proportion of the total signal and number of taxa removed in both cohorts emphasises its importance. As such, studies involving analysis of un-decontaminated GC microbiome data, especially those based on exome data could largely be analysing noise resulting from contamination or sequencing errors. Given the large proportion (251/440) of TCGA cases which had no taxa remaining following the decontamination step, one may consider whether this is reflective of the normal microbiome being very limited. However, since this pattern was not observed in the 100,000 Genomes Project cohort, it is likely that this observation is a result of the different sequencing methodologies, and that the exome data is not sufficient to determine the microbiome in lower microbial biomass cases.

The decision to add in EBV, as well as the species identified as inhabitants of the respiratory or digestive tract (with q values ≥0.05 and <0.4 from the TCGA

cohort) adds a subjective element to the include-list. This additional step was included due to concern of true tissue-resident species not being included in the mathematically generated include-list due to insufficient power. Therefore, this method was selected as a balance of attempting to capture as many recognised tissue-resident species, without manually reviewing the literature of all 1491 TCGA species.

Limitations of using an include-list such as that described in this chapter are that the statistical power of detecting species as tissue-resident is related to sample-size and the reliance upon on blood samples as a non-tissue-specific control. Therefore, it would not be feasible to develop an include-list as described in this chapter, in small studies or studies where blood samples are not available. As an alternative to using this method to produce study-specific include-lists, the include-list created through this work could be used to decontaminate data from other cohorts. In addition, the include-list described in this work could be used as a preliminary include-list and adapted by addition of specific species identified as tissue-resident through comparisons of tissue and blood from smaller datasets. This would allow project-specific species to be identified and thus ‘personalising’ the include-list, whilst reducing the limitations of smaller datasets.

2.5.2 The gastric cancer intratumoural microbiome of the two investigated cohorts reflects that of previous studies

Concordant with previous studies investigating the intratumoural GC microbiome, bacteria from the genera *Prevotella*, *Selenomonas*, *Streptococcus*, *Stomatobaculum* and *Lactobacillus* were commonly found in the GC samples across both cohorts [38,43,130]. Also consistent with previous findings [39,181], the majority of most commonly identified bacteria are recognised as human oral microbes [182–187]. This concordance with previous intratumoural GC microbiome studies supports the value of the decontamination process and the reliability of these two cohorts as snapshots of the intratumoural GC microbiome.

Prevotella are gram-negative anaerobic bacteria and may be considered oral microbes, since they have most commonly been isolated from the oral cavity. However, *Prevotella*, species have been identified from the intestinal and respiratory tracts [188,189]. Species from the *Prevotella* genus have been

associated with oral infections and bacterial vaginosis [190]. Species from the *Lactobacillus* genus are found in the vaginal microbiota and throughout the intestinal tract [191]. These gram-positive bacteria are facultatively anaerobic [191], have the ability to form biofilms [192], and grow optimally at an acidic pH [193]. Infection caused by lactobacilli is rare, and species belonging to this genus are generally thought to be protective against pathogens. Lactobacilli are used in probiotics and starter products in fermented foods [193]. Streptococci are considered as commensal bacteria on mucosal surfaces in the oral cavity, as well as bot the gastrointestinal and upper respiratory tracts [194]. Species belonging to the *Streptococcus* genus are gram-positive and predominantly facultative anaerobes; they are considered lactic acid bacill, predominantly producing lactic acid as a product of carbohydrate fermentation [194]. Streptococci may be alpha- or beta-haemolytic species. *S. pnemoniae* is an alpha-haemolytic species which causes pneumonia, meningitis, and otitis media [194]. Beta-haemolytic streptococci are classified according to the Lancefield grouping [195]. The Group A *Streptococcus*, *S. pyogenes*, causes pharyngitis, impetigo, and Scarlet fever [196]. The Group B *Streptococcus*, *S. agalecticae*, is a commensal of the gastrointestinal and genitourinary tract, although infection can cause neonatal meningitis [197]. *Stomabaculum* and *Selenomonas* are both anaerobic genera which have been isolated from dental plaques, and are associated with dental infections [198,199].

2.5.3 The microbiome of gastric cancer with microsatellite instability has greater abundance and a different microbial composition to that of microsatellite stable gastric cancer

Work presented in this study demonstrated that microbiome abundance and alpha diversity differ in GC in relation to MSI status of the tumour. The work presented in this chapter also described differences in beta diversity between MSI and MSS GC, as demonstrated by the PERMANOVA and MaAsLin2 analyses. To the author’s knowledge, no study to date has investigated the intratumoural GC microbiome with respect to PCR-determined MSI status (as was the case for the TCGA cohort in the present study).

Consistent with the findings of the present study, a recent investigation by Byrd *et al*. reported increased alpha diversity of oral genera in MSI-high patients versus MSI-low within the GC from the TCGA, when MSI status was determined by RNA-sequencing expression data [200].

Abate *et al*. compared alpha diversity of the MSI TCGA subgroup versus the CIN and GS subgroups in an un-decontaminated TCGA cohort (n=286) and a separate cohort from MSKCC. In contrast to the findings in the present study, Abate *et al*. [156] did not find a statistically significant difference in alpha diversity between the MSI subtype and other TCGA subtypes, although there was a numerically increased Shannon index in the MSI subtype relative to the other subtypes within the TCGA cohort. Abate *et al*. also investigated compositional differences between the MSI TCGA subgroup and other TCGA subgroups, both in the TCGA and MSKCC cohorts. Similar to the findings presented in this chapter, Abate *et al*. reported distinct microbial enrichment of MSI GC relative to other TCGA subtypes, in both investigated cohorts. Interestingly, there was little overlap in genera found to be enriched in MSI GC between the TCGA and MSKCC cohorts, with *Lactobacillus* being the only genus statistically enriched in both cohorts. Similarly, in the present study, there was no overlap in genera or species enriched in MSI (versus MSS) GC between the 100,000 Genomes Project and TCGA cohorts.

There are a number of possible explanations for the differences in findings between the present study and the work by Abate *et al*. First, the present study used PCR confirmation of MSI status for the TCGA cohort and SCV number to infer MSI status for the 100,000 Genomes Project cohort, whereas Abate *et al*. compared the MSI TCGA molecular subtype versus those in the CIN and GC subtypes. Second, the study by Abate *et al*. did not perform any *in silico* decontamination process. As such, there may be a certain amount of noise from contaminating microbiota, diluting any true effect in the difference in alpha-diversity of MSI versus MSS GC. Furthermore, the MSKCC cohort from the Abate *et al*. study used formalin-fixed paraffin-embedded (FFPE) tissue. FFPE tissue may be particularly affected by contamination (as discussed in detail in Chapter 3, section 3.1.3).

Both in the present study, and that by Abate *et al*., multiple genera have been found to be enriched in MSI GC versus MSS. However, there is little concordance regarding which taxa are enriched. This may be a reflection of the increased microbial abundance in MSI GC versus MSS GC as opposed to particular taxa driving an increased abundance.

Consistent with our findings in GC, MSI status has also been associated with microbial changes in colorectal cancer. Specifically, there is considerable

evidence of an association between abundance of *Fusobacterium*, and MSI-high colorectal cancer, compared to MSS colorectal cancer (51–55). There is growing evidence suggesting the role of the genus *Fusobacterium* and species *Fusobacterium nucleatum* in colorectal cancer tumourigenesis [201–203] although the significance of its increased abundance in MSI colorectal cancer is yet to be determined.

It is possible that the present study findings of a more abundant, more diverse microbiome in MSI GC is related to the slow-growing nature of MSI GC. MSI GC may be considered a good prognostic factor, more commonly occurring in lower stage, localised GC without lymph node metastases [204]. As such, it may be that slower growing MSI GCs have longer time to become colonised by microbes.

The local immune environment of MSI GC is understood to be different of MSS GC. Specifically, previous studies have reported increased tumour-infiltrating lymphocytes in MSI GC, relative to MSS GC [205,206]. It has been suggested that this relationship is a result of the large number of mutations present on MSI tumour cells resulting in increased tumour antigens and subsequent increased local immune infiltration [19,206]. The role of the microbiome in this relationship remains unclear; it seems counter-intuitive that increased tumour-infiltrating lymphocytes in MSI GC would result in a more abundant, more diverse intratumoural microbiota. However, it is not known whether such immune cells are in fact functional. Unfortunately, the investigation of immune cell fractions and immune subtype in the present study was not able to provide further insight on this. Further investigation with single-cell RNA sequencing may provide further insight upon this relationship, if resources allowed.

2.5.4 Pathological depth of invasion is associated with differences in gastric cancer intratumoural microbiome diversity and composition

In addition to MSI status, the results from the TCGA analysis suggest greater microbial abundance and alpha diversity in lower pT category, as well as a difference in microbial composition (beta diversity) between pT1/pT2 and pT3/pT4 GC.

There is consensus within the literature that the microbial alpha diversity of established GC is reduced, relative to precancerous conditions and healthy controls conditions [34,40,122,123]. However, results of studies investigating alpha diversity of early versus late GC are more mixed. Consistent with the findings of the present study, Park *et al*. reported a continued decrease in gastric microbiome alpha diversity from samples along the gastric carcinogenic process from gastritis, to gastric adenocarcinoma, to early GC, to advanced GC [125]. In this study, the difference in alpha diversity between early and late GC was not statistically significant, possibly related to small sample sizes (early GC n=36; advanced GC, n=20). In contrast to the findings of the current study, Wang *et al*. reported increased alpha diversity of the intratumoural GC microbiome of advanced GC, relative to early GC. This study was performed on samples from patients from China, exclusively of Chinese Han ethnicity, which may explain differences between this study and studies including patients of different ethnicities, from different geographical locations.

The findings in the present study may reflect a difference in host environment in that fewer taxa, or only certain specific taxa, are able to invade deeper into the pT3/pT4 tumours compared to pT1/pT2 tumours. Molecular differences according to tumour depth have previously been reported [207,208]; thus, when considering this in the content of our findings, it may be that spatial differences in tumour microenvironment influence the local microbiome. Alternatively, a more abundant, more diverse microbiome may be protective against greater tumour invasion, potentially through increased local immune surveillance, although this has not been mechanistically investigated to date.

2.5.5 Gastric cancer histological phenotype is associated with differences in intratumoural microbiome diversity and composition

To the author’s knowledge, the intratumoural GC microbiome has not previously been investigated with respect to histological phenotype. Abate *et al*. have reported microbiome findings of MSI-high versus intestinal-type and diffuse-type, however this study in fact used TCGA subtype as a surrogate for histological type.

Intratumoural microbiome differences according to histological phenotype have been studied in lung cancer. A study of lung tissue samples from patients with

lung cancer reported no difference in alpha diversity according to histological phenotype, but did report a microbial compositional (beta diversity) difference between patients with adenocarcinoma compared to squamous cell carcinoma [209]. Clearly the lung and gastric tumour microenvironments differ; however, in both organs microbiome differences are observed between histological phenotypes, emphasising the importance of the tumour microenvironment.

Gastric-specific mucosal conditions are important to consider in the contest of the microbiome-histology relationship. Intestinal-type GC typically arises on a background of multifocal atrophic gastritis, intestinal metaplasia, and dysplasia [91]. It has been suggested that *H. pylori* initiates an inflammatory cascade, leading to decreased acidity through a reduction in acid-secreting glands, which ultimately results in an environment more suitable for other bacteria [110]. The increased microbial abundance and diversity observed in intestinal-type, compared to diffuse-type GC within the present study may reflect a more hospitable microbial environment resulting from gastric atrophy and subsequent higher pH. Understanding the influence of pH upon the microbiome has potential clinical importance, with respect to prescription of medications which decrease stomach acid production, such as proton pump inhibitors (PPIs). A study by Parsons *et al*. investigating the gastric mucosal microbiome of individuals with gastric hypochlorhydria demonstrated lower microbial alpha diversity in those treated with PPIs compared to normal stomachs [210], suggesting that PPI use and the resulting change in luminal pH does not lead to increased diversity. In this study by Parsons *et al*., the alpha diversity of gastric mucosa with *H. pylori* atrophy and *H. pylori* gastritis was considerably lower than that of the PPI-treated or normal group, demonstrating dominance of *H. pylori.* Together with the findings presented in this chapter, further investigation is warranted of the impact of the individual contributions of pH, prior *H. pylori* infection, and gastric atrophy upon the microbiome alterations, e.g. through microbial culture studies.

Previous studies in GC have demonstrated evidence of molecular differences [116,166], as well as differences in the ratio of tumour to stroma [211], according to histological phenotype. Existing data from a single-cell RNA sequencing analysis of GC is suggestive of differences in the proportions of plasma cells and KLF2-expressing epithelial cells between diffuse- and intestinal-type GC [212]. It is possible that molecular variability between diffuse-type and intestinal-type GC could result in differences in microbial abundance

and alpha diversity, by influencing local immune cells, although further work is needed in this area to fully understand this relationship.

2.5.6 Use of immune data did not provide insight into relationships between clinicopathological variables and the intratumoural microbiome

It was decided to further explore a potential relationship between the GC microbiome and immune cells in the tumour microenvironment, using the publicly available immune cellular fractions and immune subtype data from the TCGA. The aim of this part of the study was to further understand the observed relationships between the microbiome and clinicopathological characteristics including MSI status, pT status, and histological phenotype.

Whilst statistically significant relationships between immune subtype and both microbiome abundance and alpha diversity were observed, caution should be taken in interpreting these findings. As demonstrated in Figure 18, these differences appear to be driven by the C3 immune subtype, where low microbial abundance and low Shannon index was observed. The C3 immune subtype corresponds to the ‘inflammatory’ subtype, defined by elevated Th17 and Th1 genes, low to moderate tumour cell proliferation, lower levels of aneuploidy, and lower levels of somatic copy number alterations. As such, there may be significant overlap with the GS TCGA subtype, also observed to have low microbiome abundance and alpha diversity.

When the relationships between the estimates of the immune cellular fractions of the six pre-determined immune cell types (lymphocytes, neutrophils, macrophages, dendritic cells, eosinophils, and mast cells) were investigated with respect to microbial abundance and Shannon index, no notable relationships were identified. This suggests that using individual cell type estimates may not help further understand the relationship between local immune cells and the intratumoural GC microbiome. It may be that the use of inferred cellular fraction estimates may not fully capture the complexity of the local immune response. In addition, the reliance on whole tumour immune data, which was not spatially orientated may have meant that important spatial differences in immune cells were not detected.

The lack of microbiome-immune relationships identified in the present study does not mean that the local immune cells are not important in understanding the observed relationships between clinicopathological factors and the GC microbiome. Instead, these findings are not sufficient to add to the current understanding of the role of immune cells in the GC microbiome.

Few studies have examined the relationship of the GC intratumoural microbiome and local immune cells to date. Ling *et al*. explored the relationship between local immune cells, specifically regulatory T cells (Tregs) and plasmacytoid dendritic cells (pDCs), and the intratumoural GC microbiome [43]. The authors reported *Stentrophomonas* and *Selenomas* were positively correlated with pDCs and Tregs respectively, and *Comamonas* and *Gaiella* were negatively correlated with pDCs and Tregs respectively. pDCs and Tregs are involved in immune suppression in tumours, therefore increased and decreased levels of these immune cells may play a role in immune evasion and antitumour activity, respectively [213,214]. Peng *et al*. reported an inverse correlation between CD8+ tissue-resident memory T (TRM) cells and the genus *Methylobacterium* [215]. This study further investigated the effect of this genus upon the GC microenvironment by demonstrating that *Methylobacterium* results in decreased CD8+ TRM cells in mouse models of GC.

Importantly, both *Gaiella* and *Methylobacterium* are recognised environmental microbes; *Gaiella* was isolated from a deep mineral water aquifer [216] and species from *Methylobacterium* are commonly found in soil [217]. Whilst *Methylobacterium* has been associated with healthcare-associated infections in immunocompromised individuals [218], it has also been identified as a contaminant in DNA extraction [219]. Caution should therefore be applied when interpreting these results.

Collectively, these results suggest that local immune cells should be investigated further to understand the interplay between the GC, microbiome, and immune cells. However, given the complexity of these three components, a thorough and systematic approach will be needed to fully explore these potentially complex relationships.

2.5.7 Clinicopathological factors associated with increased microbial abundance and diversity are associated with good prognosis

Notably, MSI (versus MSS) GC, low (versus high) pT category, and intestinal (versus diffuse) histological phenotype – all of which have been observed to have greater microbial diversity and abundance within the present study – are generally recognised to be good prognostic factors in GC [220–223]. In general, higher microbial diversity is thought to be associated with improved health outcomes, including in [51], but not limited to [224], patients with cancer. Considered in this context, the observations of work presented in this chapter of increased microbial diversity in such “good prognosis” subgroups (MSI, low pT, and intestinal histology) strongly warrants further investigation of the role of microbial abundance and diversity in GC behaviour. In particular, further research should investigate whether and through what mechanisms increased microbial diversity may contribute to the improved outcomes observed in these good prognosis groups. If results from future studies support the hypothesis that greater GC microbial abundance and diversity results in superior outcomes, this may inform development of therapeutic investigations to increase microbial abundance and diversity, with the aim of improving GC outcomes.

2.5.8 The gastric cancer microbiome was not found to be related to anatomical location

Of note, no significant differences in the GC microbiome (in terms of abundance, alpha diversity, or composition) were observed according to tumour location, in line with previous GC microbiome studies [40,124]. This is in contrast to findings in the colorectal cancer microbiome, where sidedness is strongly associated with microbiome compositional differences [225–227]. It is possible that such anatomical differences are not as influential in GC due to the proximity of the cardia to the non-cardia stomach, the liquidity of gastric contents, and the relative short transit through the stomach compared to the transit length and time of the lower digestive system.

2.5.9 Decreased microbial abundance was observed in GC from Asia

A decreased microbial abundance was observed in the TCGA cohort in GC from Asian origin, relative to GC from the rest of the world. Since the majority of previous GC microbiome studies have occurred at single centres or within

individual countries, there are little available data on differences according to continent. However, Yu *et al*. explored the GC microbiome of two patient cohorts: one from Mexico and one from China [120]. This study did not report alpha diversity findings according to geography; however, they did report that the microbiome composition (beta diversity) differed between samples from China versus samples from Mexico. The findings from the TCGA cohort in the present study did not find compositional differences according to geography (Asia versus not Asia). The difference in beta diversity findings between the present study and the study by Yu *et al*. may be a reflection of the fact that the TCGA cohort includes patients from across the globe, whereas the study by Yu *et al*. featured two more heterogeneous cohorts. Collectively, these results suggest that geography is important and should be taken into account when performing multivariate analyses (such as in the methods of the present study, whereby geography was considered the first variable in the PERMANOVA analysis). Factors underlying how the microbiome may be affected by geographical location are unclear and may be multifactorial. Possible factors could include environmental differences, such as diet and living conditions, as well as differences in antibiotic use and the prevalence of certain microbes (e.g. *H. pylori*) within the populations studied.

2.5.10 Study limitations

The work presented in this chapter has some limitations. The 100,000 Genomes Project Cohort had small sample numbers (n=89), limiting the statistical power, particularly in investigations of microbial abundance and Shannon index. The TCGA cohort was based on exome data, which has a limited ability to facilitate detection of the microbial component of the original sample, as discussed with respect to decontamination. Furthermore, the availability of metadata differed between the two cohorts limited investigation to a purely exploratory analysis.

2.5.11 Conclusion

In conclusion, using two separate sequencing databases, microbiome differences were identified in relation to depth of tumour invasion, tumour phenotype, and molecular characteristics such as presence of MSI. These findings further reinforce the notion that the relationship between the GC microbiome and clinicopathological variables is multifactorial and as such, should be considered when planning, conducting, and interpreting the results

from investigational clinical studies. Future work should focus on validation of the exploratory findings from the present chapter and functional studies. This work was conducted on the largest collection of GC microbiomes yet collated. Using this large collective dataset, the findings of previous studies (conducted in smaller cohorts) [155,158,228] were consolidated. This work represents the most comprehensive evaluation of the relationship between microbial differences and patient and tumour characteristics to date, the findings of which underline the potential clinical importance of the GC microbiome and provide a strong rationale for further investigations to improve the depth of understanding in this area.

2.5.12 Chapter summary

- Sequencing databases from the 100,000 Genomes Project and TCGA were curated and used to analyse the microbiome according to clinicopathological characteristics.

- Decontamination is important within GC microbiome studies; the decontamination process developed by Dohlman et al. [158] was adapted and applied to both databases.

- The microbiome of the 100,000 Genomes Project and TCGA cohorts largely is consistent with that of previous studies.

- MSI, low pT, non-Asian origin, and intestinal histological phenotype were all associated with increased microbial abundance and alpha diversity.

- Microbiome composition differed according to MSI status, patient sex, and pT category.

- Further functional studies are required to understand how the local environment may influence the intratumoural microbiome.

Chapter 3 Characterising the microbiome from gastric cancer tissue samples

3.1 Introduction

The work described in Chapter 2 involved using data previously collected by others. In this chapter, the described work involved prospective collection of GC samples and use of different techniques to characterise the microbiome of these samples. The work carried out in this chapter was performed across two laboratories: tissue processing and sequencing was performed in the Quirke laboratory, University of Leeds; bacterial isolations were performed in the Allen-Verco laboratory, University of Guelph, Canada.

3.1.1 Advantages of prospective tissue collection

For clarity, prospective tissue collection refers here to prospective collection of tissue with the prior intention of using for microbiome analysis, as opposed to studies where tissue may have been prospectively collected for other uses, (e.g. 100,000 Genomes Project) or retrospectively accessing tissue stored within the NHS. Prospective tissue collection for microbiome studies offers certain advantages over retrospective microbiome analysis of prospectively collected tissue for other uses.

Prospective tissue collection enables the collection, storage, and sequencing methods to be tailored specifically to facilitate microbiome analysis. For example, a protocol may be developed so that tissue is frozen shortly after collection. Use of fresh or fresh-frozen tissue may offer advantages over fixed tissue, as discussed further below (section 3.1.3). Another example is that if any microbial culture work is planned, tissue can be handled in a way that maximises viability of any present microbiota, e.g. through minimising oxygen exposure to protect anaerobic microbes.

With prospective collection, processes can be factored into the collection method to minimise contamination. This may involve use of sterile containers, equipment, and reagents used during collection, processing, and sequencing of tissue.

3.1.2 Challenges of prospective tissue collection

Whilst prospective tissue collection offers a number of advantages, there are also significant challenges, some of which may represent significant hurdles for the specific case of investigating the intratumoural GC microbiome.

Prospective collection of fresh GC tissue involves timely identification of patients undergoing procedures in which GC tissue could be obtained. This may involve close collaboration and coordination with the wider multidisciplinary clinical team, in order to adequately consent appropriate patients in advance of any procedure which would allow tissue collection.

For patients with confirmed resectable GC, the only opportunity to collect GC fresh tissue for research use may be at the point of surgical excision. Collection of tissue at the point of surgery requires coordination with the surgical team performing the operation in order to collect tissue in a timely manner.

For some patients with confirmed GC, there may also be the opportunity to collect tissue prior to surgical excision, if the patient were to undergo an endoscopy at the time of staging laparoscopy. Staging laparoscopy is recommended for patients with GC prior to commencing a curative pathway of surgery and perioperative chemotherapy, to exclude the presence of radiologically and macroscopically occult peritoneal disease [229]. If an upper gastrointestinal endoscopy is performed simultaneously, this may offer an opportunity to prospectively collect biopsies for GC microbiome analysis, although this would require collection of additional biopsies not part of routine care.

Often, the only biopsy performed in a patient's journey is the diagnostic biopsy. However, targeting the all patients undergoing an upper-gastrointestinal biopsy would have a low pick-up rate of GC cases, due to the proportion of investigated patients who do not have GC. This highlights one of the difficulties of using diagnostic biopsies as a source of prospective tissue collection: the large number of patients needed to consent and have tissue taken would be a very resource- and labour-intensive way of collecting GC tissue.

Some of these issues may be more challenging in the setting of GC than would be in the case of collecting more common tumours, such as colorectal cancer, or other tissue, such as stool for microbiome analysis. Whilst the challenges described may be managed, it is important to note that generally, such challenges of prospective tissue collection may lead to smaller study sample sizes.

3.1.3 The use of formalin-fixed, paraffin-embedded tissue to determine the microbiome

One alternative to prospective fresh-frozen tissue collection is to use routinely collected FFPE tissue blocks. Formalin-fixing involves leaving the tissue sample for a certain length of time in a solution consisting of formaldehyde, sodium phosphate monobasic, and sodium phosphate dibasic in distilled water, before processing into paraffin [230]. The process of formalin-fixation and paraffin-embedding maintains the cell, protein, and organelle structure, and enables room temperature storage, making FFPE an ideal process for storing tissue long-term, for light microscopy and other histopathological uses [231].

The DNA cross-linking which occurs in formalin fixation may be reversed, allowing DNA purification from FFPE tissue [231]; as such, FFPE tissue may be used for microbiome analysis. The advantages of using FFPE tissue over fresh or fresh-frozen tissue are 1) increased accessibility of tissue samples from a greater number of patients, since routine FFPE blocks have to be stored for long periods of time and 2) many sections can be cut from a single block for multiple different investigations. Potential disadvantages of using FFPE tissue for microbiome analysis are related to 1) contamination introduced by the fixation and embedding procedure, and 2) the impact of the fixing, embedding, and storage processes upon DNA quality and subsequent sequencing [231].

A number of small studies has investigated the use of FFPE tissue versus fresh-frozen tissue in characterising the microbiome of non-neoplastic gastric [232] and other human tissues [233,234] using 16S rRNA gene sequencing methods. The study in non-neoplastic gastric tissue of FFPE tissue versus fresh-frozen tissue reported a decreased number of operational taxonomic units (OTUs) in FFPE compared to fresh-frozen, following a decontamination process using blank FFPE controls. These findings suggest that that the FFPE processing introduces bias into the microbiome profile [232].

3.1.4 16S rRNA gene sequencing versus metagenomic sequencing

Microbiome sequencing methodologies fall into one of two broad groups: 16S rRNA gene sequencing and metagenomic sequencing.

16S rRNA gene sequencing (herein referred to as 16S rRNA sequencing) may be described as a marker gene sequencing technique as it relies upon the amplification of the 16S rRNA gene, present in all bacteria [31]. The 16S rRNA gene is a large (approximately 1600 base pairs) gene [31], containing nine hypervariable (V) regions (V1-V9), which differ between bacterial taxa. 16S rRNA sequencing relies upon primers targeting the conserved regions between the V regions, which results in the amplification of the V regions. This process enables sequencing via NGS or (less commonly) Sanger sequencing. Sequences are subsequently aligned to a reference database and assigned to the most likely taxa. Primer choice is important in 16S rRNA sequencing: the V region(s) sequenced affects the sequencing resolution. Sequencing of the V4 region only, the most commonly utilised single V region, results in sequencing data that can only be used to characterise to genus level [235], whereas longer sequences, such as V3-V4, V3-V6, or even full length 16S rRNA gene sequencing can increase the taxonomic resolution. Of note, whilst V regions of the 16S rRNA gene are only present in (and can therefore only be used in sequencing of) bacteria and archaea, other marker genes can be used for identification of other bacteria and microbes, such as the *cpn60* universal target, present across bacteria and eukaryotes [236,237]; *rpoB* gene for bacteria [238]; and internal transcribed spacer region (ITS) marker for fungi [239].

Metagenomic sequencing, as described in detail in Chapter 2, offers an alternative approach to 16S rRNA sequencing. Metagenomic sequencing offers the advantages that it can enable analysis to species level and allow identification of the whole microbiome (i.e. is not limited to bacteria and archaea). In general, fewer samples can be analysed at once in metagenomic approaches, meaning that 16S rRNA sequencing tends to be the lower cost option. As 16S rRNA sequencing is relatively inexpensive compared to metagenomic analysis, and, factors such as sample size, budget, and importance of taxonomic resolution will influence choice of microbiome sequencing methodology.

3.1.5 Microbial culture and isolation

Importantly, detection of microbial sequences through 16S rRNA sequencing or metagenomic sequencing cannot differentiate between alive or dead microbial taxa. Microbial culture techniques, whereby microbes are provided with laboratory conditions to optimise growth, can overcome this challenge. The culture of microbes from tissue demonstrates the presence of living microbes of such cultured taxa in the sampled tissue. Furthermore, culture allows isolation of individual microbes for further investigations. Microbial isolation is the process whereby pure colonies of microbial cultures are developed, typically through picking out a single colony from a mixed community and providing optimised conditions for growth of microbial biomass. Historically, isolated bacteria were characterised through light microscopy and applying specific stains to determine whether certain features were present, such as the Gram stain to identify bacteria with thick peptidoglycan cell walls [240]. Whilst such tests are still relied upon clinically for pathogen identification, molecular techniques involving sequencing are now widely integrated into microbial culture research methodologies. In addition to providing evidence of living microbiota, microbial culture offers the advantage of allowing investigation of phenotypic characteristics. Disadvantages of microbial culture are that techniques are time-consuming and certain microbes require such specific conditions for growth that they have not (yet) been successfully cultured in a laboratory [241,242].

A small number of studies have performed microbial isolations from non-cancerous gastric mucosa [243–247] and GC tumour biopsies [244,248]. Of these studies, only two [245,248] used more than three different culture conditions to perform isolations. A broad range of culture conditions, including different media types and gaseous conditions, is important to maximise captured diversity in isolations. However, performing isolations under a greater number of conditions results in an increased workload, often considered impractical with higher samples numbers. Of the two studies which have used more than three conditions, one was a study of GC tumour biopsies published in 1988 where taxa identification was performed by colony morphology, biochemical tests and gas-liquid chromatography, limiting the taxonomic resolution of isolates [248]. The other study involved isolations from patients with gastritis from Columbia, although only one anaerobic condition was included, which may have limited the ability to capture diversity of anaerobic bacteria from these samples [245]. To date, there has been no isolation study with a broad range of isolation conditions across aerobic, anaerobic, and

microaerobic conditions performed on GC samples, resulting in sequenced libraries isolates.

3.1.6 Aims & objectives

The primary aim of the work described in this chapter was to use prospectively collected GC tissue to characterise the GC microbiome through a range of different methodologies. Specific objectives were to:

- Prospectively collect fresh tumour and adjacent normal tissue from patients with GC.

- Adapt existing DNA extraction and sequencing methods to enable DNA extraction from GC tissue suitable for microbiome analysis.

- Perform low-coverage WGS of DNA extracted from fresh-frozen GC samples and FFPE GC samples and compare results of using these different sample types.

- Perform 16S rRNA sequencing of fresh-frozen GC samples and compare results to low-coverage WGS of fresh-frozen GC samples.

- Perform bacterial isolations from GC samples, using a broad range of isolation conditions.

- Use the results of the different characterisations to assess the extent to which prospectively collected GC samples are representative of current understanding of the GC microbiome.

3.2 Methods

3.2.1 Ethics and local approval

An application was made to the Health Research Authority (HRA) to collect, store, and analyse gastrointestinal tissue, though the Integrated Research Application System (IRAS), under IRAS project identifier 290019, study title: ‘Investigation of the microbiome in the gastrointestinal tract’, submitted in January 2021. The IRAS application was supported by PhD student Dr Alice Westwood, who drafted the initial participant information sheets, and patient representative Mr John Barns, who advised on project design and the

participant information sheets. This application was reviewed by the North West-Greater Manchester East Research Ethics Committee (REC), and further information and amended patient information sheets were supplied to the REC as requested.

Areas of particular ethical consideration were 1) clarification of additional risk associated with tissue collection and 2) link-anonymisation of participant data and determining who had access to this data. Additional risk associated with tissue collected referred to the collection of additional biopsies from study participants undergoing staging laparoscopy. To address this issue, discussions were had with upper gastrointestinal surgeons and a patient representative (Mr John Barns) to guide the content of a separate participant information sheet created for the situation of collecting biopsy tissue at staging laparoscopy. All participant data was link-anonymised, with only essential team members having access to the linking document. To ensure the correct identification of consenting participants, the researcher collecting tissue (Dr Mary Booth in the case of the present study) had access to both patient identifiable information as well as the research study number. For that reason, only the two individuals collecting tissue, Dr Mary Booth and Dr Alice Westwood (collecting colorectal tissue for aligned study under same ethical approval) had access to the password-protected, securely stored anonymisation linking document.

The overview of tissue collection as submitted in the REC application is shown in Figure 19. This application also included access to archival samples, although they were not utilised within the present study.

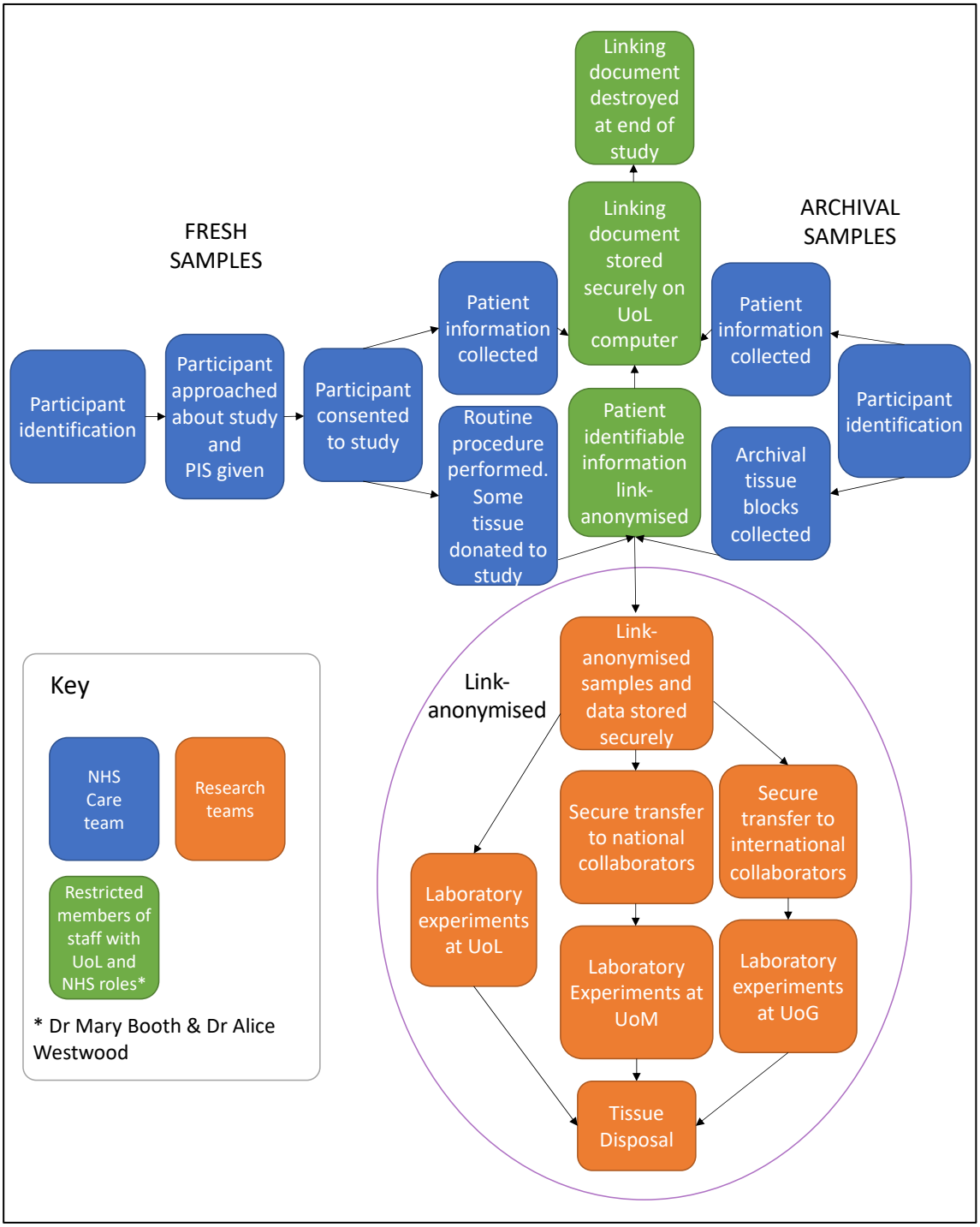


Figure 19. Flowchart of tissue collection. PIS, participant information sheet; UoG, University of Guelph; UoL, University of Leeds; UoM, University of Manchester. Colour of shapes represent which members of the team had access to data at different stages of data collection. All work within lilac oval was with link-anonymised data.

An application was created for Leeds Teaching Hospitals NHS Trust (LTHT) Capacity and Capability approval to deliver the collection of gastrointestinal tissue under the Research and Innovation number GS21/139713, with the support of Ms Catherine Moriarty (Senior Research Nurse, LTHT).

3.2.2 Patient identification and consent for tissue collection

Eligible patients had gastric or gastro-oesophageal junctional adenocarcinoma and were planned to undergo either surgical resection or staging laparoscopy with concurrent endoscopy at St James' University Hospital (SJUH), Leeds. Patients had to be at least 18 years old and able to provide informed consent without an interpreter. Patients were eligible regardless of whether or not pre-operative chemotherapy was received as part of their management. Exclusion criteria included: tumour size <2.5cm (before or after preoperative chemotherapy) and radiological appearance of fragmentation after preoperative chemotherapy. Potential study participants were identified during the weekly LTHT upper gastrointestinal multidisciplinary team meeting or during clinic visits.

Explicit consent was required for all study participants who donated fresh tissue. Eligible patients were verbally informed of the study by the clinical care team. If interested, the study was explained in more detail and the patient was offered a participant information sheet and at least 24 hours before giving consent to participate in the study. Due to a reduced number of clinician visits as a result of the COVID-19 pandemic, some patients were offered a follow-up telephone call to answer any questions, to give their consent verbally. In such cases, patients signed the consent form at home and brought the form to SJUH at the time of their surgical procedure. Paper consent forms were carried by hand in a sealed envelope from SJUH to the Clinical Sciences Building, University of Leeds, before being stored in a locked filing cabinet with restricted access.

3.2.3 Tissue retrieval

Fresh gastric tumour tissue and adjacent macroscopically uninvolved gastric mucosa was collected from patients having either a) gastrectomy or oesophagogastrectomy or staging laparoscopy with upper gastrointestinal endoscopy for the management of gastric adenocarcinoma. All fresh tissue samples were collected in the operating theatre, under the guidance of the surgical care team. Sampling protocols were discussed in advance with both histopathology and surgical departments.

For resection samples, resection specimens were opened away from the site of the tumour. Up to three samples were taken from the luminal surface of the tumour and up to three samples were taken from gastric mucosa from the

anterior stomach wall, at least 5cm away from the tumour. Samples were approximately 5mm x 5mm x 5mm in size.

Endoscopic biopsy samples collected during staging laparoscopy were taken by the clinician performing the endoscopy. Three biopsy samples were taken from the area identified endoscopically as GC and three samples were taken from an adjacent area of macroscopically normal appearing gastric mucosa. Biopsy samples were approximately 1-2mm in maximum length each.

All samples were placed in pre-labelled ultraviolet-irradiated 1.8ML CryoTube Vials™ (ThermoFisher Scientific, #377264), placed on ice in a separate container with a lid, before transporting on foot to the Wellcome Trust Brenner Building, University of Leeds, for immediate freezing. Samples were frozen by submerging cryotubes in an ethanol-dry ice slurry for one minute, before transferring into -80°C freezer for storage.

An overview of the use of the collected samples is shown in Figure 20.

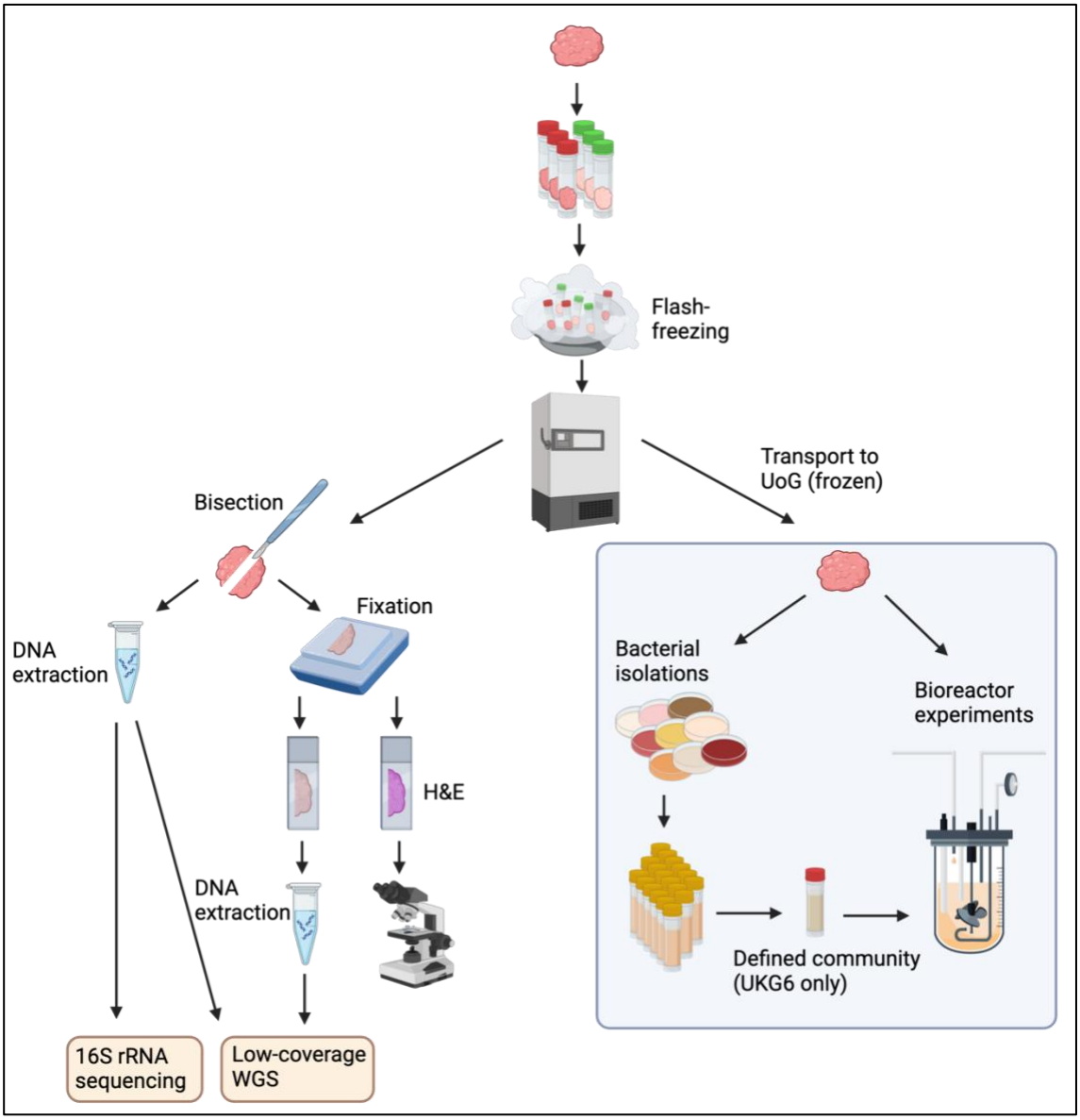


Figure 20. Overview of investigations performed using prospectively collected GC tissue. Up to three tumour and three adjacent (macroscopically normal appearing) GC samples from each study participant were flash-frozen within sample tubes and stored at -80°C. One tumour sample and one adjacent sample from each participant were bisected. Half of sample was used for fresh-frozen DNA extraction and 16S rRNA sequencing and low coverage WGS. Half of sample was formalin-fixed and paraffin-embedded and used for both H&E assessment and DNA extraction with subsequent low-coverage WGS. Additional samples from all patients were transported frozen to the University of Guelph, Canada for culture-based experiments (pink box). Samples were used to perform bacterial isolations and/or bioreactor experiments. In the case of UKG6, one GC sample was used to perform bacterial isolations and create a defined community. This defined community and a separate UKG6 GC sample were run in parallel bioreactor experiments, as described in Chapter 4. 16SrRNA, 16S ribosomal ribonucleic acid; DNA, Deoxyribonucleic acid; H&E, haematoxylin & eosin; UoG, University of Guelph; WGS, whole genome sequencing. Figure created with biorender.com.

3.2.4 Metadata collection

Relevant clinical data were collected from electronic patient records by care teams and inputted onto an electronic case report form (CRF) without patient identifiable information. The CRFs were stored using a unique research identifier. A separate document linking the NHS number with the unique research identifier for each patient was stored on drive with restricted access: only Dr Mary Booth and Dr Alice Westwood (working on an aligned lower gastrointestinal cancer project) had access to the linking document. Beyond this restricted-access linking document, no patient-identifiable information was held by the University of Leeds or any other research site.

3.2.5 Tissue fixing, embedding, and histological assessment.

Prior to DNA extraction, frozen tissue was thawed to room temperature and bisected longitudinally, on a sterile surface, using a sterile blade. One half of each sample was fixed by placing in a 2mL collection tube with 10% formalin. The other half of the sample was used for DNA extraction as described in section 3.2.6.

The half of each bisected sample which was placed in 10% formalin was used to create FFPE blocks for histological analysis and investigation of the microbiome through analysis of FFPE tissue.

Samples were fixed by leaving in 10% formalin for at least 24 hours before embedding. Embedding and sectioning was performed by Ms Gemma Hemmings, according to local Quirke laboratory protocol. In brief, samples were dehydrated by passing through increasing concentrations of ethanol, then passed through xylene as a clearing agent before passing into liquid paraffin and allowed to set. Each FFPE sample was sectioned at 5µm thick and mounted onto glass slides and dried at 37°C for 12 hours; eight slides per sample were created, each with two sections per slide.

One FFPE slide for each sample was stained with haematoxylin and eosin (H&E) to allow histological assessment. H&E staining was performed by Ms Gemma Hemmings, according to the local Quirke laboratory protocol.

Slides were digitally scanned at 40x by the Virtual Pathology Department at the University of Leeds. Digital slides were reviewed by Dr Mary Booth and Professor Heike Grabsch together, to assess for presence or absence of adenocarcinoma features.

3.2.6 DNA extraction from frozen gastric cancer samples

A number of priorities were recognised prior to determining the method of DNA extraction and will be described here. The protocol was to be suitable for processing of mucosal tissue. Since the biomass of the fresh samples was expected to be low, protocols using very small sample masses of tissue were to be avoided. Laboratory contamination was to be limited as much as possible, by using ultraclean reagents. Bead beating was to be included to improve disruption of the rigid cell wall of gram-positive bacteria.

Since no existing protocol was identified which was optimised for all of the above criteria, existing protocols were adapted to tailor the DNA extraction technique for the tissue and downstream applications. In general, the QIAamp UCP Pathogen Mini Handbook protocol [249] was followed, with the inclusion of pathogen lysis steps. The reagents and plasticware used were part of the QIAamp UCP Pathogen Mini Kit (Qiagen, #50214) unless otherwise specified. The UCP Pathogen Mini Kit was used as the reagents are certified to have no detectable contaminating microbial DNA at the point of delivery, thus reducing the risk of laboratory contamination. Manufacture's guidance using the QIAamp UCP Pathogen Mini Handbook protocol [249] was followed, except for specific adjustments as described below.

Since the QIAamp UCP Pathogen protocol does not include a protocol for the use of tissue, tissue quantities and the enzymatic digestion steps were based upon the QIAamp DNA Mini and Blood Mini Handbook 05/2016 [250]. This protocol was chosen for the enzymatic step as 1) it had been optimised to use tissue and 2) larger tissue quantities of up to 25mg tissue could be used. Specifically, up to 25mg of thawed frozen GC samples tissue from each sample was used to perform DNA extractions. Tissue was transferred into a Pathogen Lysis Tube S (Qiagen, #19091) to allow for subsequent pathogen lysis steps and manufacturer's protocol was followed for enzymatic digestion. In brief, this involved cutting up tissue to increase surface area, and enzymatic lysis through incubation with Proteinase K. The lysis incubation was performed overnight (16

hours), an option within the manufacture's protocol (QIAamp DNA Mini and Blood Mini Handbook 05/2016 [250]) and since the Quirke laboratory have previously experienced optimal lysis with an overnight incubation.

A bead beating step was included following enzymatic digestion as a mechanical lysis step to improve microbial cell lysis. For the bead beating step, the manufacture's protocol was followed, using the TissueLyser (TissueLyser II, Qiagen, #85300). In brief, this step involved vortexting the sample in the Pathogen Lysis tubes at 20Hz for 10 minutes, before transferring into fresh tubes. The remaining lysis, precipitation and washing steps were performed according to the manufacture's protocol.

DNA was subsequently quantified by fluorescence, using the Qubit 4 Fluorometer (ThermoFisher Scientific).

3.2.7 DNA extraction from formalin-fixed, paraffin-embedded GC samples

DNA was extracted from FFPE samples as previously described [251]. This cited publication, in which DNA was extracted from FFPE tissue and sequenced via NGS to detect human papillomavirus, demonstrates the ability to use DNA extracted from FFPE for detection of microbial sequences. In brief, for each sample, ten sections from five FFPE slides (i.e. two sections per slide) were deparaffinised. The QIAamp DNA Micro kit (Qiagen, #56304) was used for DNA extraction from FFPE sections, according to the manufacture's protocol. DNA was quantified by fluorescence, using the Qubit 3.0 Fluorometer (Invitrogen, #Q33216) and stored at -4°C.

3.2.8 16S rRNA sequencing of gastric cancer samples

The method used for 16S rRNA sequencing was as previously described [252]. This method is based upon The Earth Microbiome Project (EMP) 16S Illumina Amplicon Protocol [253], an open-access protocol designed for the Illumina platform, increasingly used across a number of microbiome studies, with the aim of improving standardisation. The EMP protocol had previously been adapted within the Quirke laboratory to include additional forward primers to increase high throughput capacity and optimised for use with equipment within the Quirke laboratory [252].

In brief, DNA was brought to room temperature and quantified using Quant-iT dsDNA Broad-Range Assay Kit (ThermoFisher Scientific, #Q33130), using a plate format and diluted to 20ng/µL. PCR amplification of the V4 region of the 16S rRNA gene was performed using the expanded EMP primers. Human C2020 DNA (20ng/ul) and E. coli DNA (20ng/ul) were used as negative and positive controls, respectively. PCR amplicons were quantified using the Quant-iT dsDNA Broad-Range Assay Kit (ThermoFisher Scientific, #Q33130); of 240ng of each sample was pooled, divided into 5 pools. Pools were purified using the MinElute PCR Purification kit (Qiagen, #28004) and frozen at -20°C. The protocol was followed using a 96-well thermocycler (Veriti 96-Well Thermal Cycler, Applied Biosystems, #4375786).

3.2.9 Low-coverage whole genome sequencing of gastric cancer samples

Prior to low-coverage WGS, a sonication step was included to shear DNA into short fragments. This step was included to ensure fragments are small enough for next generation sequencing. DNA was brought to room temperature and quantified using Quant-iT dsDNA Broad-Range Assay Kit (ThermoFisher Scientific, #Q33130). To 1.5mL Covaris Micro Tubes, 200ng of each DNA sample were added, and made up to a total of 50µL, using the elution Buffer EB (Qiagen, #19086). Samples were sonicated using the E220 Sonicator (Covaris, #500239) according to the following shearing parameters: 175W peak incident power; 10% duty factor; 200 cycles; 120s treatment time.

The method used for low-coverage WGS was as previously described [254], as used locally within the Quirke laboratory for metagenomic sequencing of extracted DNA. In brief, sonicated DNA samples were prepared for sequencing using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, #E7370). Performed steps included end repair, adapter ligation, size selection, PCR amplification, clean-up, quantification and quality control check. The protocol was followed using a 96-well thermocycler (Veriti 96-Well Thermal Cycler, Applied Biosystems, #4375786).

Size selection was performed using AMPure XP Beads (Beckman Coulter #A63881) to target DNA fragments of 200 base pairs. The index primers used in PCR amplification of DNA fragments were those used locally within the Quirke laboratory. These primers were based upon the NEBNext Multiple Oligos for

Illumina (New England Biolabs, #7335S/L), although further primers including more unique tags were designed by Dr Henry Wood, to allow a higher number of samples to be processed at the same time. Final clean-up was performed using AMPure XP Beads (ThermoFisher Scientific, #A63881). To a new plate, 30µL of each sample was added and stored at -20°C.

PCR library quantification and quality control was performed by running samples on TapeStation (Agilent 4200, #G2991BA), looking for peaks in the 270-400 base pair range. The final pool was created by adding 100ng of each library. This pool was submitted to the University of Leeds NGS Facility for sequencing on the Illumina NextSeq 2000.

3.2.10 Bacterial isolations

All bacterial isolations were performed in the Allen-Vercoe laboratory, University of Guelph, Canada. The overview of the bacterial isolation workflow is shown in Figure 21. The isolation procedure largely followed that of standard isolation procedures used within the Allen-Vercoe laboratory, with the main adaptations being that 1) tumour samples were first homogenised into a tissue suspension and 2) a microaerobic environment was added in addition to aerobic and anaerobic environments.

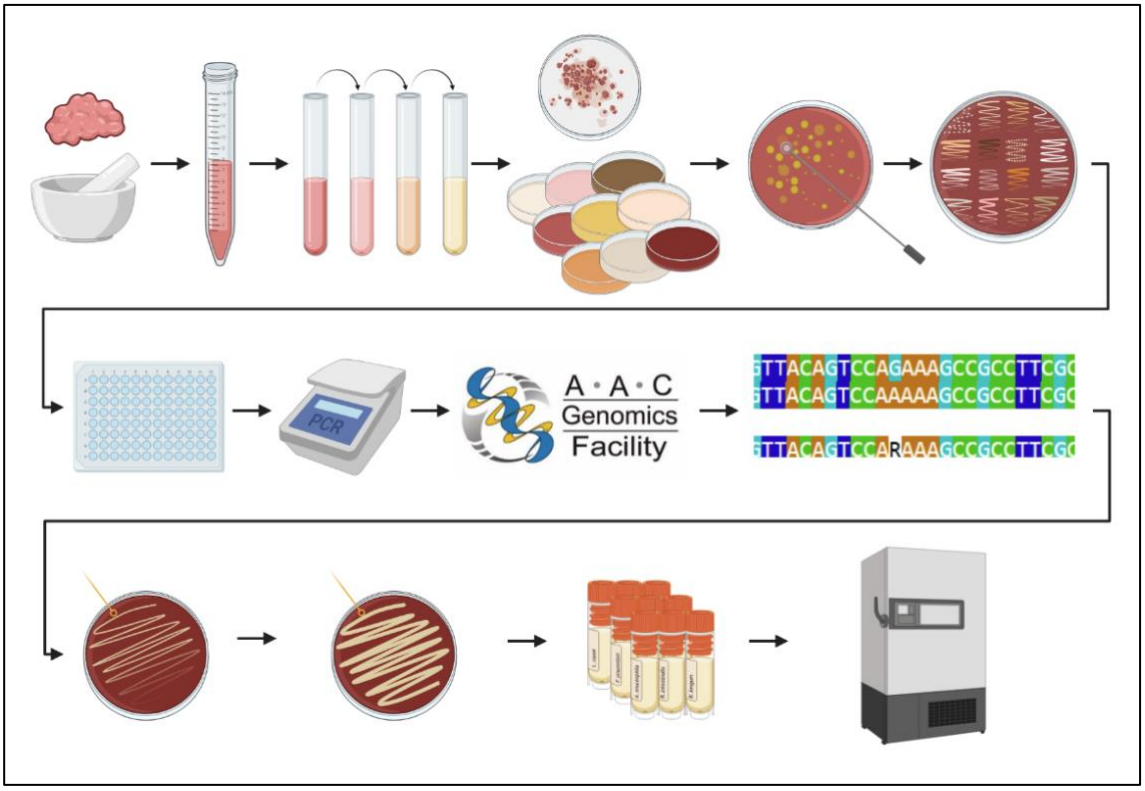


Figure 21. Workflow of bacterial isolations. Fresh-frozen GC tissue was homogenised, serially diluted before being added to agar plates across a range of conditions. Individual colonies were selected from mixed growth on agar plates and re-streaked onto patch plates. PCR and Sanger sequencing was performed on individual colonies. Unique strains were re-streaked on agar plates to maximise biomass, then stored at -80°C. AAC, Advanced Analysis Centre. Figure created using Biorender.com.

3.2.10.1 Tissue homogenisation and isolation procedure

Prior to isolations, agar plates were prepared for the required media. Media were selected with the aim of capturing as much diversity as possible from the GC samples. Details of all 11 agar media types used can be found in Table 7. For all GC sample investigated, all media types were incubated at 37°C anaerobically (under an atmosphere of 10% CO_2, 10% H_2, 80% N_2) and aerobically. In addition, Gonococci and Brucella media were incubated at 37°C with tissue suspension under microaerobic conditions, with the aim of facilitating isolation of *Helicobacter*.

Table 7. Details of media used for isolation procedure. All additions were added following autoclaving. Reagent details are described with the first use in the table. FAA, Fastidious Anaerobe Agar.

Media	**Reagents (500mL)**	**Reagent information**	**Notes**
Fastidious Anaerobe Agar (FAA)	FAA (Neogen, #NCM0014B) 23g Additions: 25mL (5%) defibrinated sheep's blood.	FAA: Neogen #NCM0014B; defibrinated sheep's blood: HemoStat laboratories, #DSB500	Enriching media; supports variety of fastidious anaerobes.
Fastidious Anaerobe Agar without blood	FAA 23g		
Fastidious Anaerobe Agar with complex carbohydrates and mucin	FAA 23g, pectin 2.5g, inulin 0.45g, soluble (potato) starch 2.5g, (porcine) mucin 2g, dextrose 1.5g. Additions: 25mL (5%) defibrinated sheep's blood.	Pectin from citrus peel: Sigma #P9135; inulin: Glentham Life Sciences #GC4090; starch: Sigma #S-2630; mucin: Sigma #M2378-500G; dextrose: Fischer Scientific #S73418-1	
Brain Heart Infusion Agar	Brain Heart Infusion broth 18.5g, agar 7.5g	Brain Heart Infusion: Becton Dickinson #237500; agar: Becton Dickinson #214010	Supports variety of fastidious and non-fastidious aerobic and anaerobic microorganisms; targeting *Prevotella.* Commonly used to isolate *Helicobacter* species.

Media	**Reagents (500mL)**	**Reagent information**	**Notes**
Reasoner's 2A agar	Agar 7.5g, yeast extract 0.25g, acid hydrolysate casein 0.25g, dextrose 0.25g, soluble starch 0.25g, K_2HPO_4 0.15g, sodium pyruvate 0.15g, peptone 0.25g, $MgSO_4$ 0.012g.	Yeast extract: Gibco #212750; casein hydrolysate: Millipore #22090; K_2HPO_4: Fisher Chemical #CAS 7778-77-0; sodium pyruvate: Sigma #S8636, peptone: Gibco #211677	Low nutrient medium; to facilitate growth of slow growing bacteria that are outcompeted on rich media.
Fastidious Anaerobe Agar with Josamycin, Vancomycin, and Norfloxacin	FAA 23g. Additions 25mL (5%) blood, josamycin 3ug/mL, vancomycin 4ug/mL, norfloxacin 1ug/mL.	Josamycin: Sigma, #59983, vancomycin: Bio Basic Inc #VB0983; norfloxacin: ThermoFisher #J62652-03	Enriching media, selective for *Fusobacterium* and *Prevotella.*
Modified peptone-yeast extract glucose	Tryptone 2.5g, peptone 2.5g, yeast extract 5g, beef extract 2.5g, dextrose 2.5g, K_2HPO_4-$3H_2O$ 1g, Tween 80 500 uL, L-Cysteine-HCl 0.25g, Bacto-agar 7.5g. 20 mL Salt solution [$CaCl_2$-$2H_2O$ 0.125g, $MgSO_4$-$7H_2O$ 0.25g, K_2HPO_4 0.5g, KH_2PO_4 0.5g, $NaHCO_3$ 5g, NaCl 1g, water 500 mL], haemin 0.5 mL, vitamin K 0.1mL, Solution brought to pH to 7.2 with NaOH prior to autoclave.	Tryptone: Becton Dickinson #211705; beef extract: Becton Dickinson #212303; K_2HPO_4-$3H_2O$: Sigma #P5504; Tween 80: FisherScientific #9005-65-6; L-Cysteine-HCl: Alfa Aesar: #A10389.22; KH_2PO_4: Fisher Chemical #CAS 7758-11-4; hemin: Sigma #51280; vitamin K: Sigma #V3501	

Media	**Reagents (500mL)**	**Reagent information**	**Notes**
Gonococci agar	7.5g peptone, 7.5g agar, 2.5g NaCl, 2g K_2HPO_4, 0.5g KH2PO4, 0.5g corn starch. Additions: 25mL (5%) blood	Corn starch: Sigma-Aldrich #S4126	Peptone-rich. Targeting fastidious organisms.
Differential clostridial agar	Clostridial Differential Broth 13.75g, agar 7.5g	Clostridial Differential Broth: Sigma-Aldrich #27544	Targets *Clostridium.*
Brucella Agar	Brucellar agar with Hemin & Vitamin K1 21.55g Additions: 25mL (5%) defibrinated sheep's blood.	Brucellar agar with Hemin & Vitamin K1: Sigma-Aldrich #B2926	Targeting fastidious anaerobes, especially Bacteroides, *Prevotella*, and *Porphyromonas;* Commonly used to isolate *Helicobacter* species.
De Man, Rogosa and Sharpe agar	De Man, Rogosa and Sharpe 26g, agar 7.5g	De Man, Rogosa and Sharpe: Oxoid #CM0361	Targets *Lactobacillus.*

Isolations were performed using one sample at a time. Tissue was first homogenised into a tissue suspension. Specifically, GC samples were thawed to 20°C in an anaerobic chamber (Anaerobe Systems) and were cut into three smaller pieces, using a sterile blade in a sterile field, before mechanically homogenising using a glass tissue grinder (Wheaton, #885450-0024). Solid tissue was handled using sterile forceps. Up to 5mL of Tryptic Soy Broth (Becton Dickinson, #211825) supplemented with 5µg/mL Hemin (Sigma #5120-5G) and 1ug/mL Menadione (Alfa Aesar #L105575) (sTSB) was added to the tissue during homogenisation to aid the process.

The homogenised tissue suspension was transferred into a sterile 15mL conical tube and transferred under anaerobic conditions into a separate anaerobe chamber (Baker Ruskinn, Bugbox) for all downstream anaerobic work. The tissue suspension was diluted serially to 10^{-4} in sTSB. Dilutions 10^{-1} to 10^{-4} were plated on all media types anaerobically, with the exception of josamycin, vancomycin, norfloxacin (JVN) medium, since this contained antibiotics and therefore the GC homogenate was plated as neat suspension to 10^{-3}. To plate tissue suspension, 100µL of the required dilution was added to the appropriate agar plate, then distributed evenly using a sterile spreader.

Microaerobic conditions were created by using microaerobic sachets (GasPak EZ Campy Container System, Becton Dickinson, #260680) and a 2.5L gas container (Mitsubishi Gas Chemical Co, ThermoFisher Scientific, #R685025), kept at 37°C. The sachet was replaced with a fresh sachet every time the microaerobic container was opened, as per manufacturer's instructions. The sachets maintained a microaerobic condition by absorbing oxygen and generating carbon dioxide. Plates kept under microaerobic conditions were plated anaerobically and passed out of the anaerobic chamber (thereby briefly being exposed to aerobic conditions), before being placed into the microaerobic container. Plates were passed between microaerobic conditions and anaerobic conditions in this manner every time they were reviewed or handled.

Homogenised tissue suspension was passed out of the anaerobic chamber for aerobic work, which was all carried out in a biosafety cabinet to protect the researcher and maintain sterile conditions. Aerobic plating was performed as for anaerobic plates, although with a reduced number of conditions and dilutions, as detailed in Table 8.

A number of additional conditions were chosen to include in the isolation procedure to maximise the potential captured diversity. An ethanol shock was included to select for spore-forming microbes and a blood-rumen fluid incubation was added as a blood-enriching condition. The ethanol shock was performed by centrifuging 1mL 10^{-1} dilution tissue homogenate at 10000g for 10 minutes, removing the supernatant and resuspending in 100% ethanol for 1h, then centrifuging at 10000g for 10 minutes. The supernatant was then removed and the pellet was resuspended in sTSB, then plated onto Fastidious Anaerobe Agar (FAA) and Fastidious Anaerobe Agar without blood (FNB), anaerobically and aerobically. The blood-rumen fluid was prepared by adding 5mL rumen fluid (Fisher Scientific, #NC1530570) to 5mL sterile sheep blood (HemoStat laboratories, #DSB500) and allowing to equilibrate at 37°C under anaerobic conditions for 12 hours. The blood-rumen fluid incubation was performed by incubating 100µL of 10^{-1} dilution tissue homogenate in 10mLs blood-rumen fluid for 7 days, before plating onto FAA and FNB, anaerobically at 37°C.

Table 8. Media and conditions used to plate tissue homogenate for microbial isolations. A total of 93 agar plates were used to plate dilutions across 29 different conditions for each isolation. Values represent concentration of tissue suspension used for plating. BHI, brain heart infusion; BR: Brucella; BRF, blood-rumen fluid; DCM: Differential clostridial agar; FAA, Fastidious Anaerobe Agar; FCC, Fastidious Anaerobe Agar with complex carbohydrates; FNB, Fastidious Anaerobe Agar with no blood; GM, Gonococci medium; JVN, josamycin, vancomycin, norfloxacin media; MPYG, Modified peptone-yeast extract glucose; MRS, De Man, Rogosa, Sharpe agar; R2A, Reasoner's 2A agar.

Media	**Anaerobic**	**Aerobic**	**Microaerobic**
FAA	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, Ethanol shock at 10^{-1}, ethanol after 24 hr, 0.45µm filter at 10^{-1}, 10^{-4} sTSB incubation: 100µl; 10µl; 1µl BRF incubation: 10^{-3}, 10^{-2}, 10^{-1}	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, Ethanol shock at 10^{-1}	
FNB	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, Ethanol shock at 10^{-1}, BRF incubation: 10^{-3}, 10^{-2}, 10^{-1}	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, Ethanol shock at 10^{-1}	
GM	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}	10^{-2}, 10^{-1}	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}
BR	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}	10^{-2}, 10^{-1}	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}
FCC	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}	10^{-2}, 10^{-1}	
BHI	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}	10^{-2}, 10^{-1}	
R2A	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}	10^{-2}, 10^{-1}	
JVN	10^{-3}, 10^{-2}, $10^{-1,}$ neat	10^{-1}, neat	
MPYG	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}	10^{-2}, 10^{-1}	
DCM	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}	10^{-2}, 10^{-1}	
MRS	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}	10^{-2}, 10^{-1}	

All agar plates were reviewed every 1-2 days and unique colonies were picked by collecting a single colony using a sterile 1µm loop and spreading onto FAA patch plates, labelled into 16 areas, as demonstrated in Figure 22. Patch plates were incubated under the same conditions as the plates from which the colonies were picked from, before characterisation by Sanger sequencing.

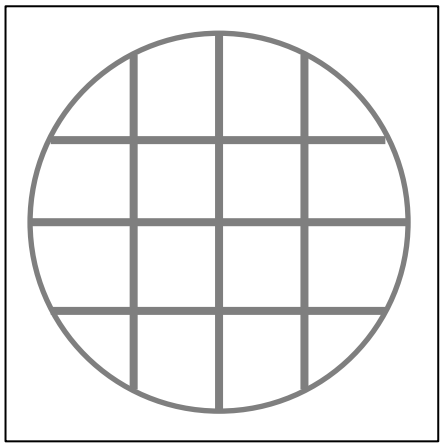


Figure 22. Patch plate layout used to streak individual colonies.

3.2.10.2 Isolate characterisation and strain library construction

Biomass from colonies was transferred into 50µL of 1x pH 8 TE buffer in 96 well plates and vortexed for 15s. Colony suspensions were boiled at 98°C for 5 minutes to disrupt cell membranes, before vortexing for a further 15s. The V3-V6 region of the 16S rRNA gene was targeted with V3kl (5'-TACGG[AG]AGGCAGCAG-3') and V6r (5'-AC[AG]ACACGAGCTGACGAC-3') primers [255] (Eurofins Genomics, Canada) and amplified as previously described [256].

Briefly, boiled colonies were placed in the thermocycler and the following PCR amplification steps were applied: 94°C for 2 minutes to denature; 30 cycles of 94˚C for 30s, 60˚C for 30s, 72˚C for 30s; then 72˚C for 5 minutes, before holding at 4°C. Amplification of products was confirmed by performing 1% agarose gel electrophoresis. Successfully amplified products were diluted 1:20 in high-performance liquid chromatography water and fluorescently tagged using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, ThermoFisher Scientific, #4337456). Since the V6r primer used in the amplification step contained the T7 tag, (5' to 3' direction), the T7 tag at the 3' end of the DNA amplification product allowed for the use of the T7 primer for fluorescent tagging in the reverse (3' to 5') direction. This step of fluorescent

tagging by the BigDye Terminator using the T7 forward primer (5'-TAATACGACTCACTATAGGG-3') involved 25 cycles of 96ºC for 5 minutes, 96ºC for 30s, 50ºC for 15s, and 60ºC for 2 minutes. Products of the BigDye tagging subsequently underwent Sanger sequencing performed by the University of Guelph Advanced Analysis Centre.

Sanger sequencing results were processed as described in section 3.2.11.3. Pure biomass of individual strains from the strain library was generated by growing pure isolates on whole agar plates then freezing at -80°C in a cryoprotectant media as previously described [257], including glycerol, dimethyl sulfoxide, and skimmed milk powder.

3.2.11 Bioinformatic and statistical analyses of microbial sequencing

3.2.11.1 Microbiome profiling of tissue sequencing

Raw reads from both 16S rRNA gene sequencing and low-coverage WGS were processed by Dr Henry Wood. For the low-coverage WGS data, this was performed using the GATK PathSeq algorithm, aligned against the default PathSeq microbial databases [168]. The PathSeq 'score' output was used for microbial sequencing reads. For the 16S rRNA sequencing, similar sequences were merged into OTUs using QIIME2 [258] (version 2021-2).

3.2.11.2 Decontamination of metagenomic sequencing

All downstream analysis was processed in R[169], using R Studio[170]. To account for potential laboratory contamination, a decontamination process was applied to the metagenomic analysis of the low-coverage WGS data of both fresh-frozen and FFPE tissue. The low sample size numbers and lack of corresponding blood samples of the collected tissue prevented a decontamination include-list being generated as described in Chapter 2. Instead, the include-list generated in Chapter 2 was applied to the metagenomic analysis of the low-coverage WGS data (include-list species described in (Supplemental) Table 19). In addition, the 50 most abundant species following this decontamination step were examined, and species identified as known contaminants or water inhabitants were removed before downstream analysis.

3.2.11.3 Isolate characterisation from Sanger sequencing of bacterial isolates

Raw sequencing reads were trimmed to remove low quality data at the ends of sequencing fragments and sequences with less than 400 base pairs or with a high amount of ambiguous bases were removed and the corresponding samples recharacterised.

Species identities were determined using NCBI BLAST (https://www.ncbi.nlm.nih.gov/BLAST/) upon trimmed sequences, using the top match as the identifier. Strain libraries were constructed based upon a workflow originally developed by Dr Sandi Yen (previous PhD student in the Allen-Vercoe laboratory) [259], in order to include any isolates with unique species classification. Where multiple isolates were characterised as the same species, a sequence alignment was performed using DECIPHER [260] upon all isolates in the strain library and the 16S V3-V4 sequences of possible new strains were manually reviewed; unique strains were added to the strain library.

3.2.11.4 Data handling and statistical analyses

Statistical analysis and figure creation throughout this chapter were conducted in R v4.3.0 [169], within R Studio v2023.06.0+421 [170], using stringr version 1.5.0 [171], dplyr v 1.1.2 [172], pheatmap v1.0.12 [175], reshape2 v1.4.4 [261] ggplot2 v 4.3.2 [176] packages. Alpha diversity was measured via Shannon index through the vegan v2.6-4 [174], package.

The results from the low-coverage WGS of fresh-frozen tissue and FFPE tissue were used to explore the differences in sequencing of these two tissue types. The most abundant genera were compared between the two processing methodologies. To determine the degree to which the microbial composition differed by processing type, PERMANOVA using Bray-Curtis dissimilarity index [178], was used to analyse beta diversity in species between subgroups (fresh-frozen versus FFPE). In addition, beta diversity was graphically displayed through principal coordinate analysis plots, using the ape v5.7-1 [262] package.

To compare the results from 16S rRNA sequencing with low-coverage WGS, heatmaps of most abundant genera were created. The most abundant bacterial

genera were compared from each methodology. For this comparison, viruses were excluded since 16S rRNA sequencing does not sequence viruses.

The bacterial strains isolated from the five GC samples were described and the proportion of taxa at both species and genus level from all five samples were displayed in bar charts. Cladograms were created at order level to demonstrate phylogenetic relationships between isolated species and to what degree taxa of the same order where isolated from the different five GC samples. The following R packages were used to create the cladogram of cultured bacterial isolates: Biostrings v2.68.1[263], msa v1.32.0 [264], ape v5.7-1 [262], phyloseq v1.44.0 [265], tidytree v0.4.2 [266], ggtreeExtra v1.10.0 [266], ggnewscale v0.4.9 [267].

Genera detected through microbial isolations were compared to those detected through 16S rRNA sequencing of DNA extracted from GC tissue. A threshold of five or more reads was set for the 16S rRNA sequencing to determine presence (versus absence) to account for read errors and to reduce the likelihood of false positives within the 16S rRNA sequencing data. Venn diagrams were created to display the overlap of bacterial at genus levels, between the two methodologies.

3.3 Results

3.3.1 Ethical and local approvals

The ethical approval for IRAS project ID 290019 was reviewed by the North West-Greater Manchester East REC on 16th February 2021, and the provisional opinion was given, accompanied by a number of required action points, which were actioned as appropriate.

A favourable ethical opinion was issued by the North West-Greater Manchester East REC and HRA Approval was confirmed on 15th April 2021. LTHT Capacity and Capability was confirmed by the Research and Innovation Team for study number GS21/139713 on 24h June 2021.

Secondary site approval was granted by the University of Guelph Research Ethics Board on 13th May 2022 to enable the investigations planned within the Allen-Vercoe laboratory in the University of Guelph, Canada.

The ethical approval documents can be found in Appendix A.

3.3.2 Sample collection & histological assessment

Samples from seven patients were collected over a period of eight months. These comprised of samples from six patients undergoing a gastrectomy or oesophagogastrectomy and one set of biopsy samples from a patient undergoing a staging laparoscopy with upper GI endoscopy.

The H&E stained FFPE slides of the samples taken for research purposes were analysed. The overview of the samples included in this study and the histology of the study FFPE H&E slides are shown in Table 8. Of note, no tumour was observed in the study H&E slides of UKG2 and UKG4 tumour samples, and some small clusters of tumour cells were observed in the UKG5 adjacent sample. As a result, subsequent bacterial isolation work focussed on the tumour samples with confirmed tumour (i.e. UKG1, UKG3, UKG5, UKG6, and UKG7).

Table 9. Overview of samples collected from Leeds Teaching Hospitals patients. PPI use and antibiotic use refers to previous 28 days. GOJ, gastro-oesophageal junction; H&E, haematoxylin and eosin; MSI, microsatellite instability; MSS, microsatellite unstable; PPI, proton pump inhibitor.

Sample ID	Sampled area	Histology according to H&E from research samples	Tumour site	Method of Sampling	PPI use	Antibiotic use	Prior chemotherapy	MSI status
UKG1	Adjacent	Normal tissue with mild inflammation, no atrophy, no intestinal metaplasia.	Cardia	Gastrectomy	Yes	No	No	MSI
	Tumour	Adenocarcinoma.						
UKG2	Adjacent	Normal tissue. Inflamed, no atrophy, no intestinal metaplasia.	Gastric antrum	Staging laparoscopy	Yes	No	No	MSS
	Tumour	Normal tissue with appearance of Brunner's glands, no other epithelium.						
UKG3	Adjacent	Normal tissue, gastric body, no atrophy, no intestinal metaplasia, no inflammation.	Gastric body	Gastrectomy	Yes	No	Yes	MSS
	Tumour	Adenocarcinoma, lots of lymphocytes and plasma cells present.						
UKG4	Adjacent	Normal tissue, no atrophy, no intestinal metaplasia.	GOJ	Oesophagectomy	Yes	No	Yes	MSS
	Tumour	Deep tissue sample from body mucosa but including submucosa and muscle. No bulk of tumour seen.						
UKG5	Adjacent	Normal body type mucosa with deep submucosa and muscularis propria, very few small clusters of tumour cells in the stroma.	GOJ	Oesophagectomy	Yes	No	No	MSS
	Tumour	Adenocarcinoma with ulceration all around the tissue piece, tubular type, well differentiated, lots of inflammation in the stroma.						
UKG6	Adjacent	Lots of fat with large vessels only, no obvious tumour cells.	Gastric body	Gastrectomy	Yes	No	No	MSS
	Tumour	Adenocarcinoma with some few normal glands.						
UKG7	Adjacent	Normal tissue with lot of lymphocytic inflammation and focal intestinal metaplasia	Pylorus	Gastrectomy	Yes	No	Yes	MSS
	Tumour	Appearance of normal duodenum (Brunner's glands) with adenocarcinoma (overall low tumour content)						

3.3.3 Decontamination of metagenomic sequencing

Prior to decontamination, the mean read count per sample was 1404 for fresh-frozen and 3897 for FFPE tissue. Following initial decontamination process using the include-list, the mean total read count per sample was 1292 for fresh and 226 for FFPE, demonstrating that 8% of reads were removed from fresh samples and 94% of reads were removed from FFPE processed samples during this decontamination step.

Prior to decontamination, the most five common species from fresh-frozen tissue were Human gammaherpesvirus 4 (EBV), *Mycobacterium species 1100029.7*, *Alloprevotella tannerae*, *Capnocytophaga sp. oral taxon_329*, and *Selenomonas sp. CM52*. The five most common species from FFPE tissue prior to decontamination were *Burkholderia sp. OLGA172*, *Caballeronia udeis*, *Burkholderia sp. AU4i*, *Sphingopyxis alaskensis*, and *Bradyrhizobium sp. CCH5.F6*.

Following this initial decontamination process, the 50 most abundant species across fresh and FFPE samples were manually reviewed and 8 species were identified as being contaminants. These species were: *Cloacibacterium normanense*, *Acidovorax sp. KKS102*, *Moraxella osloensis*, *Enhydrobacter aerosaccus*, *Enhydrobacter sp. H5*, *Tepidimonas fonticaldi*, *Xanthomonas campestris*, and *Achromobacter sp*. Following this second decontamination step, the mean total read count per sample was 1291 for fresh and 174 for FFPE, demonstrating that 8% of reads were removed from fresh samples and 96% of reads were removed from FFPE processed samples by the complete decontamination process.

Following the full decontamination process, the most five common species were Human gammaherpesvirus 4 (EBV), *Alloprevotella tannerae*, *Capnocytophaga sp. oral taxon 329, Selenomonas sp. CM52*, and *Selenomonas sputigena* from fresh-frozen tissue. *Prevotella sp.* ICM33, Leptotrichia wadei, Staphylococcus warneri, Prevotella *melaninogenica*, and *Leptotrichia shahii* were the five most common species from FFPE tissue.

3.3.4 Microbial composition of fresh-frozen gastric cancer samples using low-coverage whole genome sequencing

When all 14 fresh-frozen samples of seven paired GC and adjacent normal were sequenced by low-coverage WGS, the most common genus across the samples was *Lymphocryptovirus*, in fact accounting for 84% of all reads. This was driven almost entirely by one sample (UKG3 tumour); *Lymphocryptovirus* reads from this sample accounted for 83% of all reads across all 14 samples.

Across all samples, the ten most common bacterial genera were *Prevotella*, *Selenomonas*, *Streptococcus*, *Alloprevotella*, *Capnocytophaga*, *Rothia*, *Porphyromonas*, *Mitsuokella*, *Actinomyces*, and *Helicobacter*, in decreasing prevalence.

3.3.5 Alpha diversity of fresh-frozen gastric cancer samples using low-coverage whole genome sequencing

As seen in Figure 23, there was a spread of alpha diversity, as measured by Shannon index of low-coverage WGS of fresh samples, although limited interpretations can be made of these results due to the small sample size.

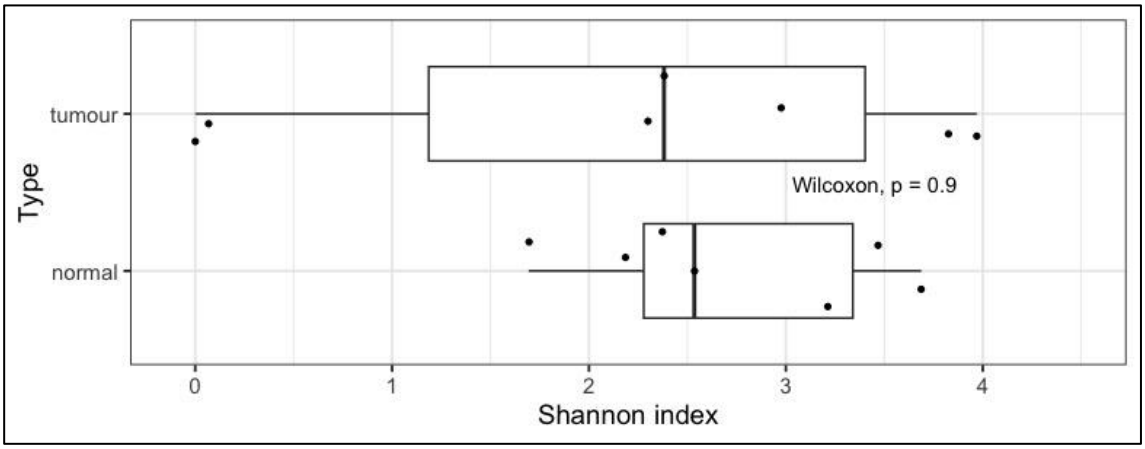


Figure 23. Boxplot of alpha diversity (Shannon index) of GC samples sequenced by low-coverage WGS analysis of fresh-frozen tissue from 7 pairs of matched GC samples collected from Leeds patients. A higher Shannon index indicates higher intra-sample diversity.

3.3.6 Comparison between metagenomic analysis of fresh-frozen and formalin-fixed, paraffin-embedded gastric cancer samples using low-coverage whole genome sequencing

To compare the results of the difference between low-coverage WGS of fresh-frozen tissue versus FFPE, first heatmaps were created to visually demonstrate the relative abundances of the most common microbiota. All work involved use of the decontaminated sequencing databases.

The 40 most abundant species were selected to balance clarity of visualisation with representation of the largest proportion of signal. The 40 most abundant species within the fresh-frozen cohort (Figure 24) represented 95% of all species reads within the sequenced samples. The 40 most abundant species within the FFPE cohort (Figure 25) represented only 64% of all species reads within the sequenced samples. This discrepancy is likely a result of the much lower read counts within the FFPE cohort following decontamination which is thus subject to noise.

The heatmap illustrating the microbiome of fresh-frozen GC tissue by low-coverage WGS (Figure 24) indicates that the both the tumour and adjacent normal tissue from UKG5 contained a more diverse microbiome than other samples. The heatmap demonstrating the microbiome of FFPE GC tissue by low-coverage WGS (Figure 25) also indicates this. Demonstrated in both heatmaps, when *H. pylori* was present (UKG7 tumour and adjacent normal; UKG2 adjacent normal), it tended to dominate the microbiota, although this is more obvious in the fresh-frozen cohort than the FFPE cohort. *H. pylori* was detected in the same samples via sequencing of both fresh-frozen and FFPE.

Of note, EBV, annotated as “Human_gammaherpesvirus_4” was dominant in the fresh-frozen samples, but did not appear in the ten most abundant species from the FFPE sequencing.

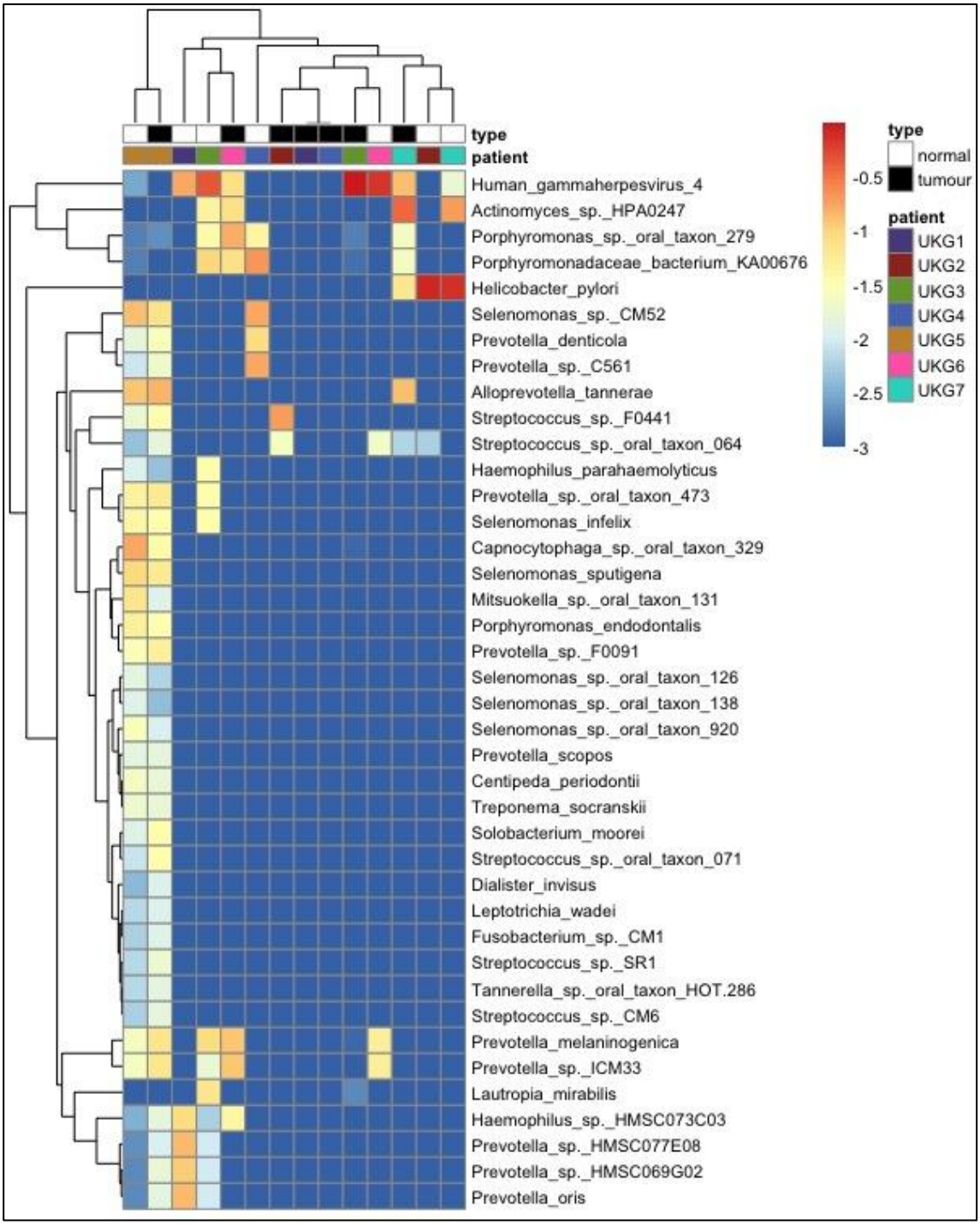


Figure 24. Heatmap of microbial abundance by low-coverage WGS analysis of fresh-frozen tissue from 7 pairs of matched GC samples collected from Leeds patients, showing 40 most abundant species. Cells are coloured according to proportion of microbial reads per sample of species (in log scale) from red (highest) to blue (lowest). Branching on the left signifies clustering species; branching at the top represents samples with similar compositions of species.

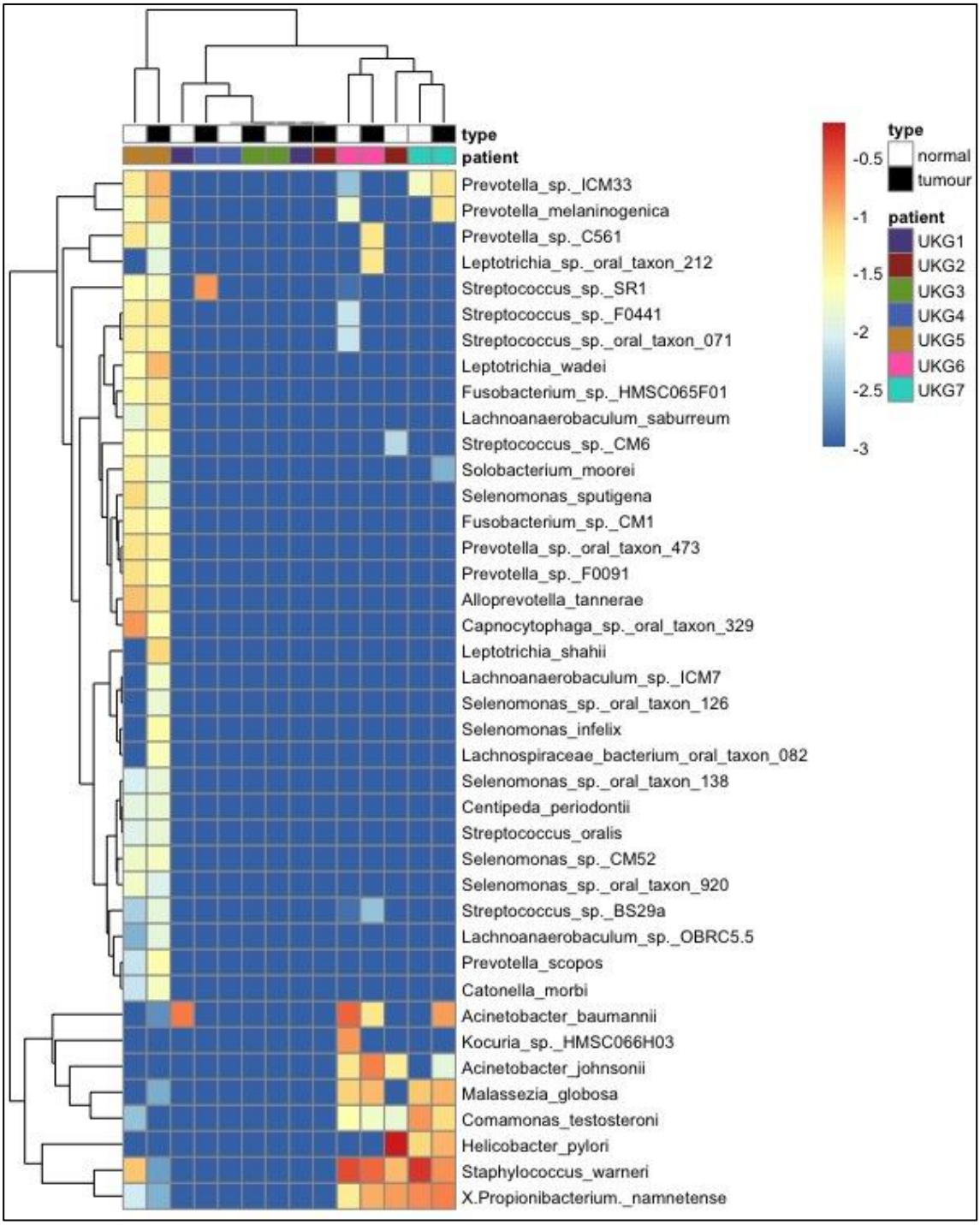


Figure 25. Heatmap of microbial abundance by low-coverage WGS analysis of FFPE tissue from 7 pairs of matched GC samples collected from Leeds patients, showing 40 most abundant species. Cells are coloured according to proportion of microbial reads per sample of species (in log scale) from red (highest) to blue (lowest). Branching on the left signifies clustering species; branching at the top represents samples with similar compositions of species.

Table 10 lists the ten most abundant genera after low-coverage WGS of fresh-frozen and FFPE of the GC samples. *Lymphocryptovirus*, the genus to which EBV belongs was the most common genus detected via sequencing of fresh-frozen tissue, however this was not present in the ten most abundant genera from sequencing of FFPE tissue. There was some overlap between the sequencing of the two different tissue methodologies, with four genera (*Prevotella*, *Streptococcus*, *Selenomonas*, and *Alloprevotella*) appearing in the ten most abundant genera for both tissue types.

Table 10. Table showing ten most abundant genera from seven pairs of matched Leeds samples by low-coverage WGS of fresh-frozen and FFPE tissue in descending order. The four genera found in the top ten bacterial genera across both sequencing methodologies are shown in blue. FFPE, formalin-fixed, paraffin-embedded.

Fresh-frozen	**FFPE**
Lymphocryptovirus	*Prevotella*
Prevotella	*Streptococcus*
Selenomonas	*Leptotrichia*
Streptococcus	*Selenomonas*
Alloprevotella	*Staphylococcus*
Capnocytophaga	*Fusobacterium*
Rothia	*Lachnoanaerobaculum*
Porphyromonas	*Acinetobacter*
Mitsuokella	*Alloprevotella*
Actinomyces	*Haemophilus*

To determine what proportion of beta diversity was attributable to type of processing, PERMANOVA analysis was performed across all low-coverage WGS samples, including fresh-frozen and FFPE. To allow this analysis, samples without any sequencing reads had to be removed. Following the decontamination process, all five samples without any species present were from the FFPE processed samples: UKG1 tumour, UKG2 tumour, UKG3 normal, UKG3 tumour, and UKG4 normal.

Table 11. PERMANOVA of low-coverage WGS from normal and adjacent GC tissue from FFPE and fresh-frozen samples. R2, the proportion of the variance in the microbiome associated with each variable; * statistically significant ($p < 0.05$).

	R^2	P value
Processing	0.05684	0.0469*
Patient	0.49005	0.0001*
Type	0.04189	0.1578
Residuals	0.41123	

The PERMANOVA analysis demonstrated statistically significant associations between both the processing (fresh-frozen versus FFPE) and patient, with the microbial composition. The R^2 values demonstrate that 49% of the variation in microbial abundance was associated with inter-individual patients, with only 5.7% of variation associated with processing differences. The beta diversity of the same set of samples (i.e. those without any species present removed) is represented visually in Figure 26, which shows that samples appear to cluster by patient (colour) more than processing method or type (tumour versus adjacent normal).

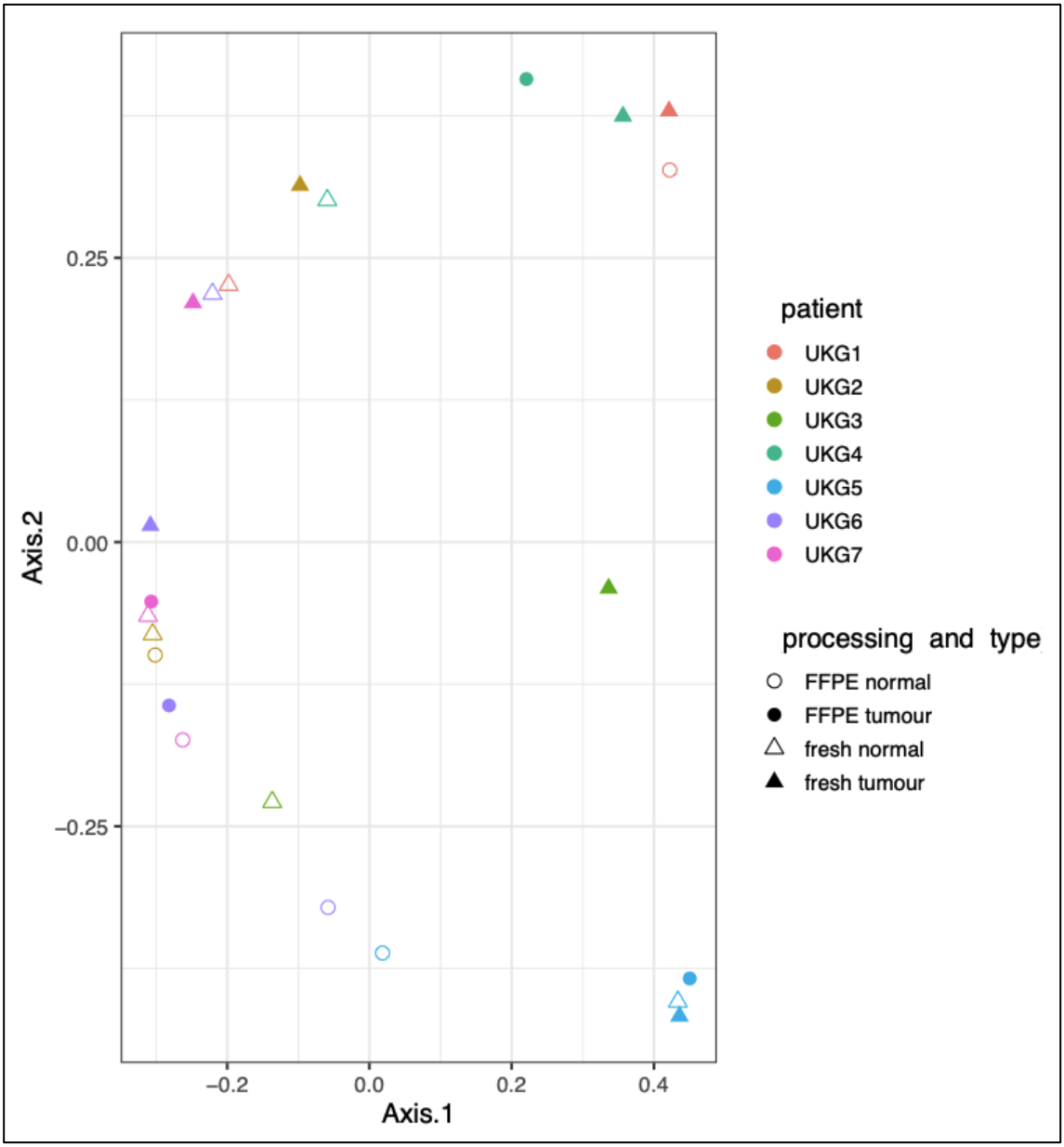


Figure 26. Principal coordinate analysis plot showing compositional differences by Bray-Curtis distances of species by method of processing and individual patient. Shapes in the figure legend correspond to the processing (fresh-frozen versus FFPE) and type (tumour versus adjacent normal); different patients are represented by different colours.

3.3.7 Comparison of 16S versus low-coverage whole genome sequencing.

Figure 27 shows a heatmap of the most abundant 25 genera detected in the GC samples following the 16S rRNA gene sequencing. Figure 28 shows a heatmap of the most abundant 25 genera detected in the GC samples following the 16S rRNA gene sequencing.

Helicobacter was identified in both UKG7 samples with both sequencing methods. 16S rRNA sequencing identified *Helicobacter* in both UKG2 samples, whereas low-coverage WGS identified *Helicobacter* in only the normal UKG2 sample, not the tumour sample. In contrast, *Rothia* was only detected in four samples via 16S rRNA sequencing, but in eight samples via low-coverage WGS. *Fusobacterium* was detected in four samples via both methodologies, although only three of these four were in the same samples.

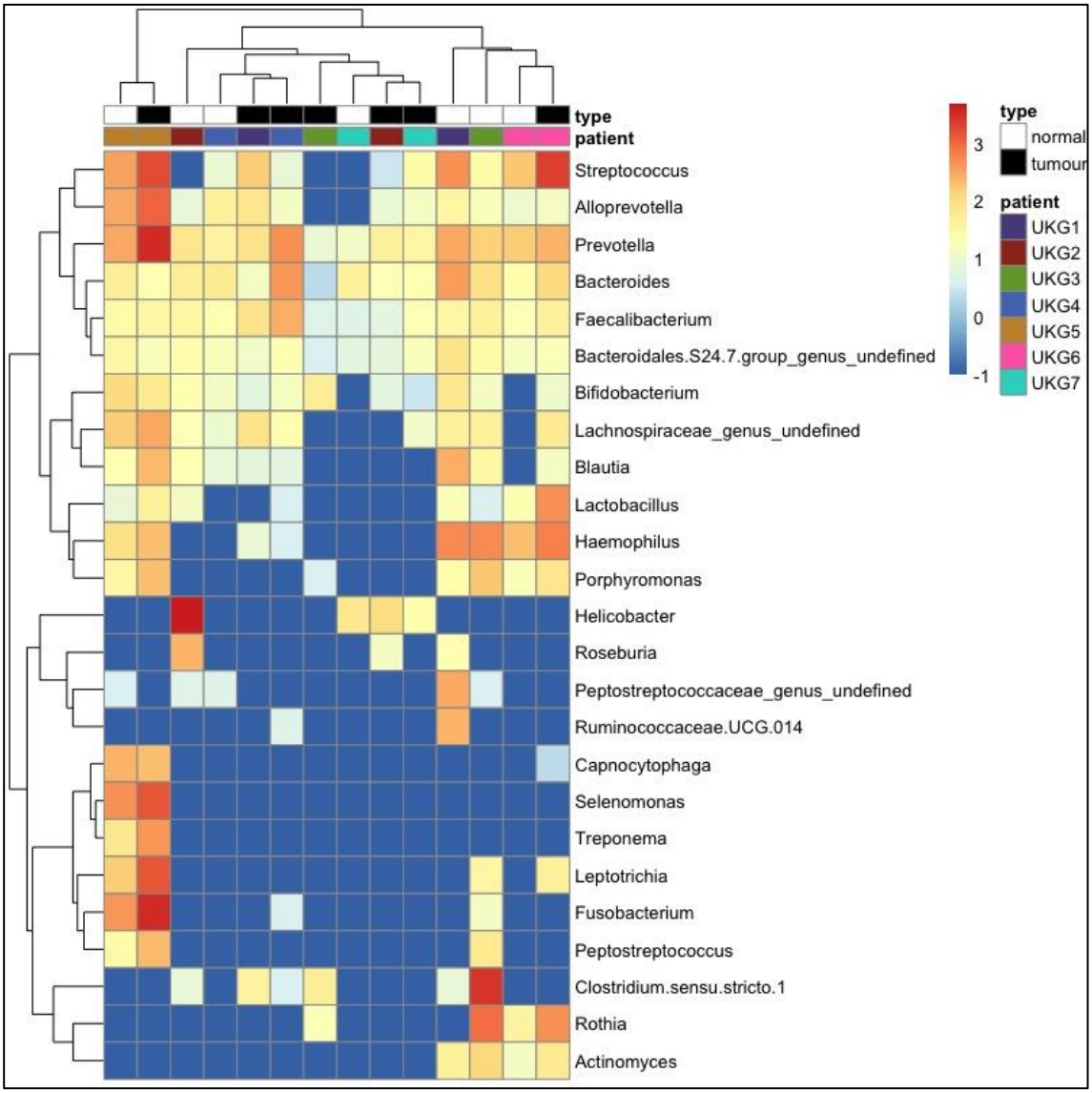


Figure 27. Heatmap of relative microbial abundance by 16S rRNA analysis in 7 pairs of matched GC samples collected from Leeds patients, showing 25 most abundant genera. Cells are coloured according to proportion of microbial reads per sample of genera (in log scale) from red (highest) to blue (lowest). Branching on the left signifies clustering genera; branching at the top represents samples with similar compositions of genera.

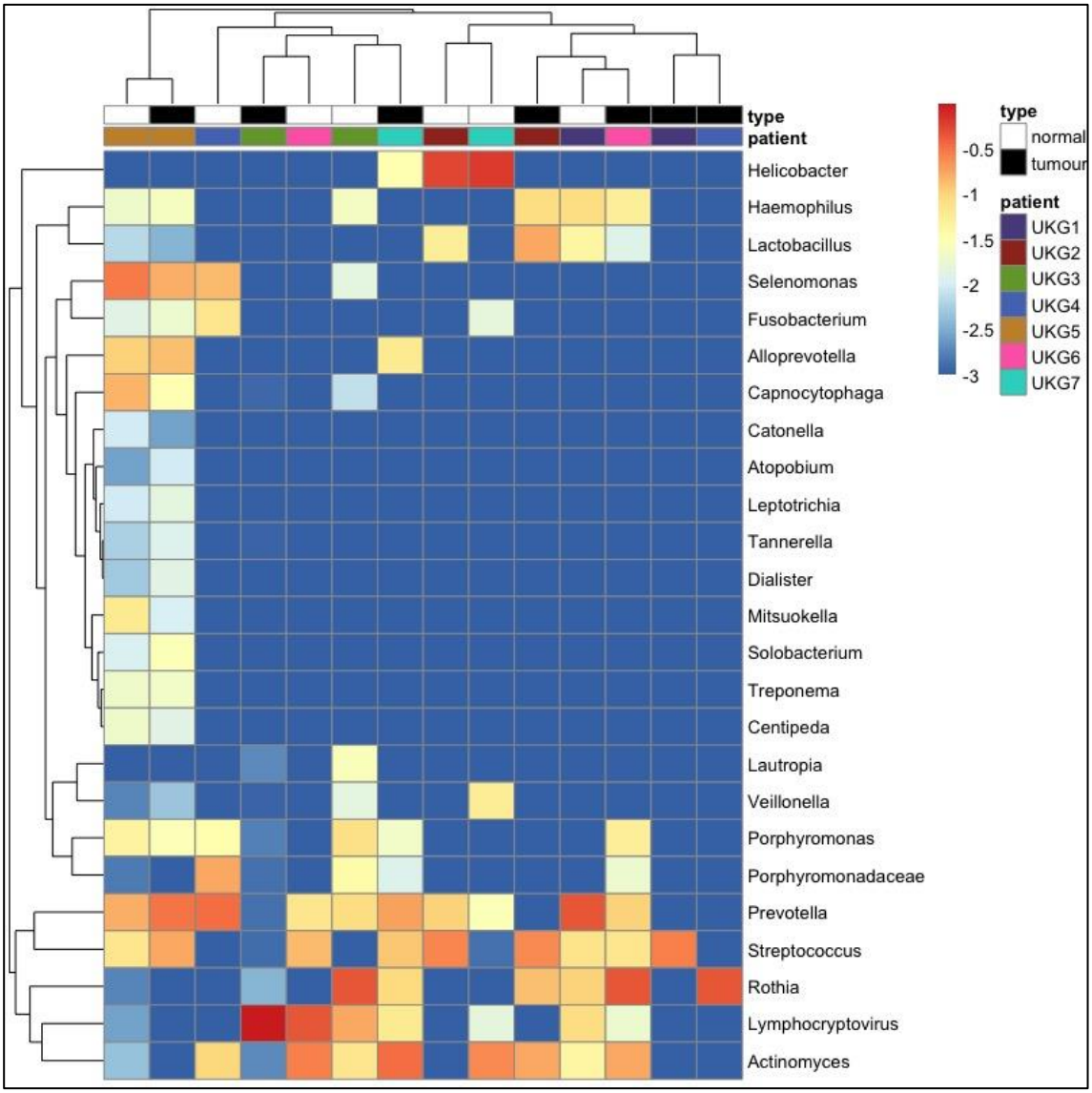


Figure 28. Heatmap of relative abundance by low-coverage WGS in 7 pairs of matched GC samples collected from Leeds patients, showing 25 most abundant genera following decontamination process. Cells are coloured according to proportion of microbial reads per sample of genera (in log scale) from red (highest) to blue (lowest). Branching on the left signifies clustering genera; branching at the top represents samples with similar compositions of genera.

Lymphocryptovirus, the genus to which EBV belongs to, was the most abundant genus within the low-coverage WGS analysis. Since 16S rRNA sequencing does not detect viruses, the most abundant ten bacterial genera were compared between 16S rRNA sequencing and low-coverage WGS. The ten most abundant genera for each sequencing methodology are shown in Table 12. Six genera were common to the top ten bacterial genera detected via both methods of sequencing.

Table 12. Table showing ten most abundant bacterial genera from seven pairs of matched Leeds samples by 16S rRNA sequencing and low-coverage WGS in descending order. The six genera found in the most abundant ten bacterial genera across both sequencing methodologies are shown in blue.

16S rRNA sequencing	**Low-coverage WGS**
Streptococcus	*Prevotella*
Prevotella	*Selenomonas*
Helicobacter	*Streptococcus*
Fusobacterium	*Alloprevotella*
Clostridium.sensu.stricto.1	*Capnocytophaga*
Haemophilus	*Rothia*
Selenomonas	*Porphyromonas*
Alloprevotella	*Mitsuokella*
Leptotrichia	*Actinomyces*
Rothia	*Helicobacter*

3.3.8 Bacterial isolations

3.3.8.1 Description of bacterial strains isolated from five gastric cancer samples.

Bacterial isolations were performed from all five GC samples taken from the tumour area, with confirmed tumour by H&E. 3, 20, 36, 49, and 5 unique bacterial strains were isolated from UKG1, UKG3, UKG5, UKG6, and UKG7, tumour samples, respectively. Collectively across the five samples, the strains represented 52 species from 26 genera, 18 families, from 4 phyla. The full list of strains identified from all five samples is detailed in Table 13.

Table 13. Bacterial strains isolated from five gastric cancer tumour samples. Species of closest match and % identity (ID) are based on the closest match from NCBI BLAST alignments. Strain name indicates which sample, media type, and condition the isolate was obtained from, e.g. UKG6_T2_FNB_8_AER: UKG6 patient, tumour sample 2, 8th colony isolated from FAA with no blood, under aerobic conditions. Seq, sequence; AER, aerobic; AN, anaerobic; MA, microaerobic.

Phylum	**Family**	***Genus***	**Species of closest match**	**Strain**	**Seq. length**	**ID (%)**
Actinomycetota	*Actinomycetaceae*	*Actinomyces*	*Actinomyces oris*	UKG6_T2_FNB_8_AER	693	99.85
			Actinomyces radicidentis	UKG6_T2_FNB_14_AER	695	99.56
			Actinomyces viscosus	UKG6_T2_BR_15_MA	601	100
		Schaalia	*Schaalia odontolytica*	UKG3_GC_5_MA	682	99.55
				UKG3_T2_BR_5_MA	578	97.75
				UKG3_T2_FAA_3_AN	682	99.12
				UKG3_T2_GC_1_MA	640	98.12
	Micrococcaceae	*Micrococcus*	*Micrococcus aloeverae*	UKG1_GC_1_MA	686	100
			Micrococcus yunnanensis	UKG3_GC_7_MA	668	99.85
		Rothia	*Rothia dentocariosa*	UKG6_T2_BR_1_AER	634	100
				UKG6_T2_FNB_15_AER	683	99.55
				UKG7_T2_BR_1_MA	672	99.85
			Rothia mucilaginosa	UKG3_T2_BR_1_MA	646	99.68
				UKG3_T2_FNB_2_AER	643	99.37
				UKG5_T2_FAA_16_AER	682	99.71
				UKG6_T2_BR_8_MA	663	99.54
				UKG6_T2_FAA_21_AER	670	99.55
				UKG7_BR_1_AER	684	99.56
	Propionibacteriaceae	*Cutibacterium*	*Cutibacterium acnes*	UKG5_T2_045_13_AN	676	99.85
				UKG7_BR_4_AN	662	100
	Bifidobacteriaceae	*Alloscardovia*	*Alloscardovia omnicolens*	UKG5_T2_FAA_14_AN	646	99.37

Phylum	**Family**	***Genus***	**Species of closest match**	**Strain**	**Seq. length**	**ID (%)**
		Bifidobacterium	*Bifidobacterium dentium*	UKG5_T2_BHI_7_AN	677	99.55
				UKG5_T2_FCC_6_AN	631	96.83
Bacillota	*Carnobacteriaceae*	*Granulicatella*	*Granulicatella adiacens*	UKG1_T2_GC_2_MA	673	99.24
				UKG3_T2_FAA_1_AN	688	99.41
				UKG5_T2_DCM_9_AN	690	99.41
				UKG5_T2_GC_10_AN	677	99.56
				UKG6_T2_FAA_7_AN	703	99.57
	Clostridiaceae	*Clostridium*	*Clostridium beijerinckii*	UKG3_T2_FNB_1_AN	671	100
	Lachnospiraceae	*Enterocloster*	*Enterocloster bolteae*	UKG6_T2_BR_12_MA	608	99.33
		Lacrimispora	*Lacrimispora xylanolytica*	UKG5_T2_BR_13_A_AN	669	94.93
				UKG5_T2_FAA_13_AN	674	94.62
		Lacticaseibacillus	*Lacticaseibacillus rhamnosus*	UKG6_T2_BHI_4_AN	701	100
				UKG6_T2_FCC_20_AN	686	99.4
		Lactobacillus	*Lactobacillus delbrueckii*	UKG6_T2_BR_3_AN	701	100
			Lactobacillus gasseri	UKG5_T2_BHI_2_AN	706	99.71
				UKG5_T2_BR_2_AER	675	98.79
				UKG5_T2_FAA_4_AN	630	97.56
				UKG5_T2_MRS_9_AER	672	99.26
			Lactobacillus paragasseri	UKG5_T2_FCC_9_AER	691	99.85
				UKG5_T2_FNB_15_AER	691	99.85
		Limosilactobacillus	*Limosilactobacillus fermentum*	UKG5_T2_FCC_37_AN	667	99.24
				UKG5_T2_FNB_8_AER	698	98.99
				UKG5_T2_R2A_8_AN	596	99.5
				UKG6_T2_DCM_12_AN	670	89.69
				UKG6_T2_FAA_3_AER	671	98.8

Phylum	Family	*Genus*	Species of closest match	Strain	Seq. length	ID (%)
				UKG6_T2_FCC_2_AER	600	98.96
				UKG6_T2_FCC_3_AN	647	98.9
				UKG6_T2_FCC_7_AN	701	100
				UKG6_T2_FCC_8_AN	674	98.96
				UKG6_T2_R2A_8_AN	673	98.96
	Leuconostocaceae	*Weissella*	*Weissella confusa*	UKG6_T2_BHI_6_AN	701	100
				UKG6_T2_FCC_1_AN	710	99.86
				UKG6_T2_FCC_21_B_AN	666	98.79
	Oscillospiraceae	*Ruminococcus*	*[Ruminococcus] gnavus*	UKG6_T2_FNB_13_AN	581	97.18
	Paenibacillaceae	*Paenibacillus*	*Paenibacillus oralis*	UKG5_T2_ETOH_FNB_1_AER	700	98
	Peptoniphilaceae	*Parvimonas*	*Parvimonas micra*	UKG5_T2_BRF_FAA_3_AN	660	100
	Staphylococcaceae	*Staphylococcus*	*Staphylococcus caprae*	UKG7_BR_1_AN	717	100
			Staphylococcus haemolyticus	UKG7_FAA_2_AN	709	99.86
		Streptococcus	*Streptococcus anginosus*	UKG5_T2_BR_6_AN	670	100
			Streptococcus constellatus	UKG3_T2_GC_2_AN	719	99.72
				UKG3_T2_sTSB_1_AN	697	99.41
			Streptococcus downii	UKG5_T2_BHI_1_AN	688	99.85
			Streptococcus gordonii	UKG3_T2_FAA_7_AN	695	100
				UKG6_T2_R2A_3_AN	694	100
			Streptococcus intermedius	UKG5_T2_FCC_14_AN	713	99.15
			Streptococcus koreensis	UKG6_T2_DCM_9_AN	694	99.42
				UKG6_T2_FCC_5_AN	707	99.29
			Streptococcus lutetiensis	UKG6_T2_FAA_1_AER	708	100
				UKG6_T2_sTSB_1_AN	703	98.57
			Streptococcus mitis	UKG5_T2_BHI_4_AN	703	100

Phylum	Family	*Genus*	Species of closest match	Strain	Seq. length	ID (%)
				UKG5_T2_GC_12_AN	700	100
			Streptococcus mutans	UKG6_T2_GC_4_AN	719	100
			Streptococcus oralis	UKG5_T2_FAA_33_AN	705	99.86
				UKG5_T2_FNB_9_AN	715	100
				UKG6_T2_BR_4_AN	694	99.71
				UKG6_T2_BR_5_MA	658	99.23
				UKG6_T2_FAA_12_AN	694	99.86
				UKG6_T2_FNB_2_AN	703	100
				UKG6_T2_GC_5_MA	702	99.71
			Streptococcus parasanguinis	UKG6_T2_GC_4_MA	716	99.43
			Streptococcus periodonticum	UKG5_T2_BHI_9_AN	668	99.69
				UKG5_T2_BR_5_AN	616	97.56
				UKG5_T2_MPYG_5_AN	655	99.84
				UKG5_T2_R2A_5_AN	504	90.34
				UKG6_T2_BRF_FNB_8_AN	542	99.62
			Streptococcus rubneri	UKG3_T2_BR_2_AER	653	94.3
				UKG3_T2_GC_1_AN	709	99.71
				UKG6_T2_FCC_2_AN	703	100
			Streptococcus salivarius	UKG5_T2_BR_1_AN	671	99.85
				UKG6_T2_BHI_1_AN	708	99.71
				UKG6_T2_FAA_4_AER	694	100
				UKG3_T2_BHI_3_AER	701	99.85
				UKG3_T2_BR_3_AER	694	99.86
				UKG6_T2_BHI_11_AN	693	100
				UKG6_T2_DCM_10_AN	655	89.45

Phylum	**Family**	***Genus***	**Species of closest match**	**Strain**	**Seq. length**	**ID (%)**
			Streptococcus symci	UKG3_T2_BHI_2_AN	708	99.71
			Streptococcus timonensis	UKG3_T2_BR_5_AN	717	99.86
				UKG3_T2_DCM_4_AN	692	99.86
			Streptococcus vicugnae	UKG5_T2_STSB_7_AN	694	100
				UKG6_T2_BRF_FAA_10a_AN	704	100
			Streptococcus vulneris	UKG1_T2_BRF_FNB_2_AN	687	100
				UKG3_T2_R2A_2_AER	669	100
				UKG5_T2_DCM_1_AN	675	100
	Veillonellaceae	*Veillonella*	*Veillonella atypica*	UKG6_T2_DCM_11_AN	685	99.12
				UKG6_T2_BRF_FAA_3_AN	677	98.52
Bacteroidota	*Flavobacteriaceae*	*Capnocytophaga*	*Capnocytophaga gingivalis*	UKG6_T2_FNB_12_AN	681	92.36
			Capnocytophaga granulosa	UKG6_T2_FNB_15_AN	673	99.4
	Prevotellaceae	*Prevotella*	*Prevotella corporis*	UKG6_T2_BR_1_MA	681	99.41
			Prevotella denticola	UKG5_T2_FNB_17_AN	715	99.43
	Rikenellaceae	*Alistipes*	*Alistipes senegalensis*	UKG6_T2_BR_4_MA	677	99.85
	Tannerellaceae	*Parabacteroides*	*Parabacteroides merdae*	UKG6_T2_BR_11_MA	702	99.48
Mycoplasmatota	*Metamycoplasmataceae*	*Metamycoplasma*	*Metamycoplasma salivarium*	UKG5_T2_045_3_AN	677	99.85

Figure 29 and Figure 30 demonstrate the proportion of bacterial taxa identified in each of the five GC samples from which isolations were performed at species and genus level, respectively. As demonstrated in these bar charts, a much greater alpha diversity was observed in samples UKG5 and UKG6 than other samples. *Streptococcus* is the genus which isolated strains most commonly belonged to; strains from *Limosilactobacillus*, *Rothia*, and *Lactobacillus* were also commonly isolated across the samples. At species level, the majority of bacteria identified are known inhabitants of the oral and gastrointestinal tract.

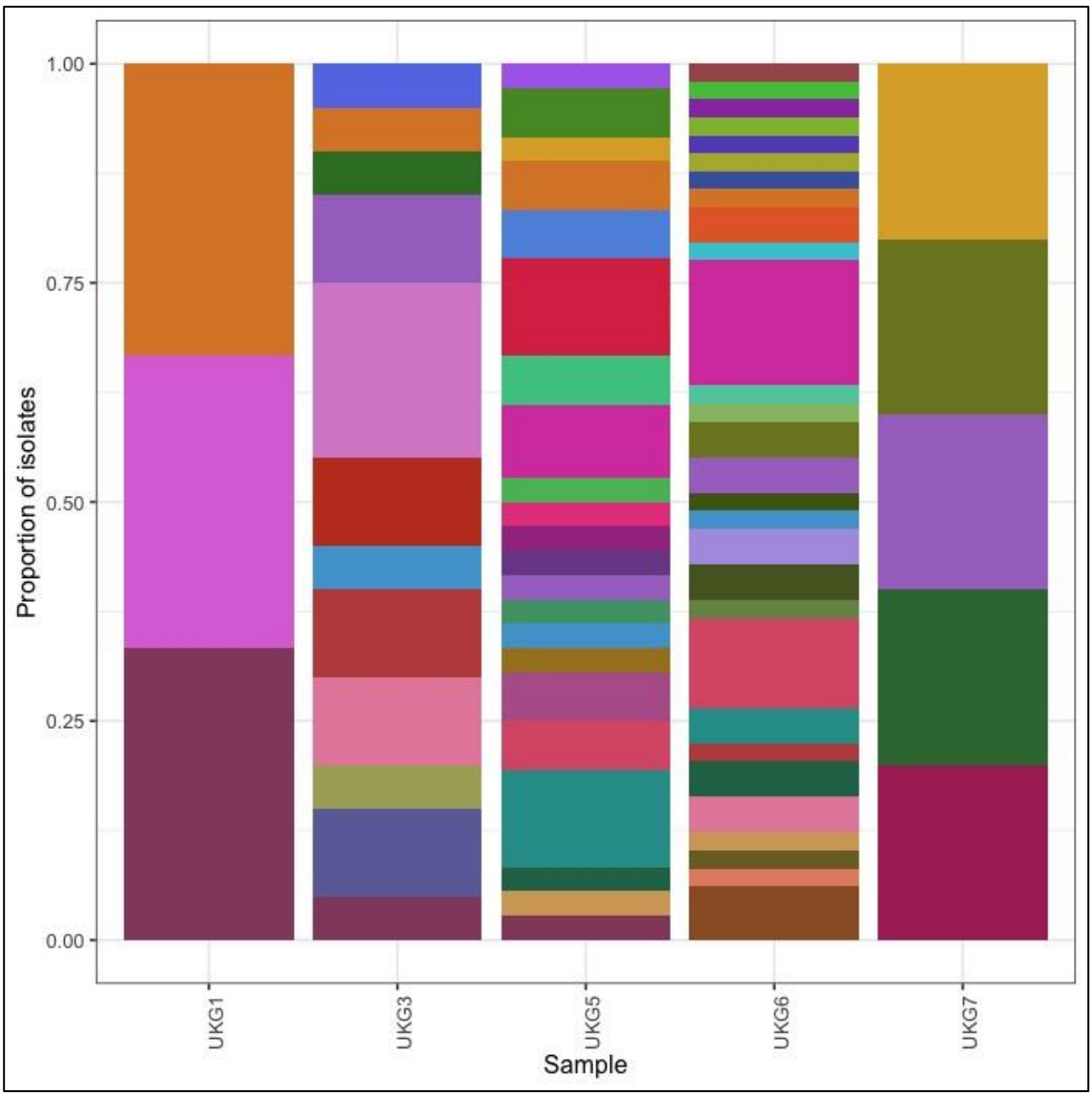


Figure 29. Bar chart of proportion of isolates cultured according to sample, coloured by species. Colours represent different species; there are too many species (52) to identify by inclusion of a legend in figure (legend available in supplementary data). A greater number of species were isolated from UKG5 and UKG6 than other samples.

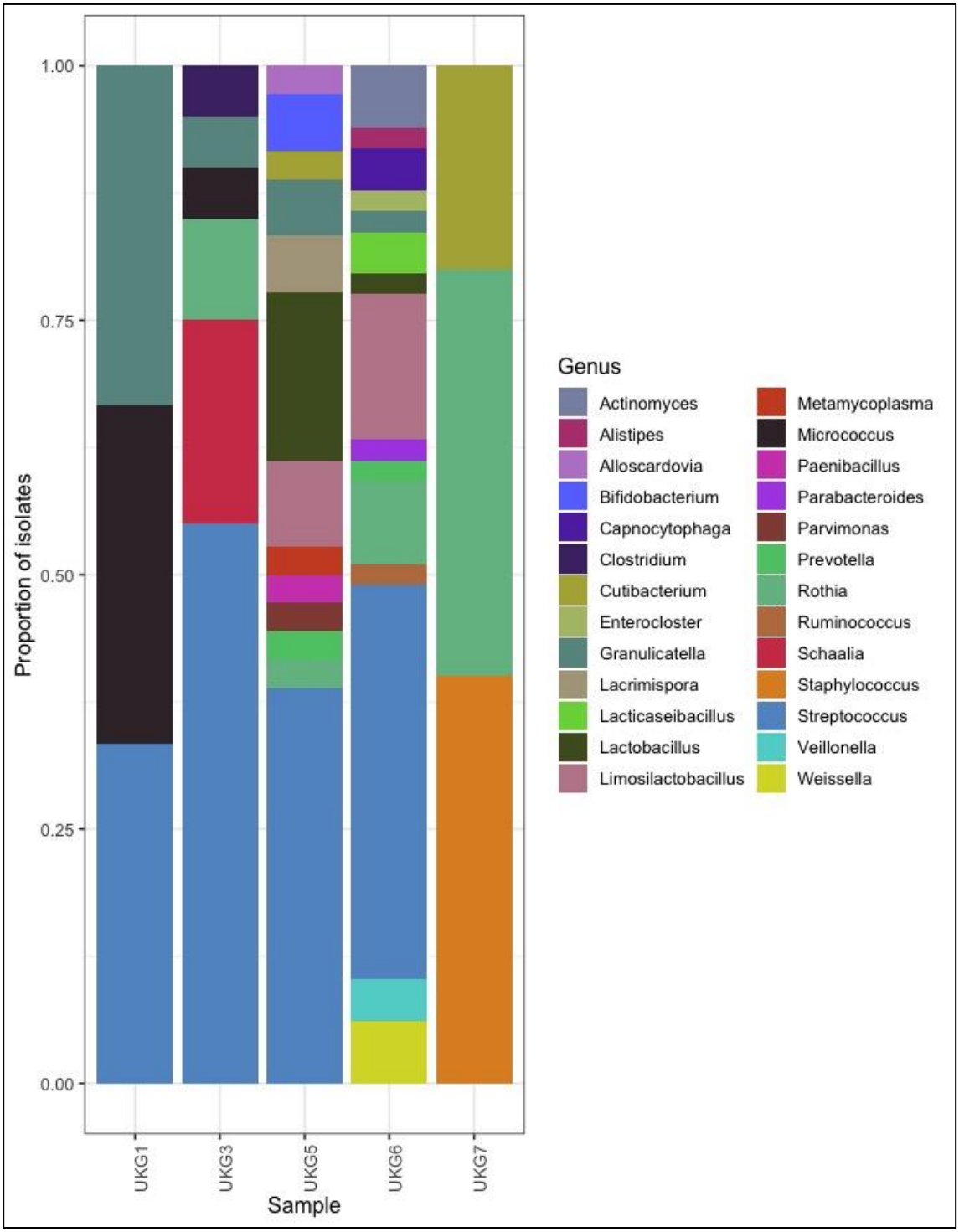


Figure 30. Bar chart of proportion of isolates cultured according to sample, coloured by genus. A greater number of genera were isolated from UKG5 and UKG6 than other samples.

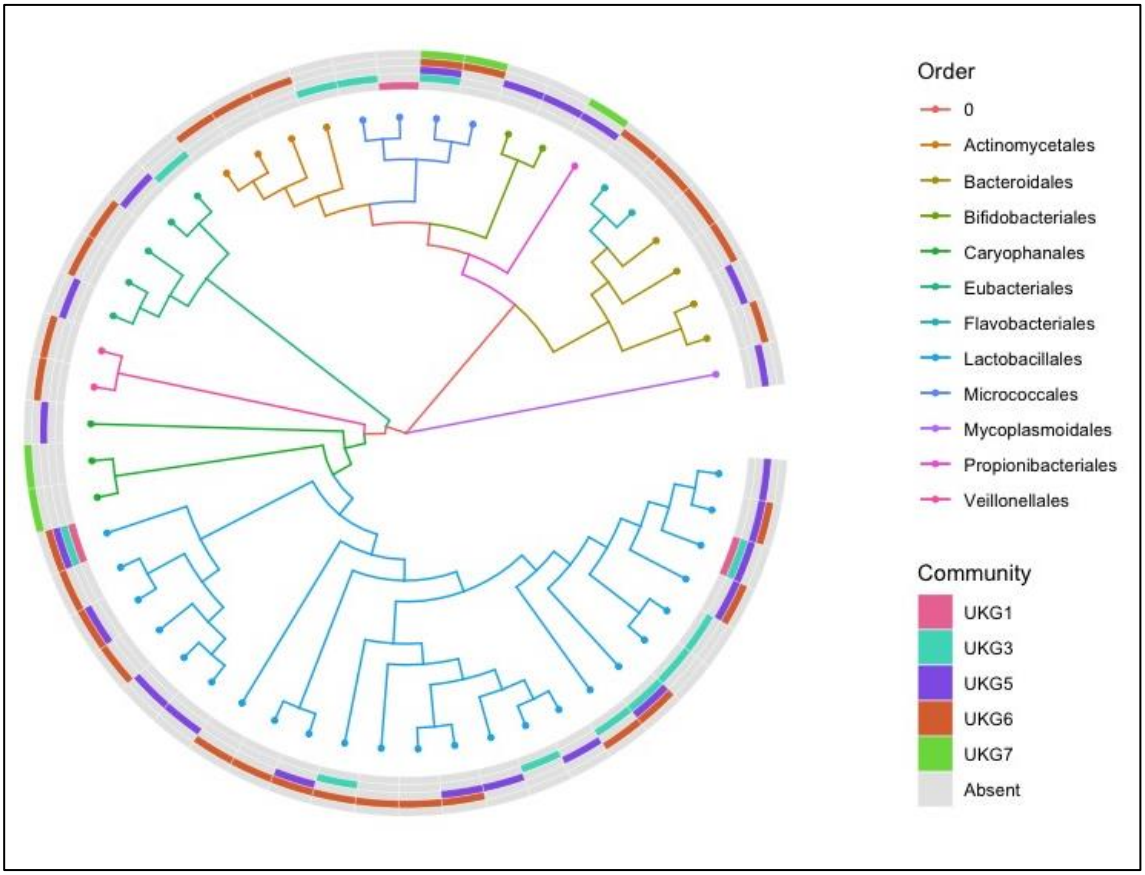


Figure 31. Cladogram visualising the diversity identified through bacterial isolations across the five GC samples showing taxa reconstructed from bacterial isolates. The cladogram was built using 16S rRNA gene V3-V6 variable region Sanger sequences, MUSCLE v3.8.31 [268]; a neighbour-joining tree was created with the ape package v5.7-1 [262], using a Mycoplasmoidales sequence as the outgroup. Colour of spokes signifies order; colours of outer rim signify samples from which each order was isolated from. Lactobacillales is the most common order to which isolates belong.

As demonstrated in Figure 31, Lactobacillales is the order which isolated strains most commonly belonged to. This figure also demonstrates that, even at order taxonomic level, many taxa were only identified in one of the five samples.

3.3.8.2 Comparison of bacterial isolates to tissue sequencing data

The genera isolated from the GC tumour samples were compared to the genera detected through 16S rRNA sequencing of DNA extracted from adjacent GC tumour samples. Here, taxa undefined at genus level through 16S rRNA sequencing were removed. Samples with a greater number of taxa identified through sequencing of tissue tended to have a greater number of taxa identified through sequencing of bacterial isolations. In all samples, there were some

genera identified through tissue sequencing alone and some genera identified through sequencing of bacterial isolations alone.

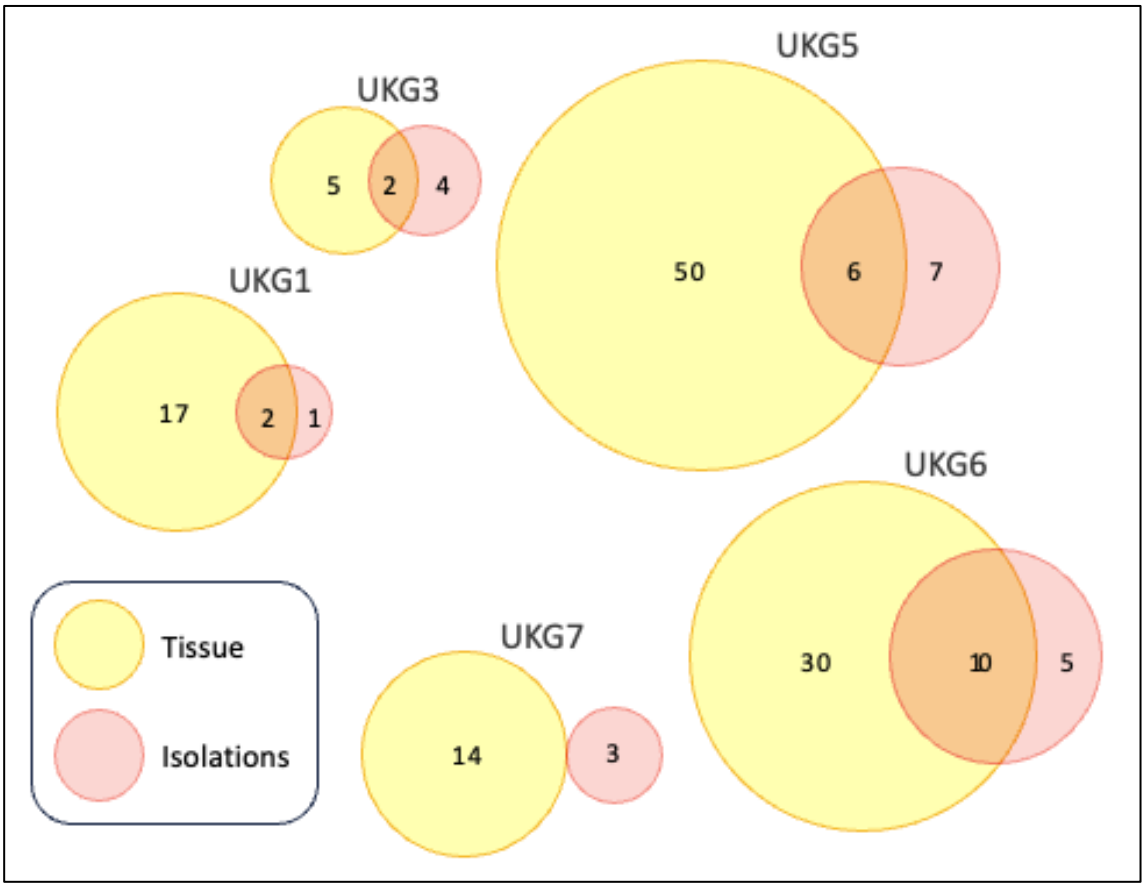


Figure 32. Venn diagrams of number of genera identified through sequencing of tissue and isolations. Yellow circles represent genera identified through 16S rRNA V4 amplicon sequencing of DNA extracted from the tissue sample. Pink circles represent genera identified through Sanger sequencing of the 16S rRNA gene V3-V6 regions of bacterial isolations. Overlapping areas represent genera identified through both methodologies.

3.4 Discussion

3.4.1 The microbiota of the freshly collected gastric cancer samples is representative of the current knowledge of the gastric cancer microbiome

In this chapter, both primary tumour and adjacent samples from seven patients with GC were obtained and the microbiome was analysed via a number of different methodologies. In order to evaluate how representative the microbiome

of samples were of GC, the commonly identified taxa may be compared to previously published studies.

Across all fresh-frozen GC samples, the ten most common bacterial genera identified through metagenomic analysis using low-coverage WGS all belonged to the four most abundant phyla identified by both Bik *et al*. [82] and Liu *et al*. [130] in previous gastric and GC microbiome characterisation studies: Pseudomonadata, Bacillota, Bacteroidota, and Actinomycetota. All of these genera were also detected within the 40 most abundant genera within the TCGA GC as described in Chapter 2. In the largest dedicated microbiome study of GC using 16S rRNA sequencing to date by Liu *et al*. [130], the most commonly detected genera were *Helicobacter*, *Halomonas*, *Streptococcus*, *Prevotella*, and *Shewanella*, in descending order. Of these five genera, *Helicobacter*, *Streptococcus*, and *Prevotella* were all found in the ten most abundant genera in both the low-coverage WGS (post decontamination) and 16S rRNA sequencing of fresh tissue. Interestingly, *Halomonas* and *Shewanella*, the two genera not identified in the ten most abundant genera in the present study, are both recognised as bacterial genera typically found in water habitats [269,270] and it is possible they may represent contaminants in the study by Liu *et al*.

Many of the commonly identified taxa through low-coverage WGS of both fresh-frozen and FFPE tissue, as well as through 16S rRNA sequencing of fresh-frozen tissue have been previously reported in intratumoural GC microbiome [82,130]. Interestingly, all of the ten most abundant genera from both fresh-frozen and FFPE are recognised as oral inhabitants or implicated in oral disease [182,183,185,271–282].

When considering individual microbes of interest, such as *H. pylori* and EBV, the sequencing results of the prospectively collected GC samples is also typical of what would be expected. *H. pylori* is recognised to often dominate the microbiota when present [210], which was observed in the present study findings. A recent systematic review and meta-analysis estimating the global burden of EBV-associated GC described the prevalence of EBV in tumour cells as 7.5% of GC. The one case with a very high abundance of EBV may be expected, although it is interesting that, in addition, a number of other samples also demonstrated high levels of EBV, when considered as a proportion of total microbial reads.

The purpose of this study was not to examine the differences between the GC tumour microbiome and adjacent mucosal microbiome. Other studies have previously investigated this in cohorts with larger sample sizes. Such studies have found differences in microbial beta diversity between tumour and adjacent mucosa stages [34,40,46,48]. As such, the difference between tumour and adjacent microbiomes was not the focus of this study, due to low sample size and statistical power. The PERMANOVA analysis presented in this chapter (Figure 26) did not identify microbial beta diversity differences according to sample type (tumour vs adjacent mucosa). This observation may reflect a true finding of little or no difference between the microbiome of the stomach, although this study was insufficiently powered to determine that. An interesting observation is the presence of *H. pylori* at tumour and respective adjacent tumour sites. Interestingly, a study of 255 *H. pylori*-infected GC cases identified synchronous GC in 7.5% and a subsequent GC in 16.9% [283]. Collectively, these observations may indicate that the presence of *H. pylori* across different anatomical areas of the stomach does not necessarily result in uniform development of neoplastic mucosal changes, although multifocal disease may occur.

Collectively, the comparison of the sequencing results of the prospectively collected GC samples suggest that these samples may be considered representative of what is currently known about GC microbiota and therefore appropriate samples to perform microbial culture studies upon.

3.4.2 Decontamination is required to remove contaminating sequences from metagenomic sequencing

Application of the include-list to the fresh-frozen and FFPE low-coverage WGS sequencing of GC tissue resulted in removal of 94% of reads from the FFPE samples, and removal of only 8% of reads from the fresh-frozen samples. This disparity emphasises the importance of a post-sequencing decontamination step, particularly when FFPE GC tissue has been sequenced, to avoid results being dominated by contaminants.

Following the include-list decontamination step, manual review of the 50 most abundant species and exclusion of identified contaminants was still felt to be necessary to provide a sufficiently strict decontamination process. This demonstrates the importance of manual review of results following a pre-determined decontamination process, to ascertain whether a stricter process is necessary.

Whilst this combined approach to decontamination allows for a potentially more thorough decontamination, the disadvantages of the addition of the manual step are that it is more time-intensive and less reproducible when applied to other data-sets.

3.4.3 Metagenomic sequencing of formalin-fixed, paraffin-embedded gastric cancer samples may not detect microbes at lower abundance

In order to compare fresh-frozen and FFPE, it was assumed that fresh-frozen represented the “gold standard.” However, as it is recognised that FFPE tissue is much easier to process, store, and access retrospectively, the important methodological question to ask is not necessarily whether sequencing of FFPE leads to the same results as sequencing of fresh-frozen, but rather to determine the degree to which results differ, and in what way results differ.

The PERMANOVA analysis of low-coverage WGS data (post decontamination) demonstrated that only 5.7% of variation in microbial composition was due to the difference between using fresh-frozen versus FFPE tissue. This low degree of variation supports the use of FFPE tissue as a suitable alternative to fresh or fresh-frozen tissue. However, due to the mathematical requirements of PERMANOVA, samples where there were no taxa (n=5 of total 14) were removed, which may therefore have distorted the true influence of tissue processing upon microbial composition. It is also important to note that following decontamination, the mean read count of FFPE samples was only 174 reads as opposed to 1291 reads for fresh-frozen samples. This may reflect a lower sensitivity in detecting tissue-resident taxa using the FFPE method than the fresh-frozen approach. This is consistent with the previous study of 16S rRNA sequencing of FFPE versus fresh tissue in gastric tissue, which reported a lower number of OTUs in FFPE tissue compared to fresh-frozen [232].

Collectively, these findings suggest that there are significant disadvantages to be aware of when considering sequencing FFPE tissue. This may be a result of the low-biomass nature of GC. With a strict decontamination process, as performed here, the bacteria identified through low-coverage WGS of FFPE tissue are likely to reflect true residents of the GC microbiome, however, the significant disadvantage is that this method may not be sufficiently sensitive to

detect microbes at lower abundance, which is particularly important in low-biomass tissue, such as in intratumoural GC environment.

3.4.4 16S rRNA sequencing and low-coverage whole genome sequencing of fresh-frozen gastric cancer samples results in similar findings

Although it is difficult to make statistical comparisons between 16S rRNA sequencing and low-coverage WGS due to the different forms of sequencing data produced, the tables of most common genera and heatmaps demonstrate similar findings between the two methodologies. Given that species-level phylogeny cannot be determined from 16S rRNA V4 amplicon sequencing, this should be considered when choosing sequencing method if that level of information is required.

As demonstrated in the heatmaps, certain genera such as *Helicobacter* were detected in more samples through the 16S rRNA methodology, as opposed to the low-coverage WGS. This may be a reflection of the selective amplification of the V4 region, employed in 16S rRNA sequencing, leading to an increased detection of low-abundance microbes. Furthermore, biases can occur in 16S rRNA sequencing resulting from the V regions of different taxa being differentially amplified during PCR, resulting in unbalanced sequencing results [147]. In contrast, some microbes may be below the level of detection by low-coverage WGS, but amplified through targeted PCR in 16S rRNA sequencing. Despite this, overall, at genus levels the general characterisation of the microbiome of GC samples was similar through both methodologies.

Whilst 16S rRNA sequencing has not previously been compared to metagenomic approaches in GC, it has within the setting of the gut microbiome [284]. Within this study, the authors constructed an artificial bacterial population which was subsequently tested via 16S rRNA sequencing and metagenomic approaches. Both methodologies performed similarly in determining relative abundance, compared to the known input bacterial population.

It should be noted that the accuracy of both methods may be influenced by both the laboratory sequencing methodology and the bioinformatics pipelines. Using primers targeting a larger section of the 16S rRNA gene and increasing the

sequencing depth of WGS will result in improved microbiome characterisation of fresh-frozen samples.

It is therefore acceptable that the choice of 16S rRNA sequencing versus low-coverage WGS will largely depend on the characteristics of the study; in studies with greater numbers of samples, it may be more appropriate to use 16S rRNA sequencing with lower costs, whereas in smaller studies and those where species-level information is required, it may be more appropriate to employ metagenomic approaches.

3.4.5 Use of an extensive set of isolation conditions resulted a large number of bacterial isolates from fresh-frozen gastric cancer samples

The conditions used within the bacterial isolations presented in this chapter represent the greatest number of conditions used in any microbial isolations performed from GC and gastric mucosa to date [243–248]. Through this approach, under aerobic, anaerobic and microaerobic conditions, a total of 113 bacterial strains from 26 genera were isolated from just five GC tumour samples. This represents a greater number of genera than detected through any previous studies of the GC tumour [243–248], including those where over 100 samples were used [246,247]. This includes some genera which have not previously been cultured from GC or gastric mucosa (*Alistipes*, *Alloscardovia*, *Bifidobacterium*, *Enterocloster*, *Parabacteroides*, *Parvimonas*, *Ruminococcus*, and *Shaalia*) [243–248].

Samples with a greater number of different genera identified through 16S rRNA sequencing tended to result in a greater number of bacterial isolates. As expected, several taxa identified through sequencing studies were not isolated through bacterial culture. A notable example of this is *H. pylori*, which was not identified from the UKG7 tumour sample despite being detected through tissue sequencing. This may be a result of a number of reasons, described as follows. Some bacteria are known to be difficult to culture or have not previously been cultured [241,242]. Some of the sequenced bacteria may have dead at the time of sampling. Some bacterial may not have survived the freezing and thawing process. Some bacteria preferentially grow in communities, reliant upon metabolites produced by other members of the community [285–288]. Finally,

there may be a true difference in bacteria present in the sample used for DNA extraction versus the adjacent sample used for bacterial isolations.

Some bacterial taxa identified through isolations were not identified through 16S rRNA sequencing. This may be a result of certain bacteria being present in the GC sample at levels too low to detect through 16S rRNA sequencing, yet able to multiply within an appropriate environment. These differences could also be a result of microbial differences in adjacent samples.

Several bacterial taxa isolated are of particular interest given what is known about such taxa in GC or other environments. *Parvimonas micra* (*P. micra*) was isolated from one of the GC tumour samples (UKG5), concordant with sequencing results from this sample. *P. micra* (formerly known as *Peptostreptococcus micros*) is an obligate anaerobe, commonly found within the human oral cavity [289]. This species has been observed to be enriched in stools from patients with colorectal cancer, relative to stool from patients with colorectal adenocarcinoma [290] and healthy individuals [291], as well as in colon tissue biopsies from patients with colorectal cancer, versus healthy individuals [291]. As such, it has been investigated as a potential oncogenic microbe, with *in vitro* evidence of impacting cell proliferation, inflammation and impact upon epithelial–mesenchymal transition pathways [292]. *P. micra* has been previously reported to be enriched in the microbiome of GC mucosa [40] and to the author's knowledge, has not been previously isolated from the GC environment.

Ruminococcus gnavus (*R. gnavus*) is another species identified from the GC isolations which has not previously been cultured from gastric tissue. *R. gnavus* has been implicated in colorectal cancer [293] and hepatocellular carcinoma [294], as well as non-cancer diseases such as inflammatory bowel disease [295]. *R. gnavus* has not previously been investigated in the context of GC, although the performed culture and isolation of this strain could enable further investigation of this species in the context of GC, should other findings warrant this.

A number of studies have reported increased relative abundance of *Lactobacillus* in GC relative to non-cancerous controls [296]. Lactobacilli ferment glucose, predominantly producing lactic acid as the end product of this

fermentation [297]. Whilst the role of lactobacilli in GC is currently not well understood, it is noteworthy that many members of this genus frequently appear in probiotics. Future characterisation and mechanistic studies individual *Lactobacillus* species may help better understand the role of this genus in GC.

The fact that the present study, with the broadest range of GC microbiota isolation conditions performed to date has resulted in the greatest number of genera to date demonstrates the importance of incorporating a variety of conditions to support maximal diversity. However, even the extensive range of conditions used in this study will not cater for all possible isolates and the balance of time and resources should be evaluated when performing GC isolations. The species isolated which were not identified through sequencing methodologies and the ability to perform phenotype characterising and functional studies of GC isolates highlight the value of GC microbial isolations. As such, bacterial isolations from GC samples may be used in preference to, or in conjunction with sequencing investigations, depending upon the specific research question being asked.

3.4.6 Chapter summary

- Fresh tumour samples were prospectively collected from GC patients which may be considered representative of the intratumoural GC microbiome.

- Metagenomic sequencing of GC tissue should include a decontamination process, particularly when using FFPE tissue.

- Metagenomic sequencing of FFPE GC tissue results may be considered as an alternative to using fresh-frozen tissue, although may not detect microbes at lower abundance.

- Both 16S rRNA sequencing and low-coverage WGS of GC tissue provide useful microbiome information and choice of methodology should be guided by sample size, budget, and required taxonomic resolution.

- Bacterial isolations using a broad range of conditions to target GC taxa resulted in successful isolations from all five GC tumour samples, representing the greatest number of genera isolated from GC samples to date.

- Different sequencing methods and culture methods play complementary roles in the investigation of the GC microbiome.

Chapter 4 Development of a bioreactor model to culture gastric cancer microbes

4.1 Introduction

In this chapter, the development of a bioreactor model in which GC microbiota can be investigated in continuous culture is described. The work described in this chapter was carried out in the Allen-Vercoe laboratory, University of Guelph, Canada, during a six-month study placement.

4.1.1 Principles of continuous *in vitro* culture of microbial communities

The work described in previous chapters involved interrogation of microbial sequencing of tissue and of isolation of individual microbes. The former work gives insight into a cross-section of microbial community and the latter gives insight into living microbes, as well as allowing investigation of individual microbial strains. *In vitro* culture of microbial communities offers some advantages over either community tissue sequencing or microbial isolations. Culturing microbial communities *in vitro* can have the advantage of supporting diverse microbiota which may rely upon synergistic or asymmetric relationships, where certain microbial taxa benefit from cross-feeding (with mutual benefit) or parasitism (where one taxon benefits at the cost of another) [285–288]. *In vitro* culture of microbial communities has been performed through the “intestine-on-a-chip” system, whereby human epithelial cells are co-cultured with microbial communities using a microfluidic device [49], and through bioreactor models, as described further below.

4.1.2 Existing bioreactor models

Bioreactors are vessels which may be used to support microbial communities in an aqueous environment, with fixed conditions such as oxygen exposure, pH, temperature, and agitation. Bioreactors may take the form of “batch” bioreactors, i.e. closed systems where media is not replenished, or “chemostat” bioreactors, where media is replaced at a fixed flow rate, allowing for continual removal of waste and addition of nutrients and allowing a stable equilibrium to be formed [298]. Chemostat bioreactors allow continuous culture to be take place often for weeks at a time, and are considered more physiologically relevant models than batch bioreactors [50]. Single-stage bioreactor models

include the Proximal Environmental Control System for Intestinal Microbiota (P-ECSIM) [299] and the Allen-Vercoe Robogut model [300]. In addition to single-stage bioreactors, multi-stage bioreactor models have been used to culture the human gastrointestinal microbiome. These multi-stage bioreactors comprise individual vessels to model the different compartments of the gastrointestinal tract, such as the mucus simulator of the human intestinal microbial ecosystem (M-SHIME) [301] and the SIMulator Gastro-Intestinal (SIMGI) model [302]. The M-SHIME, SIMGI, as well as the TNO Gastro-Intestinal Model (TIM) platform [303] have gastric compartments to their models, although neither of these use the respective gastric compartments to model the gastric microbiome, since the stomach compartments are upstream of any added microbial inocula.

To date, no bioreactor model has been used to model either the gastric luminal contents nor the intratumoural GC microbiome.

4.1.3 Defined and complex microbial communities

In addition to the bioreactor system used to perform microbial culture experiments, the type of inoculum used to seed the system can be selected in order to best answer the desired research question, guided by available resources. Inoculating bioreactors with faecal samples or homogenised mucosal tissue allows direct inoculation of the whole community from the environment which is being studied and may allow the expansion of low-abundance microbes which are not easily detectable within the sample. Communities derived from such a method may be referred to as “complex” communities.

In contrast, “defined” microbial communities, also known as “synthetic” microbial communities, are microbial mixtures made from isolates previously purified from an individual microbial environment, such as a faecal sample or mucosal biopsy [304,305]. Through this method, many theoretically identical vials of individual defined communities can be stored frozen, with viability maintained, for subsequent use. Therefore, compared to complex communities, defined communities offer the advantage of reproducibility in that biological replicates can be performed at later timepoints after an initial study.

4.1.4 pH in the gastric environment

A key advantage of the use of bioreactors is the ability to maintain constant conditions, such as temperature, oxygen exposure, and pH. In the case of bioreactors to culture the faecal microbiome, pH 7 is commonly used [306,307]. In the context of developing a bioreactor to model GC microbiota, pH is an important consideration. In normal physiology, the pH of the stomach luminal contents (i.e. gastric fluid) is acidic, varying over time; it can be as low as pH 1-2, raising to 3-5 with feeding [308,309]. The gastric fluid pH may be affected in the short term, by swallowed food and drink. Certain medication, e.g. PPIs, commonly prescribed in GC, result in a more consistently raised pH through decreasing gastric acid secretion. It has previously been demonstrated that PPI use is associated with increased bacterial abundance in gastric fluid [310,311]. However, when considering the culture of gastric tumour environment, it is important to appreciate that the intratumoural microenvironment is distinct from the luminal environment. Gastric acid is secreted into the gastric lumen and the gastric mucosa is protected from the acidity by a mucus layer. The pH of parietal cells has been reported as 7.4 [312]. To the author's knowledge, the pH of the GC microenvironment is not known. However, the intracellular and extracellular pH of cancer (not GC specific) has been investigated, where the pH of tumour tissue was found to be higher than the corresponding non-cancerous tissue [313]. Together, these findings suggest that the pH of the GC microenvironment in which the GC microbiota resides is more likely to be closer to a neutral pH than a strong acidic pH; it may therefore be appropriate to use pH 7 as a baseline for GC microbiota model development.

4.2 Aim and objectives

The primary aim of the work described in this chapter was to develop a bioreactor system capable of supporting and amplifying GC microbiota. Specific objectives were to:

- Adapt the distal gut chemostat medium to one reflective of the GC environment and suitable for supporting GC microbiota.
- Pilot a bioreactor run in a GC sample (featuring low biodiversity).
- Explore the differences between microbiota identified in the GC bioreactor and previously used methodologies.

- Create a GC defined community and explore the relationship between the defined community and complex community.

- Explore the feasibility of running the GC bioreactor at a lower pH, and the impact of doing so upon the GC microbiota.

- Investigate the metabolites produced by GC microbial bioreactor communities.

4.3 Methods

4.3.1 Bioreactor model

The *in vitro* bioreactor used for the GC bioreactor experiments was based upon the Robogut model described by Gianetto-Hill *et al.* [314]. Briefly, the model used was a continuous culture system, whereby GC microbiota were inoculated into glass vessels, provided with medium through a continuous feed (with waste removal at a matching rate), under anaerobic conditions, with controlled pH and temperature. Figure 33 demonstrates the layout of a single bioreactor vessel. For both bioreactor experiments, twin vessels were run in parallel, enabling biological replicates (UKG7) or comparison of different inocula (UKG6), as described in detail below. Due to the expectedly low microbial diversity (relative to colorectal cancer microbiota) an anaerobic gas mixture (10% CO_2, 10% H_2, 80% N_2) was used, as opposed to pure nitrogen, to provide gases that may be required as co-factors or substrates for growth. An anaerobic (as opposed to aerobic) environment was chosen since a greater number of bacterial taxa were isolated through anaerobic culture than through aerobic culture. Figure 34 shows a photograph of the twin vessel setup as used for the GC bioreactor.

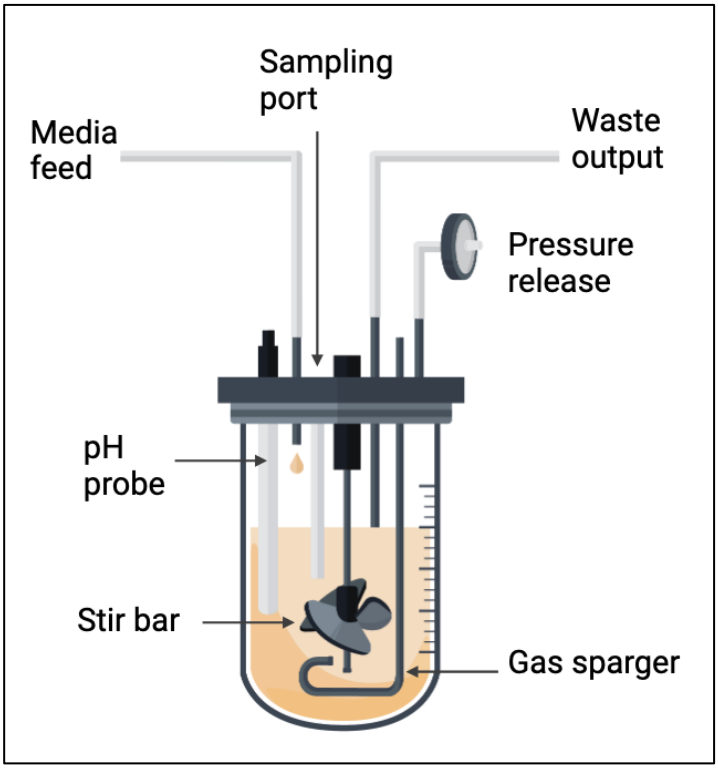


Figure 33. Single vessel bioreactor schematic. Original image produced by Ms Judy Yen and commissioned by the Allen-Vercoe group. Figure has been amended through the addition of labels, using Biorender.com.

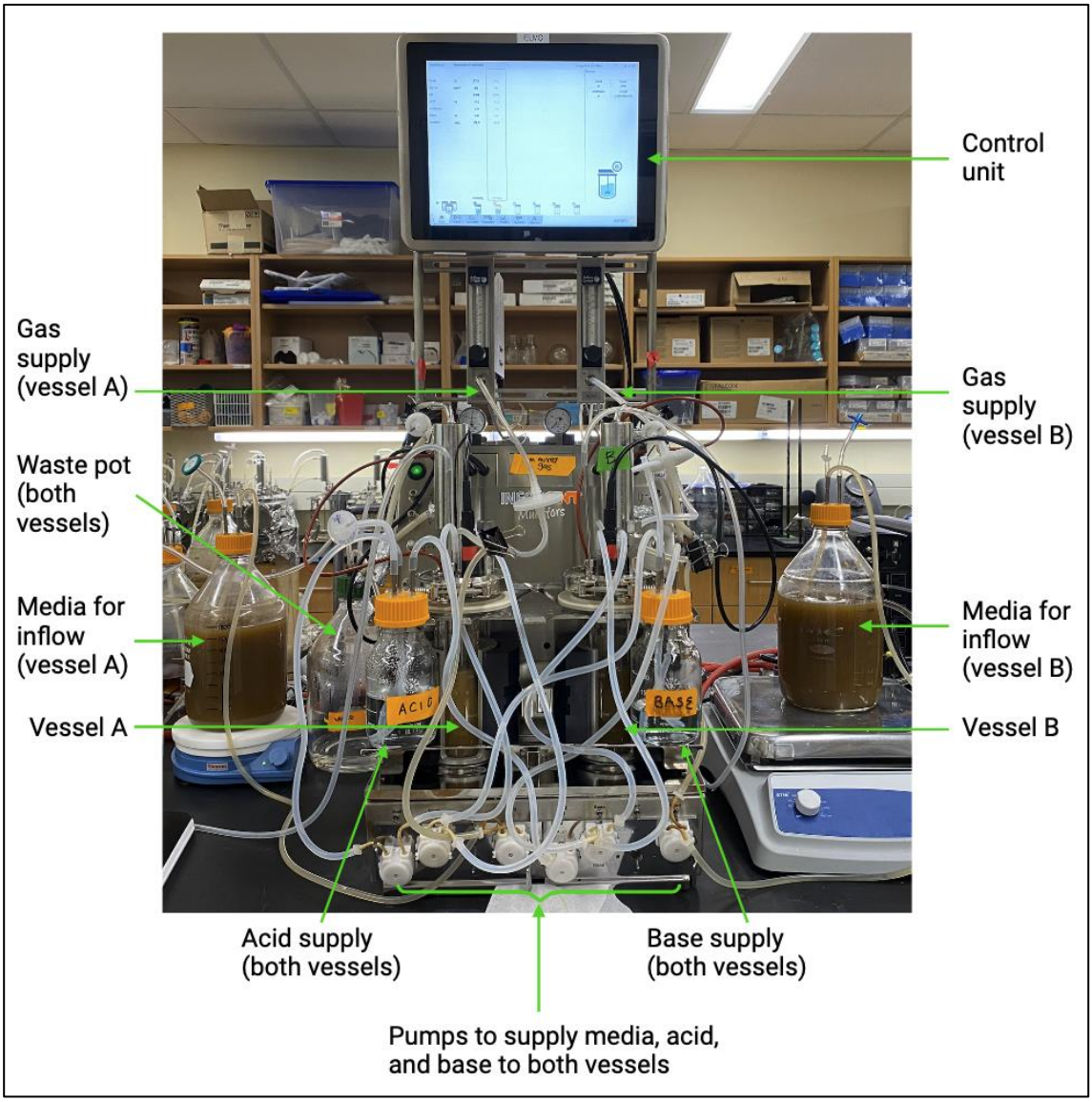


Figure 34. Photograph of twin vessel GC bioreactor setup. The control unit (top) allows adjustment of parameters including pH, temperature, stir rate, and gas supply. The media lines are clamped in this image as the photograph was taken during batch phase. Labels added using Biorender.com.

A detailed description of the materials, preparation and assembly of the bioreactor system is described by Gianetto-Hill *et al*. [314], but is described in brief as follows. pH probes were calibrated prior to each bioreactor run. All bioreactor vessels and associated tubing and filters were assembled and autoclaved whilst containing water. Media bottles were made by adding dry ingredients to 2L glass bottles, adding water and appropriate tubing, then autoclaving, before adding any non-autoclavable additions under sterile conditions in a biosafety cabinet. The water within the autoclaved bioreactor vessels was then replaced with sterile media. Gas sparging was initiated at least 12h before inoculation to ensure an anaerobic environment within the bioreactor vessels. Bioreactor conditions were set to pH 7.0 (+/-0.1) and 37°C

as standard. Immediately prior to bioreactor inoculation, bioreactors were evaluated to determine there were no visible signs of contamination as judged by a lightening of the medium, and that the pH was stable over a minimum of 12 hours. To confirm there was no contamination of media, samples were taken prior to inoculation for DNA extraction. Sterility was later confirmed by absence of detectable DNA by fluorometry (Qubit broad-range dsDNA assay (Invitrogen, #Q32853) and absence of 16S V3-V4 amplification bands on electrophoresis, which were submitted as controls for Illumina sequencing.

Vessels were maintained under batch conditions until there was visible evidence of microbial growth within the vessels, and base was being utilised. Demand for base represented a drop in acidity, since the control unit was set to demand base in order to maintain a pH of 7 (24-72h). Following this, media was supplied at a flow rate of 500mL per vessel per day with matched volume of waste removal, equivalent to a 24h retention period, as per previous work [300]. Although the retention period of the stomach is shorter (i.e. faster transit time), this parameter was not changed from the standard protocol used for colorectal luminal contents, due to the potential consequence of microbial washout with a very fast media flow-rate. In addition, the environment being simulated was the GC environment, not the stomach lumen, therefore a very short retention rate (e.g. 4h or less) may not be reflective of this environment. Vessels were continuously stirred at a rate of 50 revolutions per minute to enable homogenisation of vessel contents.

Sampling of vessels was performed daily under sterile conditions, via a self-sealing silicon septum, as described by Gianetto-Hill *et al*. [314]. In brief, the sampling port area was sprayed with 70% ethanol prior to unscrewing the sampling port cap. The septum was subsequently cleaned with a cotton swab dipped in 10% bleach. Up to 6mL was sampled from each vessel, using a 10mL syringe attached to a sterile needle, and placed in prelabelled 2mL sterile sampling tubes. The septum was cleaned again with 10% bleach, before reattaching the sampling port cap. Withdrawn samples were stored immediately at -80°C.

4.3.2 Bioreactor Media

The Allen-Vercoe group typically uses the “distal gut chemostat medium” developed by McDonald *et al*. [300] to support the growth of diverse

gastrointestinal microbial communities in bioreactor vessels, which was designed to simulate the luminal contents of the distal colon. This medium includes sugars, starches, peptones, and polysaccharides, to reflect the result of dietary components that have undergone digestion within the upper gastrointestinal tract; bile salts, produced in the liver; and mucin to reflect the mucin secreted by the intestinal epithelium. The full details of the distal full chemostat medium are shown in Table 15. For the current work, the distal gut chemostat medium recipe was adapted to reflect the environment around the gastric tumour. The GC bioreactor model was not intended to mimic the luminal contents of the stomach, but instead the environment around the GC. However, the medium was adapted upon the assumption that small molecules present within the luminal contents of the stomach would be similar to those within the GC environment.

The following steps were taken in order to adapt the distal gut chemostat medium to the GC bioreactor medium. First, the UK diet was analysed using data from the UK Total Diet Study [315]. From these data, the 20 most abundant food subgroups were categorised into five food groups, excluding lipids: dairy products (milk and other dairy products); cereals & starch products (miscellaneous cereals, bread and potatoes), vegetables & fruit (vegetables, fresh fruit, fruit products, canned vegetables, green vegetables); meat and eggs (meat products, carcase meat, poultry, fish and eggs); and sugar and preservatives. Lipids were excluded from the medium recipe as they are digested downstream of the stomach, therefore would likely have minimal impact on the microbiota of the GC environment, as well as because of technical issues that would be incurred by including them in the bioreactor system. Using the estimated average weight of food eaten per person per day and the estimated water content of each of the food items (using reference [316]), the proportion of dry weight of each food group was calculated. Table 14 shows the proportion of estimated dry weight per food group. Then, using representative foods from each of the five food groups to reflect that eaten in a typical UK diet, based upon the UK Total Diet Study, the proportion of different sugars, proteins, starches, fibre, and fats were calculated, as depicted in Figure 35. These proportions (with the exclusion of fats) were used to adapt the original distal gut chemostat medium to the GC bioreactor medium, as demonstrated in Table 15. The mucus content of the GC bioreactor medium was kept the same as the distal gut chemostat medium to reflect the mucus layer present across the gastric epithelium. Bile salts were not included in the GC medium as bile salts are produced downstream of the stomach.

Table 14. Food groups as a proportion of total diet by dry weight used to calculate medium composition.

Food group	% dry weight
Cereals & starch products	41%
Dairy products	21%
Meat and eggs	19%
Sugar and sugar products	11%
Vegetables & fruit	9%

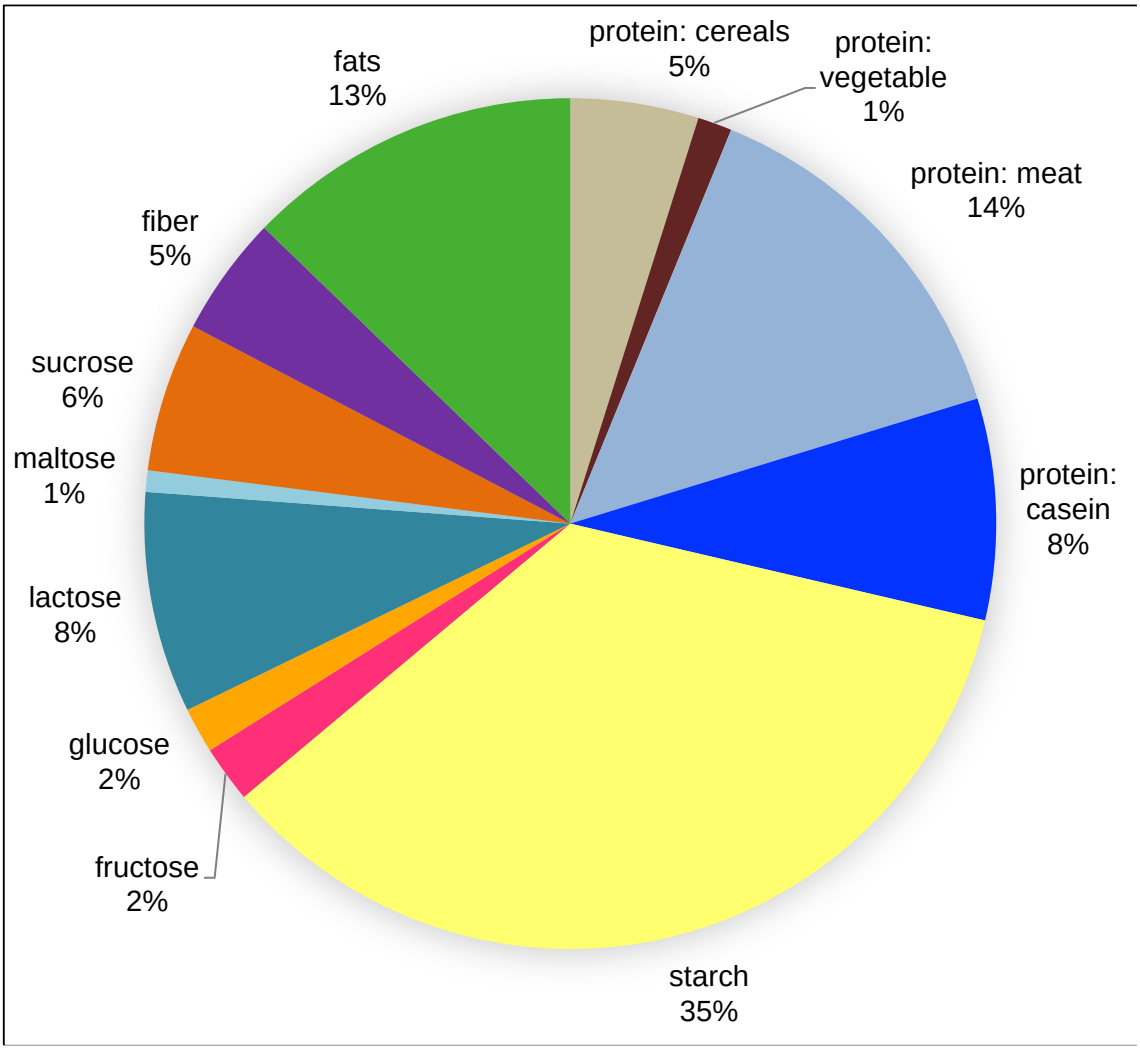


Figure 35. Pie chart demonstrating proportional composition of sugars, proteins, starches, and fats within the typical food groups of a UK diet.

Table 15. Composition of bioreactor media formulations. Composition of the original distal gut chemostat medium (“Original”), as described by McDonald *et al*. [300], and the adapted gastric cancer (GC) bioreactor medium used for the present studies. Colour of cell represents the proportion of a reagent in the GC bioreactor medium, relative to the distal gut chemostat medium: pink, increased in GC bioreactor medium; yellow, no difference between media; green, reduced in GC bioreactor medium.

Reagent	**Original (g/L)***	**GC (g/L)**
Peptone water	**2**	**2.47**
Wheat peptone	**0**	**0.43**
Potato peptone	**0**	**0.43**
Pea peptone	**0**	**0.23**
Casein	**3**	**1.47**
Yeast extract	**2**	**2**
$NaHCO_3$	**2**	**2**
Pectin	**2**	**2**
Xylo-oligosaccharide	**2**	**2**
Arabinoglactan	**2**	**2**
Starch (wheat)	**5**	**3.08**
Starch (potato)	**0**	**3.08**
Inulin	**1**	**1**
NaCl	**1**	**1**
Fructose	**0**	**0.38**
Glucose	**0**	**0.3**
Lactose	**0**	**1.47**
Maltose	**0**	**0.14**
Sucrose	**0**	**0.99**
Mucin (porcine gastric)	**4**	**4**
Bile salts	**0.5**	**0**
L-cysteine HCl	**0.5**	**0.5**
K_2HPO_4	**0.4**	**0.4**
KH_2PO_4	**0.4**	**0.4**
$MgSO_4$	**0.1**	**0.1**
$CaCl_2$	**0.01**	**0.01**
Hemin	**0.005**	**0.005**
Menadione	**0.001**	**0.001**

4.3.3 Pilot of GC bioreactor system to culture low biomass gastric cancer sample

A GC sample (UKG7) was used to pilot the GC bioreactor system. UKG7 was selected since it was demonstrated to have *H. pylori* present, according to prior 16S rRNA sequencing analysis. One UKG7 tumour sample was used for both the standard isolation procedure, as described in Chapter 3, and the GC bioreactor pilot run. Specifically, tissue was homogenised as per the protocol for the tissue isolations (Chapter 3) into a volume of approximately 6mL. The resulting tissue suspension was split as follows: 1.2mL used for tissue isolations and 2mL as inoculum for each of vessels A and B, whilst maintaining anaerobic conditions throughout. The inoculum samples were diluted in filter-sterilised sTSB to a total volume of 10mL each, then introduced into the bioreactor through the waste pipe (prior to any waste passing through the waste pipe) and followed by flushing with 10mL sTSB, to reduce loss of microbial cells as a result of this process. Vessels A and B were run under identical conditions as biological replicates. Three 2mL samples were obtained from each vessel daily for the entirety of the ten-day bioreactor run. In addition to the tissue isolations, vessel content was harvested on day 10 from the bioreactor vessel B, and used as a starting point for isolations as per the standard isolation procedure described in Chapter 3 (described henceforth as “bioreactor isolations”). Figure 36 shows an overview of the workflow using this UKG7 GC tissue sample.

The primary aim of the UKG7 bioreactor run was the investigation of feasibility of the GC bioreactor system to support and amplify microbiota from GC samples. The taxa which were supported to grow within the GC bioreactor system were investigated through community sequencing, as described in detail in section 4.3.7. In addition, results of bioreactor isolations were compared with both the bioreactor sequencing and the tissue isolations.

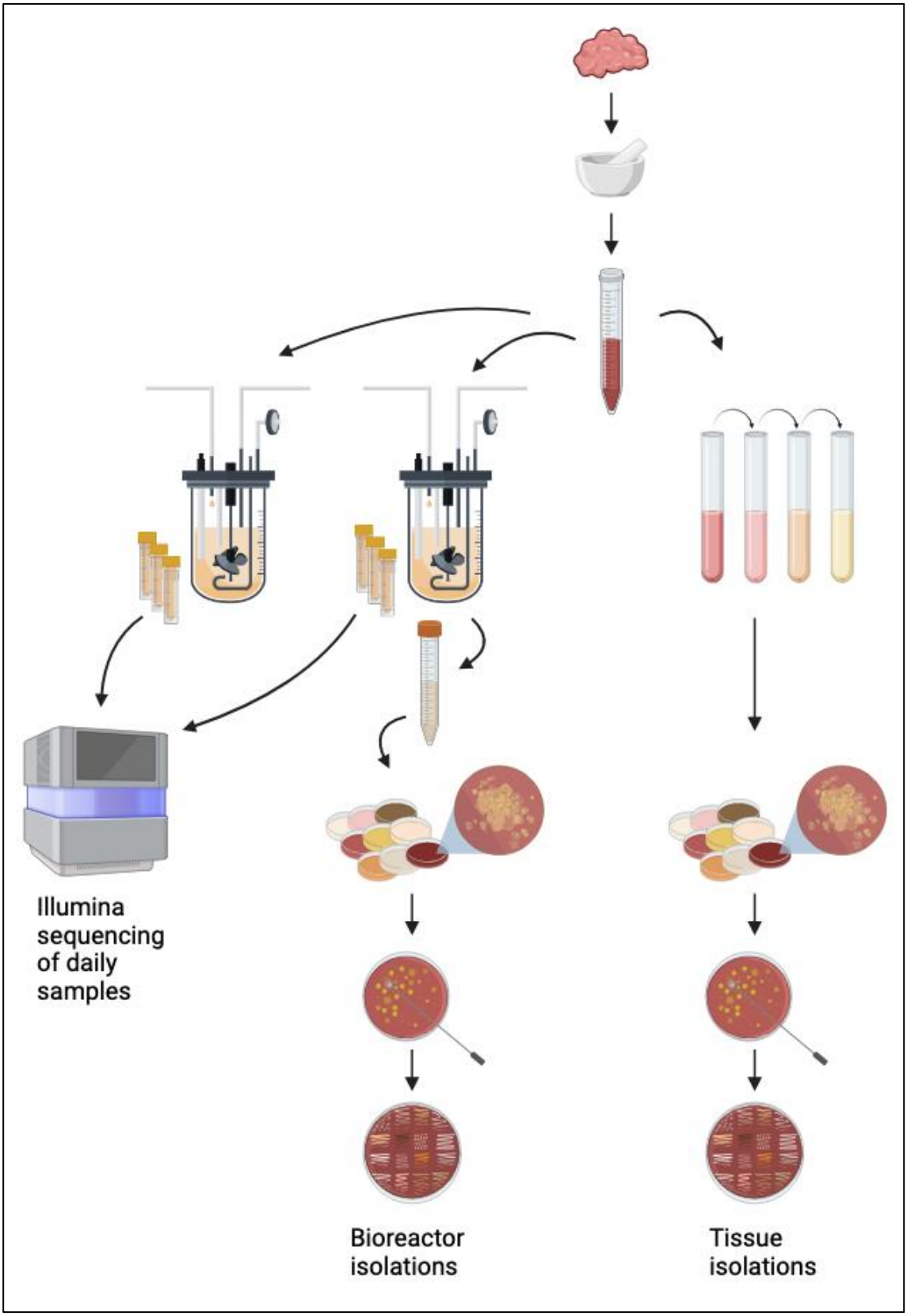


Figure 36. Workflow of UKG7 sample. One GC sample was homogenised and used for two bioreactor vessels as and tissue isolations. Daily samples were taken from the bioreactor for 16S rRNA sequencing using the Illumina platform. After ten days, bioreactor content was harvested from one bioreactor and used to perform bioreactor isolations. Figure created using Biorender.com.

4.3.4 Defined and complex gastric cancer communities

Two separate UKG6 GC tumour samples were used to compare bioreactor microbiota of defined and complex GC communities. These two samples were both taken from the same GC tumour, adjacent to each other. Vessel A was inoculated with a UKG6 defined community from a UKG6 tumour sample (described in full in section 4.3.4.1) and vessel B was inoculated with a UKG6 tumour sample taken adjacent to that used to generate the defined community, homogenised in sTSB. Each inoculum was added to its respective bioreactor vessel on day 0 (following a media sterility check). As a result of differences in inocula volumes, the two vessels were inoculated via different routes. The defined community inoculum (2mL volume) was added to vessel A through the sampling port, using a sterile needle and syringe. The complex community (10mL volume) was added to vessel B as described in the UKG7 bioreactor set-up, via the unused wastepipe, followed by a flush with 10mL sTSB.

Both vessels were run in batch, i.e. without flow of feed and removal of waste until visible evidence of microbial growth. The defined community remained in batch for 24hr and the complex community was left in batch for 48hr. Three 2mL samples were obtained from each vessel daily for the entirety of the twenty-day bioreactor run. Figure 37 shows an overview of the workflow performed using the UKG6 defined community and tissue sample.

The primary aims of the UKG6 bioreactor run were to 1) compare the defined and complex communities and 2) investigate the impact of pH on the GC communities (described in section 4.3.5).

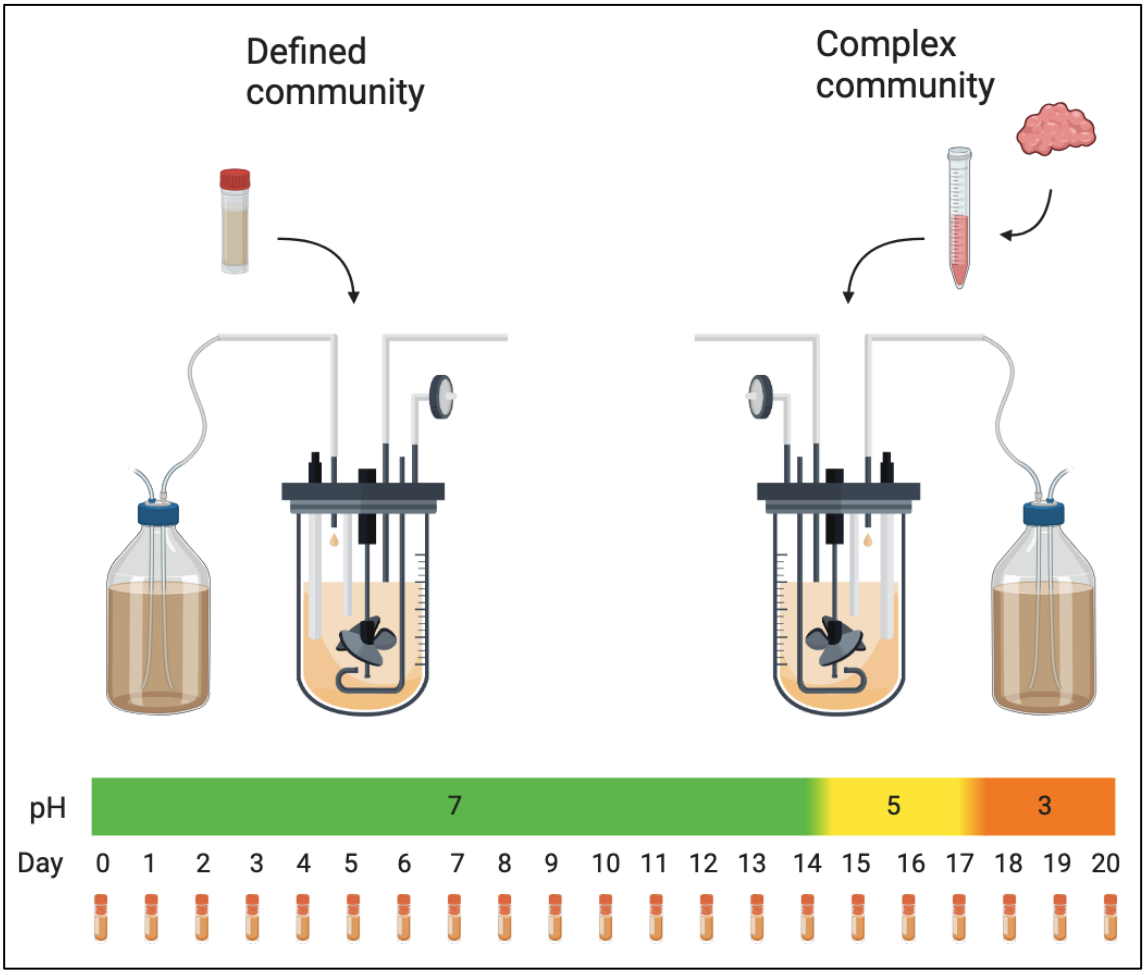


Figure 37. Workflow of UKG6 bioreactor run, comparing defined and complex GC communities and exploring the impact of pH upon community composition. Twin bioreactors were run in parallel, under identical conditions. Vessel A (left) contained the UKG6 defined community; vessel B contained a complex (homogenised whole tumour) community, from a directly adjacent tumour sample to that used to generate the UKG6 defined community. The pH of the vessels over time is represented by the colour bar at the bottom. Daily samples were taken from both vessels. Figure created using Biorender.com.

4.3.4.1 Defined community assembly

A defined GC community was assembled using bacterial strains previously isolated from a UKG6 tumour sample isolations, as described in Chapter 3. All unique strains were grown anaerobically (for strains isolated under anaerobic or microaerobic conditions) or aerobically (for strains isolated under aerobic conditions), at 37°C on FAA plates for 48h, re-sequenced via 16S rRNA gene V3-V6 Sanger sequencing, and comparisons made against NCBI database, as described in Chapter 3, to confirm purity and identity. All strains were re-streaked onto one FAA plate per strain and incubated for 24 hours under

conditions as above to grow biomass; except for strains which were identified as being slow-growing, where two plates were prepared per strain.

Following incubation, all agar plates were transferred into the anaerobic chamber and all biomass was collected by scraping from agar plates and combined by resuspending into a single tube containing 5mL sterile pre-reduced (anaerobic) sTSB. The contents were then transferred into a conical tube containing 80mL freezing media (12% weight/volume skimmed milk; 2% glycerol; sterile H_20), vortexed, and aliquoted into 40 x 2mL cryotubes before freezing immediately at -80°C. Figure 38 demonstrates the workflow used to create the defined community.

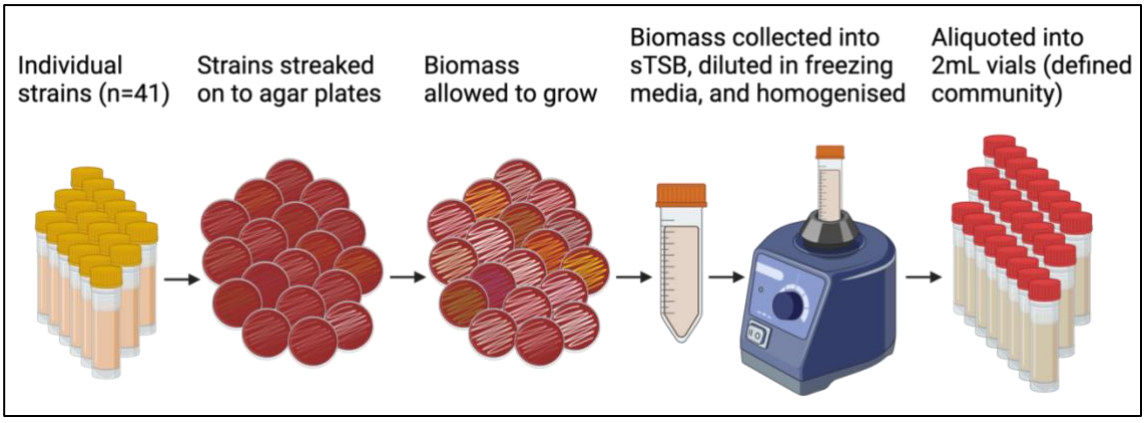


Figure 38. Workflow of UKG6 defined community assembly. Biomass from unique UKG6 bacterial strains was grown on agar plates, combined in sTSB (supplemented tryptic soy broth), diluted in freezing media and homogenised, before aliquoting into separate 2mL vials of UKG6 defined community. Not shown: sequencing of individual strains to confirm identity and purity. Figure created using Biorender.com.

4.3.5 Investigation of pH upon gastric cancer microbiota

The pH of the stomach may vary greatly within and between individuals although the impact of pH upon the GC microbiota is largely unknown. Bioreactor experiments with differing pH levels may help to understand the impact of pH changes upon the GC microbiota. Furthermore, controlled bioreactor experiments may help understand the impact of pH altering medication (such as PPIs), and gastric achlorhydria (as a consequence of *H. pylori* infection) upon the GC microbiome. Within the UKG6 bioreactor run, the pH was altered in order to 1) assess the feasibility of running the bioreactors at a low pH and 2) investigate the impact of lowering the pH on the GC microbial community. Since the bioreactor work is resource- and labour-intensive, the

investigation of pH was performed in the same bioreactor run as the UKG6 comparison of defined versus complex GC microbiota, after the communities were assumed to have reached steady-state.

The pH of both UKG6 vessels was maintained at 7 for the first 14 days, by which point the communities were assumed to be at steady state, based upon previous bioreactor investigations [256]. On day 14, the pH was lowered from 7 to 5 over a period of 10h. The pH then remained at 5 until day 17, when the pH was lowered from 5 to 3 over a period of 10h. The changes in pH are shown in Figure 37. Practically, these pH changes were achieved by amending the target pH on the bioreactor control unit by 0.2 pH units every hour, over 10h. All pH values are provided within 0.1 unit, since the pH control works in a way such that when the pH probe senses a pH >0.1 from the target pH, acid or base is pumped to the system until the pH reading lies within 0.1 from the target pH.

4.3.6 Bioreactor vessel community genomic DNA extraction and Illumina sequencing

The taxonomic composition of the bioreactor vessel microbial communities was assessed through extraction of genomic DNA and 16S rRNA gene V3-V4 Illumina sequencing.

Genomic DNA was extracted from bioreactor vessel content samples with the QIAamp PowerFecal Pro DNA kit (Qiagen, #51804) with additional lysing steps. In brief, 1mL of bioreactor sample was centrifuged at 16000g for 5 minutes at 4°C. Supernatant was discarded and cell pellets were resuspended in 200µL 1 x phosphate buffered saline, which was then added to the PowerBead Pro Tube. After bead beating, samples were boiled at 95°C for 10 minutes and sonicated for 5 minutes. The supernatant from the PowerBead Pro tube was treated with 20µL of 20mg/mL Proteinase K (Invitrogen #AM2548) at 70°C for 10 minutes, prior to proceeding with manufacturer's instructions. DNA was quantified with the Qubit broad-range dsDNA assay (Invitrogen, #Q32853), according to manufacturer's instructions.

Microbial communities were analysed by 16S rRNA gene V3-V4 sequencing. DNA was normalized to 10ng/µL and amplified in a single-step PCR reaction with ultramers including: the 341f and 805r locus-specific primers [317], Illumina

overhang sequences, combinatorial barcodes, and Flowcell adapter sequences. The PCR reaction included 12.5µL GoTaq Hot Start Master Mix (Promega, #M5132), 5µL of each primer at a concentration of 1µM, and 2.5µL normalized DNA. The PCR cycler conditions included an initial denaturation at 95°C for 3 min, followed by 30 cycles of 95°C for 30s, 55°C for 30s, 72°C for 30s, and a final extension at 72°C for 5 minutes. PCR products were cleaned with 1.5x volume of GB-Magic DNA Magnetic Beads (Genebio, #N441), following manufacturer's instructions. Purified PCR products were submitted to the Genomics facility at the University of Guelph Advanced Analysis Centre, and prepared for Illumina MiSeq 2 x 300 base pair sequencing.

4.3.7 Community 16S rRNA gene sequence processing

16S rRNA sequencing results were processed according to standard Allen-Vercoe laboratory protocols. Raw Illumina reads from chemostat samples were processed in R v4.2.3 by Ms Sarah Vancuren. Adapter sequences were first removed from reads using cutadapt v4.0 [318]. Trimmed reads were then imported into DADA2 v1.26.0 [319] for error modelling and to form amplicon sequence variants (ASVs). ASVs were first classified using DECIPHER v2.26.0 [260] to the Silva SSU rRNA database v138, and lower taxonomic classification to genus and species levels were conducted with USEARCH v11.0.667 [320] to the NCBI 16S RefSeq database v217.

The results ASV table was then used for further microbiome analysis as described in section 4.3.9.

4.3.8 One-dimensional ^{1}H nuclear magnetic resonance spectroscopy

One-dimensional ^{1}H nuclear magnetic resonance (NMR) spectroscopy was performed on selected samples from both the UKG7 (low biodiversity) and UKG6 (defined versus complex with pH change) bioreactor experiments. The aims of this investigation were to 1) assess the feasibility of investigating metabolites produced by bioreactor microbial communities, 2) explore differences in metabolites produced by the defined and complex bioreactor communities, and 3) explore the impact of low pH on the metabolites produced.

Vessel content samples were thawed from -80°C in a stepwise fashion, at -20°C, then 4°C prior to preparation. Samples were prepared as previously

described [321]. Briefly, samples were centrifuged at 14000g for 10 minutes at 4°C. Supernatants were filtered with 0.2µm polyethersulfone syringe filters, to which 5mM IS-2 Chenomx Internal Standard, DSS-d6 (Chenomx Inc.) was added at 10% (volume/volume) and well mixed prior to loading in 5mm NMR tubes. Spectra were acquired on a 600MHz Bruker NMR spectrometer using TopSpin 4.0 software at the NMR facility of the University of Guelph's Advanced Analysis Centre. Spectra were processed and profiled by Ms Sarah Vancuren using the Chenomx software suite v8.6 to the Chenomx 600 MHz compound reference library. The resulting metabolite concentrations were analysed as described in section 4.3.9.

4.3.9 Bioinformatics, statistical analysis and Images

Statistical analysis and figure creation throughout the chapter were conducted in R v4.3. [169] within R Studio v2023.06.0+421 0 [170], using pheatmap v1.0.12 [175], phyloseq v1.44.0 [265], tidytree v 0.4.4 [322], treeio v 1.242 [323], ggtree v3.8.0 [266], ape [262], msa v1.32.0 [264], ggnewscale v0.4.9 [289], dplyr v 1.1.2 [172], tidytree v0.4.2 [266], and ggplot2 v4.3.2 [176] packages.

The cladogram of UKG6 taxa from bioreactor vessels, tissue sequencing and tumour isolations, was constructed from 16S rRNA gene sequencing data. Representative sequences from each unique species in each condition was selected and used in a multiple sequence alignment with MUSCLE v3.8.31 [268], a neighbour-joining tree was created with the ape package v5.7-1 [262], using one of the methanogen sequences (*Amylibacter kogurei*) as the outgroup.

Alpha diversity was measured via Shannon index through the phyloseq v1.44.0 package [265] and t-test test was used to compare the differences in Shannon index between communities. Beta-diversity was demonstrated using principal coordinate analysis of weighted UniFrac distances using the phyloseq v1.44.0 [265]. Weighted UniFrac distances were chosen as this takes into account the relative abundance of taxa shared between samples.

4.3.10 Gram staining and microscopy

Gram staining was performed on both UKG6 bioreactor vessel content samples on day 10 and day 18. Using a sterile loop, approximately 10µL of required sample was added to a glass slide and allowed to air dry, before passing

through a flame to heat-fix. The following reagents were applied to the slide, with water rinses following each stage: Crystal violet (stains all bacteria) for 60s, iodine (enhances Crystal violet stain) for 60s, Gram's decolouriser (acetone-ethanol mixture which results in primary stain being washed away from gram-negative bacteria) for 5-10s, and safranin (counterstains colourless gram-negative bacteria) for 30s. The Gram stain protocol results in purple staining of gram-positive bacteria with thick peptidoglycan walls which retain the Crystal Violet stain despite the application of Gram's decolouriser. Gram-negative bacteria are stained red/pink since their thin layers of peptidoglycan allow the Crystal Violet stain to be washed out with Gram's decolouriser and are therefore identified by safranin counterstain.

Slides were imaged with an Axio microscope (Zeiss) and images were captured with an Axiocam 305 color camera (Zeiss). Images were produced using ZenBlue software (Zeiss).

4.4 Results

4.4.1 Bioreactor model of low-biodiversity gastric cancer microbial community

Batch conditions were maintained for both UKG7 vessels for 72 hours, at which point there was visible evidence of microbial growth and base was being demanded in order to maintain a pH of 7 within both vessels. Both vessels were run for a total of ten days.

The tissue and isolation results presented in Chapter 3 demonstrated the low microbial diversity of UKG7, relative to other investigated samples. Sequencing data from both vessels demonstrate that the GC bioreactor system is capable of supporting and amplifying bacteria from this low-biodiversity sample over ten days.

The relative abundance of taxa identified through 16S rRNA sequencing at genus and species levels is shown in Figure 39 and Figure 40, respectively. Whilst caution should be taken with interpreting species-level taxonomy from this V3-V4 sequencing, the species-level data suggests that the initially dominant species (*Streptococcus salivarius*) became proportionally less

abundant in both vessels over time. At the species level, the bioreactor communities appear to be dominated by six taxa: *Streptococcus salivarius*, *Streptococcus mutans*, *Lancefieldella rimae*, *Limosilactobacillus fermentum*, *Lactobacillus gasseri*, and *Scardovia wiggsiae*. In both vessels, *Streptococcus mutans* dominated initially, with other species becoming relative more abundant with time. Both vessels showed similar community changes over time, with the exception that *Lancefieldella rimae* became dominant in vessel A as the experiment progressed (and was not present in vessel B), and *Streptococcus mutans* became dominant in vessel B as the experiment progressed (and was not present in vessel A).

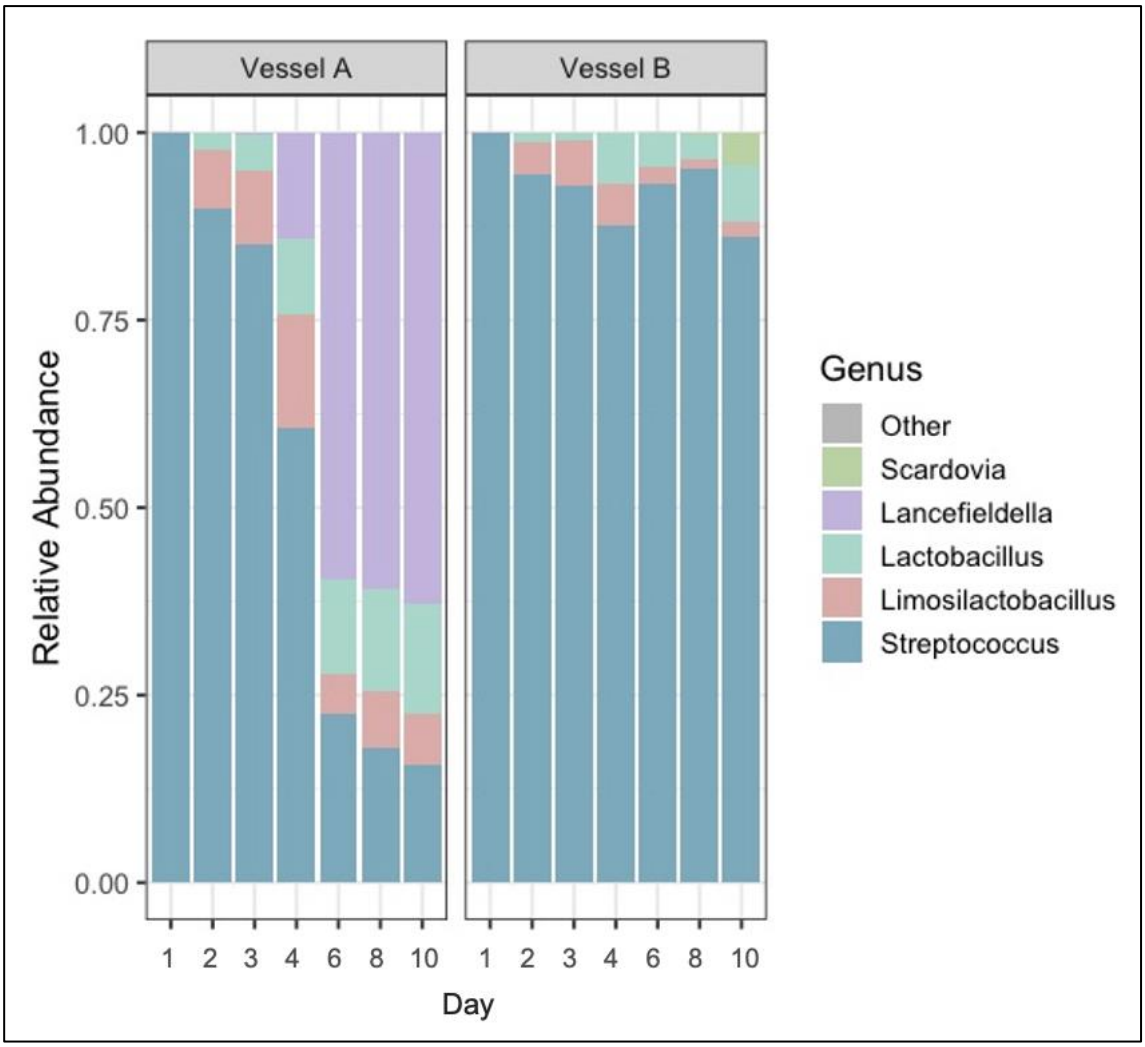


Figure 39. Relative abundance of genera over time in UKG7 bioreactor vessels. In both vessels, *Streptococcus* dominated initially, with other genera increasing in relative abundance over time.

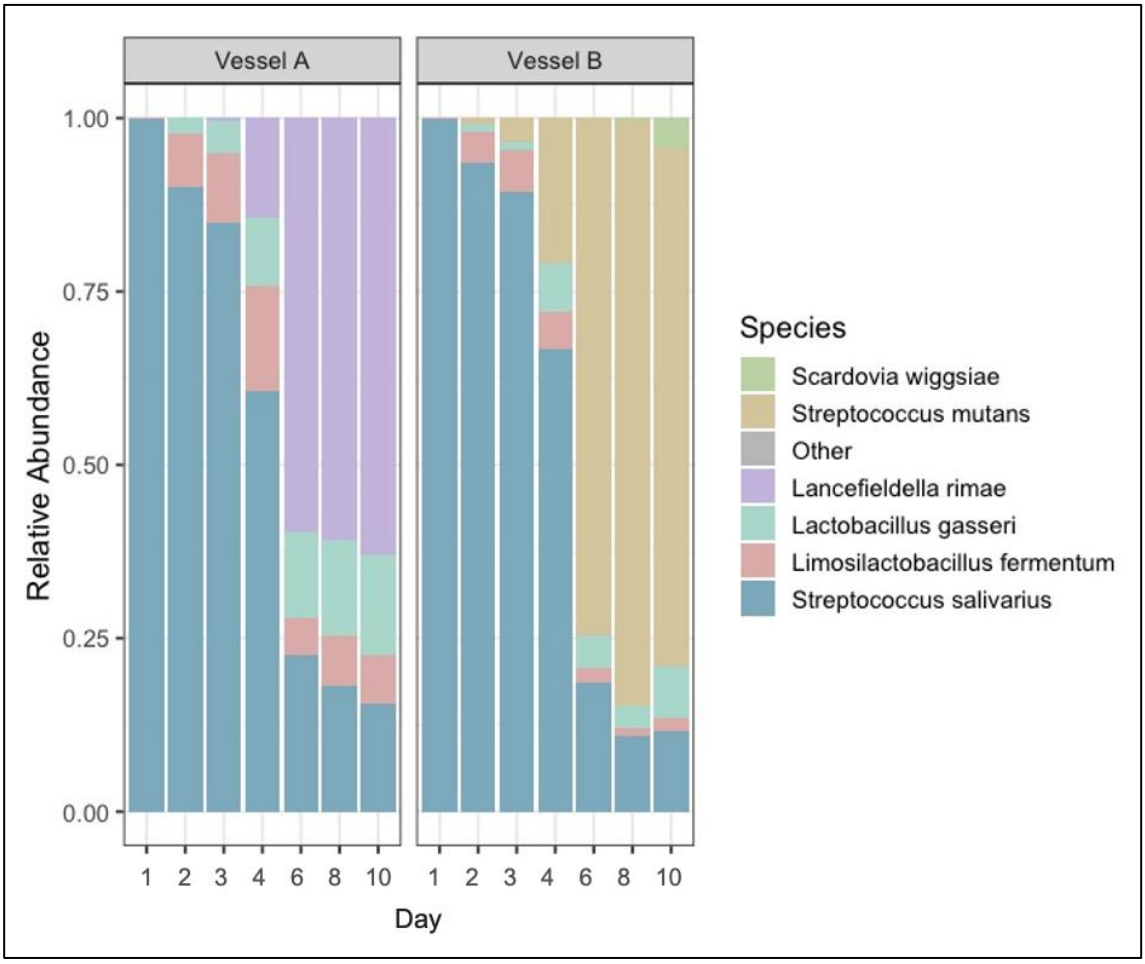


Figure 40. Relative abundance of species over time in UKG7 bioreactor vessels. In both vessels, *Streptococcus salivarius* dominated initially, then decreased in abundance over time, relative to other species.

The UKG7 GC bioreactor communities were compared to the tissue sequencing of an adjacent UKG7 GC sample. In the tissue V4 16S rRNA sequencing of UKG7, 11 genera were identified as present using a cut-off value of 5 or more reads to define presence (versus absence), representing 76% of all read counts. This excluded three undefined taxa groups where the genus could not be defined through sequencing due to sequences aligning to multiple genera. In the Illumina V3-V4 16S rRNA sequencing of the twin bioreactors, five genera were identified as present using a cut-off value of 500 or more reads across read counts of both bioreactor vessels over ten days to define presence (versus absence), representing >99.9% of all reads. Of note, none of the genera with read counts of 1-4 from tissue V4 16S rRNA sequencing overlapped with genera identified through bioreactor sequencing. These data are presented in Table 16.

Table 16. Table showing genera identified through sequencing of DNA from tissue and from bioreactor samples from UKG7. V4 16S rRNA sequencing was used to sequence tissue sample. V3-V4 16S rRNA sequencing was used to sequence bioreactor samples. Species of closest match and identity are based on the closest match from NCBI BLAST alignments. Genera are ordered in decreasing abundance (across all samples for bioreactor sequencing). *Streptococcus*, the only genus identified through both methodologies, is shown in blue.

Genera identified through tissue sequencing	**Genera identified through bioreactor sequencing**
Prevotella	*Streptococcus*
Streptococcus	*Lancefieldella*
Helicobacter	*Lactobacillus*
Bacteroides	*Limosilactobacillus*
Faecalibacterium	*Scardovia*
Succinivibrio	
Alloprevotella	
Prevotellaceae.UCG.003	
Acidaminococcus	
Sutterella	
Methanobrevibacter	

Isolations performed from samples harvested from the bioreactor (vessel B) resulted in the isolation of the following five species: *Lactobacillus paragasseri*, *Limosilactobacillus fermentum*, *Streptococcus mutans*, and *Streptococcus salivarius*, by Sanger sequencing of the V3-V6 16S rRNA gene. These species were all identified through the V3-V4 16S rRNA Illumina sequencing of the chemostat samples, with the exception of *Lactobacillus paragasseri*. However, when the sequence identified as *Lactobacillus paragasseri* from the V3-V6 Sanger sequence was aligned to the *Lactobacillus gasseri*, sequence, as identified through Illumina sequencing of the bioreactor vessels, using NCBI BLAST, there was 99.77% identity, suggesting that these likely represent the same species.

UKG7 tissue isolation and bioreactor isolation results were compared. The closest identity at species level of both tissue and bioreactor isolations are

shown in Table 17. At the species level, there was no overlap in bacterial strains isolated from the UKG7 tissue and UKG7 bioreactor isolations.

Table 17. Species identities of bacterial strains isolated from UKG7 GC tissue sample and from bioreactor vessel contents. Species of closest match and identity are based on the closest match from NCBI BLAST alignments following Sanger sequencing of V3-V6 16S rRNA gene. Species are ordered alphabetically.

Tissue isolations	**Bioreactor isolations**
Cutibacterium acnes	*Lactobacillus paragasseri*
Rothia dentocariosa	*Limosilactobacillus fermentum*
Rothia mucilaginosa	*Streptococcus mutans*
Staphylococcus caprae	*Streptococcus salivarius*
Staphylococcus haemolyticus	

4.4.2 Defined community

The UKG6 tumour defined community was assembled from unique strains obtained from the isolation procedure described in Chapter 3. Table 18 details the 44 strains included within the defined community; these represent 21 distinct species, from 11 separate genera. Forty individual 2mL aliquots of the UKG6 defined community were created in total.

Table 18. Details of 44 bacterial strains comprising the UKG6 defined community. Species of closest match and % identity are based on the closest match from NCBI BLAST alignments.

Strain	Species of closest match	Identity (%)	Colony morphology
UKG6_BHI_1_AN	*Streptococcus salivarius*	99.71	Haemolytic; brown when gathered
UKG6_BHI_11_AN	*Streptococcus sanguinis*	100	Creamy texture; tan colour when gathered
UKG6_BHI_4_AN	*Limosilactobacillus fermentum*	100	Matt grey colonies on plate; medium texture; irregular margins of colonies
UKG6_BHI_6_AN	*Weissella confusa*	99.85	Watery; white
UKG6_BR_1_AER	*Rothia dentocariosa*	100	Creamy colour, medium texture
UKG6_BR_1_MA	*Streptococcus timonensis*	99.72	Irregular sized colonies; haemolysis; medium texture; brown when gathered
UKG6_BR_11_MA	*Weissella confusa*	98.98	Translucent on plate; creamy/watery texture; white & semi-translucent when gathered
UKG6_BR_12_MA	*Weissella confusa*	98.83	Tiny colonies; watery; off-white when gathered
UKG6_BR_15_MA	*Actinomyces oris*	99.43	Very textured, hard to gather; slower grower (required 2 plates)
UKG6_BR_3_AN	*Lactobacillus delbrueckii*	99.86	Creamy, easy to gather texture; white on plate; easy grower
UKG6_BR_4_AN	*Streptococcus oralis*	99.57	Hard texture; brown colour
UKG6_BR_4_MA	*Streptococcus vicugnae*	100	Watery; brown
UKG6_BR_5_MA	*Streptococcus oralis*	99.43	Brown when gathered; medium smoothe texture;
UKG6_BRF_FAA_3_AN	*Veillonella nakazawae*	97.46	White colour; creamy texture; streaky growth
UKG6_BRF_FNB_8_AN	*Streptococcus oralis*	99.71	Medium texture; brown when gathered; haemolytic
UKG6_DCM_10_AN	*Streptococcus sanguinis*	99.71	Moderate texture; creamy/tan colour
UKG6_DCM_12_AN	*Limosilactobacillus fermentum*	100	Matt grey colonies on plate; medium texture; irregular colony margins
UKG6_DCM_9_AN	*Streptococcus ilei*	99.55	Off-white; hard to gather as a lawn/easy to gather from isolates
UKG6_FAA_1_AER	*Streptococcus vicugnae*	100	Off-white
UKG6_FAA_12_AN	*Streptococcus timonensis*	100	Easy to gather; brown, shiny
UKG6_FAA_21_AER	*Rothia mucilaginosa*	99.28	Beaded texture; required 2 passes to collect biomass
UKG6_FAA_4_AER	*Streptococcus salivarius*	100	Grey colour; loose, creamy texture
UKG6_FAA_7_AN	*Granulicatella adiacens*	99.43	Small colonies; smooth texture; translucent on plate; white on loop

Strain	Species of closest match	Identity (%)	Colony morphology
UKG6_FCC_1_AN	*Weissella confusa*	100	White, creamy texture, easy to gather on loop
UKG6_FCC_2_AER	*Weissella confusa*	99.71	Loose texture; grey/translucent colour
UKG6_FCC_2_AN	*Streptococcus rubneri*	99.43	Medium texture; tan on loop; white on plate
UKG6_FCC_20_AN	*Lacticaseibacillus rhamnosus*	100	Shiny on plate; sticky; tan when gathered
UKG6_FCC_21_B_AN	*Weissella confusa*	100	Watery; white
UKG6_FCC_3_AN	*Limosilactobacillus fermentum*	100	Watery; white
UKG6_FCC_5_AN	*Streptococcus ilei*	99.08	Off-white; hard to gather as a lawn/easy to gather from isolates
UKG6_FCC_7_AN	*Limosilactobacillus fermentum*	100	Liquid texture; white on plate; small colonies
UKG6_FCC_8_AN	*Limosilactobacillus fermentum*	100	Liquid texture; small colonies
UKG6_FNB_13_AN	*Prevotella melaninogenica*	100	Off-white; pink when gathered; creamy texture/easy to pick
UKG6_FNB_14_AER	*Actinomyces radicidentis*	99.57	White; creamy loose texture
UKG6_FNB_15_AER	*Streptococcus mutans*	99.86	Hard with liquid layer (looks dry); required 2 plates
UKG6_FNB_15_AN	*Capnocytophaga granulosa*	99.42	Red/orange on plate, peach colour on loop (required 2 plates)
UKG6_FNB_2_AN	*Streptococcus oralis*	99.85	Texture: semi-hard; dark beige colour; easy to collect
UKG6_FNB_8_AER	*Actinomyces oris*	99.86	Dry, hard texture; white; required 2 plates
UKG6_GC_4_AN	*Streptococcus mutans*	100	Matt appearance; looks hard texture but watery to gather (required 2 plates)
UKG6_GC_4_MA	*Streptococcus rubneri*	99.85	White on plate; medium texture; tan when gathered
UKG6_GC_5_MA	*Streptococcus timonensis*	99.71	Medium texture; tan on loop
UKG6_R2A_3_AN	*Streptococcus gordonii*	100	Small colonies; white on plate, tan when gathered; irregular haemolysis
UKG6_R2A_8_AN	*Limosilactobacillus fermentum*	100	Watery; white
UKG6_STSB_1_AN	*Streptococcus mutans*	99.86	Matt on plate; watery to gather (looks hard); single colonies: yellow tint

4.4.3 Community differences between defined and complex GC sample

Both UKG6 bioreactor vessels supported microbial growth over the 20 day bioreactor run. As can be seen in Figure 41, there was overlap of genera detected in both communities. Many genera detected in the complex community were not detected in the defined community, reflecting that more taxa were maintained *in vitro* then were cultivated by traditional isolations used to formulate the synthetic community. In both vessels *Streptococcus* was relatively the most abundant genera initially. The changes seen when pH was altered are discussed in section 4.4.4.

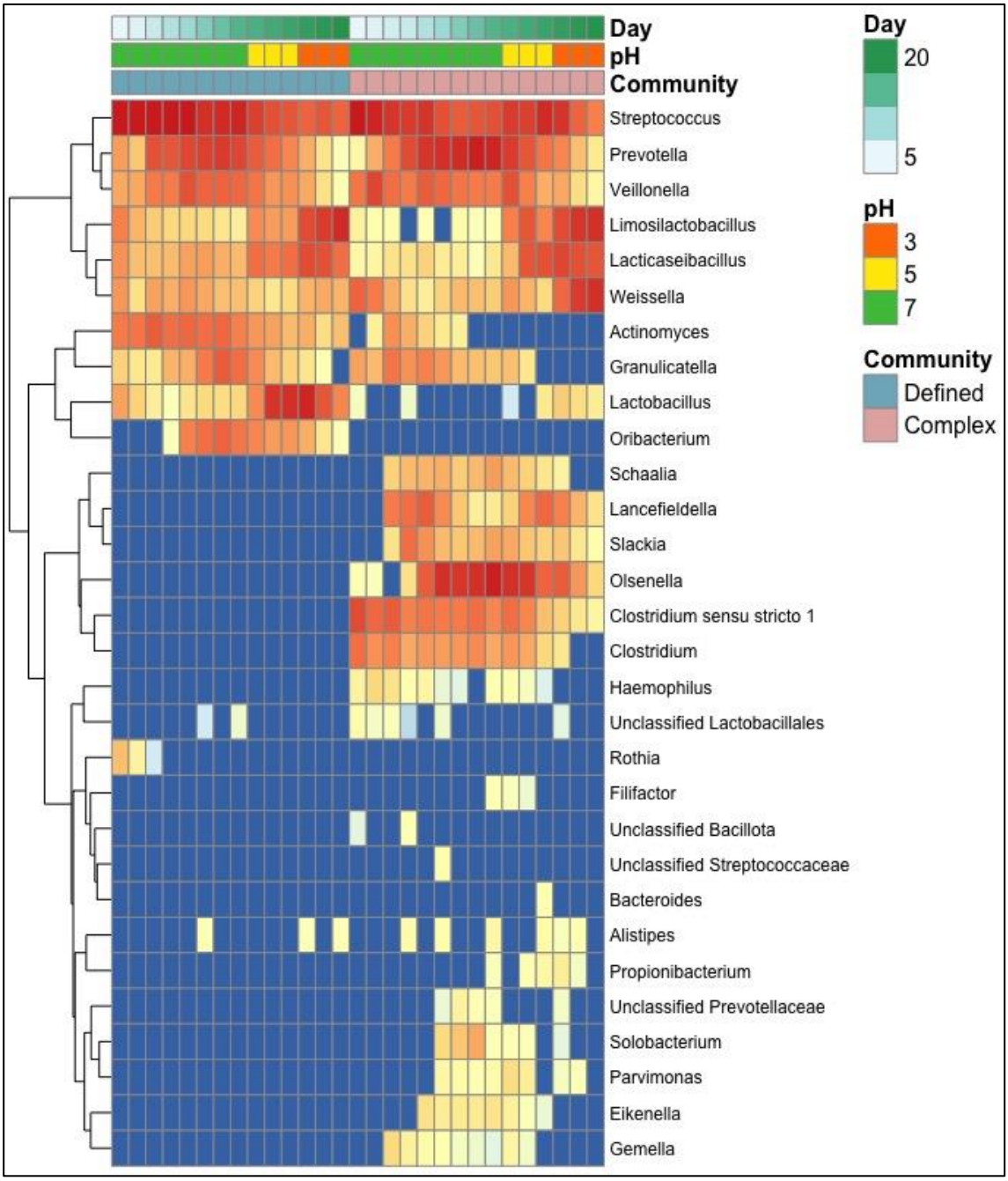


Figure 41. Heatmap of microbial genera detected in defined and complex GC bioreactor communities. The thirty most abundant genera are shown. Cells are coloured according to proportion of microbial reads per sample (in log scale). Branching on the left signifies clustering of similar species. There was overlap with genera detected in both communities; many genera detected in the complex community were not detected in the defined community.

The complex bioreactor community supported more microbial diversity than the defined community by Shannon index; t-test p = 0.022. The difference in Shannon index between the two communities is shown in Figure 42. As can be seen in Figure 43, for both the complex and defined communities, the diversity was low at day 1, and generally followed a similar pattern in both communities, except for a lower Shannon index on day 6 in the defined community.

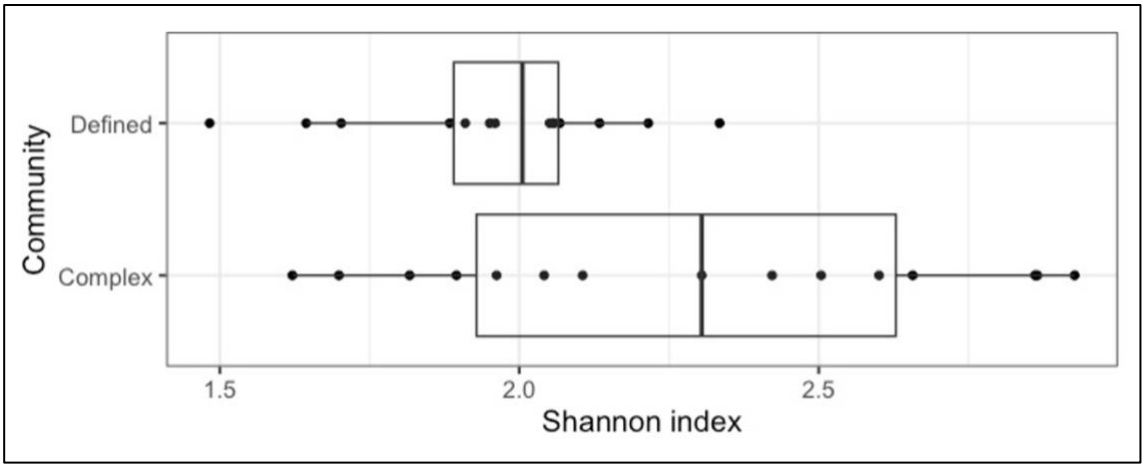


Figure 42. Boxplot of Shannon index of UKG6 bioreactor samples, according to community (defined versus complex).

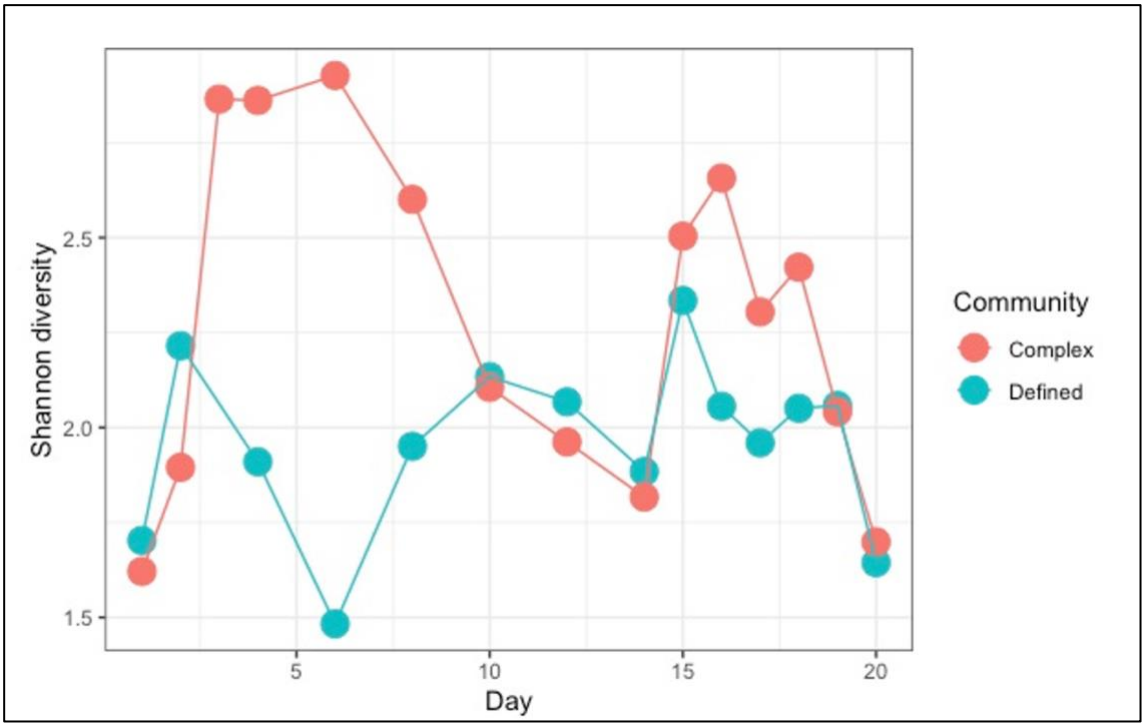


Figure 43. Line plot showing change in Shannon index of UKG6 bioreactor samples over time, according to community (defined versus complex).

The principal co-ordinate analysis plot (Figure 44) demonstrates compositional differences between the defined and complex communities. The clustering of points according to community (indicated by colour) demonstrates the compositional differences between these communities. The observed clustering by date (point size) demonstrates similar compositional changes occurring with time.

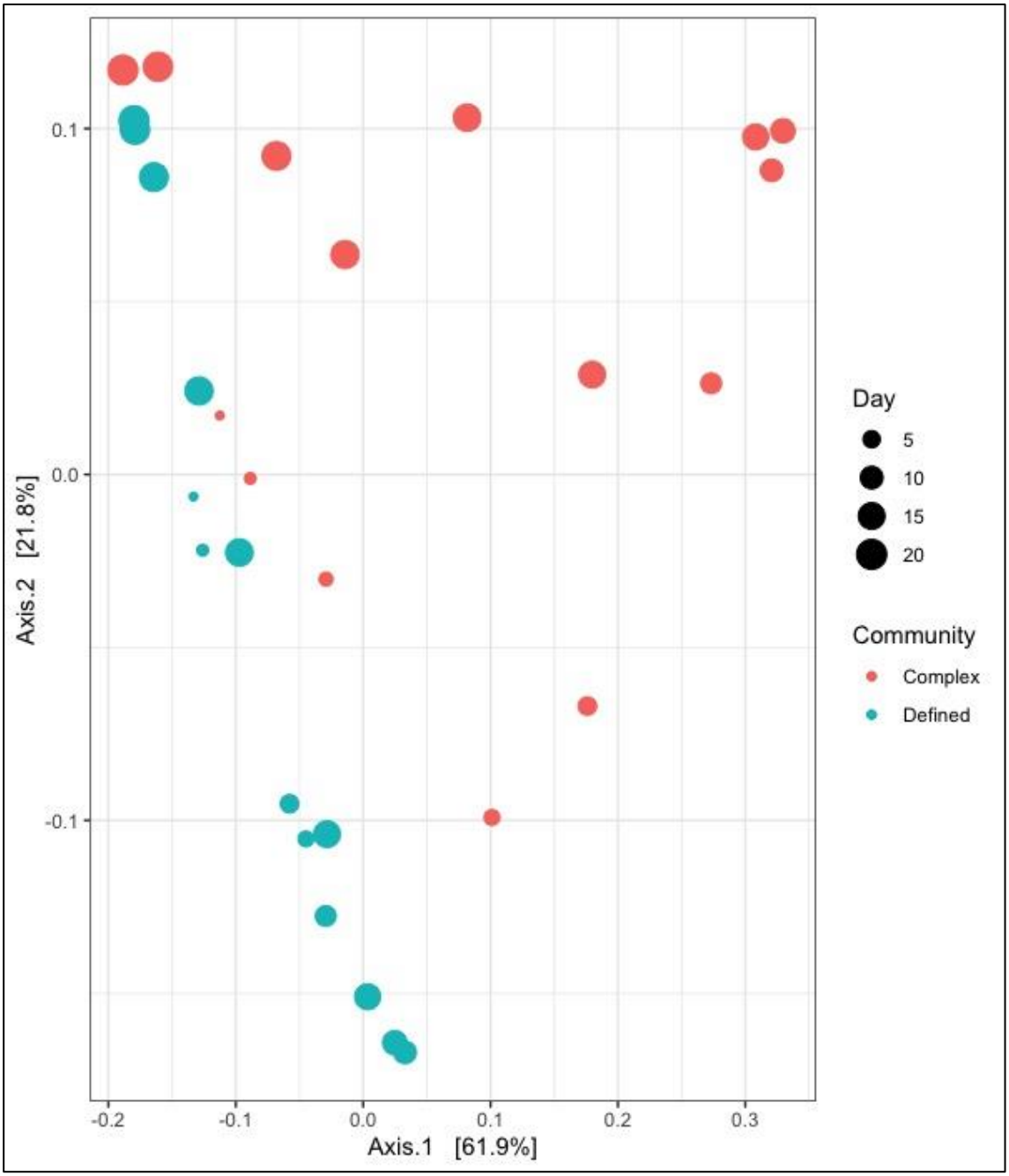


Figure 44. Principal co-ordinate analysis plot compositional differences by unweighted UniFrac distances of species by bioreactor community (complex versus defined). Colour represents community; point size represents day of sampling. The percentage variation associated with the plotted principal coordinates is indicated in the axis legend.

When combined with the tissue sequencing and isolations described in Chapter 3, there was incomplete overlap in taxa identified through the different methodologies. This is demonstrated in Figure 45, where detection of species is shown, according to the research methodologies employed. This figure demonstrates that, whilst there was some overlap, the different methodologies resulted in detection of different species. Many species identified through sequencing of tumour tissue were not identified through any *in vitro* culture-based methodology, in particular, taxa from the Pseudomonadota phylum.

Bacillota, Bacteroidota, and Actinomycetota appear to be the dominant phyla across all methodologies. Pseudomonadota and Fusobacteria were less commonly detected in through culture methodologies than tissue sequencing.

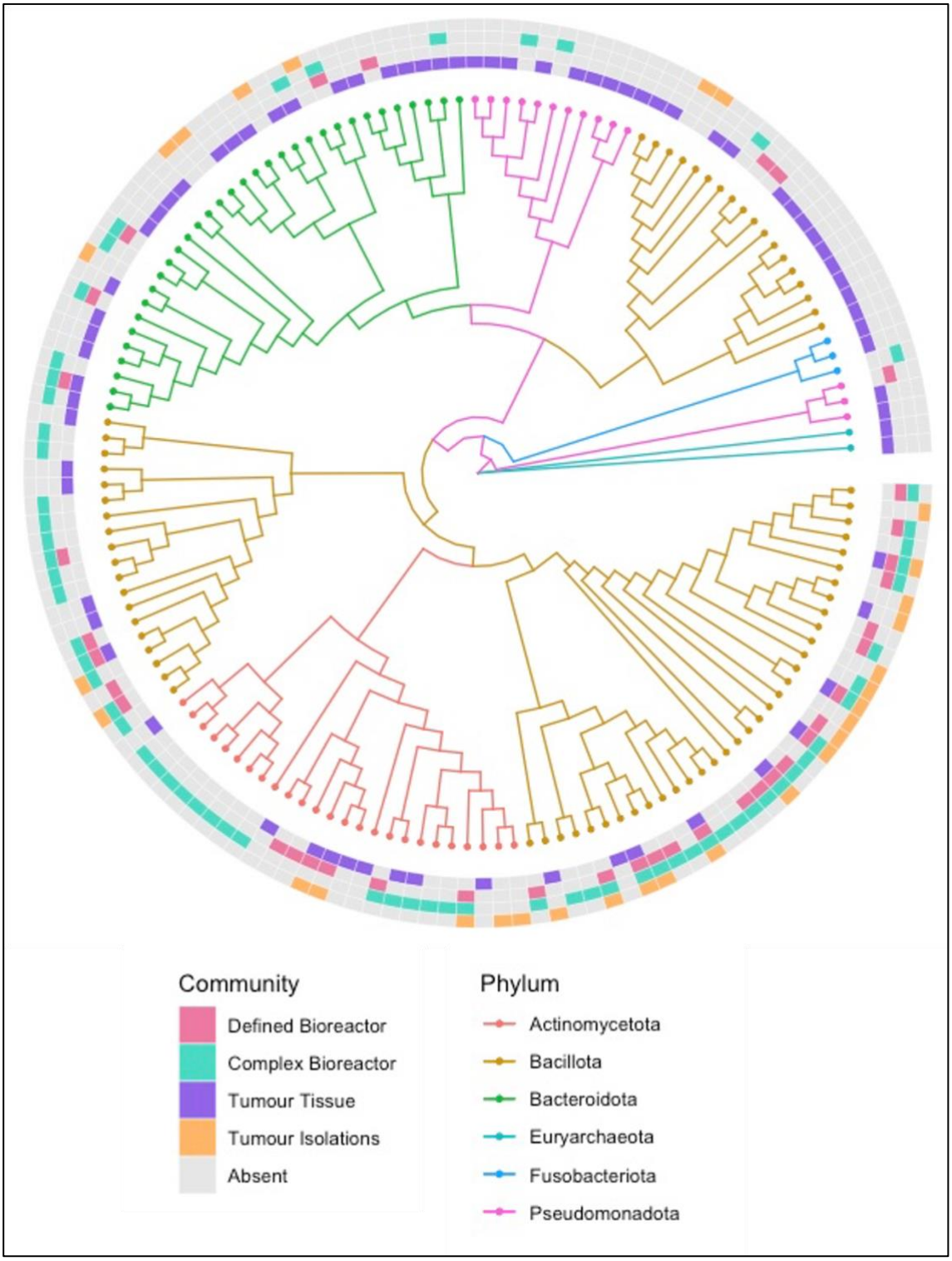


Figure 45. Cladogram showing UKG6 taxa from defined and complex bioreactor vessels, tissue sequencing and tumour isolations, constructed from 16S rRNA gene sequencing data. Branching represents evolutionary similarity. Colour of branches and terminal nodes represents phyla. Colour of circular heatmap demonstrates in which community each species were identified.

4.4.4 Effect of lowering the pH on gastric cancer microbial communities

The pH of both bioreactor vessels was lowered on day 14 (from 7 to 5), and on day 17 (from 5 to 3). The system was able to support this low pH without any equipment issues.

In both the defined and complex communities, many genera reduced in abundance as the pH was lowered (Figure 41). However, the lactic acid bacteria *Limosilactobacillus*, *Lacticaseibacillus*, *Weissella*, and *Lactobacillus* all appeared to increase in abundance with the lowering of pH, in both vessels.

In both communities, the Shannon index appeared to increase in diversity at day 15 (i.e. 24h after the pH was decreased to 5), followed by a steep decline in Shannon index following the lowering of the pH to 3 on day 17 (Figure 43).

The changes in the microbial communities following pH can be seen in the Gram stain images taken on day 10 (pH 7) and day 18 (pH 3) of the defined (Figure 46) and complex (Figure 47) communities. These Gram stain images indicate lower diversity and a change in composition between days 10 and 18 in both bioreactor communities.

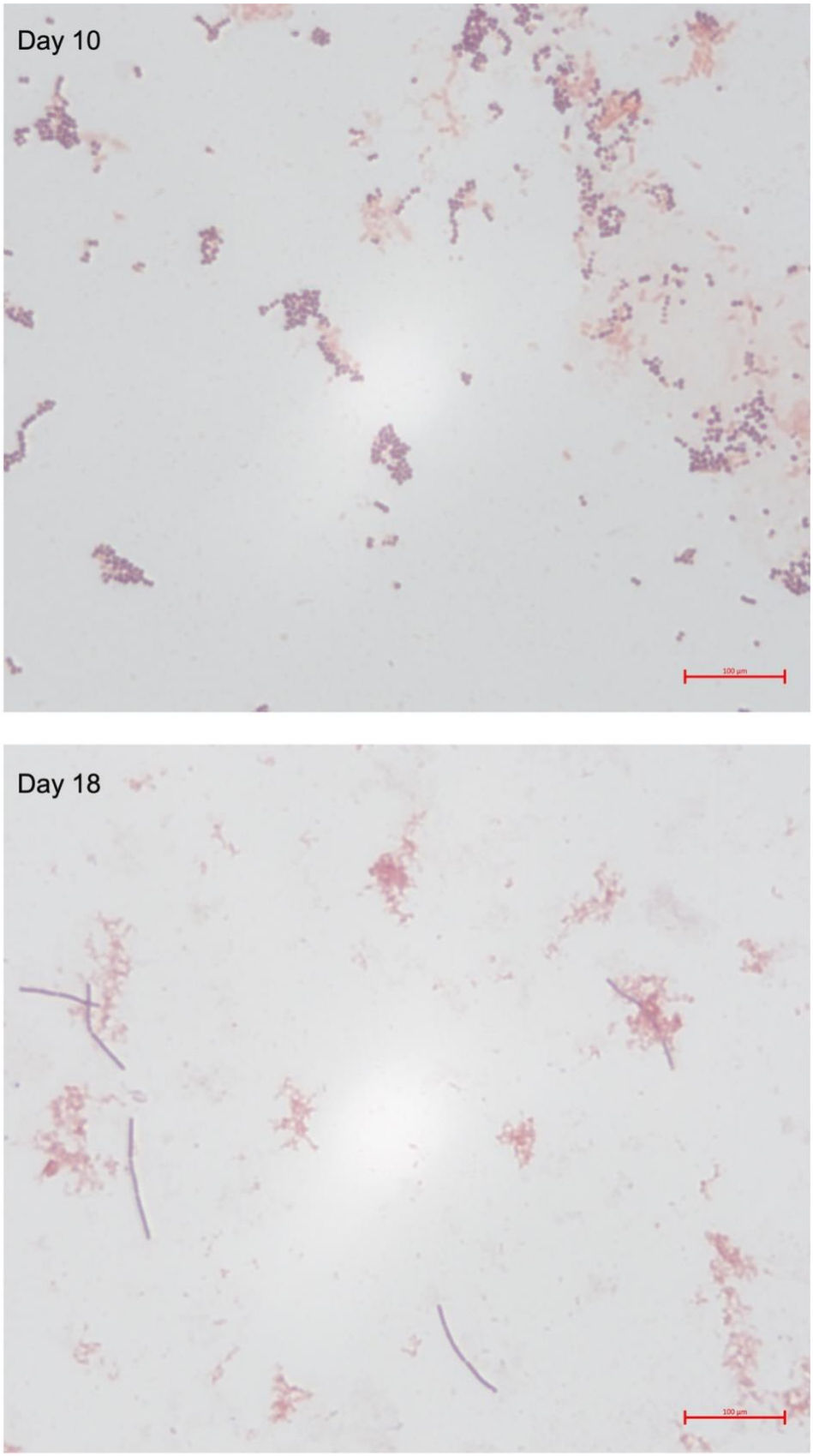


Figure 46. Gram stain images of UKG6 defined community vessel samples on day 10 (upper) and day 18 (lower). A greater proportion of gram-positive rods or chain-forming cocci is observed on day 18 than day 10. Red scale bar indicates 100µm.

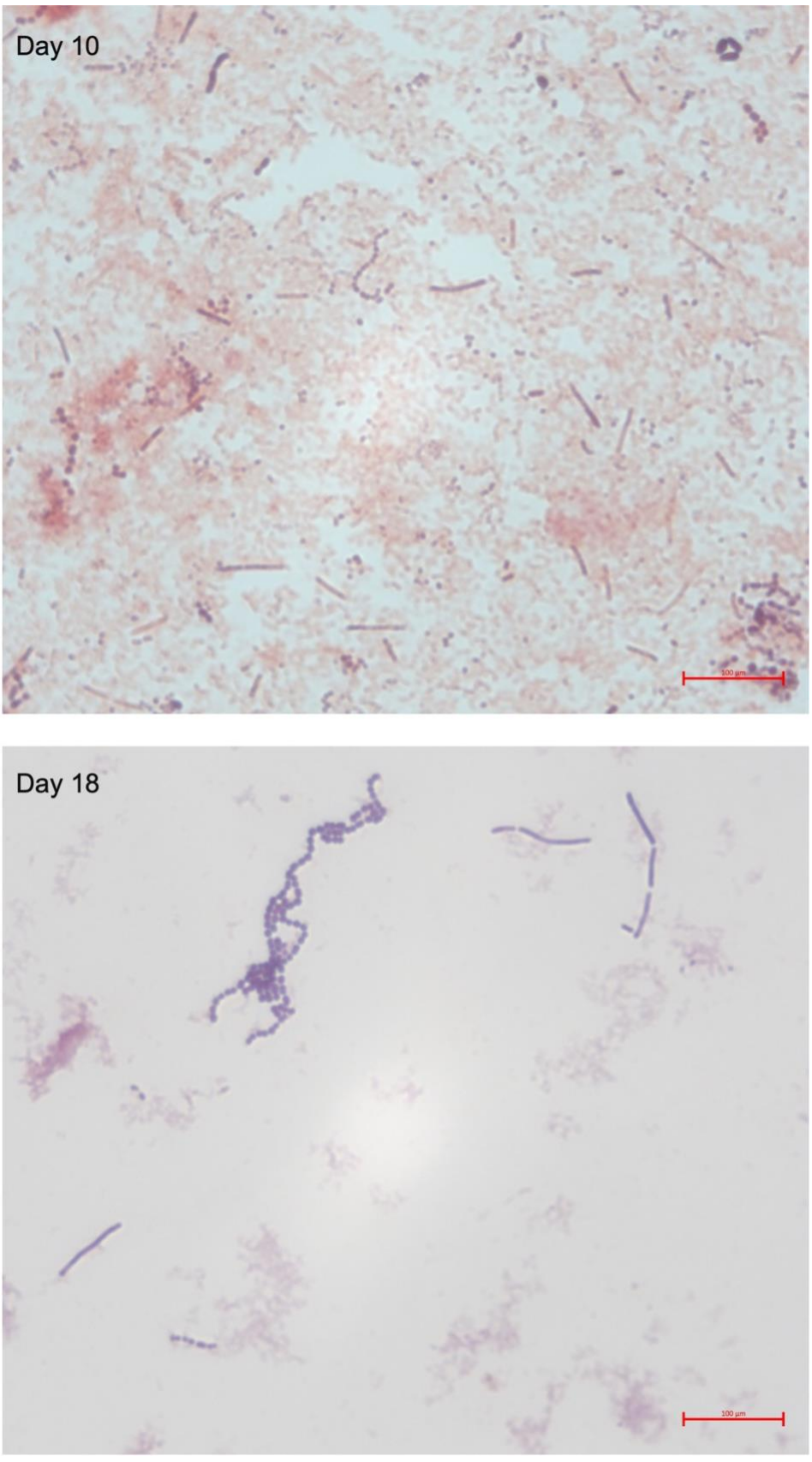


Figure 47. Gram stain images of UKG6 complex community vessel samples on day 10 (upper) and day 18 (lower). A greater proportion of gram-positive rods and gram-positive cocci is observed on day 18 than day 10. Red scale bar indicates 100µm.

Upon lowering the pH from 5 to 3, a visible layer of biomass was observed on the walls of both vessels, indicated by a paler cream coloured deposit on the vessel walls (Figure 48).

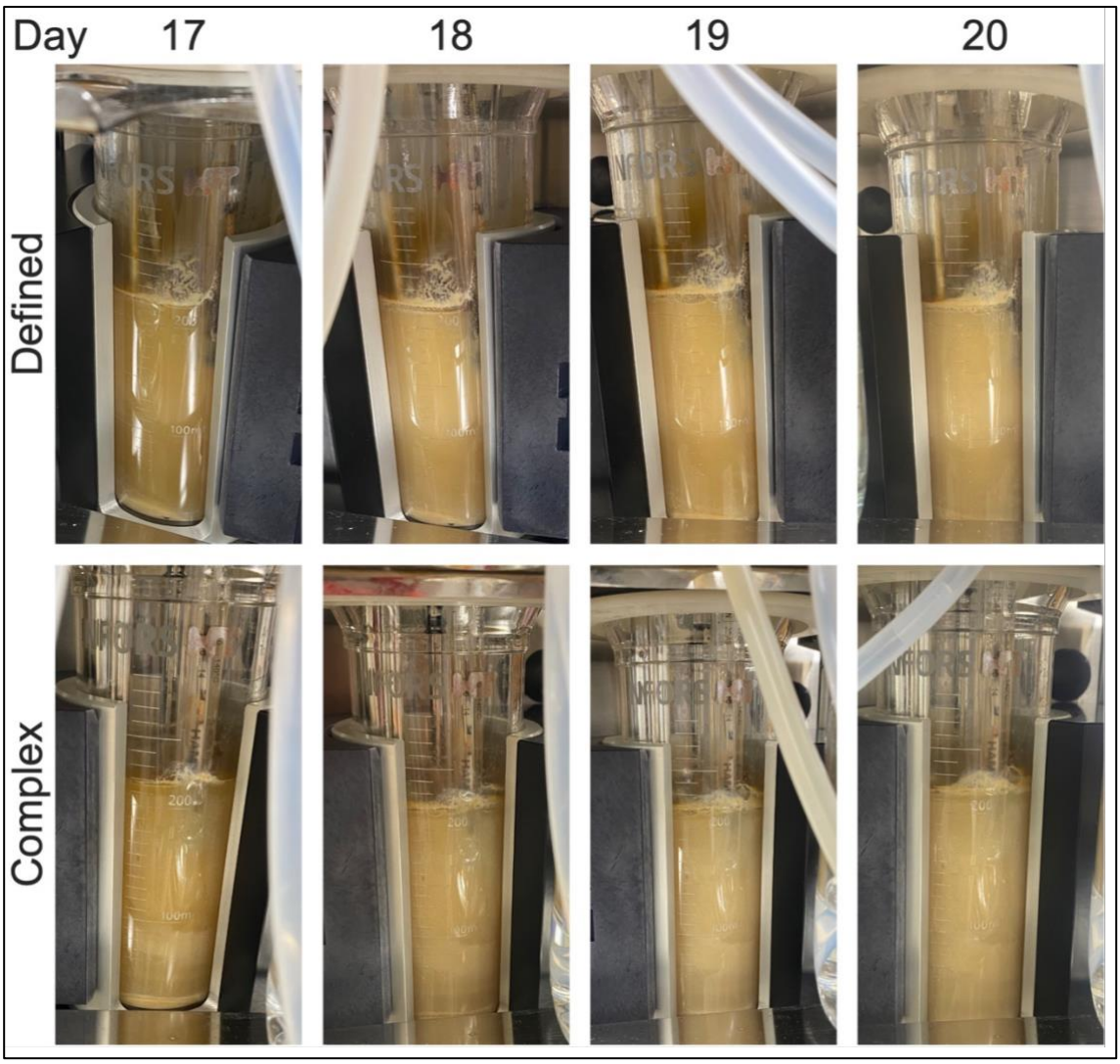


Figure 48. Photograph of biofilm development on walls of bioreactor vessels days 17 to 20, following pH lowering from 5 to 3. Increase in biofilm is demonstrated by paler coloured cloudy appearance on walls of vessels. The top row shows the defined community vessel; the bottom row shows the complex community vessel.

4.4.5 One-dimensional ^{1}H NMR Spectroscopy

One-dimensional 1H NMR Spectroscopy was performed on samples from both bioreactor experiments.

Within the UKG7 (low-biodiversity) vessels, acetate, ethanol, formate, and lactate were the major metabolites detected. Butyrate was not detected in either vessel. In both vessels, acetate concentrations increased after day one, although was maintained at a higher level in vessel B than vessel A. Ethanol levels increased following day one then plateaued in both vessels. Formate concentration was higher on day 4 in both vessels than day one; in vessel B the formate concentration continued to steadily rise after day 4, whereas the formate concentration dropped after day 4 in vessel B. Lactate was highest on day 1 in both vessels, with a similar pattern seen in both vessels. Figure 49 demonstrates the changes in concentrations in both vessels over time.

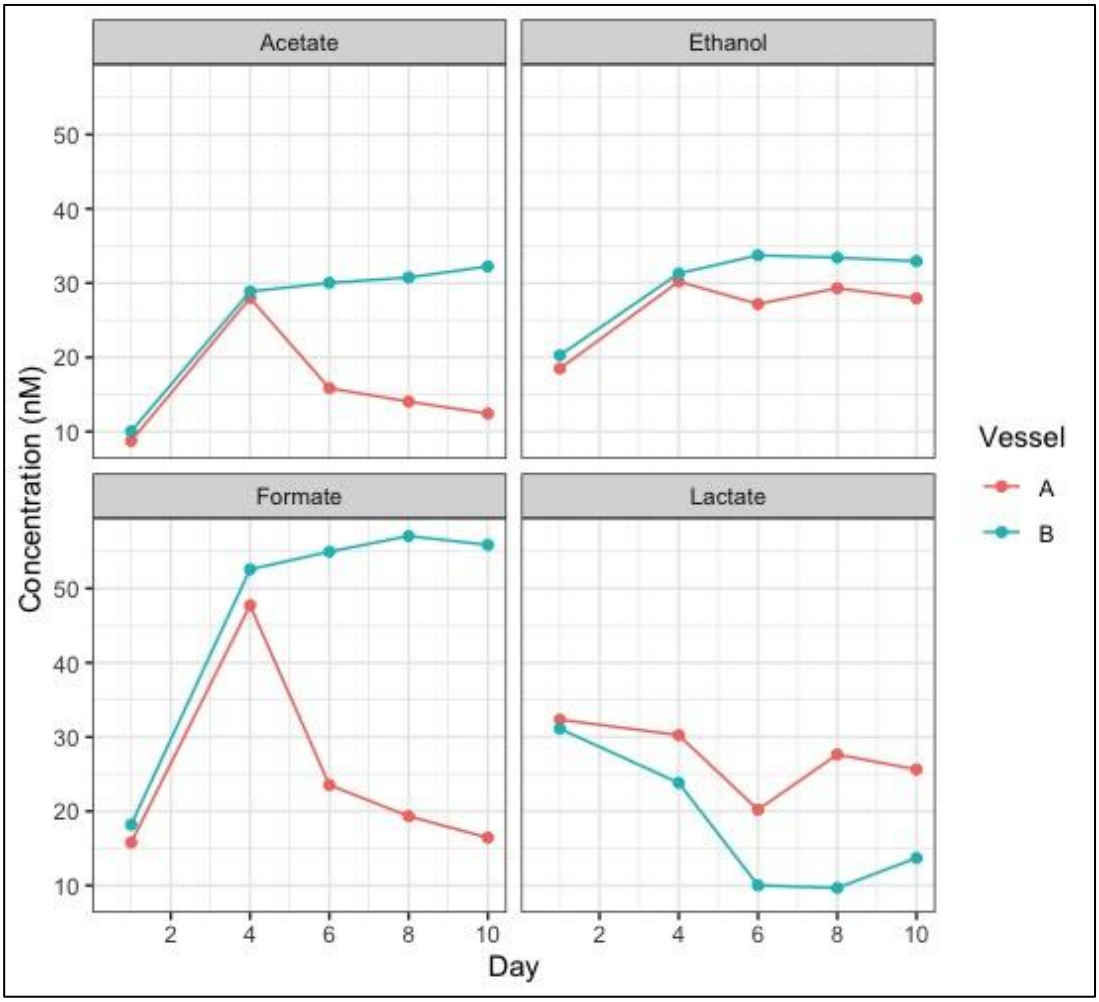


Figure 49. Metabolite concentrations detected in UKG7 bioreactor communities over time. Both vessels demonstrated similar patterns of change in metabolites, although vessel A produced more formate than vessel B.

Within the UKG6 vessels, acetate, butyrate, ethanol, formate, lactate, and propionate were the major metabolites detected. In both vessels, acetate concentrations rose sharply after day one, and maintained fairly constant until after day 14, when the concentration lowered. Butyrate was only detected in the complex community vessel and the concentration diminished after day 14. Ethanol levels remained fairly constant in both vessels, with a similar pattern of a lowering from around day 15, with a slight increase in concentration by day 20. Formate concentration rose following day 1 in both vessels; the formate concentration was higher throughout in the defined community than the complex community; in both communities the concentration dropped substantially after day 15. Lactate was highest on day 1, before dropping to much lower/negligible concentrations in both vessels until day 15, where there was a second (smaller) peak until day 17. Propionate concentrations were similar in both vessels, although initially higher in the complex community than the defined community; concentrations dropped in both vessels from day 15. Figure 50 demonstrates the changes in concentrations in both vessels over time.

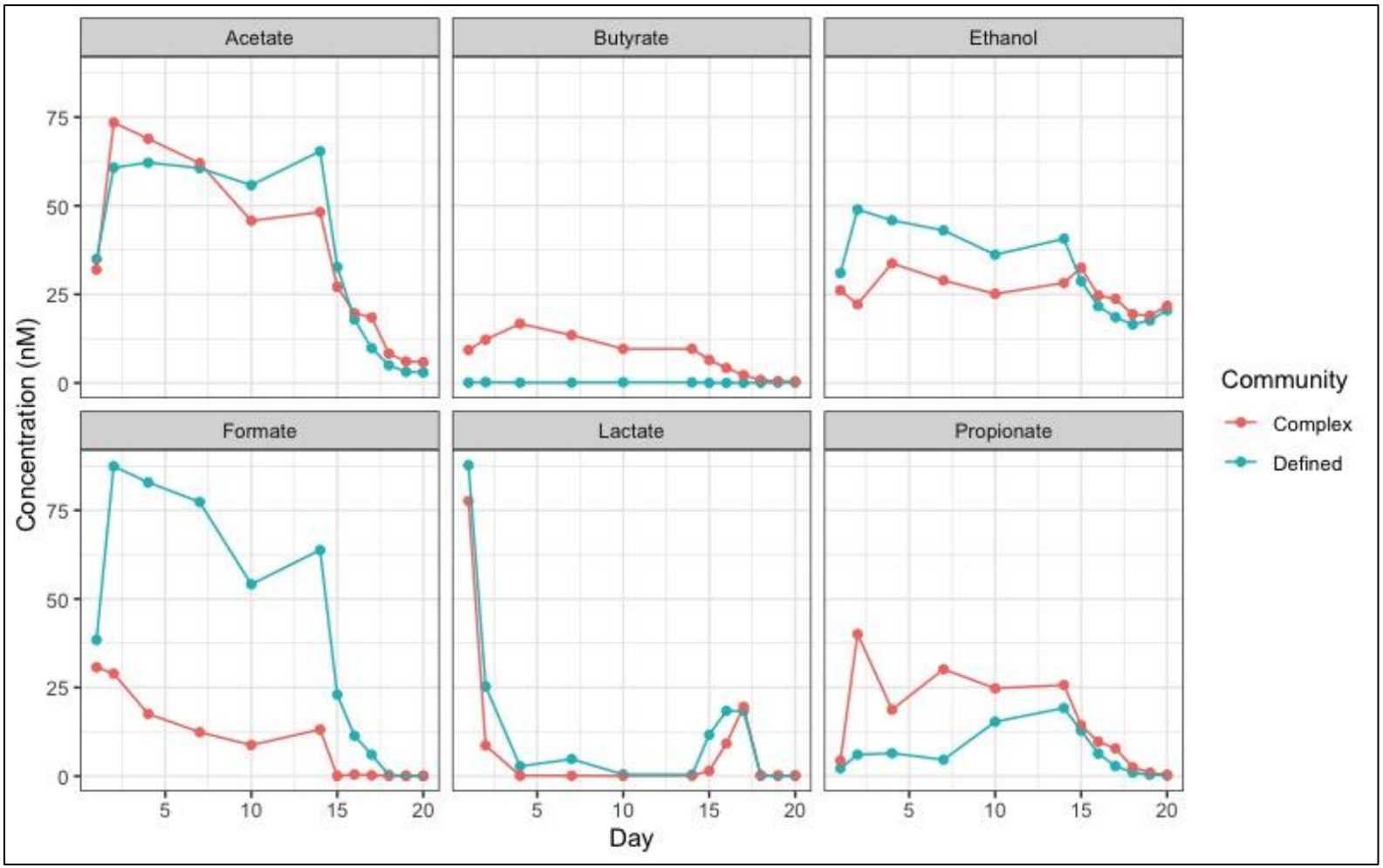


Figure 50. Metabolite concentrations detected in UKG6 bioreactor communities over time. Bioreactor community (defined versus complex) is indicated by colour. Butyrate was only produced in the complex community; higher concentrations of formate were produced in the defined community. Lactate was produced in batch phase and at pH 5 in both bioreactor communities.

4.5 Discussion

4.5.1 Development of a bioreactor model to investigate GC microbiota

The primary aim of the work presented in this chapter was to develop a bioreactor system capable of supporting and amplifying GC microbiota. The work presented here demonstrates that the described GC bioreactor model was able to support and amplify the microbiota from both a low-biodiversity GC sample, as well as both complex and defined communities from a more biodiverse GC sample. This represents the first bioreactor model used to investigate the GC microbiota.

4.5.2 Assembly and use of a gastric cancer microbial community

A defined GC community of 44 bacterial strains was assembled, which was demonstrated to reflect that of the corresponding complex community, albeit with reduced alpha diversity. In line with what has been observed in defined communities from other microbial environments [145], not all bacterial genera observed in the complex communities were observed in the defined communities. The similar patterns of community shift seen with the change in pH, as well as the broadly similar patterns of metabolite concentrations between the defined and complex communities support the use of the UKG6 defined community as a standard defined community for further research use. This defined community holds great potential value for future microbiome investigation of the GC microbiome. For example, testing of the impact of probiotics and antibiotics, and other drugs upon community composition; investigation of the impact of diet upon community function through modulation of the media feed; and further investigation of pH.

Whilst use of the defined community offers reproducibility and does not depend upon access to fresh tissue, advantages of using complex communities exist. First, as demonstrated by the heatmap of defined and complex communities, some taxa present in the complex community were not present in the defined community. This is a recognised disadvantage of using defined communities [145]. Second, the use of complex communities allows the selection of specific GC tissue which may best answer a specific research question, for example, if tissue sequencing has previously identified a particular microbe or microbes of

interest, an adjacent GC tumour sample may be used to investigate the community through use of a complex community as described in this chapter.

With the exception of the genus *Oribacterium*, genera identified through sequencing of the defined community bioreactor were genera deliberately introduced into the bioreactor through the defined community. *Oribacterium* was not included in the defined community, and therefore unexpected within this bioreactor community. This phenomenon has been previously noted and may be explained by the presence of 'contaminating' strains originating from the same environment from which the defined community was generated from, existing undetected in low levels in the assumed pure stocks and subsequently allowed to bloom to detectable levels in the bioreactor [324].

When the sequencing results were compared across the different methodologies, including tissue sequencing, bioreactor sampling, and tissue isolations, there was minimal overlap in identified taxa. Bacteria identified through culture techniques that were not identified through tissue sequencing may have been present at lower concentrations than the sequencing threshold. Of the five phyla commonly in the GC microbiome, in the present study Bacillota, Bacteroidota, and Actinomycetota appear to be the dominant phyla across all sequencing methodologies, whereas Pseudomonadota and Fusobacteria were less commonly detected in through culture methodologies than sequencing. The discrepancies between methodologies could be either a result of technical difficulties in culturing certain taxa, or true differences in live bacteria (i.e. the sequencing methodologies have detected non-viable bacteria).

4.5.3 Investigating the impact of pH upon gastric cancer microbial communities using the gastric cancer bioreactor model

The importance of pH in GC microbiota is not fully understood. As described in the introduction of this chapter, the low pH of the gastric fluid does not necessarily result in a lower pH of the microenvironment which GC microbiota inhabit. However, since the exact pH of the colonised GC environment is not known, and because certain factors (e.g. PPIs, atrophic gastritis) can impact the gastric fluid pH and gastric fluid microbiota [310,311], it is reasonable to explore to what extent pH may impact the GC intratumoural microbiota. It should be noted that this work was exploratory in nature, first investigating whether the

bioreactor could be used to investigate pH in such a way, and second to observe patterns of community shift with a change in pH.

The bioreactor system supported a pH change from 7 to 5, and from 5 to 3, in both vessels. The four genera which increased in relative abundance across both vessels as pH was reduced, *Limosilactobacillus*, *Lacticaseibacillus*, *Weissella*, and *Lactobacillus,* are all lactic acid bacteria, which produce lactic acid as the major end product of carbohydrate fermentation [325]. Lactate concentration was observed to increase in both vessels when pH was 5, corresponding with the relative increase in these lactic acid bacteria, although it was not detected on days 18-20, at pH 3. Interestingly, even though negligible lactic acid concentrations were produced on days 18-20, the relative abundance of *Limosilactobacillus* and *Weissella* still appeared to increase, despite a pH of 3.

Whilst not cultured in this bioreactor, it has been postulated that *H. pylori* infection results in GC through the reduction in stomach acidity, leading to a new dominant microbiota [110]. It is recognised that *H. pylori* infection leads to atrophic gastritis in some individuals, likely resulting from of chronic inflammation [91]. This gastric atrophy is defined by loss of glandular tissue, including the parietal cells responsible for gastric acid production [326]. de Assumpçäo *et al*. [110] have hypothesised that the resulting hypochlorhydria changes alters the microenvironment in a way that allows colonisation of a new carcinogenic microbiota. de Assumpçäo's hypothesis has been difficult to assess, due to the inherent difficulties of defining what the so-called "carcinogenic microbiota" is, particularly due to its potential transient nature. However, tissue sequencing studies have demonstrated compositional differences between patients with chronic gastritis and *H. pylori* induced atrophic gastritis, with healthy controls [124,210]. The GC bioreactor model offers the opportunity to investigate the impact of changes such as pH on the GC microbiota. Furthermore, the indication that certain lactic acid bacteria genera were able to withstand a lower pH within this community and environment raises the question of whether they may play a protective role against GC development. The value of the bioreactor as a research tool is not in assessing carcinogenic potential (for which GC cell lines, organoids, or animal models would be required), but instead in investigating how microbial communities may be altered by changes to host environment, thereby generating hypotheses which may be tested in further *in vitro* or *in vivo* laboratory studies.

4.5.4 Understanding microbial community differences in low-biodiversity sample bioreactor replicates

The results from the UKG7 (low-biodiversity) sample demonstrated some differences between the two vessels acting as biological replicates. The genus *Lancefieldella* was present in only vessel A, and the species *Streptococcus mutans* was present only in vessel B. This differences between vessels may have been a result of incomplete homogenisation of the tissue used to inoculate the vessels. To improve this, a more thorough homogenisation process, by increasing the time of vortexing, could be employed. Alternatively, these species may have been in very low abundance, resulting in their presence in only one inoculum.

As expected, the bioreactor isolations did not result in identification of any taxa in addition to those identified in the bioreactor sequencing of daily samples. However, it did result in isolation of different species from those obtained from the UKG7 tissue sample. Species identified through bioreactor culture which were not identified through tissue isolations may be a result of the bioreactor providing an optimum environment for bacteria at low concentrations to allow them to be amplified sufficiently to allow detection through sequencing and/or isolation. Species identified through tissue isolations which were not identified through bioreactor isolations may not have had the optimum conditions in the bioreactor for growth, or may have been outcompeted by other species.

4.5.5 Metabolic outputs from bioreactor communities

One-dimensional ^{1}H NMR Spectroscopy was used to generate metabolic profiles across both bioreactor runs. Across both UKG7 and UKG6, the major metabolic products were the short-chain fatty acids (SCFA) acetate, formate, butyrate, and propionate, as well as ethanol and lactate. Acetate, butyrate, and propionate are the three main metabolites produced by colonic bacterial fermentation of dietary fibres and resistant starch [327]. Interestingly, butyrate was only produced in the complex UKG6 community, not the defined community nor either of the UKG7 vessels. Butyrate is known to act as an energy source for colonocytes (along with acetate and propionate), as well as decreasing intestinal inflammation [328] and may be preventative against colorectal cancer [329]. Presence of butyrate-producing bacteria in the oral cavity is associated with periodontitis [330]. The role of butyrate in GC is not fully understood, however there is *in vitro* evidence suggesting butyrate may

inhibit proliferation and induce apoptosis in GC cell lines [331], potentially warranting further investigation of the role of this SCFA in GC.

Previous investigation of the colorectal microbiota has not found formate to be produced at high concentrations [305]. In contrast, formate was found to be a major metabolite produced in the present study GC bioreactor experiments. In colorectal cancer, microbiome-derived formate has been shown to drive tumour invasion in *in vitro* experiments [332]. Collectively, these observations warrant further investigation of the role of microbiome-derived metabolites in the role of GC development and progression.

4.5.6 Advantages and limitations of the gastric cancer bioreactor model

The GC bioreactor model offers a number of advantages. Through the use of either a complex (tissue) community, or through a defined community, community dynamics may be studied over time. The impact of changing environmental conditions may be studied, as demonstrated by changing pH over time. Furthermore, the twin system can be utilised to compare controlled conditions beyond those employed in the present study, for example the impact of diet through different media or the impact of antibiotics of interest. As demonstrated in the cladogram showing UKG6 taxa from defined and complex vessels, tissue sequencing, and tumour isolations, different bacteria are identified through the different characterisation methodologies. As such, they may have a complementary role in characterising present bacteria. Another advantage of the GC bioreactor system is the ability to study metabolites produced by the microbial community, as demonstrated by the NMR results presented in this chapter. Further applications of the GC bioreactor model could involve investigation of the impact of diet, antibiotics, chemotherapy, and PPIs upon the GC microbiome, as well as investigation of particular microbes of interest through inclusion or exclusion of microbes of interest in defined communities.

There are a number of limitations to the GC bioreactor model. The GC bioreactor model was adapted from the Robogut model, initially designed to study colorectal luminal contents [300]. However, the aim of the GC bioreactor model was to simulate the intratumoural environment, not the luminal contents of the stomach. Therefore, the model lacks the cellular complexity and tissue

structure of the intratumoural environment. To try to simulate the intratumoural environment entirely would be extremely complex and resource-intensive. However, future adaptations may be considered to make the model more representative of the intratumoural environment, such as incorporating mucus beads to encourage a mucus-rich environment or integrating with an organoid system to study interactions with GC cell lines. The GC medium was developed by adapting the standard distal gut chemostat medium, informed by a study assessing the components of the UK diet. This medium has not been validated, although to do so would be technically difficult as there would be no standard against which to validate it.

4.5.7 Chapter summary

- The Robogut bioreactor model and adult chemostat media were adapted to support GC microbiota in continuous *in vitro* culture.

- A defined GC community was assembled, which was reflective of the complex GC community in terms of community composition and metabolite production.

- The GC bioreactor model was capable of supporting a low-biodiversity GC community, as well as complex and defined communities from a more biodiverse GC sample.

- The GC bioreactor model can be used to investigate the impact of low pH on GC microbiota. At low pH, lactic acid bacteria increase in relative abundance.

- The GC bioreactor model can be used to investigate metabolic outputs. Formate was found to be a major metabolic product within the GC bioreactor experiments.

Chapter 5 Summary and Future Directions

In this chapter, the key findings of the work presented in this thesis are presented and future research directions in the field of the GC microbiome are explored.

5.1 Summary of research findings

5.1.1 Characterising the gastric cancer microbiome using in silico data

A multitude of studies have previously been published comparing the GC microbiome to healthy controls, precancerous conditions, or adjacent tissue. The aim of the *in silico* work was not to repeat these studies, but instead, to define the intratumoural microbiome of established GC and explore relationships between clinicopathological features and the GC microbiome.

Using sequencing data from the 100,000 Genomes Project and TCGA, the intratumoural microbiome was described. Common genera included *Prevotella*, *Selenomonas*, *Streptococcus*, *Stomatobaculum* and *Lactobacillus*. The majority of most commonly identified bacteria are recognised as human oral microbes.

MSI, low pT, non-Asian origin, and intestinal histological phenotype were all associated with increased microbial abundance and alpha diversity. Beta diversity analyses suggested that the GC microbial composition differed according to MSI status, patient sex, and pT category.

5.1.2 Characterising the microbiome from gastric cancer tissue samples

Samples were prospectively collected from seven patients with GC. Characterisation of the GC samples through metagenomic sequencing, 16S rRNA sequencing, and bacterial isolations was considered representative of GC microbiota.

Comparisons of sequencing methodology demonstrated the importance of performing a decontamination step in metagenomic approaches, particularly in

the use of FFPE GC tissue. It was demonstrated that the include-list developed in Chapter 2 could be used to perform decontamination sequencing data from low-coverage WGS of FFPE GC samples which resulted in a similar microbial composition that of low-coverage WGS of fresh-frozen GC samples, although there was substantial loss of microbial reads. At genus level, results of metagenomic sequencing and 16S rRNA sequencing were similar; choice of sequencing approach should be guided by sample size, cost, and required taxonomic resolution.

Bacterial isolations were performed on five GC samples, using the most extensive set of isolation conditions performed to date. This resulted in isolation 113 bacterial strains belonging to 26 genera, representing the greatest number of genera isolated from any single GC isolation study date [243–248]. This work provided evidence for an underlying assumption of many observational sequencing studies: the intratumoural GC environment is capable of supporting a viable microbiota.

5.1.3 Development of a bioreactor model to culture gastric cancer microbes

A GC bioreactor model was developed by adapting the Robogut model [314]. This involved adapting the medium supplied to be more reflective of the intratumoural GC environment. In addition, a defined GC community was developed from the UKG6 GC sample isolations. The GC bioreactor model was capable of supporting and amplifying GC microbiota from a low-biodiversity sample, as well as complex and defined communities from a more biodiverse GC sample. The GC bioreactor was used to investigate the impact of low pH on GC microbiota, which resulted in an increase in relative abundance of lactic acid bacilli.

One-dimensional 1H NMR Spectroscopy was used to investigate the metabolic outputs of the GC microbial communities in the GC bioreactor model. Different profiles of metabolites were observed to be produced by GC microbiota, compared to what has previously been observed in colonic fermentation of dietary material. In particular, formate was found to be a major metabolite produced in the GC bioreactor experiments, in contrast to the colorectal microbiota, which does not produce formate is not a major metabolic product [305].

5.2 Future research directions

Future work investigating the GC microbiome should be grounded in improving outcomes for patients with GC. The work presented in this thesis has further reinforced that GC is a heterogeneous disease, and future GC research, including microbiome research should take this into account.

5.2.1 Further investigating the relationship between clinicopathological factors and the gastric cancer microbiome

MSI, low pT, and intestinal histological phenotype are all recognised as good prognostic factors in GC [220–223] and were associated with increased microbial abundance and diversity. Future research is therefore warranted on the role of microbial abundance and diversity in GC behaviour, and whether increased microbial abundance and greater alpha diversity is contributing to this effect.

The findings presented in this thesis demonstrated a more abundant, more diverse microbiome in MSI GC, compared to MSS GC. This finding is particularly interesting given that MSI GC responds better to immunotherapy than MSS GC to immunotherapy [28]. The current understanding is that this is a result of an increased number of neoantigens present on MSI GC cells [19], although the role of the microbiome has not been investigated in with respect to the MSI-immune response relationship to date. Further investigation of local immune cells and the microbiome is warranted, particularly in light of the relationships between MSI GC and the microbiome and the current lack of research into the relationships between the GC microbiome and local immune cells. Use of the inferred immune cellular fractions and immune subtype data did not reveal immune-microbiome relationships that progress our understanding how the microbiome differs according to clinicopathological factors. Future research, therefore may require more sophisticated approaches, such as taking into account spatial location of immune cells, and a more detailed consideration of specific immune cell subtypes.

5.2.2 The role of the gastric cancer microbiome in response to immunotherapy

The role of the GC microbiome in response to immunotherapy is an area of great potential importance, yet remains unexplored to date. Over the past five to ten years, immunotherapy has emerged as an effective treatment option for selected patients with advanced GC [333–335]. However, response rates vary, even in biomarker-selected patients. In other cancer types, there is suggestion that the microbiome may be modulated to improve response to immunotherapy. A phase I study by Davar *et al*. investigated the effect of faecal microbiota transplantation (FMT) and anti-programmed cell death protein-1 (PD1) therapy from anti-PD1 responders to patients with melanoma refractory to anti-PD1 therapy [336]. This study reported objective responses in three and stable disease in a further three patients out of a total of 15 patients. Riquelme *et al*. performed human-to-mice FMT experiments from patients with pancreatic cancer into orthotopic mouse models [51]. This study demonstrated a reduction in tumour growth in mice receiving FMT from long term survivors of pancreatic cancer relative to mice receiving FMT from likely-short term survivors or healthy controls. Collectively, these studies suggest that the faecal microbiome may influence tumour behaviour and response to immunotherapy. Given the observed relationship between MSI and the intratumoural GC microbiome in the present study, investigation of the microbiota (intratumoural and faecal) of responders and non-responders to immunotherapy would be of great potential importance. If a “responder microbiome” or “non-responder microbiome” were identified, investigation should subsequently focus on modulation of the GC microbiome with the aim of improving response rates.

5.2.3 Investigation of the gastric cancer fluid microbiome

The microbiome of GC fluid remains relatively understudied. Although not as immediately accessible as stool samples, obtaining gastric fluid samples is less invasive than obtaining mucosal biopsies. In the largest investigation of the gastric fluid of patients with GC fluid to date, a random forest classifier model including 13 genera was constructed to differentiate between superficial gastritis and GC [119]. The receiving operational curve analyses demonstrated a performance of 83% and 89% in the test and validation cohort analyses, respectively, in patients from Japan. If these findings are generalisable across other patient cohorts, this may represent an opportunity to screen patients through the analysis of gastric fluid. Further investigation in other geographical areas is therefore warranted, as may be clinical trials of diagnostic testing or

screening. This study also reported an increase in pH value of gastric fluid in patients with GC, relative to superficial gastritis, providing an argument for the investigation of the impact of pH upon both the GC fluid and mucosal microbiome. It is currently unknown whether the gastric fluid may drive mucosal microbiome changes, representing another potential research avenue.

5.2.4 Optimisation of sample collection and processing

Future research using prospectively collected samples should ideally be performed according to standardised protocols with the aim of minimising contamination. In addition to external contamination, careful attention should be paid to contamination from other body cavities, for example during endoscopic collection of gastric tissue. In addition, where adjacent, non-tumour tissue is collected, the anatomical location of the adjacent tissue should be recorded, to allow analysis of anatomical variation. Collection of clinical metadata, such as PPI use, antibiotic use, and chemotherapy, should be collected as standard since these factors have the potential to impact the microbiome and thus may confound results if not accounted for appropriately.

5.2.5 Preclinical models are required to test hypotheses generated from observational studies

The majority of GC microbiota studies performed to date are observational, limiting interpretation of these results to associations alone. With the exception of *H. pylori*, the role of the microbiota in GC development remains unknown. To test hypotheses generated from observational studies and to understand potential mechanisms of cancer development, functional models are required. The GC bioreactor model described in Chapter 4 represents one such model. However, in order to fully appreciate the complexity of microbiome-cancer relationships, a variety of models, including cell lines and animal models are required, with the specific model guided by the individual research questions.

Such functional studies should investigate how the local microenvironment may affect and be affected by the GC microbiome, how the microbiome may affect GC phenotype, and ultimately whether microbiome manipulation could affect outcomes of patients with GC and/or response to specific therapeutic interventions.

5.2.6 Future research should be multi-disciplinary

The work within this thesis has demonstrated that there is no single “best” way to investigate the GC microbiome. Each methodological approach has its own advantages and disadvantages. To further progress understanding in this area, defined research objectives are required with methods and technologies selected to answer those specific questions. Large-scale sequencing studies and culture-based investigations should be used in order to identify and test hypotheses. This multi-disciplinary research and collaborative approach is required to further progress the field of GC microbiome research to ultimately improve outcomes for patients with GC.

List of References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209–49. https://doi.org/10.3322/caac.21660.

[2] Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. The Lancet 2018;391:1023–75. https://doi.org/10.1016/S0140-6736(17)33326-3.

[3] de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. Lancet Glob Health 2020;8:e180–90. https://doi.org/10.1016/S2214-109X(19)30488-7.

[4] Laurén P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. Acta Pathologica Microbiologica Scandinavica 1965;64:31–49. https://doi.org/10.1111/apm.1965.64.1.31.

[5] WHO Classification of Tumours Editorial Board. WHO Classification of Tumours, 5th edition: Digestive System Tumours. 5th ed. France: IARC; 2019.

[6] Yamada T, Yoshikawa T, Taguri M, Hayashi T, Aoyama T, Sue-Ling HM, et al. The survival difference between gastric cancer patients from the UK and Japan remains after weighted propensity score analysis considering all background factors. Gastric Cancer 2016;19:479–89. https://doi.org/10.1007/s10120-015-0480-5.

[7] Wu H, Rusiecki JA, Zhu K, Potter J, Devesa SS. Stomach carcinoma incidence patterns in the United States by histologic type and anatomic site. Cancer Epidemiology Biomarkers and Prevention 2009;18:1945–52. https://doi.org/10.1158/1055-9965.EPI-09-0250.

[8] Kaneko S, Yoshimura T. Time trend analysis of gastric cancer incidence in Japan by histological types, 1975-1989. Br J Cancer 2001;84:400–5. https://doi.org/10.1054/bjoc.2000.1602.

[9] Hoyo C, Cook MB, Kamangar F, Freedman ND, Whiteman DC, Bernstein L, et al. Body mass index in relation to oesophageal and oesophagogastric junction adenocarcinomas: a pooled analysis from the International BEACON Consortium. Int J Epidemiol 2012;41:1706–18. https://doi.org/10.1093/IJE/DYS176.

[10] Ye W, Chow WH, Lagergren J, Yin L, Nyrén O. Risk of adenocarcinomas of the esophagus and gastric cardia in patients with gastroesophageal reflux diseases and after antireflux surgery. Gastroenterology 2001;121:1286–93. https://doi.org/10.1053/GAST.2001.29569.

[11] International Agency for Research on Cancer. Agents Classified by the IARC Monographs, Volumes 1–129 – IARC Monographs on the Identification of Carcinogenic Hazards to Humans n.d. https://monographs.iarc.who.int/agents-classified-by-the-iarc/ (accessed April 19, 2021).

[12] Webb PM, Law M, Varghese C, Forman D. Gastric cancer and Helicobacter pylori: a combined analysis of 12 case control studies nested within prospective cohorts. Gut 2001;49:347–53. https://doi.org/10.1136/GUT.49.3.347.

[13] Colquhoun A, Arnold M, Ferlay J, Goodman KJ, Forman D, Soerjomataram I. Global patterns of cardia and non-cardia gastric cancer incidence in 2012. Gut 2015;64:1881–8. https://doi.org/10.1136/GUTJNL-2014-308915.

[14] Wang Q, Liu G, Hu C. Molecular Classification of Gastric Adenocarcinoma. Gastroenterology Res 2019;12:275–82. https://doi.org/10.14740/gr1187.

[15] Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, et al. Comprehensive Molecular Characterization of Gastric Adenocarcinoma Supplementary Materials. Nature 2014.

[16] Gu L, Chen M, Guo D, Zhu H, Zhang W, Pan J, et al. PD-L1 and gastric cancer prognosis: A systematic review and meta-analysis. PLoS One 2017;12. https://doi.org/10.1371/journal.pone.0182692.

[17] Cristescu R, Lee J, Nebozhyn M, Kim K-M, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 2015;21:449–56. https://doi.org/10.1038/nm.3850.

[18] Garrido-Ramos MA. Satellite DNA: An Evolving Topic. Genes 2017, Vol 8, Page 230 2017;8:230. https://doi.org/10.3390/GENES8090230.

[19] Puliga E, Corso S, Pietrantonio F, Giordano S. Microsatellite instability in Gastric Cancer: Between lights and shadows. Cancer Treat Rev 2021;95:102175. https://doi.org/10.1016/J.CTRV.2021.102175.

[20] Baretti M, Le DT. DNA mismatch repair in cancer. Pharmacol Ther 2018;189:45–62. https://doi.org/10.1016/J.PHARMTHERA.2018.04.004.

[21] Patil DT, Bronner MP, Portier BP, Fraser CR, Plesec TP, Liu X. A five-marker panel in a multiplex pcr accurately detects microsatellite instability-high colorectal tumors without control DNA. Diagnostic Molecular Pathology 2012;21:127–33. https://doi.org/10.1097/PDM.0B013E3182461CC3.

[22] Bacher JW, Flanagan LA, Smalley RL, Nassif NA, Burgart LJ, Halberg RB, et al. Development of a fluorescent multiplex assay for detection of MSI-High tumors. Dis Markers 2004;20:237–50.

[23] Vanderwalde A, Spetzler D, Xiao N, Gatalica Z, Marshall J. Microsatellite instability status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348 patients. Cancer Med 2018;7:746–56. https://doi.org/10.1002/CAM4.1372.

[24] Choi YY, Kim H, Shin SJ, Kim HY, Lee J, Yang HK, et al. Microsatellite Instability and Programmed Cell Death-Ligand 1 Expression in Stage II/III Gastric Cancer: Post Hoc Analysis of the CLASSIC Randomized Controlled study. Ann Surg 2019;270:309–16. https://doi.org/10.1097/SLA.0000000000002803.

[25] Smyth EC, Wotherspoon A, Peckitt C, Gonzalez D, Hulkki-Wilson S, Eltahir Z, et al. Mismatch Repair Deficiency, Microsatellite Instability, and Survival: An Exploratory Analysis of the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC) Trial. JAMA Oncol 2017;3:1197–203. https://doi.org/10.1001/JAMAONCOL.2016.6762.

[26] Lordick F. Chemotherapy for resectable microsatellite instability-high gastric cancer? Lancet Oncol 2020;21:203. https://doi.org/10.1016/S1470-2045(20)30012-7.

[27] Smyth EC. Chemotherapy for resectable microsatellite instability-high gastric cancer? Lancet Oncol 2020;21:204. https://doi.org/10.1016/S1470-2045(20)30025-5.

[28] Pietrantonio F, Randon G, Di Bartolomeo M, Luciani A, Chao J, Smyth EC, et al. Predictive role of microsatellite instability for PD-1 blockade in patients with advanced gastric cancer: a meta-analysis of randomized clinical trials. ESMO Open 2021;6. https://doi.org/10.1016/j.esmoop.2020.100036.

[29] Cohen R, Pudlarz T, Garcia-Larnicol M, Vernerey D, Dray X, Clavel L, et al. Localized MSI/dMMR gastric cancer patients, perioperative immunotherapy instead of chemotherapy: The GERCOR NEONIPIGA phase II study is opened to recruitment. Bull Cancer 2020;107:438–46. https://doi.org/10.1016/J.BULCAN.2019.11.016.

[30] Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol 2016;14:e1002533. https://doi.org/10.1371/journal.pbio.1002533.

[31] Patel J. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. Molecular Diagnosis 2001;6:313–21. https://doi.org/10.1054/modi.2001.29158.

[32] Wang WL, Xu SY, Ren ZG, Tao L, Jiang JW, Zheng S Sen. Application of metagenomics in the human gut microbiome. World Journal of Gastroenterology : WJG 2015;21:803. https://doi.org/10.3748/WJG.V21.I3.803.

[33] Yang I, Woltemate S, Piazuelo MB, Bravo LE, Yepez MC, Romero-Gallo J, et al. Different gastric microbiota compositions in two human populations with high and low gastric cancer risk in Colombia. Sci Rep 2016;6. https://doi.org/10.1038/srep18594.

[34] Ferreira RM, Pereira-Marques J, Pinto-Ribeiro I, Costa JL, Carneiro F, MacHado JC, et al. Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. Gut 2018;67:226–36. https://doi.org/10.1136/gutjnl-2017-314205.

[35] Eun CS, Kim BK, Han DS, Kim SY, Kim KM, Choi BY, et al. Differences in gastric mucosal microbiota profiling in patients with chronic gastritis, intestinal metaplasia, and gastric cancer using pyrosequencing methods. Helicobacter 2014;19:407–16. https://doi.org/10.1111/hel.12145.

[36] Gantuya B, El-Serag HB, Matsumoto T, Ajami NJ, Oyuntsetseg K, Azzaya D, et al. Gastric Microbiota in Helicobacter pylori-Negative and-Positive Gastritis Among High Incidence of Gastric Cancer Area 2019. https://doi.org/10.3390/cancers11040504.

[37] Gunathilake M, Lee JH, Choi IJ, Kim Y Il, Kim JS. Effect of the Interaction between Dietary Patterns and the Gastric Microbiome on the Risk of Gastric Cancer. Nutrients 2021, Vol 13, Page 2692 2021;13:2692. https://doi.org/10.3390/NU13082692.

[38] Nikitina D, Lehr K, Vilchez-Vargas R, Jonaitis LV, Urba M, Kupcinskas J, et al. Comparison of genomic and transcriptional microbiome analysis in gastric cancer patients and healthy individuals. World J Gastroenterol 2023;29:1202–18. https://doi.org/10.3748/wjg.v29.i7.1202.

[39] Castaño-Rodríguez N, Goh KL, Fock KM, Mitchell HM, Kaakoush NO. Dysbiosis of the microbiome in gastric carcinogenesis. Sci Rep 2017;7. https://doi.org/10.1038/s41598-017-16289-2.

[40] Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G, et al. Mucosal microbiome dysbiosis in gastric carcinogenesis. Gut 2018;67:1024–32. https://doi.org/10.1136/gutjnl-2017-314281.

[41] Jo HJ, Kim J, Kim N, Park JH, Nam RH, Seok Y-J, et al. Analysis of Gastric Microbiota by Pyrosequencing: Minor Role of Bacteria Other Than Helicobacter pylori in the Gastric Carcinogenesis. Helicobacter 2016;21:364–74. https://doi.org/10.1111/hel.12293.

[42] Sung J, Kim N, Kim J, Jo HJ, Park JH, Nam RH, et al. Comparison of Gastric Microbiota Between Gastric Juice and Mucosa by Next Generation Sequencing Method. J Cancer Prev 2016;21:60–5. https://doi.org/10.15430/jcp.2016.21.1.60.

[43] Ling Z, Shao L, Liu X, Cheng Y, Yan C, Mei Y, et al. Regulatory T cells and plasmacytoid dendritic cells within the tumor microenvironment in gastric cancer are correlated with gastric microbiota dysbiosis: A preliminary study. Front Immunol 2019;10. https://doi.org/10.3389/fimmu.2019.00533.

[44] Park CH, Lee A, Lee Y, Eun CS, Lee SK, Han DS. Evaluation of gastric microbiome and metagenomic function in patients with intestinal metaplasia using 16S rRNA gene sequencing. Helicobacter 2018;24:336–48. https://doi.org/10.1111/hel.12547.

[45] Wang L, Zhou J, Xin Y, Geng C, Tian Z, Yu X, et al. Bacterial overgrowth and diversification of microbiota in gastric cancer. Eur J Gastroenterol Hepatol 2016;28:261–6. https://doi.org/10.1097/MEG.0000000000000542.

[46] Hu Y-L, Pang W, Huang Y, Zhang Y, Zhang C-J. The Gastric Microbiome Is Perturbed in Advanced Gastric Adenocarcinoma Identified Through Shotgun Metagenomics. Front Cell Infect Microbiol 2018;8:433. https://doi.org/10.3389/fcimb.2018.00433.

[47] Sarhadi V, Mathew B, Kokkola A, Karla T, Tikkanen M, Rautelin H, et al. Gut microbiota of patients with different subtypes of gastric cancer and gastrointestinal stromal tumors. Gut Pathog 2021;13:11. https://doi.org/10.1186/s13099-021-00403-x.

[48] Zhao Y, Gao X, Guo J, Yu D, Xiao Y, Wang H, et al. Helicobacter pylori infection alters gastric and tongue coating microbial communities. Helicobacter 2019;24. https://doi.org/10.1111/hel.12567.

[49] Jalili-Firoozinezhad S, Gazzaniga FS, Calamari EL, Camacho DM, Fadel CW, Bein A, et al. A complex human gut microbiome cultured in an anaerobic intestine-on-a-chip. Nature Biomedical Engineering 2019 3:7 2019;3:520–31. https://doi.org/10.1038/s41551-019-0397-0.

[50] Guzman-Rodriguez M, McDonald JAK, Hyde R, Allen-Vercoe E, Claud EC, Sheth PM, et al. Using bioreactors to study the effects of drugs on the human microbiota. Methods 2018;149:31–41. https://doi.org/10.1016/j.ymeth.2018.08.003.

[51] Riquelme E, Zhang Y, Zhang L, Montiel M, Zoltan M, Dong W, et al. Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes. Cell 2019;178:795-806.e12. https://doi.org/10.1016/j.cell.2019.07.008.

[52] Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, Yen HR, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. Nat Med 2009;15:1016–22. https://doi.org/10.1038/nm.2015.

[53] Camacho-Ortiz A, Gutiérrez-Delgado EM, Garcia-Mazcorro JF, Mendoza-Olazarán S, Martínez-Meléndez A, Palau-Davila L, et al. Randomized clinical trial to evaluate the effect of fecal microbiota transplant for initial clostridium difficile infection in intestinal

microbiome. PLoS One 2017;12:e0189768. https://doi.org/10.1371/journal.pone.0189768.

[54] van Ruissen MCE, Bos LD, Dickson RP, Dondorp AM, Schultsz C, Schultz MJ. Manipulation of the microbiome in critical illness—probiotics as a preventive measure against ventilator-associated pneumonia. Intensive Care Med Exp 2019;7:37. https://doi.org/10.1186/s40635-019-0238-1.

[55] Callewaert C, Knödlseder N, Karoglan A, Güell M, Paetzold B. Skin microbiome transplantation and manipulation: Current state of the art. Comput Struct Biotechnol J 2021;19:624–31. https://doi.org/10.1016/j.csbj.2021.01.001.

[56] Denny L, Kuhn L, Hu CC, Tsai WY, Wright TC. Human papillomavirus-based cervical cancer prevention: Long-term results of a randomized screening trial. J Natl Cancer Inst 2010;102:1557–67. https://doi.org/10.1093/jnci/djq342.

[57] Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nature Publishing Group 2012. https://doi.org/10.1038/nrg3182.

[58] Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. Nature 2007 449:7164 2007;449:804–10. https://doi.org/10.1038/nature06244.

[59] Degruttola AK, Low D, Mizoguchi A, Mizoguchi E. Current understanding of dysbiosis in disease in human and animal models. Inflamm Bowel Dis 2016;22:1137. https://doi.org/10.1097/MIB.0000000000000750.

[60] Lynch S V., Pedersen O. The Human Intestinal Microbiome in Health and Disease. New England Journal of Medicine 2016;375:2369–79. https://doi.org/10.1056/NEJMRA1600266/SUPPL_FILE/NEJMRA1600266_DISCLOSURES.PDF.

[61] Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Fölsch UR, et al. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. Gut 2004;53:685–93. https://doi.org/10.1136/GUT.2003.025403.

[62] Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The Treatment-Naive Microbiome in New-

Onset Crohn’s Disease. Cell Host Microbe 2014;15:382–92. https://doi.org/10.1016/J.CHOM.2014.02.005.

[63] Ciofi C, De Boer M, Bruford MW, Hecht MK, Wallace B, Macintyre RJ, et al. Human gut microbes associated with obesity. Nature 2006 444:7122 2006;444:1022–3. https://doi.org/10.1038/4441022a.

[64] Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nature Medicine 2015 21:8 2015;21:895–905. https://doi.org/10.1038/nm.3914.

[65] Radjabzadeh D, Bosch JA, Uitterlinden AG, Zwinderman AH, Ikram MA, van Meurs JBJ, et al. Gut microbiome-wide association study of depressive symptoms. Nature Communications 2022 13:1 2022;13:1–10. https://doi.org/10.1038/s41467-022-34502-3.

[66] Paller AS, Kong HH, Seed P, Naik S, Scharschmidt TC, Gallo RL, et al. The microbiome in patients with atopic dermatitis. Journal of Allergy and Clinical Immunology 2019;143:26–35. https://doi.org/10.1016/J.JACI.2018.11.015.

[67] Petricevic L, Domig KJ, Nierscher FJ, Sandhofer MJ, Fidesser M, Krondorfer I, et al. Characterisation of the vaginal Lactobacillus microbiota associated with preterm delivery. Scientific Reports 2014 4:1 2014;4:1–6. https://doi.org/10.1038/srep05136.

[68] Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov 2022;12:31–46. https://doi.org/10.1158/2159-8290.CD-21-1059.

[69] Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks + E. coli. Nature 2020;580:269–73. https://doi.org/10.1038/s41586-020-2080-8.

[70] Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature 2020;579:567–74. https://doi.org/10.1038/s41586-020-2095-1.

[71] Geller LT, Barzily-Rokni M, Danino T, Jonas OH, Shental N, Nejman D, et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. Science (1979) 2017;357:1156–60. https://doi.org/10.1126/science.aah5043.

[72] Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets T v., et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science (1979) 2018;359:97–103. https://doi.org/10.1126/science.aan4236.

[73] Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science (1979) 2018;359:91–7. https://doi.org/10.1126/science.aan3706.

[74] Yu TC, Guo F, Yu Y, Sun T, Ma D, Han J, et al. Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. Cell 2017;170:548-563.e16. https://doi.org/10.1016/j.cell.2017.07.008.

[75] El Tekle G, Garrett WS. Bacteria in cancer initiation, promotion and progression. Nature Reviews Cancer 2023 2023:1–19. https://doi.org/10.1038/s41568-023-00594-2.

[76] Di Domenico EG, Cavallo I, Pontone M, Toma L, Ensoli F. Biofilm Producing Salmonella Typhi: Chronic Colonization and Development of Gallbladder Cancer. International Journal of Molecular Sciences 2017, Vol 18, Page 1887 2017;18:1887. https://doi.org/10.3390/IJMS18091887.

[77] Gopalakrishnan V, Helmink BA, Spencer CN, Reuben A, Wargo JA. The Influence of the Gut Microbiome on Cancer, Immunity, and Cancer Immunotherapy. Cancer Cell 2018;33:570–80. https://doi.org/10.1016/j.ccell.2018.03.015.

[78] Garrett WS. Cancer and the microbiota. Science (1979) 2015;348:80–6. https://doi.org/10.1126/science.aaa4972.

[79] Savage DC. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 1977;31:107–33. https://doi.org/10.1146/ANNUREV.MI.31.100177.000543.

[80] Marshall BJ, Warren JR. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. The Lancet 1983;321:1273–5.

[81] Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. The Lancet 1984;323:1311–5. https://doi.org/10.1016/S0140-6736(84)91816-6.

[82] Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F, et al. Molecular analysis of the bacterial microbiota in the human

stomach. Proceedings of the National Academy of Sciences 2006;103:732–7.

[83] Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P, Engstrand L. Comparative Analysis of Human Gut Microbiota by Barcoded Pyrosequencing. PLoS One 2008;3:e2836. https://doi.org/10.1371/journal.pone.0002836.

[84] Maldonado-Contreras A, Goldfarb KC, Godoy-Vitorino F, Karaoz U, Contreras M, Blaser MJ, et al. Structure of the human gastric bacterial community in relation to Helicobacter pylori status. ISME Journal 2011;5:574–9. https://doi.org/10.1038/ismej.2010.149.

[85] Oren A, Garrity GM. Valid publication of the names of forty-two phyla of prokaryotes. Int J Syst Evol Microbiol 2021;71:005056. https://doi.org/10.1099/IJSEM.0.005056/CITE/REFWORKS.

[86] Li XX, Wong GLH, To KF, Wong VWS, Lai LH, Chow DKL, et al. Bacterial microbiota profiling in gastritis without Helicobacter pylori infection or non-steroidal anti-inflammatory drug use. PLoS One 2009;4. https://doi.org/10.1371/journal.pone.0007985.

[87] Engstrand L, Lindberg M. Helicobacter pylori and the gastric microbiota. Best Pract Res Clin Gastroenterol 2013;27:39–45. https://doi.org/10.1016/j.bpg.2013.03.016.

[88] McColl KEL, El-Omar E, Gillen D. HELICOBACTER PYLORI GASTRITIS AND GASTRIC PHYSIOLOGY. Gastroenterol Clin North Am 2000;29:687–703. https://doi.org/10.1016/S0889-8553(05)70138-2.

[89] Kuipers EJ, Uyterlinde AM, Pena AS, Roosendaale R, Pals G, Nelis EF, et al. Long-term sequelae of Helicobacter pylori gastritis. Lancet 1995;345:1525–8.

[90] Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, et al. Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. Gastroenterology 2017;153:420–9. https://doi.org/10.1053/j.gastro.2017.04.022.

[91] Correa P, Piazuelo MB. The gastric precancerous cascade. J Dig Dis 2012;13:2–9. https://doi.org/10.1111/j.1751-2980.2011.00550.x.

[92] Ford AC, Yuan Y, Moayyedi P. Helicobacter pylori eradication therapy to prevent gastric cancer: systematic review and meta-analysis. Gut 2020:1–9. https://doi.org/10.1136/gutjnl-2020-320839.

[93] Peek RM, Blaser MJ. Helicobacter pylori and gastrointestinal tract adenocarcinomas. Nat Rev Cancer 2002;2:28–37. https://doi.org/10.1038/nrc703.

[94] Blaser MJ, Perez-Perez GI, Kleanthous H, Cover TL, Peek RM, Chyou PH, et al. Infection with Helicobacter pylori Strains Possessing cagA Is Associated with an Increased Risk of Developing Adenocarcinoma of the Stomach. Cancer Res 1995;55.

[95] Parsonnet J, Friedman GD, Orentreich N, Vogelman H. Risk for gastric cancer in people with CagA positive or CagA negative Helicobacter pylori infection. Gut 1997;40:297–301. https://doi.org/10.1136/gut.40.3.297.

[96] Cover TL. Helicobacter pylori diversity and gastric cancer risk. mBio 2016;7:1–9. https://doi.org/10.1128/mBio.01869-15.

[97] Bento-Miranda M, Figueiredo C. Helicobacter heilmannii sensu lato: An overview of the infection in humans. World Journal of Gastroenterology : WJG 2014;20:17779. https://doi.org/10.3748/WJG.V20.I47.17779.

[98] Fischer W. Assembly and molecular mode of action of the Helicobacter pylori Cag type IV secretion apparatus. FEBS Journal 2011;278:1203–12. https://doi.org/10.1111/j.1742-4658.2011.08036.x.

[99] Wang F, Meng W, Wang B, Qiao L. Helicobacter pylori-induced gastric inflammation and gastric cancer. Cancer Lett 2014;345:196–202. https://doi.org/10.1016/J.CANLET.2013.08.016.

[100] Peek RM, Miller GG, Tham KT, Perez-Perez GI, Zhao X, Atherton JC, et al. Heightened inflammatory response and cytokine expression in vivo to cagA+ Helicobacter pylori strains. Laboratory Investigation 1995;73:760–70.

[101] Kuipers EJ, Pérez-pérez GI, Meuwissen SGM, Blaser MJ. Helicobacter pylori and atrophic gastritis: Importance of the cagA status. J Natl Cancer Inst 1995;87:1777–80. https://doi.org/10.1093/jnci/87.23.1777.

[102] pormohammad A, Ghotaslo R, Leylabadlo HE, Nasiri MJ, Dabiri H, Hashemi A. Risk of gastric cancer in association with Helicobacter pylori different virulence factors: A systematic review and meta-analysis. Microb Pathog 2018;118:214–9. https://doi.org/10.1016/j.micpath.2018.03.004.

[103] Palframan SL, Kwok T, Gabriel K. Vacuolating cytotoxin A (VacA), a key toxin for Helicobacter pylori pathogenesis. Front Cell Infect Microbiol 2012;2:92. https://doi.org/10.3389/fcimb.2012.00092.

[104] Yahiro K, Akazawa Y, Nakano M, Suzuki H, Hisatune J, Isomoto H, et al. Helicobacter pylori VacA induces apoptosis by accumulation of connexin 43 in autophagic vesicles via a Rac1/ERK-dependent pathway. Cell Death Discov 2015;1:15035. https://doi.org/10.1038/cddiscovery.2015.35.

[105] Sewald X, Gebert-Vogl B, Prassl S, Barwig I, Weiss E, Fabbri M, et al. Integrin Subunit CD18 Is the T-Lymphocyte Receptor for the Helicobacter pylori Vacuolating Cytotoxin. Cell Host Microbe 2008;3:20–9. https://doi.org/10.1016/j.chom.2007.11.003.

[106] Torres VJ, VanCompernolle SE, Sundrud MS, Unutmaz D, Cover TL. Helicobacter pylori Vacuolating Cytotoxin Inhibits Activation-Induced Proliferation of Human T and B Lymphocyte Subsets . The Journal of Immunology 2007;179:5433–40. https://doi.org/10.4049/jimmunol.179.8.5433.

[107] Weiss G, Forster S, Irving A, Tate M, Ferrero RL, Hertzog P, et al. Helicobacter pylori VacA suppresses Lactobacillus acidophilus-induced interferon beta signaling in macrophages via alterations in the endocytic pathway. mBio 2013;4. https://doi.org/10.1128/mBio.00609-12.

[108] Foegeding NJ, Caston RR, McClain MS, Ohi MD, Cover TL. An Overview of Helicobacter pylori VacA Toxin Biology. Toxins 2016, Vol 8, Page 173 2016;8:173. https://doi.org/10.3390/TOXINS8060173.

[109] Hsieh Y-Y, Tung S-Y, Pan H-Y, Yen C-W, Xu H-W, Lin Y-J, et al. Increased Abundance of Clostridium and Fusobacterium in Gastric Microbiota of Patients with Gastric Cancer in Taiwan. Scientific Reports 2017 8:1 2018;8:1–11. https://doi.org/10.1038/s41598-017-18596-0.

[110] de Assumpção PP, Araújo TMT, de Assumpção PB, Barra WF, Khayat AS, Assumpção CB, et al. Suicide journey of H. pylori through gastric carcinogenesis: the role of non-H. pylori microbiome and potential consequences for clinical practice. European Journal of Clinical Microbiology and Infectious Diseases 2019;38:1591–7. https://doi.org/10.1007/s10096-019-03564-5.

[111] Lieberman PM. Epstein-Barr virus turns 50. Science (1979) 2014;343:1323–5. https://doi.org/10.1126/science.1252786.

[112] Yuen ST, Chung LP, Leung SY, Luk ISC, Chan SY, Ho J. In situ detection of Epstein-Barr virus in gastric and colorectal adenocarcinomas. American Journal of Surgical Pathology 1994;18:1158–63. https://doi.org/10.1097/00000478-199411000-00010.

[113] Harn HJ, Chang JY, Wang MW, Ho LI, Lee HS, Chiang JH, et al. Epstein-Barr virus-associated gastric adenocarcinoma in Taiwan. Hum Pathol 1995;26:267–71. https://doi.org/10.1016/0046-8177(95)90056-X.

[114] Shibata D, Hawes D, Stemmermann GN, Weiss LM. Epstein-Barr virus-associated gastric adenocarcinoma among Japanese Americans in Hawaii. Cancer Epidemiology and Prevention Biomarkers 1993;2.

[115] Murphy G, Pfeiffer R, Camargo MC, Rabkin CS. Meta-analysis Shows That Prevalence of Epstein-Barr Virus-Positive Gastric Cancer Differs Based on Sex and Anatomic Location. Gastroenterology 2009;137:824–33. https://doi.org/10.1053/j.gastro.2009.05.001.

[116] The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202–9. https://doi.org/10.1038/nature13480.

[117] Constanza Camargo M, Kim WH, Chiaravalli AM, Kim KM, Corvalan AH, Matsuo K, et al. Improved survival of gastric cancer with tumour Epstein-Barr virus positivity: An international pooled analysis. Gut 2014;63:236–43. https://doi.org/10.1136/gutjnl-2013-304531.

[118] Zhao Y, Zhang J, Cheng ASL, Yu J, To KF, Kang W. Gastric cancer: genome damaged by bugs. Oncogene 2020;39:3427–42. https://doi.org/10.1038/s41388-020-1241-4.

[119] He C, Peng C, Shu X, Wang H, Zhu Z, Ouyang Y, et al. Convergent dysbiosis of gastric mucosa and fluid microbiome during stomach carcinogenesis. Gastric Cancer 2022;25:837–49. https://doi.org/10.1007/S10120-022-01302-Z/FIGURES/7.

[120] Yu G, Torres J, Hu N, Medrano-Guzman R, Herrera-Goepfert R, Humphrys MS, et al. Molecular characterization of the human

stomach microbiota in Gastric Cancer Patients. Front Cell Infect Microbiol 2017;7:302. https://doi.org/10.3389/fcimb.2017.00302.

[121] Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure, function and diversity of the healthy human microbiome. Nature 2012 486:7402 2012;486:207–14. https://doi.org/10.1038/nature11234.

[122] Gunathilake MN, Lee J, Choi IJ, Kim Y Il, Ahn Y, Park C, et al. Association between the relative abundance of gastric microbiota and the risk of gastric cancer: a case-control study. Sci Rep 2019;9:1–11. https://doi.org/10.1038/s41598-019-50054-x.

[123] Aviles-Jimenez F, Vazquez-Jimenez F, Medrano-Guzman R, Mantilla A, Torres J. Stomach microbiota composition varies between patients with non-atrophic gastritis and patients with intestinal type of gastric cancer. Sci Rep 2014;4:1–11. https://doi.org/10.1038/srep04202.

[124] Wang Z, Gao X, Zeng R, Wu Q, Sun H, Wu W, et al. Changes of the Gastric Mucosal Microbiome Associated With Histological Stages of Gastric Carcinogenesis. Front Microbiol 2020;11:997. https://doi.org/10.3389/fmicb.2020.00997.

[125] Park JY, Seo H, Kang CS, Shin TS, Kim JW, Park JM, et al. Dysbiotic change in gastric microbiome and its functional implication in gastric carcinogenesis. Scientific Reports 2022 12:1 2022;12:1–11. https://doi.org/10.1038/s41598-022-08288-9.

[126] Park CH, Hong C, Lee A reum, Sung J, Hwang TH. Multi-omics reveals microbiome, host gene expression, and immune landscape in gastric carcinogenesis. iScience 2022;25:103956. https://doi.org/10.1016/J.ISCI.2022.103956.

[127] Bang EJ, Choi HS, Lee JM, Kim ES, Keum B, Jeen YT, et al. Analysis of the gastric microbiome in gastric cancer patients and in non-tumorous patients. Https://DoiOrg/101200/JCO2022404_suppl349 2022;40:349–349. https://doi.org/10.1200/JCO.2022.40.4_SUPPL.349.

[128] Kim HN, Kim MJ, Jacobs JP, Yang HJ. Altered Gastric Microbiota and Inflammatory Cytokine Responses in Patients with Helicobacter pylori-Negative Gastric Cancer. Nutrients 2022;14:4981. https://doi.org/10.3390/NU14234981/S1.

[129] Liu C, Ng SK, Ding Y, Lin Y, Liu W, Wong SH, et al. Meta-analysis of mucosal microbiota reveals universal microbial signatures and dysbiosis in gastric carcinogenesis. Oncogene 2022 41:28 2022;41:3599–610. https://doi.org/10.1038/s41388-022-02377-9.

[130] Liu X, Shao L, Liu X, Ji F, Mei Y, Cheng Y, et al. Alterations of gastric mucosal microbiota across different stomach microhabitats in a cohort of 276 patients with gastric cancer. EBioMedicine 2019;40:336–48. https://doi.org/10.1016/j.ebiom.2018.12.034.

[131] Lehr K, Vilchez-Vargas R, Nikitina D, Thon C, Steponaitienė R, Skiecevičienė J, et al. Gastric Cancer Mucosal Microbiome is Associated with Prognosis of Gastric Cancer Patients. United European Gastroenterol J, vol. 10, Wiley; 2022, p. 185–472. https://doi.org/10.1002/UEG2.12294.

[132] Dai D, Yang Y, Yu J, Dang T, Qin W, Teng L, et al. Interactions between gastric microbiota and metabolites in gastric cancer. Cell Death & Disease 2021 12:12 2021;12:1–11. https://doi.org/10.1038/s41419-021-04396-y.

[133] Chen XH, Wang A, Chu AN, Gong YH, Yuan Y. Mucosa-associated microbiota in gastric cancer tissues compared with non-cancer tissues. Front Microbiol 2019;10. https://doi.org/10.3389/fmicb.2019.01261.

[134] Wang L, Xin Y, Zhou J, Tian Z, Liu C, Yu X, et al. Gastric Mucosa-Associated Microbial Signatures of Early Gastric Cancer. Front Microbiol 2020;11:1548. https://doi.org/10.3389/fmicb.2020.01548.

[135] Fu BC, Randolph TW, Lim U, Monroe KR, Cheng I, Wilkens LR, et al. Characterization of the Gut Microbiome in Epidemiologic Studies: The Multiethnic Cohort Experience. Ann Epidemiol 2016;26:373. https://doi.org/10.1016/J.ANNEPIDEM.2016.02.009.

[136] Wu GD, Lewis JD, Hoffmann C, Chen YY, Knight R, Bittinger K, et al. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol 2010;10:1–14. https://doi.org/10.1186/1471-2180-10-206/FIGURES/6.

[137] Sze MA, Schloss PD. The Impact of DNA Polymerase and Number of Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data. mSphere 2019;4. https://doi.org/10.1128/MSPHERE.00163-19/ASSET/4F8C6776-4829-48DF-8E4E-

8D97B3F48BAD/ASSETS/GRAPHIC/MSPHERE.00163-19-F0006.JPEG.

[138] Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol 2017;35:1077. https://doi.org/10.1038/NBT.3981.

[139] Kulakov LA, McAlister MB, Ogden KL, Larkin MJ, O'Hanlon JF. Analysis of bacteria contaminating ultrapure water in industrial systems. Appl Environ Microbiol 2002;68:1548–55. https://doi.org/10.1128/AEM.68.4.1548-1555.2002/ASSET/BBFB5B66-A42C-4C27-8B4F-32FD94BA1F09/ASSETS/GRAPHIC/AM0421837002.JPEG.

[140] Shen H, Rogelj S, Kieft TL. Sensitive, real-time PCR detects low-levels of contamination by Legionella pneumophila in commercial reagents. Mol Cell Probes 2006;20:147–53. https://doi.org/10.1016/j.mcp.2005.09.007.

[141] Newsome T, Li B-J, Zou N, Lo S-C. Presence of Bacterial Phage-Like DNA Sequences in Commercial Taq DNA Polymerase Reagents. J Clin Microbiol 2004;42:2264–7. https://doi.org/10.1128/JCM.42.5.2264-2267.2004.

[142] Motley T, Picuri JM, Crowder CD, Minich JJ, Hofstadler SA, Eshoo MW. Improved Multiple Displacement Amplification (iMDA) and Ultraclean Reagents. 2014.

[143] Mohammadi T, Reesink HW, Vandenbroucke-Grauls CMJE, Savelkoul PHM. Removal of contaminating DNA from commercial nucleic acid extraction kit reagents 2004. https://doi.org/10.1016/j.mimet.2004.11.018.

[144] Shanahan ER, Zhong L, Talley NJ, Morrison M, Holtmann G. Characterisation of the gastrointestinal mucosa-associated microbiota: A novel technique to prevent cross-contamination during endoscopic procedures. Aliment Pharmacol Ther 2016;43:1186–96. https://doi.org/10.1111/apt.13622.

[145] Bayer G, Ganobis CM, Allen-Vercoe E, Philpott DJ. Defined gut microbial communities: promising tools to understand and combat disease. Microbes Infect 2021;23:104816. https://doi.org/10.1016/J.MICINF.2021.104816.

[146] Ganobis C. Characterizing the Mouse Gut Microbiome and Improving Mouse Gut-Derived Microbial Communities for Mouse Model Studies. University of Guelph, 2023.

[147] Brooks JP, Edwards DJ, Harwich MD, Rivera MC, Fettweis JM, Serrano MG, et al. The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies Ecological and evolutionary microbiology. BMC Microbiol 2015;15:1–14. https://doi.org/10.1186/S12866-015-0351-6/TABLES/5.

[148] Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 2017;35:833–44. https://doi.org/10.1038/nbt.3935.

[149] Thermes C. Ten years of next-generation sequencing technology. Trends in Genetics 2014;30:418–26. https://doi.org/10.1016/J.TIG.2014.07.001.

[150] Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100 000 Genomes Project: Bringing whole genome sequencing to the NHS. BMJ (Online) 2018;361. https://doi.org/10.1136/bmj.k1687.

[151] Cornish AJ, Gruber AJ, Kinnersley B, Chubb D, Frangou A, Caravagna G, et al. Whole genome sequencing of 2,023 colorectal cancers reveals mutational landscapes, new driver genes and immune interactions. bioRxiv 2022:2022.11.16.515599. https://doi.org/10.1101/2022.11.16.515599.

[152] The Cancer Genome Atlas Program (TCGA) - NCI n.d. https://www.cancer.gov/ccg/research/genome-sequencing/tcga (accessed May 4, 2023).

[153] The Cancer Genome Atlas (TCGA) n.d. https://www.genome.gov/Funded-Programs-Projects/Cancer-Genome-Atlas (accessed May 4, 2023).

[154] NIH National Cancer Institute. Genomic Data Commons Data Portal n.d. https://portal.gdc.cancer.gov/ (accessed June 5, 2021).

[155] Rodriguez RM, Hernandez BY, Menor M, Deng Y, Khadka VS. The landscape of bacterial presence in tumor and adjacent normal tissue across 9 major cancer types using TCGA exome sequencing. Comput Struct Biotechnol J 2020;18:631–41. https://doi.org/10.1016/j.csbj.2020.03.003.

[156] Abate M, Vos E, Gonen M, Janjigian YY, Schattner M, Laszkowska M, et al. A Novel Microbiome Signature in Gastric Cancer: A Two Independent Cohort Retrospective Analysis. Ann Surg 2022;276:605–15. https://doi.org/10.1097/SLA.0000000000005587.

[157] Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog 2016. https://doi.org/10.1186/s13099-016-0103-7.

[158] Dohlman AB, Arguijo Mendoza D, Ding S, Gao M, Dressman H, Iliev ID, et al. The cancer microbiome atlas: a pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants. Cell Host Microbe 2021;29:281-298.e5. https://doi.org/10.1016/J.CHOM.2020.12.001.

[159] Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 2018;6. https://doi.org/10.1186/s40168-018-0605-2.

[160] Wang D, Zhang T, Lu Y, Wang C, Wu Y, Li J, et al. Helicobacter pylori infection affects the human gastric microbiome, as revealed by metagenomic sequencing. FEBS Open Bio 2022;12:1188–96. https://doi.org/10.1002/2211-5463.13390.

[161] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles HHS Public Access. Nat Methods 2015;12:453–7. https://doi.org/10.1038/nmeth.3337.

[162] Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. Immunity 2018;48:812-830.e14. https://doi.org/10.1016/j.immuni.2018.03.023.

[163] Genomics England. The National Genomics Research and Healthcare Knowledgebase v5 2019.

[164] Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell 2012;149:994–1007. https://doi.org/10.1016/j.cell.2012.04.023.

[165] Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative Molecular Analysis of Gastrointestinal

Adenocarcinomas. Cancer Cell 2018;33:721–35. https://doi.org/10.1016/j.ccell.2018.03.010.

[166] Hewitt LC, Saito Y, Wang T, Matsuda Y, Oosting J, S Silva AN, et al. KRAS status is related to histological phenotype in gastric cancer: results from a large multicentre study. Gastric Cancer 2019;22:1193–203. https://doi.org/10.1007/s10120-019-00972-6.

[167] Ni Huang M, McPherson JR, Cutcutache I, Teh BT, Tan P, Rozen SG. MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. Scientific Reports 2015 5:1 2015;5:1–10. https://doi.org/10.1038/srep13321.

[168] Walker MA, Pedamallu CS, Ojesina AI, Bullman S, Sharpe T, Whelan CW, et al. GATK PathSeq: A customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. Bioinformatics 2018;34:4287–9. https://doi.org/10.1093/bioinformatics/bty501.

[169] R Core Team. R: A Language and Environment for Statistical Computing 2020.

[170] RStudio Team. RStudio: Integrated Development Environment for R 2021.

[171] Wickham H. stringr: Simple, Consistent Wrappers for Common String Operations 2019.

[172] Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. R package version 1.0.8. 2022.

[173] Storey JD, Bass AJ, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery rate control 2021.

[174] Simpson JO and FGB and MF and RK and PL and DM and PRM and RBO and GL and PS and MHHS and ES and HW. vegan: Community Ecology Package 2020.

[175] Kolde R. pheatmap: Pretty Heatmaps 2019.

[176] Wickham H. ggplot2: Elegant Graphics for Data Analysis 2016.

[177] Shannon CE, Weaver W. The Mathematical Theory of Communication. Urbana: The University of Illinois Press; 1949.

[178] Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecol Monogr 1957;27:325–49. https://doi.org/10.2307/1942268.

[179] Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable Association Discovery in Population-scale Meta-omics Studies. PLoS Comput Biol 2021;17:e1009442. https://doi.org/10.1101/2021.01.20.427420.

[180] Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, et al. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330. https://doi.org/10.1038/NATURE11252.

[181] Rajilic-Stojanovic M, Figueiredo C, Smet A, Hansen R, Kupcinskas J, Rokkas T, et al. Systematic review: gastric microbiota in health and disease. Aliment Pharmacol Ther 2020;51:582–602. https://doi.org/10.1111/apt.15650.

[182] Könönen E, Fteita D, Gursoy UK, Gursoy M. Prevotella species as oral residents and infectious agents with potential impact on systemic conditions. Https://DoiOrg/101080/20002297202220798142022;14. https://doi.org/10.1080/20002297.2022.2079814.

[183] Gonçalves LFH, Fermiano D, Feres M, Figueiredo LC, Teles FRP, Mayer MPA, et al. Levels of Selenomonas species in generalized aggressive periodontitis. J Periodontal Res 2012;47:711–8. https://doi.org/10.1111/J.1600-0765.2012.01485.X.

[184] Sizova M V., Muller P, Panikov N, Mandalakis M, Hohmann T, Hazen A, et al. Stomatobaculum longum gen. nov., sp. nov., an obligately anaerobic bacterium from the human oral cavity. Int J Syst Evol Microbiol 2013;63:1450–6. https://doi.org/10.1099/ijs.0.042812-0.

[185] Abranches J, Zeng L, Kajfasz JK, Palmer SR, Chakraborty B, Wen ZT, et al. Biology of Oral Streptococci. Microbiol Spectr 2018;6. https://doi.org/10.1128/MICROBIOLSPEC.GPP3-0042-2018/ASSET/34F06DE8-4503-417B-8F9A-6AFB6C081F64/ASSETS/GRAPHIC/GPP3-0042-2018-FIG3.GIF.

[186] Badet C, Thebaud NB. Ecology of Lactobacilli in the Oral Cavity: A Review of Literature. vol. 2. 2008.

[187] Meehan CJ, Beiko RG. A Phylogenomic View of Ecological Specialization in the Lachnospiraceae, a Family of Digestive Tract-Associated Bacteria. Genome Biol Evol 2014;6:703–13. https://doi.org/10.1093/GBE/EVU050.

[188] Shah HN, Chattaway MA, Rajakurana L, Gharbia SE. Prevotella. Bergey's Manual of Systematics of Archaea and Bacteria 2015:1–25. https://doi.org/10.1002/9781118960608.GBM00249.

[189] Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. mBio 2015;6. https://doi.org/10.1128/MBIO.00037-15/ASSET/B3454C18-07C9-428A-9A36-C9432B835006/ASSETS/GRAPHIC/MBO0011521950005.JPEG.

[190] Tett A, Pasolli E, Masetti G, Ercolini D, Segata N. Prevotella diversity, niches and interactions with the human host. Nature Reviews Microbiology 2021 19:9 2021;19:585–99. https://doi.org/10.1038/s41579-021-00559-y.

[191] Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, et al. A taxonomic note on the genus Lactobacillus: Description of 23 novel genera, emended description of the genus Lactobacillus beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. Int J Syst Evol Microbiol 2020;70:2782–858. https://doi.org/10.1099/IJSEM.0.004107/CITE/REFWORKS.

[192] Lin XB, Wang T, Stothard P, Corander J, Wang J, Baines JF, et al. The evolution of ecological facilitation within mixed-species biofilms in the mouse gastrointestinal tract. ISME J 2018;12:2770. https://doi.org/10.1038/S41396-018-0211-0.

[193] Hammes WP, Hertel C. Lactobacillus. Bergey's Manual of Systematics of Archaea and Bacteria, Wiley; 2015, p. 1–76. https://doi.org/10.1002/9781118960608.gbm00604.

[194] Whiley RA, Hardie JM. Streptococcus. Bergey's Manual of Systematics of Archaea and Bacteria, Wiley; 2015, p. 1–86. https://doi.org/10.1002/9781118960608.gbm00612.

[195] Lancefield RC. A SEROLOGICAL DIFFERENTIATION OF HUMAN AND OTHER GROUPS OF HEMOLYTIC STREPTOCOCCI. Journal of Experimental Medicine 1933;57:571–95. https://doi.org/10.1084/JEM.57.4.571.

[196] Brouwer S, Rivera-Hernandez T, Curren BF, Harbison-Price N, De Oliveira DMP, Jespersen MG, et al. Pathogenesis, epidemiology and control of Group A Streptococcus infection. Nature Reviews

Microbiology 2023 21:7 2023;21:431–47. https://doi.org/10.1038/s41579-023-00865-7.

[197] Johri AK, Paoletti LC, Glaser P, Dua M, Sharma PK, Grandi G, et al. Group B Streptococcus: global incidence and vaccine development. Nature Reviews Microbiology 2006 4:12 2006;4:932–42. https://doi.org/10.1038/nrmicro1552.

[198] Sizova M V., Muller P, Panikov N, Mandalakis M, Hohmann T, Hazen A, et al. Stomatobaculum longum gen. nov., sp. nov., an obligately anaerobic bacterium from the human oral cavity. Int J Syst Evol Microbiol 2013;63:1450. https://doi.org/10.1099/IJS.0.042812-0.

[199] Shouche YS, Dighe AS, Dhotre DP, Patole MS, Ranade DR. Selenomonas . Bergey's Manual of Systematics of Archaea and Bacteria, Wiley; 2015, p. 1–12. https://doi.org/10.1002/9781118960608.gbm00705.

[200] Byrd DA, Fan W, Greathouse KL, Wu MC, Xie H, Wang X. The intratumor microbiome is associated with microsatellite instability. JNCI: Journal of the National Cancer Institute 2023;115:989–93. https://doi.org/10.1093/JNCI/DJAD083.

[201] Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. Cell Host Microbe 2013;14:207–15. https://doi.org/10.1016/J.CHOM.2013.07.007.

[202] Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, et al. Binding of the Fap2 Protein of Fusobacterium nucleatum to Human Inhibitory Receptor TIGIT Protects Tumors from Immune Cell Attack. Immunity 2015;42:344–55. https://doi.org/10.1016/J.IMMUNI.2015.01.010.

[203] Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/β-Catenin Signaling via its FadA Adhesin. Cell Host Microbe 2013;14:195–206. https://doi.org/10.1016/J.CHOM.2013.07.012.

[204] Polom K, Marano L, Marrelli D, De Luca R, Roviello G, Savelli V, et al. Meta-analysis of microsatellite instability in relation to clinicopathological characteristics and overall survival in gastric

cancer. British Journal of Surgery 2018;105:159–67. https://doi.org/10.1002/BJS.10663.

[205] Giampieri R, Maccaroni E, Mandolesi A, Del Prete M, Andrikou K, Faloppi L, et al. Mismatch repair deficiency may affect clinical outcome through immune response activation in metastatic gastric cancer patients receiving first-line chemotherapy. Gastric Cancer 2017;20:156–63. https://doi.org/10.1007/S10120-016-0594-4/FIGURES/5.

[206] Grogg KL, Lohse CM, Pankratz VS, Halling KC, Smyrk TC. Lymphocyte-Rich Gastric Cancer: Associations with Epstein-Barr Virus, Microsatellite Instability, Histology, and Survival. Modern Pathology 2003;16:641–51. https://doi.org/10.1097/01.MP.0000076980.73826.C0.

[207] Naruke A, Azuma M, Takeuchi A, Kenji •, Chikatoshi Katada I•, Sasaki T, et al. Comparison of site-specific gene expression levels in primary tumors and synchronous lymph node metastases in advanced gastric cancer. Gastric Cancer 2015;18:262–70. https://doi.org/10.1007/s10120-014-0357-z.

[208] Sundar R, Liu DH, Hutchins GGG, Slaney HL, Silva AN, Oosting J, et al. Spatial profiling of gastric cancer patient-matched primary and locoregional metastases reveals principles of tumour dissemination. Gut 2020;0:gutjnl-2020-320805. https://doi.org/10.1136/gutjnl-2020-320805.

[209] Kim O-H, Choi B-Y, Kim DK, Kim NH, Rho JK, Jun W, et al. The microbiome of lung cancer tissue and its association with pathological and clinical parameters. Am J Cancer Res 2022;12:2350.

[210] Parsons BN, Ijaz UZ, D'Amore R, Burkitt MD, Eccles R, Lenzi L, et al. Comparison of the human gastric microbiota in hypochlorhydric states arising as a result of Helicobacter pylori-induced atrophic gastritis, autoimmune atrophic gastritis and proton pump inhibitor use. PLoS Pathog 2017;13:e1006653. https://doi.org/10.1371/journal.ppat.1006653.

[211] Aoyama T, Hutchins G, Arai T, Sakamaki K, Miyagi Y, Tsuburaya A, et al. Identification of a high-risk subtype of intestinal-type Japanese gastric cancer by quantitative measurement of the luminal tumor proportion. Cancer Med 2018;7:4914–23. https://doi.org/10.1002/cam4.1744.

[212] Kumar V, Ramnarayanan K, Sundar R, Padmanabhan N, Srivastava S, Koiwa M, et al. Single-Cell Atlas of Lineage States, Tumor Microenvironment, and Subtype-Specific Expression Programs in Gastric Cancer. Cancer Discov 2022;12:670–91. https://doi.org/10.1158/2159-8290.CD-21-0683/673848/AM/SINGLE-CELL-ATLAS-OF-LINEAGE-STATES-TUMOR.

[213] Facciabene A, Motz GT, Coukos G. T-Regulatory Cells: Key Players in Tumor Immune Escape and Angiogenesis. Cancer Res 2012;72:2162–71. https://doi.org/10.1158/0008-5472.CAN-11-3687.

[214] Gerlini G, Urso C, Mariotti G, Di Gennaro P, Palli D, Brandani P, et al. Plasmacytoid dendritic cells represent a major dendritic cell subset in sentinel lymph nodes of melanoma patients and accumulate in metastatic nodes. Clinical Immunology 2007;125:184–93. https://doi.org/10.1016/J.CLIM.2007.07.018.

[215] Peng R, Liu S, You W, Huang Y, Hu C, Gao Y, et al. Gastric Microbiome Alterations Are Associated with Decreased CD8+ Tissue-Resident Memory T Cells in the Tumor Microenvironment of Gastric Cancer. Cancer Immunol Res 2022;10:1224–40. https://doi.org/10.1158/2326-6066.CIR-22-0107.

[216] Albuquerque L, França L, Rainey FA, Schumann P, Nobre MF, Da Costa MS. Gaiella occulta gen. nov., sp. nov., a novel representative of a deep branching phylogenetic lineage within the class Actinobacteria and proposal of Gaiellaceae fam. nov. and Gaiellales ord. nov. Syst Appl Microbiol 2011;34:595–9. https://doi.org/10.1016/J.SYAPM.2011.07.001.

[217] Green PN, Ardley JK. Review of the genus Methylobacterium and closely related organisms: A proposal that some Methylobacterium species be reclassified into a new genus, Methylorubrum gen. nov. Int J Syst Evol Microbiol 2018;68:2727–48. https://doi.org/10.1099/IJSEM.0.002856/CITE/REFWORKS.

[218] Kovaleva J, Degener JE, Van Der Mei HC. Methylobacterium and its role in health care-associated infection. J Clin Microbiol 2014;52:1317–21. https://doi.org/10.1128/JCM.03561-13/ASSET/36CAA628-B43C-4F56-B4D2-B3BEF4D77DD4/ASSETS/GRAPHIC/ZJM9990932320001.JPEG.

[219] Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact

sequence-based microbiome analyses. BMC Biol 2014;12:1–12. https://doi.org/10.1186/S12915-014-0087-Z/FIGURES/4.

[220] Choi YY, Bae JM, An JY, Kwon IG, Cho I, Shin HB, et al. Is microsatellite instability a prognostic marker in gastric cancer?: A systematic review with meta-analysis. J Surg Oncol 2014;110:129–35. https://doi.org/10.1002/JSO.23618.

[221] Siewert JR, Bottcher K, Stein HJ, Roder JD, Gastric G, Group CS, et al. Relevant prognostic factors in gastric cancer: ten-year results of the German Gastric Cancer Study. Ann Surg 1998;228:449. https://doi.org/10.1097/00000658-199810000-00002.

[222] Matsuo K, Lee SW, Tanaka R, Imai Y, Honda K, Taniguchi K, et al. T stage and venous invasion are crucial prognostic factors for long-term survival of patients with remnant gastric cancer: a cohort study. World J Surg Oncol 2021;19:1–7. https://doi.org/10.1186/S12957-021-02400-5/TABLES/3.

[223] Petrelli F, Berenato R, Turati L, Mennitto A, Steccanella F, Caporale M, et al. Prognostic value of diffuse versus intestinal histotype in patients with gastric cancer: a systematic review and meta-analysis. J Gastrointest Oncol 2017;8:148. https://doi.org/10.21037/JGO.2017.01.10.

[224] Claesson MJ, Jeffery IB, Conde S, Power SE, O'connor EM, Cusack S, et al. Gut microbiota composition correlates with diet and health in the elderly. Nature 2012;488:178–84. https://doi.org/10.1038/nature11319.

[225] Gao Z, Guo B, Gao R, Zhu Q, Qin H, Fortini Grenfell Queiroz RE, et al. Microbiota disbiosis is associated with colorectal cancer 2015. https://doi.org/10.3389/fmicb.2015.00020.

[226] Zhang Y, Hoffmeister M, Weck MN, Chang-Claude J, Brenner H. Original Contribution Helicobacter pylori Infection and Colorectal Cancer Risk: Evidence From a Large Population-based Case-Control Study in Germany. Am J Epidemiol 2012;175:441–50. https://doi.org/10.1093/aje/kwr331.

[227] Kohoutova D, Smajs D, Moravkova P, Cyrany J, Moravkova M, Forstlova M, et al. Escherichia coli strains of phylogenetic group B2 and D and bacteriocin production are associated with advanced colorectal neoplasia. BMC Infect Dis 2014;14:1–8. https://doi.org/10.1186/S12879-014-0733-7/TABLES/4.

[228] Wang J, Wang Y, Li Z, Gao X, Huang D. Global Analysis of Microbiota Signatures in Four Major Types of Gastrointestinal Cancer. Front Oncol 2021;11. https://doi.org/10.3389/fonc.2021.685641.

[229] Lordick F, Carneiro F, Cascinu S, Fleitas T, Haustermans K, Piessen G, et al. Gastric cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Annals of Oncology 2022;33:1005–20. https://doi.org/10.1016/J.ANNONC.2022.07.004/ATTACHMENT/04C7EB99-242E-44D5-8E30-5A5A8B8464E3/MMC1.PDF.

[230] Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. American Journal of Pathology 2002;161:1961–71. https://doi.org/10.1016/S0002-9440(10)64472-0.

[231] Cruz-Flores R, López-Carvallo JA, Cáceres-Martínez J, Dhar AK. Microbiome analysis from formalin-fixed paraffin-embedded tissues: Current challenges and future perspectives. J Microbiol Methods 2022;196. https://doi.org/10.1016/j.mimet.2022.106476.

[232] Pinto-Ribeiro I, Ferreira RM, Pereira-Marques J, Pinto V, Macedo G, Carneiro F, et al. Evaluation of the Use of Formalin-Fixed and Paraffin-Embedded Archive Gastric Tissues for Microbiota Characterization Using Next-Generation Sequencing. International Journal of Molecular Sciences Article 2020;21:1096. https://doi.org/10.3390/ijms21031096.

[233] Borgognone A, Serna G, Noguera-Julian M, Alonso L, Parera M, Català-Moll F, et al. Performance of 16S Metagenomic Profiling in Formalin-Fixed Paraffin-Embedded versus Fresh-Frozen Colorectal Cancer Tissues. Cancers (Basel) 2021;13:5421. https://doi.org/10.3390/cancers13215421.

[234] Emery DC, Shoemark DK, Batstone TE, Waterfall CM, Coghill JA, Cerajewska TL, et al. 16S rRNA next generation sequencing analysis shows bacteria in Alzheimer's Post-Mortem Brain. Front Aging Neurosci 2017;9:261413. https://doi.org/10.3389/FNAGI.2017.00195/BIBTEX.

[235] Janda JM, Abbott SL. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. J Clin Microbiol 2007;45:2761. https://doi.org/10.1128/JCM.01228-07.

[236] Hill JE, Paccagnella A, Law K, Melito PL, Woodward DL, Price L, et al. Identification of Campylobacter spp. and discrimination from Helicobacter and Arcobacter spp. by direct sequencing of PCR-amplified cpn60 sequences and comparison to cpnDB, a chaperonin reference sequence database. J Med Microbiol 2006;55:393–9. https://doi.org/10.1099/JMM.0.46282-0/CITE/REFWORKS.

[237] Links MG, Demeke T, Gräfenhan T, Hill JE, Hemmingsen SM, Dumonceaux TJ. Simultaneous profiling of seed-associated bacteria and fungi reveals antagonistic interactions between microorganisms within a shared epiphytic microbiome on Triticum and Brassica seeds. New Phytologist 2014;202:542–53. https://doi.org/10.1111/NPH.12693.

[238] Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microbiol 2007;73:278–88. https://doi.org/10.1128/AEM.01177-06/ASSET/A3A83AFC-709C-4D41-9D8D-A367782282DF/ASSETS/GRAPHIC/ZAM0010773950004.JPEG.

[239] Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A 2012;109:6241–6. https://doi.org/10.1073/PNAS.1117018109/SUPPL_FILE/SD01.XLS.

[240] Beveridge TJ. Use of the Gram stain in microbiology. Biotechnic and Histochemistry 2001;76:111–8. https://doi.org/10.1080/bih.76.3.111.118.

[241] Stewart EJ. Growing unculturable bacteria. J Bacteriol 2012;194:4151–60. https://doi.org/10.1128/JB.00345-12/ASSET/E8A88DE6-116B-4524-81EB-E9C97B7C4E8F/ASSETS/GRAPHIC/ZJB9990917380003.JPEG.

[242] Renwick S, Ganobis CM, Elder RA, Gianetto-Hill C, Higgins G, Robinson A V., et al. Culturing Human Gut Microbiomes in the Laboratory. Https://DoiOrg/101146/Annurev-Micro-031021-084116 2021;75:49–69. https://doi.org/10.1146/ANNUREV-MICRO-031021-084116.

[243] Delgado S, Cabrera-Rubio R, Mira A, Suárez A, Mayo B. Microbiological Survey of the Human Gastric Ecosystem Using Culturing and Pyrosequencing Methods. Microb Ecol 2013;65:763–72. https://doi.org/10.1007/s00248-013-0192-5.

[244] Spiegelhauer MR, Kupcinskas J, Johannesen TB, Urba M, Skieceviciene J, Jonaitis L, et al. Transient and Persistent Gastric Microbiome: Adherence of Bacteria in Gastric Cancer and Dyspeptic Patient Biopsies after Washing. J Clin Med 2020;9:1–23. https://doi.org/10.3390/JCM9061882.

[245] Mannion A, Sheh A, Shen Z, Dzink-Fox J, Piazuelo MB, Wilson KT, et al. Shotgun Metagenomics of Gastric Biopsies Reveals Compositional and Functional Microbiome Shifts in High- and Low-Gastric-Cancer-Risk Populations from Colombia, South America. Gut Microbes 2023;15. https://doi.org/10.1080/19490976.2023.2186677/SUPPL_FILE/KGMI_A_2186677_SM5315.PPTX.

[246] Khosravi Y, Dieye Y, Poh BH, Ng CG, Loke MF, Goh KL, et al. Culturable bacterial microbiota of the stomach of helicobacter pylori positive and negative gastric disease patients. Scientific World Journal 2014;2014. https://doi.org/10.1155/2014/610421.

[247] Hu Y, He LH, Xiao D, Liu GD, Gu YX, Tao XX, et al. Bacterial flora concurrent with Helicobacter pylori in the stomach of patients with upper gastrointestinal diseases. World Journal of Gastroenterology : WJG 2012;18:1257. https://doi.org/10.3748/WJG.V18.I11.1257.

[248] Sjostedt S, Kager L, Heimdahl A, Nordt CE. Microbial Colonization of Tumors in Relation to the Upper Gastrointestinal Tract in Patients with Gastric Carcinoma. Ann Surg 1988;207:341–6.

[249] QIAGEN. QIAamp UCP Pathogen Mini Handbook - QIAGEN. 2021.

[250] QIAGEN. QIAamp DNA Mini and Blood Mini Handbook - EN - QIAGEN. 5th ed. 2016.

[251] Conway C, Chalkley R, High A, MacLennan K, Berri S, Chengot P, et al. Next-generation sequencing for simultaneous determination of human papillomavirus load, subtype, and associated genomic copy number changes in tumors. Journal of Molecular Diagnostics 2012;14:104–11. https://doi.org/10.1016/j.jmoldx.2011.10.003.

[252] Young C, Wood HM, Balaguer AF, Bottomley D, Gallop N, Wilkinson L, et al. Microbiome analysis of over 2000 NHS Bowel

Cancer Screening Programme (NHSBCSP) samples shows the potential to improve screening accuracy. Clin Cancer Res 2021;27:2246. https://doi.org/10.1158/1078-0432.CCR-20-3807.

[253] 16S Illumina Amplicon Protocol : earthmicrobiome n.d. https://earthmicrobiome.org/protocols-and-standards/16s/ (accessed June 2, 2023).

[254] Kanatas A, Chengot P, Ong TK, Ho MW, Sethi N, Taylor M, et al. Genomic analysis to assess disease progression and recurrence in patients with oral squamous cell carcinoma: – a preliminary study. British Journal of Oral and Maxillofacial Surgery 2018;56:198–205. https://doi.org/10.1016/j.bjoms.2018.01.010.

[255] Gloor GB, Hummelen R, Macklaim JM, Dickson RJ, Fernandes AD, MacPhee R, et al. Microbiome Profiling by Illumina Sequencing of Combinatorial Sequence-Tagged PCR Products. PLoS One 2010;5:e15406. https://doi.org/10.1371/JOURNAL.PONE.0015406.

[256] Carlucci C, Jones CS, Oliphant K, Yen S, Daigneault M, Carriero C, et al. Effects of defined gut microbial ecosystem components on virulence determinants of Clostridioides difficile. Sci Rep 2019;9. https://doi.org/10.1038/s41598-018-37547-x.

[257] Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res 2012;22:299–306. https://doi.org/10.1101/GR.126516.111.

[258] Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37:852–7. https://doi.org/10.1038/S41587-019-0209-9.

[259] Sandi Yen. Microbial and Metabolic Relationships of the Preterm Gut Microbiome in Response to Antibiotic and Dietary Interventions. PhD thesis, University of Guelph, 2019.

[260] Wright ES. DECIPHER: Harnessing local sequence context to improve protein multiple sequence alignment. BMC Bioinformatics 2015;16:1–14. https://doi.org/10.1186/S12859-015-0749-Z/FIGURES/6.

[261] Wickham H. Reshaping Data with the reshape Package. J Stat Softw 2007;21:1–20.

[262] Paradis E, Schliep K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 2019;35:526–8. https://doi.org/10.1093/BIOINFORMATICS/BTY633.

[263] Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings 2023.

[264] Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. Msa: An R package for multiple sequence alignment. Bioinformatics 2015;31:3997–9. https://doi.org/10.1093/BIOINFORMATICS/BTV494.

[265] McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 2013;8. https://doi.org/10.1371/JOURNAL.PONE.0061217.

[266] Yu G. Data Integration, Manipulation and Visualization of Phylogenetic Trees. Data Integration, Manipulation and Visualization of Phylogenetic Trees 2022. https://doi.org/10.1201/9781003279242.

[267] Campitelli E. ggnewscale: Multiple Fill and Colour Scales in “ggplot2” 2023.

[268] Edgar RC. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping n.d. https://doi.org/10.1101/2021.06.20.449169.

[269] Kim KK, Lee JS, Stevens DA. Microbiology and epidemiology of Halomonas species. Http://DxDoiOrg/102217/Fmb13108 2013;8:1559–73. https://doi.org/10.2217/FMB.13.108.

[270] Hau HH, Gralnick JA. Ecology and Biotechnology of the Genus Shewanella. Https://DoiOrg/101146/AnnurevMicro61080706093257 2007;61:237–58. https://doi.org/10.1146/ANNUREV.MICRO.61.080706.093257.

[271] Downes J, Dewhirst FE, Tanner ACR, Wade WG. Description of Alloprevotella rava gen. nov., sp. nov., isolated from the human oral cavity, and reclassification of Prevotella tannerae Moore et al. 1994 as Alloprevotella tannerae gen. nov., comb. nov. Int J Syst Evol Microbiol 2013;63:1214. https://doi.org/10.1099/IJS.0.041376-0.

[272] Frandsen EVG, Poulsen K, Könönen E, Kilian M. Diversity of Capnocytophaga species in children and description of

Capnocytophaga leadbetteri sp. nov. and Capnocytophaga genospecies AHN8471. Int J Syst Evol Microbiol 2008;58:324–36. https://doi.org/10.1099/IJS.0.65373-0/CITE/REFWORKS.

[273] Tsuzukibashi O, Uchibori S, Kobayashi T, Umezawa K, Mashimo C, Nambu T, et al. Isolation and identification methods of Rothia species in oral cavities. J Microbiol Methods 2017;134:21–6. https://doi.org/10.1016/J.MIMET.2017.01.005.

[274] Guilloux CA, Lamoureux C, Beauruelle C, Héry-Arnaud G. Porphyromonas: A neglected potential key genus in human microbiomes. Anaerobe 2021;68:102230. https://doi.org/10.1016/J.ANAEROBE.2020.102230.

[275] Haapasalo M, Ranta H, Shah H. Mitsuokella dentalis sp. nov. from dental root canals. Int J Syst Bacteriol 1986;36:566–8. https://doi.org/10.1099/00207713-36-4-566/CITE/REFWORKS.

[276] Dioguardi M, Crincoli V, Laino L, Alovisi M, Sovereto D, Muzio L Lo, et al. Prevalence of Bacteria of Genus Actinomyces in Persistent Extraradicular Lesions—Systematic Review. Journal of Clinical Medicine 2020, Vol 9, Page 457 2020;9:457. https://doi.org/10.3390/JCM9020457.

[277] Eribe ERK, Olsen I. Leptotrichia species in human infections. Anaerobe 2008;14:131–7. https://doi.org/10.1016/J.ANAEROBE.2008.04.004.

[278] Smith AJ, Jackson MS, Bagg J. The ecology of staphylococcus species in the oral cavity. J Med Microbiol 2001;50:940–6. https://doi.org/10.1099/0022-1317-50-11-940/CITE/REFWORKS.

[279] Han YW. Fusobacterium nucleatum: a commensal-turned pathogen. Curr Opin Microbiol 2015;23:141–7. https://doi.org/10.1016/J.MIB.2014.11.013.

[280] Hedberg ME, Moore ERB, Svensson-Stadler L, Hörstedt P, Baranov V, Hernell O, et al. Lachnoanaerobaculum gen. nov., a new genus in the Lachnospiraceae: characterization of Lachnoanaerobaculum umeaense gen. nov., sp. nov., isolated from the human small intestine, and Lachnoanaerobaculum orale sp. nov., isolated from saliva, and reclassification of Eubacterium saburreum (Prévot 1966) Holdeman and Moore 1970 as Lachnoanaerobaculum saburreum comb. nov. Int J Syst Evol Microbiol 2012;62:2685. https://doi.org/10.1099/IJS.0.033613-0.

[281] Slots J, Feik D, Rams TE. Prevalence and antimicrobial susceptibility of Enterobacteriaceae, Pseudomonadaceae and Acinetobacter in human periodontitis. Oral Microbiol Immunol 1990;5:149–54. https://doi.org/10.1111/J.1399-302X.1990.TB00413.X.

[282] Willis JR, Gabaldón T. The Human Oral Microbiome in Health and Disease: From Sequences to Ecosystems. Microorganisms 2020, Vol 8, Page 308 2020;8:308. https://doi.org/10.3390/MICROORGANISMS8020308.

[283] Mizuguchi A, Takai A, Shimizu T, Matsumoto T, Kumagai K, Miyamoto S, et al. Genetic features of multicentric/multifocal intramucosal gastric carcinoma. Int J Cancer 2018;143:1923–34. https://doi.org/10.1002/IJC.31578.

[284] Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. Front Microbiol 2016;7:180723. https://doi.org/10.3389/FMICB.2016.00459/BIBTEX.

[285] Hoek MJA van, Merks RMH. Emergence of microbial diversity due to cross-feeding interactions in a spatial model of gut microbial metabolism. BMC Syst Biol 2017;11:1–18. https://doi.org/10.1186/S12918-017-0430-4/FIGURES/12.

[286] Rakoff-Nahoum S, Foster KR, Comstock LE. The evolution of cooperation within the gut microbiota. Nature 2015 533:7602 2016;533:255–9. https://doi.org/10.1038/nature17626.

[287] Catlett JL, Carr S, Cashman M, Smith MD, Walter M, Sakkaff Z, et al. Metabolic Synergy between Human Symbionts Bacteroides and Methanobrevibacter . Microbiol Spectr 2022;10. https://doi.org/10.1128/spectrum.01067-22.

[288] Coyte KZ, Rakoff-Nahoum S. Understanding Competition and Cooperation within the Mammalian Gut Microbiome. Current Biology 2019;29:R538–44. https://doi.org/10.1016/j.cub.2019.04.017.

[289] Rams TE, Feik D, Listgarten MA, Slots J. Peptostreptococcus micros in human periodontitis. Oral Microbiol Immunol 1992;7:1–6. https://doi.org/10.1111/J.1399-302X.1992.TB00011.X.

[290] Shah MS, DeSantis TZ, Weinmaier T, McMurdie PJ, Cope JL, Altrichter A, et al. Leveraging sequence-based faecal microbial

community survey data to identify a composite biomarker for colorectal cancer. Gut 2018;67:882–91. https://doi.org/10.1136/GUTJNL-2016-313189.

[291] Zhao L, Zhang X, Zhou Y, Fu K, Lau HCH, Chun TWY, et al. Parvimonas micra promotes colorectal tumorigenesis and is associated with prognosis of colorectal cancer patients. Oncogene 2022;41:4200–10. https://doi.org/10.1038/s41388-022-02395-7.

[292] Hatta MNA, Mohamad Hanif EA, Chin S-F, Low TY, Neoh H. Parvimonas micra infection enhances proliferation, wound healing, and inflammation of a colorectal cancer cell line . Biosci Rep 2023;43. https://doi.org/10.1042/bsr20230609.

[293] Hong B young, Ideta T, Lemos BS, Igarashi Y, Tan Y, DiSiena M, et al. Characterization of Mucosal Dysbiosis of Early Colonic Neoplasia. Npj Precision Oncology 2019 3:1 2019;3:1–10. https://doi.org/10.1038/s41698-019-0101-6.

[294] Komiyama S, Yamada T, Takemura N, Kokudo N, Hase K, Kawamura YI. Profiling of tumour-associated microbiota in human hepatocellular carcinoma. Scientific Reports 2021 11:1 2021;11:1–9. https://doi.org/10.1038/s41598-021-89963-1.

[295] Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, et al. A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. Genome Med 2017;9:1–12. https://doi.org/10.1186/S13073-017-0490-5/FIGURES/4.

[296] Li ZP, Liu JX, Lu LL, Wang LL, Xu L, Guo ZH, et al. Overgrowth of Lactobacillus in gastric cancer. World J Gastrointest Oncol 2021;13:1099. https://doi.org/10.4251/WJGO.V13.I9.1099.

[297] O'Hanlon DE, Moench TR, Cone RA. In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. BMC Infect Dis 2011;11. https://doi.org/10.1186/1471-2334-11-200.

[298] Renwick S, Ganobis CM, Elder RA, Gianetto-Hill C, Higgins G, Robinson A V, et al. Culturing Human Gut Microbiomes in the Laboratory. Annu Rev Microbiol 2021;75:49–69. https://doi.org/10.1146/annurev-micro-031021.

[299] Feria-Gervasio D, Denis S, Alric M, Brugère JF. In vitro maintenance of a human proximal colon microbiota using the continuous fermentation system P-ECSIM. Appl Microbiol

Biotechnol 2011;91:1425–33. https://doi.org/10.1007/S00253-011-3462-5/TABLES/2.

[300] McDonald JAK, Schroeter K, Fuentes S, Heikamp-deJong I, Khursigara CM, de Vos WM, et al. Evaluation of microbial community reproducibility, stability and composition in a human distal gut chemostat model. J Microbiol Methods 2013;95:167–74. https://doi.org/10.1016/J.MIMET.2013.08.008.

[301] Van Den Abbeele P, Grootaert C, Marzorati M, Possemiers S, Verstraete W, Gérard P, et al. Microbial community development in a dynamic gut model is reproducible, colon region specific, and selective for bacteroidetes and Clostridium cluster IX. Appl Environ Microbiol 2010;76:5237–46. https://doi.org/10.1128/AEM.00759-10.

[302] Cueva C, Gil-Sánchez I, Tamargo A, Miralles B, Crespo J, Bartolomé B, et al. Gastrointestinal digestion of food-use silver nanoparticles in the dynamic SIMulator of the GastroIntestinal tract (simgi®). Impact on human gut microbiota. Food and Chemical Toxicology 2019;132. https://doi.org/10.1016/J.FCT.2019.110657.

[303] Verhoeckx K, Cotter P, López-Expósito I, Kleiveland C, Lea T, Mackie A, et al. The Impact of Food Bioactives on Health. Cham: Springer International Publishing; 2015. https://doi.org/10.1007/978-3-319-16104-4.

[304] Petrof EO, Gloor GB, Vanner SJ, Weese SJ, Carter D, Daigneault MC, et al. Stool substitute transplant therapy for the eradication of Clostridium difficile infection: “RePOOPulating” the gut. Microbiome 2013;1:1–12. https://doi.org/10.1186/2049-2618-1-3/FIGURES/5.

[305] Yen S, McDonald JAK, Schroeter K, Oliphant K, Sokolenko S, Blondeel EJM, et al. Metabolomic analysis of human fecal microbiota: A comparison of feces-derived communities and defined mixed communities. J Proteome Res 2015;14:1472–82. https://doi.org/10.1021/pr5011247.

[306] Attai H, Wilde J, Liu R, Chopyk J, Garcia AG, Allen-Vercoe E, et al. Bacteriophage-Mediated Perturbation of Defined Bacterial Communities in an In Vitro Model of the Human Gut . Microbiol Spectr 2022;10. https://doi.org/10.1128/SPECTRUM.01135-22/SUPPL_FILE/SPECTRUM.01135-22-S0004.XLSX.

[307] Ruiz-Rico M, Renwick S, Vancuren SJ, Robinson A V., Gianetto-Hill C, Allen-Vercoe E, et al. Impact of food preservatives based on

immobilized phenolic compounds on an in vitro model of human gut microbiota. Food Chem 2023;403:134363. https://doi.org/10.1016/J.FOODCHEM.2022.134363.

[308] Koziolek M, Schneider F, Grimm M, Mode C, Seekamp A, Roustom T, et al. Intragastric pH and pressure profiles after intake of the high-caloric, high-fat meal as used for food effect studies. Journal of Controlled Release 2015;220:71–8. https://doi.org/10.1016/j.jconrel.2015.10.022.

[309] Hill MJ. Factors Controlling the Microflora of the Healthy Upper Gastrointestinal Tract. CRC Press; 2020. https://doi.org/10.1201/9781003068020-2.

[310] Sanduleanu S, Jonkers D, De Bruine A, Hameeteman W, Stockbrügger RW. Non-Helicobacter pylori bacterial flora during acid-suppressive therapy: Differential findings in gastric juice and gastric mucosa. Aliment Pharmacol Ther 2001;15:379–88. https://doi.org/10.1046/j.1365-2036.2001.00888.x.

[311] Tsuda A, Suda W, Morita H, Takanashi K, Takagi A, Koga Y, et al. Influence of Proton-Pump Inhibitors on the Luminal Microbiota in the Gastrointestinal Tract. Clin Transl Gastroenterol 2015;6:89. https://doi.org/10.1038/ctg.2015.20.

[312] Schreiber S, Garten D, Nguyen TH, Konradt M, Bücker R, Scheid P. In situ measurement of pH in the secreting canaliculus of the gastric parietal cell and adjacent structures. Cell Tissue Res 2007;329:313–20. https://doi.org/10.1007/s00441-007-0427-1.

[313] Hao G, Xu ZP, Li L. Manipulating extracellular tumour pH: an effective target for cancer therapy. RSC Adv 2018;8:22182–92. https://doi.org/10.1039/C8RA02095G.

[314] Gianetto-Hill CM, Vancuren SJ, Daisley B, Renwick S, Wilde J, Schroeter K, et al. The Robogut: A Bioreactor Model of the Human Colon for Evaluation of Gut Microbial Community Ecology and Function. Curr Protoc 2023;3:e737. https://doi.org/10.1002/CPZ1.737.

[315] Rose M, Baxter M, Brereton N, Baskaran C. Dietary exposure to metals and other elements in the 2006 UK Total Diet Study and some trends over the last 30 years. Http://DxDoiOrg/101080/194400492010496794 2010;27:1380–404. https://doi.org/10.1080/19440049.2010.496794.

[316] Popkin BM, D'Anci KE, Rosenberg IH. Water, Hydration and Health. Nutr Rev 2010;68:439. https://doi.org/10.1111/J.1753-4887.2010.00304.X.

[317] Herlemann DPR, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. The ISME Journal 2011 5:10 2011;5:1571–9. https://doi.org/10.1038/ismej.2011.41.

[318] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 2011;17:10–2. https://doi.org/10.14806/EJ.17.1.200.

[319] Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13:581. https://doi.org/10.1038/nmeth.3869.

[320] Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26:2460–1. https://doi.org/10.1093/bioinformatics/btq461.

[321] Ganobis CM, Al-Abdul-Wahid MS, Renwick S, Yen S, Carriero C, Aucoin MG, et al. 1D 1H NMR as a Tool for Fecal Metabolomics. Curr Protoc Chem Biol 2020;12:e83. https://doi.org/10.1002/CPCH.83.

[322] Yu G. A Tidy Tool for Phylogenetic Tree Data Manipulation [R package tidytree version 0.4.5] 2023.

[323] Wang LG, Lam TTY, Xu S, Dai Z, Zhou L, Feng T, et al. Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Mol Biol Evol 2020;37:599–603. https://doi.org/10.1093/MOLBEV/MSZ240.

[324] Oliphant K, Parreira VR, Cochrane K, Allen-Vercoe E. Drivers of human gut microbial community assembly: coadaptation, determinism and stochasticity. The ISME Journal 2019 13:12 2019;13:3080–92. https://doi.org/10.1038/s41396-019-0498-5.

[325] Axelsson L. Lactic Acid Bacteria: Microbiological and Functional Aspects. Third Edition. Marcel Dekker, Inc; 2004.

[326] Genta RM. Helicobacter pylori, Inflammation, Mucosal Damage, and Apoptosis: Pathogenesis and Definition of Gastric Atrophy. Gastroenterology 1997;113:S51–5. https://doi.org/10.1016/S0016-5085(97)80012-1.

[327] Silva YP, Bernardi A, Frozza RL. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. Front Endocrinol (Lausanne) 2020;11:508738. https://doi.org/10.3389/FENDO.2020.00025/BIBTEX.

[328] Koh A, De Vadder F, Kovatcheva-Datchary P, Bäckhed F. From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. Cell 2016;165:1332–45. https://doi.org/10.1016/J.CELL.2016.05.041.

[329] Wu X, Wu Y, He L, Wu L, Wang X, Liu Z. Effects of the intestinal microbial metabolite butyrate on the development of colorectal cancer. J Cancer 2018;9:2510. https://doi.org/10.7150/JCA.25324.

[330] Evans M, Murofushi T, Tsuda H, Mikami Y, Zhao N, Ochiai K, et al. Combined effects of starvation and butyrate on autophagy-dependent gingival epithelial cell death. J Periodontal Res 2017;52:522–31. https://doi.org/10.1111/jre.12418.

[331] Zhang K, Ji X, Song Z, Wu F, Qu Y, Jin X, et al. Butyrate Inhibits Gastric Cancer Cells by Inducing Mitochondriamediated Apoptosis. Comb Chem High Throughput Screen 2022;26:630–8. https://doi.org/10.2174/1386207325666220720114642.

[332] Ternes D, Tsenkova M, Pozdeev VI, Meyers M, Koncina E, Atatri S, et al. The gut microbial metabolite formate exacerbates colorectal cancer progression. Nat Metab 2022;4:458–75. https://doi.org/10.1038/s42255-022-00558-0.

[333] Kang Y-K, Boku N, Satoh T, Ryu M-H, Chao Y, Kato K, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. The Lancet 2017;390:2461–71. https://doi.org/10.1016/S0140-6736(17)31827-5.

[334] Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. The Lancet 2021;398:27–40. https://doi.org/10.1016/S0140-6736(21)00797-2.

[335] Rha SY, Wyrwicz LS, Weber PEY, Bai Y, Ryu MH, Lee J, et al. VP1-2023: Pembrolizumab (pembro) plus chemotherapy (chemo) as first-line therapy for advanced HER2-negative gastric or gastroesophageal junction (G/GEJ) cancer: Phase III KEYNOTE-859 study. Annals of Oncology 2023;34:319–20. https://doi.org/10.1016/j.annonc.2023.01.006.

[336] Davar D, Dzutsev AK, McCulloch JA, Rodrigues RR, Chauvin JM, Morrison RM, et al. Fecal microbiota transplant overcomes resistance to anti-PD-1 therapy in melanoma patients. Science (1979) 2021;371:595–602. https://doi.org/10.1126/science.abf3363.

Appendix A: Ethical Approvals

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Prof Phil Quirke
Pathology and Data Analytics, Leeds Institute of Medical Research at St. James'
Wellcome Trust Brenner Building University of Leeds
LS9 7TFN/A

Email: approvals@hra.nhs.uk
HCRW.approvals@wales.nhs.uk

15 April 2021

Dear Prof Quirke

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Investigation of the microbiome in the gastrointestinal tract**
IRAS project ID:	**290019**
Protocol number:	**N/A**
REC reference:	**21/NW/0042**
Sponsor	**University of Leeds**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The standard conditions document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **290019**. Please quote this on all correspondence.

Yours sincerely,
Amber Ecclestone

Approvals Specialist

Email: approvals@hra.nhs.uk

Copy to: *Dr Jean Uniacke*

(Supplementary) Figure 51. Health Research Authority and Health and Care Research Wales Approval letter

UNIVERSITY of GUELPH

RESEARCH ETHICS BOARDS
Certification of Ethical Acceptability of Research Involving Human Participants

APPROVAL PERIOD:	May 13, 2022
EXPIRY DATE:	May 12, 2023
REB:	NPES
REB NUMBER:	22-04-020
TYPE OF REVIEW:	Delegated
PRINCIPAL INVESTIGATOR:	Allen-Vercoe, Emma (eav@uoguelph.ca)
DEPARTMENT:	Molecular & Cellular Biology
SPONSOR(S):	Cancer Research UK, Grand Challenges 2019; University of Leeds discretionary funding
TITLE OF PROJECT:	Isolation of microbiota from stomach cancer biopsies

The members of the University of Guelph Research Ethics Board have examined the protocol which describes the participation of the human participants in the above-named research project and considers the procedures, as described by the applicant, to conform to the University's ethical standards and the Tri-Council Policy Statement, 2nd Edition.

The REB requires that researchers:

- Adhere to the protocol as last reviewed and **approved** by the REB.
- Receive approval from the REB for any **modifications** before they can be implemented.
- Report any **change in the source of funding.**
- Report **unexpected events or incidental findings** to the REB as soon as possible with an indication of how these events affect, in the view of the Principal Investigator, the safety of the participants, and the continuation of the protocol.
- Are responsible for **ascertaining and complying with all applicable legal and regulatory requirements** with respect to consent and the protection of privacy of participants in the jurisdiction of the research project.

The Principal Investigator must:

- Ensure that the ethical guidelines and approvals of facilities or institutions involved in the research are obtained and filed with the REB prior to the initiation of any research protocols.
- Submit an **Annual Renewal** to the REB upon completion of the project. If the research is a multi-year project, a status report must be submitted annually prior to the expiry date. Failure to submit an annual status report will lead to your study being suspended and potentially terminated.

The approval for this protocol terminates on the **EXPIRY DATE**, or the term of your appointment or employment at the University of Guelph whichever comes first.

Signature: Date: May 13, 2022

W. G. Pyle
Chair, Research Ethics Board-NPES

Page **1** of **1**

(Supplementary) Figure 52. University of Guelph Research Ethics Boards approval

Appendix B: Supplementary data

(Supplemental) Table 19. Species present on include-list used to perform decontamination. Cohort(s) through which each species was identified as tissue-resident is demonstrated. TCGA, The Cancer Genome Atlas.

Species	Cohort through which species identified as tissue-resident
Actinomyces_graevenitzii	both
Actinomyces_naeslundii	both
Actinomyces_sp._HPA0247	both
Aggregatibacter_segnis	both
Aggregatibacter_sp._oral_taxon_458	both
Alloprevotella_rava	both
Alloprevotella_tannerae	both
Campylobacter_concisus	both
Capnocytophaga_sp._oral_taxon_329	both
Catonella_morbi	both
Centipeda_periodontii	both
Dialister_invisus	both
Dialister_pneumosintes	both
Filifactor_alocis	both
Gemella_haemolysans	both
Haemophilus_parainfluenzae	both
Helicobacter_pylori	both
Human_betaherpesvirus_5	both
Johnsonella_ignava	both
Lachnoanaerobaculum_saburreum	both
Lachnospiraceae_bacterium_oral_taxon_082	both
Lachnospiraceae_bacterium_oral_taxon_500	both
Lactobacillus_crispatus	both
Lactobacillus_gasseri	both
Lactobacillus_reuteri	both
Lactobacillus_salivarius	both
Lactobacillus_ultunensis	both
Lactobacillus_vaginalis	both
Leptotrichia_wadei	both
Megasphaera_micronuciformis	both
Mitsuokella_sp._oral_taxon_131	both
Olsenella_uli	both
Oribacterium_sp._oral_taxon_078	both
Parvimonas_micra	both
Peptoanaerobacter_stomatis	both
Peptostreptococcus_stomatis	both
Porphyromonas_endodontalis	both
Porphyromonas_gingivalis	both
Prevotella_aurantiaca	both
Prevotella_baroniae	both
Prevotella_buccae	both

Species	Cohort through which species identified as tissue-resident
Prevotella_denticola	both
Prevotella_disiens	both
Prevotella_enoeca	both
Prevotella_fusca	both
Prevotella_histicola	both
Prevotella_intermedia	both
Prevotella_maculosa	both
Prevotella_marshii	both
Prevotella_melaninogenica	both
Prevotella_multiformis	both
Prevotella_multisaccharivorax	both
Prevotella_nanceiensis	both
Prevotella_nigrescens	both
Prevotella_oris	both
Prevotella_oulorum	both
Prevotella_pallens	both
Prevotella_pleuritidis	both
Prevotella_salivae	both
Prevotella_scopos	both
Prevotella_sp._C561	both
Prevotella_sp._F0091	both
Prevotella_sp._HJM029	both
Prevotella_sp._HMSC069G02	both
Prevotella_sp._ICM33	both
Prevotella_sp._oral_taxon_306	both
Prevotella_veroralis	both
Selenomonas_flueggei	both
Selenomonas_infelix	both
Selenomonas_sp._CM52	both
Selenomonas_sp._oral_taxon_126	both
Selenomonas_sp._oral_taxon_136	both
Selenomonas_sp._oral_taxon_149	both
Selenomonas_sp._oral_taxon_892	both
Selenomonas_sp._oral_taxon_920	both
Selenomonas_sputigena	both
Slackia_exigua	both
Slackia_sp._CM382	both
Solobacterium_moorei	both
Stomatobaculum_longum	both
Streptococcus_mitis	both
Streptococcus_oralis	both
Streptococcus_sp._oral_taxon_071	both
Tannerella_forsythia	both
Treponema_denticola	both
Treponema_maltophilum	both
Veillonella_dispar	both
Veillonella_parvula	both
Veillonella_sp._3_1_44	both

Species	Cohort through which species identified as tissue-resident
Veillonella_sp._6_1_27	both
Veillonella_sp._oral_taxon_158	both
Abiotrophia_defectiva	100,000 Genomes Project only
Abiotrophia_sp._HMSC24B09	100,000 Genomes Project only
Achromobacter_insuavis	100,000 Genomes Project only
Achromobacter_ruhlandii	100,000 Genomes Project only
Achromobacter_sp.	100,000 Genomes Project only
Acidovorax_sp._KKS102	100,000 Genomes Project only
Acidovorax_temperans	100,000 Genomes Project only
Acinetobacter_baumannii	100,000 Genomes Project only
Acinetobacter_bohemicus	100,000 Genomes Project only
Acinetobacter_guillouiae	100,000 Genomes Project only
Acinetobacter_johnsonii	100,000 Genomes Project only
Acinetobacter_sp._BMW17	100,000 Genomes Project only
Acinetobacter_sp._CIP_64.7	100,000 Genomes Project only
Acinetobacter_sp._ETR1	100,000 Genomes Project only
Acinetobacter_sp._Root1280	100,000 Genomes Project only
Actinomyces_gerencseriae	100,000 Genomes Project only
Actinomyces_israelii	100,000 Genomes Project only
Actinomyces_odontolyticus	100,000 Genomes Project only
Actinomyces_oris	100,000 Genomes Project only
Actinomyces_sp._HMSC035G02	100,000 Genomes Project only
Actinomyces_sp._HMSC075C01	100,000 Genomes Project only
Actinomyces_sp._HMSC08A09	100,000 Genomes Project only
Actinomyces_sp._ICM39	100,000 Genomes Project only
Actinomyces_sp._ICM47	100,000 Genomes Project only
Actinomyces_sp._ICM54	100,000 Genomes Project only
Actinomyces_sp._ICM58	100,000 Genomes Project only
Actinomyces_sp._Marseille.P2825	100,000 Genomes Project only
Actinomyces_sp._oral_taxon_171	100,000 Genomes Project only
Actinomyces_sp._oral_taxon_172	100,000 Genomes Project only
Actinomyces_sp._oral_taxon_175	100,000 Genomes Project only
Actinomyces_sp._oral_taxon_180	100,000 Genomes Project only
Actinomyces_sp._oral_taxon_181	100,000 Genomes Project only
Actinomyces_sp._oral_taxon_448	100,000 Genomes Project only
Actinomyces_sp._ph3	100,000 Genomes Project only
Actinomyces_sp._S6.Spd3	100,000 Genomes Project only
Actinomyces_viscosus	100,000 Genomes Project only
Aggregatibacter_aphrophilus	100,000 Genomes Project only
Alistipes_putredinis	100,000 Genomes Project only
Alloscardovia_omnicolens	100,000 Genomes Project only
Anaeroglobus_geminatus	100,000 Genomes Project only
Atopobium_parvulum	100,000 Genomes Project only
Atopobium_rimae	100,000 Genomes Project only
Atopobium_sp._BS2	100,000 Genomes Project only
Atopobium_sp._HMSC064B08	100,000 Genomes Project only
Atopobium_sp._ICM42b	100,000 Genomes Project only
Atopobium_sp._oral_taxon_199	100,000 Genomes Project only

Species	Cohort through which species identified as tissue-resident
Bacteroidetes_oral_taxon_274	100,000 Genomes Project only
Bifidobacterium_adolescentis	100,000 Genomes Project only
Bifidobacterium_dentium	100,000 Genomes Project only
Bifidobacterium_longum	100,000 Genomes Project only
Bifidobacterium_sp._12_1_47BFAA	100,000 Genomes Project only
Bifidobacterium_sp._MSTE12	100,000 Genomes Project only
Bosea_sp._WAO	100,000 Genomes Project only
Bulleidia_extructa	100,000 Genomes Project only
Campylobacter_sp._10_1_50	100,000 Genomes Project only
Candida_albicans	100,000 Genomes Project only
Capnocytophaga_gingivalis	100,000 Genomes Project only
Cloacibacterium_normanense	100,000 Genomes Project only
Clostridiales_bacterium_KLE1615	100,000 Genomes Project only
Clostridium_perfringens	100,000 Genomes Project only
Comamonas_testosteroni	100,000 Genomes Project only
Corynebacterium_kroppenstedtii	100,000 Genomes Project only
Corynebacterium_matruchotii	100,000 Genomes Project only
Criibacterium_bergeronii	100,000 Genomes Project only
Cryptobacterium_curtum	100,000 Genomes Project only
Delftia_sp._670	100,000 Genomes Project only
Dolosigranulum_pigrum	100,000 Genomes Project only
Eggerthia_catenaformis	100,000 Genomes Project only
Enhydrobacter_aerosaccus	100,000 Genomes Project only
Enhydrobacter_sp._H5	100,000 Genomes Project only
Enterococcus_cecorum	100,000 Genomes Project only
Escherichia_sp._3_2_53FAA	100,000 Genomes Project only
Faecalibacterium_prausnitzii	100,000 Genomes Project only
Fusicatenibacter_saccharivorans	100,000 Genomes Project only
Fusobacterium_hwasookii	100,000 Genomes Project only
Fusobacterium_massiliense	100,000 Genomes Project only
Fusobacterium_nucleatum	100,000 Genomes Project only
Fusobacterium_periodonticum	100,000 Genomes Project only
Fusobacterium_sp._CM1	100,000 Genomes Project only
Fusobacterium_sp._CM21	100,000 Genomes Project only
Fusobacterium_sp._CM22	100,000 Genomes Project only
Fusobacterium_sp._HMSC064B11	100,000 Genomes Project only
Fusobacterium_sp._HMSC064B12	100,000 Genomes Project only
Fusobacterium_sp._HMSC065F01	100,000 Genomes Project only
Fusobacterium_sp._OBRC1	100,000 Genomes Project only
Fusobacterium_sp._oral_taxon_370	100,000 Genomes Project only
Gemella_morbillorum	100,000 Genomes Project only
Gemella_sanguinis	100,000 Genomes Project only
Gemella_sp._Marseille.P3249	100,000 Genomes Project only
Gemella_sp._oral_taxon_928	100,000 Genomes Project only
Granulicatella_adiacens	100,000 Genomes Project only
Granulicatella_elegans	100,000 Genomes Project only
Granulicatella_sp._HMSC30F09	100,000 Genomes Project only
Granulicatella_sp._HMSC31F03	100,000 Genomes Project only

Species	Cohort through which species identified as tissue-resident
Haemophilus_aegyptius	100,000 Genomes Project only
Haemophilus_haemolyticus	100,000 Genomes Project only
Haemophilus_influenzae	100,000 Genomes Project only
Haemophilus_parahaemolyticus	100,000 Genomes Project only
Haemophilus_paraphrohaemolyticus	100,000 Genomes Project only
Haemophilus_sp._C1	100,000 Genomes Project only
Haemophilus_sp._CCUG_60358	100,000 Genomes Project only
Haemophilus_sp._CCUG_66565	100,000 Genomes Project only
Haemophilus_sp._HMSC061E01	100,000 Genomes Project only
Haemophilus_sp._HMSC066A11	100,000 Genomes Project only
Haemophilus_sp._HMSC068C11	100,000 Genomes Project only
Haemophilus_sp._HMSC071C11	100,000 Genomes Project only
Haemophilus_sp._HMSC073C03	100,000 Genomes Project only
Haemophilus_sp._HMSC61B11	100,000 Genomes Project only
Haemophilus_sp._HMSC71H05	100,000 Genomes Project only
Haemophilus_sp._oral_taxon_851	100,000 Genomes Project only
Haemophilus_sputorum	100,000 Genomes Project only
Helicobacter_acinonychis	100,000 Genomes Project only
Human_betaherpesvirus_6B	100,000 Genomes Project only
Human_betaherpesvirus_7	100,000 Genomes Project only
Kocuria_sp._HMSC066H03	100,000 Genomes Project only
Lachnoanaerobaculum_sp._ICM7	100,000 Genomes Project only
Lachnoanaerobaculum_sp._MSX33	100,000 Genomes Project only
Lachnoanaerobaculum_sp._OBRC5.5	100,000 Genomes Project only
Lactobacillus_acidophilus	100,000 Genomes Project only
Lactobacillus_alimentarius	100,000 Genomes Project only
Lactobacillus_antri	100,000 Genomes Project only
Lactobacillus_casei	100,000 Genomes Project only
Lactobacillus_coleohominis	100,000 Genomes Project only
Lactobacillus_fermentum	100,000 Genomes Project only
Lactobacillus_frumenti	100,000 Genomes Project only
Lactobacillus_gallinarum	100,000 Genomes Project only
Lactobacillus_gastricus	100,000 Genomes Project only
Lactobacillus_helveticus	100,000 Genomes Project only
Lactobacillus_hominis	100,000 Genomes Project only
Lactobacillus_jensenii	100,000 Genomes Project only
Lactobacillus_johnsonii	100,000 Genomes Project only
Lactobacillus_kefiranofaciens	100,000 Genomes Project only
Lactobacillus_mucosae	100,000 Genomes Project only
Lactobacillus_oris	100,000 Genomes Project only
Lactobacillus_panis	100,000 Genomes Project only
Lactobacillus_paracollinoides	100,000 Genomes Project only
Lactobacillus_phage_phiadh	100,000 Genomes Project only
Lactobacillus_sp._HMSC066G01	100,000 Genomes Project only
Lactobacillus_sp._HMSC24D01	100,000 Genomes Project only
Lactobacillus_sp._Marseille.P3519	100,000 Genomes Project only
Lactobacillus_taiwanensis	100,000 Genomes Project only
Lactococcus_lactis	100,000 Genomes Project only

Species	Cohort through which species identified as tissue-resident
Lautropia_mirabilis	100,000 Genomes Project only
Leptotrichia_buccalis	100,000 Genomes Project only
Leptotrichia_hofstadii	100,000 Genomes Project only
Leptotrichia_shahii	100,000 Genomes Project only
Leptotrichia_sp._Marseille.P3007	100,000 Genomes Project only
Leptotrichia_sp._oral_taxon_212	100,000 Genomes Project only
Leptotrichia_sp._oral_taxon_215	100,000 Genomes Project only
Leptotrichia_sp._oral_taxon_225	100,000 Genomes Project only
Leptotrichia_sp._oral_taxon_847	100,000 Genomes Project only
Leptotrichia_sp._oral_taxon_879	100,000 Genomes Project only
Leptotrichia_trevisanii	100,000 Genomes Project only
Limnohabitans_sp._Rim47	100,000 Genomes Project only
Malassezia_globosa	100,000 Genomes Project only
Microbacterium_azadirachtae	100,000 Genomes Project only
Microbacterium_sp._MRS.1	100,000 Genomes Project only
Microbacterium_sp._oral_taxon_186	100,000 Genomes Project only
Micrococcus_aloeverae	100,000 Genomes Project only
Mogibacterium_sp._CM50	100,000 Genomes Project only
Mogibacterium_timidum	100,000 Genomes Project only
Moraxella_osloensis	100,000 Genomes Project only
Mycobacterium_sp._1554424.7	100,000 Genomes Project only
Neisseria_mucosa	100,000 Genomes Project only
Neisseria_sp._HMSC061B04	100,000 Genomes Project only
Neisseria_sp._HMSC063B05	100,000 Genomes Project only
Neisseria_sp._HMSC065C04	100,000 Genomes Project only
Neisseria_sp._HMSC066H01	100,000 Genomes Project only
Neisseria_sp._HMSC073G10	100,000 Genomes Project only
Neisseria_sp._HMSC075C12	100,000 Genomes Project only
Neisseria_sp._HMSC078C12	100,000 Genomes Project only
Neisseria_sp._HMSC15G01	100,000 Genomes Project only
Neisseria_sp._HMSC31F04	100,000 Genomes Project only
Neisseria_sp._oral_taxon_014	100,000 Genomes Project only
Olsenella_profusa	100,000 Genomes Project only
Oribacterium_sinus	100,000 Genomes Project only
Oribacterium_sp._oral_taxon_108	100,000 Genomes Project only
Parabacteroides_merdae	100,000 Genomes Project only
Paracoccus_sp._228	100,000 Genomes Project only
Paracoccus_yeei	100,000 Genomes Project only
Parascardovia_denticolens	100,000 Genomes Project only
Parvimonas_sp._oral_taxon_110	100,000 Genomes Project only
Parvimonas_sp._oral_taxon_393	100,000 Genomes Project only
Peptoniphilus_indolicus	100,000 Genomes Project only
Peptoniphilus_lacrimalis	100,000 Genomes Project only
Peptoniphilus_sp._BV3C26	100,000 Genomes Project only
Peptostreptococcaceae_bacterium_oral_taxon_113	100,000 Genomes Project only
Peptostreptococcus_anaerobius	100,000 Genomes Project only
Peptostreptococcus_sp._MV1	100,000 Genomes Project only
Porphyromonadaceae_bacterium_KA00676	100,000 Genomes Project only

Species	Cohort through which species identified as tissue-resident
Porphyromonas_sp._KLE_1280	100,000 Genomes Project only
Porphyromonas_sp._oral_taxon_279	100,000 Genomes Project only
Prevotella_amnii	100,000 Genomes Project only
Prevotella_bergensis	100,000 Genomes Project only
Prevotella_bivia	100,000 Genomes Project only
Prevotella_buccalis	100,000 Genomes Project only
Prevotella_conceptionensis	100,000 Genomes Project only
Prevotella_copri	100,000 Genomes Project only
Prevotella_corporis	100,000 Genomes Project only
Prevotella_dentalis	100,000 Genomes Project only
Prevotella_ihumii	100,000 Genomes Project only
Prevotella_phocaeensis	100,000 Genomes Project only
Prevotella_saccharolytica	100,000 Genomes Project only
Prevotella_shahii	100,000 Genomes Project only
Prevotella_sp._BV3P1	100,000 Genomes Project only
Prevotella_sp._HMSC073D09	100,000 Genomes Project only
Prevotella_sp._HMSC077E08	100,000 Genomes Project only
Prevotella_sp._HMSC077E09	100,000 Genomes Project only
Prevotella_sp._HUN102	100,000 Genomes Project only
Prevotella_sp._MSX73	100,000 Genomes Project only
Prevotella_sp._oral_taxon_299	100,000 Genomes Project only
Prevotella_sp._oral_taxon_317	100,000 Genomes Project only
Prevotella_sp._oral_taxon_472	100,000 Genomes Project only
Prevotella_sp._S7.1.8	100,000 Genomes Project only
Prevotella_timonensis	100,000 Genomes Project only
Propionibacterium_acidifaciens	100,000 Genomes Project only
Pseudomonas_syringae_pv._coryli	100,000 Genomes Project only
Pseudomonas_veronii	100,000 Genomes Project only
Psychrobacter_sp._SHUES1	100,000 Genomes Project only
Rhodococcus_sp._1139	100,000 Genomes Project only
Romboutsia_timonensis	100,000 Genomes Project only
Roseburia_intestinalis	100,000 Genomes Project only
Roseburia_inulinivorans	100,000 Genomes Project only
Rothia_mucilaginosa	100,000 Genomes Project only
Rothia_sp._HMSC061C12	100,000 Genomes Project only
Rothia_sp._HMSC061D12	100,000 Genomes Project only
Rothia_sp._HMSC061E04	100,000 Genomes Project only
Rothia_sp._HMSC062F03	100,000 Genomes Project only
Rothia_sp._HMSC062H08	100,000 Genomes Project only
Rothia_sp._HMSC064F07	100,000 Genomes Project only
Rothia_sp._HMSC065B04	100,000 Genomes Project only
Rothia_sp._HMSC065C03	100,000 Genomes Project only
Rothia_sp._HMSC065C12	100,000 Genomes Project only
Rothia_sp._HMSC066G02	100,000 Genomes Project only
Rothia_sp._HMSC066G07	100,000 Genomes Project only
Rothia_sp._HMSC068E02	100,000 Genomes Project only
Rothia_sp._HMSC068F09	100,000 Genomes Project only
Rothia_sp._HMSC069C04	100,000 Genomes Project only

Species	Cohort through which species identified as tissue-resident
Rothia_sp._HMSC069C10	100,000 Genomes Project only
Rothia_sp._HMSC069D01	100,000 Genomes Project only
Rothia_sp._HMSC071B01	100,000 Genomes Project only
Rothia_sp._HMSC071C12	100,000 Genomes Project only
Rothia_sp._HMSC072B03	100,000 Genomes Project only
Rothia_sp._HMSC072B04	100,000 Genomes Project only
Rothia_sp._HMSC072E10	100,000 Genomes Project only
Rothia_sp._HMSC075F09	100,000 Genomes Project only
Rothia_sp._HMSC076D04	100,000 Genomes Project only
Rothia_sp._HMSC078H08	100,000 Genomes Project only
Ruminococcus_faecis	100,000 Genomes Project only
Selenomonas_sp._oral_taxon_478	100,000 Genomes Project only
Shuttleworthia_satelles	100,000 Genomes Project only
Shuttleworthia_sp._MSX8B	100,000 Genomes Project only
Sphingomonas_melonis	100,000 Genomes Project only
Staphylococcus_warneri	100,000 Genomes Project only
Streptococcus_anginosus	100,000 Genomes Project only
Streptococcus_australis	100,000 Genomes Project only
Streptococcus_constellatus	100,000 Genomes Project only
Streptococcus_cristatus	100,000 Genomes Project only
Streptococcus_gordonii	100,000 Genomes Project only
Streptococcus_infantis	100,000 Genomes Project only
Streptococcus_intermedius	100,000 Genomes Project only
Streptococcus_lutetiensis	100,000 Genomes Project only
Streptococcus_macedonicus	100,000 Genomes Project only
Streptococcus_massiliensis	100,000 Genomes Project only
Streptococcus_mutans	100,000 Genomes Project only
Streptococcus_parasanguinis	100,000 Genomes Project only
Streptococcus_peroris	100,000 Genomes Project only
Streptococcus_phage_EJ.1	100,000 Genomes Project only
Streptococcus_pneumoniae	100,000 Genomes Project only
Streptococcus_pseudopneumoniae	100,000 Genomes Project only
Streptococcus_salivarius	100,000 Genomes Project only
Streptococcus_sanguinis	100,000 Genomes Project only
Streptococcus_sinensis	100,000 Genomes Project only
Streptococcus_sp._1004_SSPC	100,000 Genomes Project only
Streptococcus_sp._1171_SSPC	100,000 Genomes Project only
Streptococcus_sp._2_1_36FAA	100,000 Genomes Project only
Streptococcus_sp._263_SSPC	100,000 Genomes Project only
Streptococcus_sp._343_SSPC	100,000 Genomes Project only
Streptococcus_sp._400_SSPC	100,000 Genomes Project only
Streptococcus_sp._449_SSPC	100,000 Genomes Project only
Streptococcus_sp._A12	100,000 Genomes Project only
Streptococcus_sp._ACS2	100,000 Genomes Project only
Streptococcus_sp._AS20	100,000 Genomes Project only
Streptococcus_sp._BS29a	100,000 Genomes Project only
Streptococcus_sp._C150	100,000 Genomes Project only
Streptococcus_sp._C300	100,000 Genomes Project only

Species	Cohort through which species identified as tissue-resident
Streptococcus_sp._CCH5.D3	100,000 Genomes Project only
Streptococcus_sp._CCH8.C6	100,000 Genomes Project only
Streptococcus_sp._CCH8.G7	100,000 Genomes Project only
Streptococcus_sp._CCH8.H5	100,000 Genomes Project only
Streptococcus_sp._CCUG_49591	100,000 Genomes Project only
Streptococcus_sp._CM6	100,000 Genomes Project only
Streptococcus_sp._CM7	100,000 Genomes Project only
Streptococcus_sp._DD04	100,000 Genomes Project only
Streptococcus_sp._DD10	100,000 Genomes Project only
Streptococcus_sp._F0441	100,000 Genomes Project only
Streptococcus_sp._F0442	100,000 Genomes Project only
Streptococcus_sp._FDAARGOS_146	100,000 Genomes Project only
Streptococcus_sp._GMD3S	100,000 Genomes Project only
Streptococcus_sp._GMD4S	100,000 Genomes Project only
Streptococcus_sp._GMD5S	100,000 Genomes Project only
Streptococcus_sp._HMSC034A12	100,000 Genomes Project only
Streptococcus_sp._HMSC034B03	100,000 Genomes Project only
Streptococcus_sp._HMSC034B05	100,000 Genomes Project only
Streptococcus_sp._HMSC034E03	100,000 Genomes Project only
Streptococcus_sp._HMSC034E12	100,000 Genomes Project only
Streptococcus_sp._HMSC056C01	100,000 Genomes Project only
Streptococcus_sp._HMSC057E02	100,000 Genomes Project only
Streptococcus_sp._HMSC057G03	100,000 Genomes Project only
Streptococcus_sp._HMSC061D01	100,000 Genomes Project only
Streptococcus_sp._HMSC061D10	100,000 Genomes Project only
Streptococcus_sp._HMSC061E03	100,000 Genomes Project only
Streptococcus_sp._HMSC062D07	100,000 Genomes Project only
Streptococcus_sp._HMSC062H02	100,000 Genomes Project only
Streptococcus_sp._HMSC063B03	100,000 Genomes Project only
Streptococcus_sp._HMSC064D12	100,000 Genomes Project only
Streptococcus_sp._HMSC064H09	100,000 Genomes Project only
Streptococcus_sp._HMSC065C01	100,000 Genomes Project only
Streptococcus_sp._HMSC065E03	100,000 Genomes Project only
Streptococcus_sp._HMSC065H07	100,000 Genomes Project only
Streptococcus_sp._HMSC066E07	100,000 Genomes Project only
Streptococcus_sp._HMSC066F01	100,000 Genomes Project only
Streptococcus_sp._HMSC067A03	100,000 Genomes Project only
Streptococcus_sp._HMSC067H01	100,000 Genomes Project only
Streptococcus_sp._HMSC068F04	100,000 Genomes Project only
Streptococcus_sp._HMSC070B10	100,000 Genomes Project only
Streptococcus_sp._HMSC071D03	100,000 Genomes Project only
Streptococcus_sp._HMSC072C09	100,000 Genomes Project only
Streptococcus_sp._HMSC072D03	100,000 Genomes Project only
Streptococcus_sp._HMSC072D05	100,000 Genomes Project only
Streptococcus_sp._HMSC072D07	100,000 Genomes Project only
Streptococcus_sp._HMSC072G04	100,000 Genomes Project only
Streptococcus_sp._HMSC073A12	100,000 Genomes Project only
Streptococcus_sp._HMSC073D05	100,000 Genomes Project only

Species	Cohort through which species identified as tissue-resident
Streptococcus_sp._HMSC073F11	100,000 Genomes Project only
Streptococcus_sp._HMSC074B11	100,000 Genomes Project only
Streptococcus_sp._HMSC074F05	100,000 Genomes Project only
Streptococcus_sp._HMSC076C08	100,000 Genomes Project only
Streptococcus_sp._HMSC076C09	100,000 Genomes Project only
Streptococcus_sp._HMSC077D04	100,000 Genomes Project only
Streptococcus_sp._HMSC077F03	100,000 Genomes Project only
Streptococcus_sp._HMSC078D09	100,000 Genomes Project only
Streptococcus_sp._HMSC078H03	100,000 Genomes Project only
Streptococcus_sp._HMSC078H12	100,000 Genomes Project only
Streptococcus_sp._HMSC10A01	100,000 Genomes Project only
Streptococcus_sp._HMSC10E12	100,000 Genomes Project only
Streptococcus_sp._HMSC34B10	100,000 Genomes Project only
Streptococcus_sp._HMSC36C04	100,000 Genomes Project only
Streptococcus_sp._HPH0090	100,000 Genomes Project only
Streptococcus_sp._HSISS1	100,000 Genomes Project only
Streptococcus_sp._HSISS3	100,000 Genomes Project only
Streptococcus_sp._I.G2	100,000 Genomes Project only
Streptococcus_sp._I.P16	100,000 Genomes Project only
Streptococcus_sp._M143	100,000 Genomes Project only
Streptococcus_sp._M334	100,000 Genomes Project only
Streptococcus_sp._OBRC6	100,000 Genomes Project only
Streptococcus_sp._oral_taxon_056	100,000 Genomes Project only
Streptococcus_sp._oral_taxon_058	100,000 Genomes Project only
Streptococcus_sp._oral_taxon_064	100,000 Genomes Project only
Streptococcus_sp._oral_taxon_431	100,000 Genomes Project only
Streptococcus_sp._SK140	100,000 Genomes Project only
Streptococcus_sp._SK643	100,000 Genomes Project only
Streptococcus_sp._SR1	100,000 Genomes Project only
Streptococcus_sp._SR4	100,000 Genomes Project only
Streptococcus_thermophilus	100,000 Genomes Project only
Streptococcus_timonensis	100,000 Genomes Project only
Streptococcus_vestibularis	100,000 Genomes Project only
Tannerella_sp._oral_taxon_HOT.286	100,000 Genomes Project only
Tepidimonas_fonticaldi	100,000 Genomes Project only
Tepidimonas_taiwanensis	100,000 Genomes Project only
Thermosipho_affectus	100,000 Genomes Project only
Tyzzerella_nexilis	100,000 Genomes Project only
Veillonella_atypica	100,000 Genomes Project only
Veillonella_sp._ACP1	100,000 Genomes Project only
Veillonella_sp._AS16	100,000 Genomes Project only
Veillonella_sp._HPA0037	100,000 Genomes Project only
Veillonella_sp._ICM51a	100,000 Genomes Project only
Veillonella_tobetsuensis	100,000 Genomes Project only
X.Eubacterium._brachy	100,000 Genomes Project only
X.Eubacterium._eligens	100,000 Genomes Project only
X.Eubacterium._hallii	100,000 Genomes Project only
X.Eubacterium._infirmum	100,000 Genomes Project only

Species	Cohort through which species identified as tissue-resident
X.Eubacterium._nodatum	100,000 Genomes Project only
X.Eubacterium._rectale	100,000 Genomes Project only
X.Eubacterium._saphenum	100,000 Genomes Project only
X.Eubacterium._sulci	100,000 Genomes Project only
X.Hallella._seregens	100,000 Genomes Project only
X.Propionibacterium._namnetense	100,000 Genomes Project only
Xanthomonas_campestris	100,000 Genomes Project only
Human_gammaherpesvirus_4	manually identified
Campylobacter_rectus	TCGA only
Campylobacter_showae	TCGA only
Corynebacterium_durum	TCGA only
Eubacterium_infirmum	TCGA only
Exophiala_oligosperma	TCGA only
Oribacterium_parvum	TCGA only
Prevotella_loescheii	TCGA only
Prevotella_sp._oral_taxon_473	TCGA only
Prevotellaceae_bacterium_Marseille.P2826	TCGA only
Selenomonas_sp._oral_taxon_138	TCGA only
Treponema_lecithinolyticum	TCGA only
Treponema_medium	TCGA only
Treponema_socranskii	TCGA only
Treponema_sp._OMZ_838	TCGA only
Treponema_vincentii	TCGA only

(Supplemental) Table 20. Microbial abundance and Shannon index according to clinical and pathological characteristics from the 100,000 Genomes Project. CIN, chromosomal instability; EBV, Epstein-Barr virus-positive; GS, genomically stable; KW, Kruskal-Wallis chi-squared; MSI, microsatellite instability; MSS, microsatellite stabile; pN, pathological lymph node status; pT, Pathological depth of invasion; S, Spearman's rank coefficient; TCGA, The Cancer Genome Atlas subtype; W, Wilcoxon rank sum.

		Microbial abundance (microbes per human cell)			**Shannon index**		
		Median	Test statistic	p value	Median	Test statistic	p value
Age		-	S = 123922	0.610	-	S = 102056	0.220
Sex	Male	0.00479	W=781	0.374	2.16	W=73	0.871
	Female	0.00200			2.21		
MSI status	MSI	0.08279	W=455	0.061	2.71	W=376	0.460
	MSS	0.00361			2.03		
Tumour location	Cardia	0.00463	W=568	0.764	2.64	W=661	0.433
	Non-cardia	0.00568			2.01		
pT	T1/T2	0.00375	W=522	0.519	2.03	W=587	0.937
	T3/T4	0.00361			1.98		
pTN	positive	0.00395	W=625	0.786	2.27	W=563	0.685
	negative	0.00327			1.85		
Histological phenotype	Diffuse	0.00340	KW = 3.5204	0.172	2.98	KW = 1.574	0.455
	Intestinal	0.00789			1.98		
	Mixed	0.03039			1.56		
TCGA	EBV	0.01468	KW = 4.5026	0.212	2.92	KW = 3.0112	0.390
	MSI	0.08279			2.71		
	GS	0.00340			1.94		
	CIN	0.00375			2.50		

(Supplemental) Table 21. Bacterial strains isolated from five gastric cancer tumour samples. Species of closest match and identity (ID) are based on the closest match from NCBI BLAST alignments. Strain name indicates which sample, media type, and condition the isolate was obtained from. Seq, sequence.

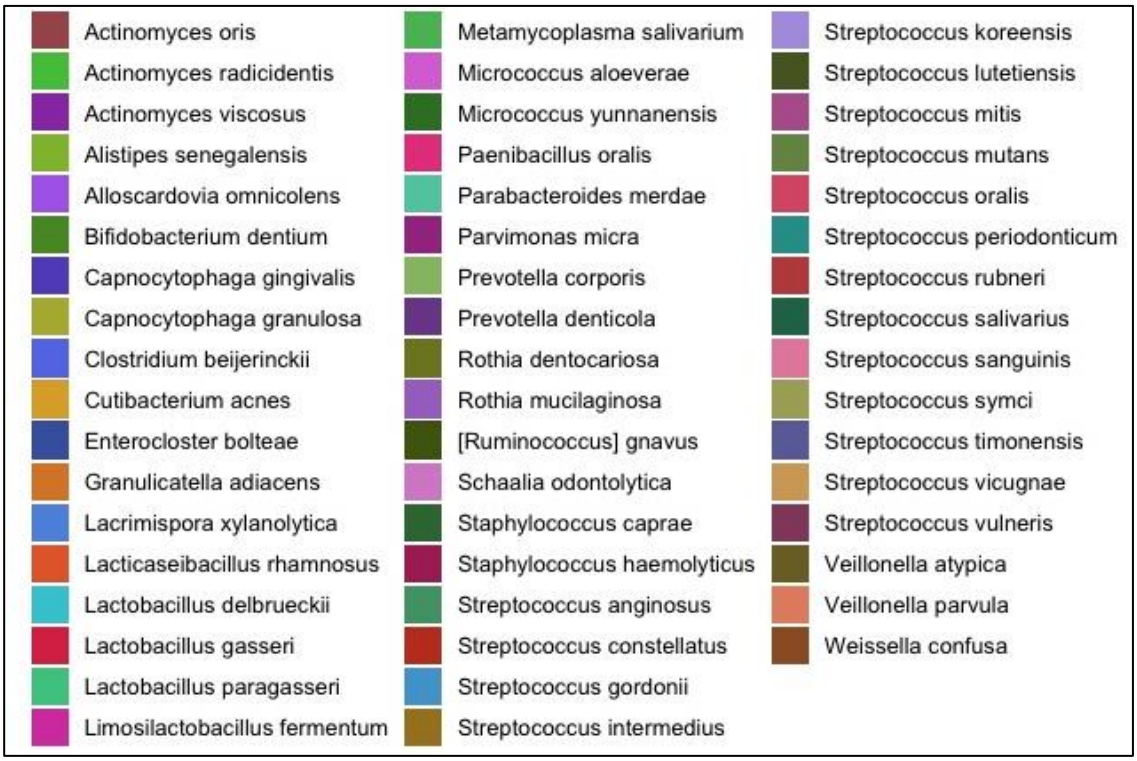
Phylum	**Family**	***Genus***	**Species of closest match**	**Strain**	**Seq. length**	**ID (%)**
Actinomycetota	*Actinomycetaceae*	*Actinomyces*	*Actinomyces oris*	UKG6_T2_FNB_8_AER	693	99.85
			Actinomyces radicidentis	UKG6_T2_FNB_14_AER	695	99.56
			Actinomyces viscosus	UKG6_T2_BR_15_MA	601	100
		Schaalia	*Schaalia odontolytica*	UKG3_GC_5_MA	682	99.55
				UKG3_T2_BR_5_MA	578	97.75
				UKG3_T2_FAA_3_AN	682	99.12
				UKG3_T2_GC_1_MA	640	98.12
	Micrococcaceae	*Micrococcus*	*Micrococcus aloeverae*	UKG1_GC_1_MA	686	100
			Micrococcus yunnanensis	UKG3_GC_7_MA	668	99.85
		Rothia	*Rothia dentocariosa*	UKG6_T2_BR_1_AER	634	100
				UKG6_T2_FNB_15_AER	683	99.55
				UKG7_T2_BR_1_MA	672	99.85
			Rothia mucilaginosa	UKG3_T2_BR_1_MA	646	99.68
				UKG3_T2_FNB_2_AER	643	99.37
				UKG5_T2_FAA_16_AER	682	99.71
				UKG6_T2_BR_8_MA	663	99.54
				UKG6_T2_FAA_21_AER	670	99.55
				UKG7_BR_1_AER	684	99.56
	Propionibacteriaceae	*Cutibacterium*	*Cutibacterium acnes*	UKG5_T2_045_13_AN	676	99.85
				UKG7_BR_4_AN	662	100
Bacillota	*Carnobacteriaceae*	*Granulicatella*	*Granulicatella adiacens*	UKG1_T2_GC_2_MA	673	99.24
				UKG3_T2_FAA_1_AN	688	99.41

Phylum	Family	*Genus*	Species of closest match	Strain	Seq. length	ID (%)
				UKG5_T2_DCM_9_AN	690	99.41
				UKG5_T2_GC_10_AN	677	99.56
				UKG6_T2_FAA_7_AN	703	99.57
	Clostridiaceae	*Clostridium*	*Clostridium beijerinckii*	UKG3_T2_FNB_1_AN	671	100
	Lachnospiraceae	*Enterocloster*	*Enterocloster bolteae*	UKG6_T2_BR_12_MA	608	99.33
		Lacrimispora	*Lacrimispora xylanolytica*	UKG5_T2_BR_13_A_AN	669	94.93
				UKG5_T2_FAA_13_AN	674	94.62
		Lacticaseibacillus	*Lacticaseibacillus rhamnosus*	UKG6_T2_BHI_4_AN	701	100
				UKG6_T2_FCC_20_AN	686	99.4
		Lactobacillus	*Lactobacillus delbrueckii*	UKG6_T2_BR_3_AN	701	100
			Lactobacillus gasseri	UKG5_T2_BHI_2_AN	706	99.71
				UKG5_T2_BR_2_AER	675	98.79
				UKG5_T2_FAA_4_AN	630	97.56
				UKG5_T2_MRS_9_AER	672	99.26
			Lactobacillus paragasseri	UKG5_T2_FCC_9_AER	691	99.85
				UKG5_T2_FNB_15_AER	691	99.85
		Limosilactobacillus	*Limosilactobacillus fermentum*	UKG5_T2_FCC_37_AN	667	99.24
				UKG5_T2_FNB_8_AER	698	98.99
				UKG5_T2_R2A_8_AN	596	99.5
				UKG6_T2_DCM_12_AN	670	89.69
				UKG6_T2_FAA_3_AER	671	98.8
				UKG6_T2_FCC_2_AER	600	98.96
				UKG6_T2_FCC_3_AN	647	98.9
				UKG6_T2_FCC_7_AN	701	100
				UKG6_T2_FCC_8_AN	674	98.96

Phylum	Family	*Genus*	Species of closest match	Strain	Seq. length	ID (%)
				UKG6_T2_R2A_8_AN	673	98.96
	Leuconostocaceae	*Weissella*	*Weissella confusa*	UKG6_T2_BHI_6_AN	701	100
				UKG6_T2_FCC_1_AN	710	99.86
				UKG6_T2_FCC_21_B_AN	666	98.79
	Oscillospiraceae	*Ruminococcus*	*[Ruminococcus] gnavus*	UKG6_T2_FNB_13_AN	581	97.18
	Paenibacillaceae	*Paenibacillus*	*Paenibacillus oralis*	UKG5_T2_ETOH_FNB_1_AER	700	98
	Peptoniphilaceae	*Parvimonas*	*Parvimonas micra*	UKG5_T2_BRF_FAA_3_AN	660	100
	Staphylococcaceae	*Staphylococcus*	*Staphylococcus caprae*	UKG7_BR_1_AN	717	100
			Staphylococcus haemolyticus	UKG7_FAA_2_AN	709	99.86
		Streptococcus	*Streptococcus anginosus*	UKG5_T2_BR_6_AN	670	100
			Streptococcus constellatus	UKG3_T2_GC_2_AN	719	99.72
				UKG3_T2_sTSB_1_AN	697	99.41
			Streptococcus downii	UKG5_T2_BHI_1_AN	688	99.85
			Streptococcus gordonii	UKG3_T2_FAA_7_AN	695	100
				UKG6_T2_R2A_3_AN	694	100
			Streptococcus intermedius	UKG5_T2_FCC_14_AN	713	99.15
			Streptococcus koreensis	UKG6_T2_DCM_9_AN	694	99.42
				UKG6_T2_FCC_5_AN	707	99.29
			Streptococcus lutetiensis	UKG6_T2_FAA_1_AER	708	100
				UKG6_T2_sTSB_1_AN	703	98.57
			Streptococcus mitis	UKG5_T2_BHI_4_AN	703	100
				UKG5_T2_GC_12_AN	700	100
			Streptococcus mutans	UKG6_T2_GC_4_AN	719	100
			Streptococcus oralis	UKG5_T2_FAA_33_AN	705	99.86
				UKG5_T2_FNB_9_AN	715	100

Phylum	Family	*Genus*	Species of closest match	Strain	Seq. length	ID (%)
				UKG6_T2_BR_4_AN	694	99.71
				UKG6_T2_BR_5_MA	658	99.23
				UKG6_T2_FAA_12_AN	694	99.86
				UKG6_T2_FNB_2_AN	703	100
				UKG6_T2_GC_5_MA	702	99.71
			Streptococcus parasanguinis	UKG6_T2_GC_4_MA	716	99.43
			Streptococcus periodonticum	UKG5_T2_BHI_9_AN	668	99.69
				UKG5_T2_BR_5_AN	616	97.56
				UKG5_T2_MPYG_5_AN	655	99.84
				UKG5_T2_R2A_5_AN	504	90.34
				UKG6_T2_BRF_FNB_8_AN	542	99.62
			Streptococcus rubneri	UKG3_T2_BR_2_AER	653	94.3
				UKG3_T2_GC_1_AN	709	99.71
				UKG6_T2_FCC_2_AN	703	100
			Streptococcus salivarius	UKG5_T2_BR_1_AN	671	99.85
				UKG6_T2_BHI_1_AN	708	99.71
				UKG6_T2_FAA_4_AER	694	100
				UKG3_T2_BHI_3_AER	701	99.85
				UKG3_T2_BR_3_AER	694	99.86
				UKG6_T2_BHI_11_AN	693	100
				UKG6_T2_DCM_10_AN	655	89.45
			Streptococcus symci	UKG3_T2_BHI_2_AN	708	99.71
			Streptococcus timonensis	UKG3_T2_BR_5_AN	717	99.86
				UKG3_T2_DCM_4_AN	692	99.86
			Streptococcus vicugnae	UKG5_T2_STSB_7_AN	694	100

Phylum	**Family**	***Genus***	**Species of closest match**	**Strain**	**Seq. length**	**ID (%)**
				UKG6_T2_BRF_FAA_10a_AN	704	100
			Streptococcus vulneris	UKG1_T2_BRF_FNB_2_AN	687	100
				UKG3_T2_R2A_2_AER	669	100
				UKG5_T2_DCM_1_AN	675	100
	Veillonellaceae	*Veillonella*	*Veillonella atypica*	UKG6_T2_DCM_11_AN	685	99.12
				UKG6_T2_BRF_FAA_3_AN	677	98.52
Bacteroidota	*Flavobacteriaceae*	*Capnocytophaga*	*Capnocytophaga gingivalis*	UKG6_T2_FNB_12_AN	681	92.36
			Capnocytophaga granulosa	UKG6_T2_FNB_15_AN	673	99.4
	Prevotellaceae	*Prevotella*	*Prevotella corporis*	UKG6_T2_BR_1_MA	681	99.41
			Prevotella denticola	UKG5_T2_FNB_17_AN	715	99.43
	Rikenellaceae	*Alistipes*	*Alistipes senegalensis*	UKG6_T2_BR_4_MA	677	99.85
	Tannerellaceae	*Parabacteroides*	*Parabacteroides merdae*	UKG6_T2_BR_11_MA	702	99.48
Mycoplasmatota	*Metamycoplasmataceae*	*Metamycoplasma*	*Metamycoplasma salivarium*	UKG5_T2_045_3_AN	677	99.85
Bifidobacteriales	*Bifidobacteriaceae*	*Alloscardovia*	*Alloscardovia omnicolens*	UKG5_T2_FAA_14_AN	646	99.37
		Bifidobacterium	*Bifidobacterium dentium*	UKG5_T2_BHI_7_AN	677	99.55
				UKG5_T2_FCC_6_AN	631	96.83



(Supplementary) Figure 53. Species legend for Figure 29.

Appendix C: Presentations

Characterisation of gastric tumour microbiota using isolation culture methodology and a novel gastric tumour bioreactor system. Poster presented at the 25th World Congress on Gastrointestinal Cancer. 2023 June 28 – July 1; (Barcelona, Spain).

The Gastric Cancer Microbiome Varies According to Patient Sex and Molecular Characteristics. Poster presented at Embracing Complexity in Modern Pathology - 4th Joint Meeting - Winter Meeting of the Pathological Society and the Royal Society of Medicine. 2022 Jan 25-26; (UK, virtual).