Development of anaerobic culturing methods to study the evolution and transmission of the human intestinal microbiota

Hilary Patrick Browne

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the

University of London

March 2018

Department of Pathogen Molecular Biology

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the Medical Research Council.

Research group affiliation: Lawley Faculty. Host-Microbiota Interactions Laboratory. Wellcome Trust Sanger Institute

Declaration

The work presented in this thesis is my own unless otherwise stated. It was carried out under the supervision of Dr. Trevor Lawley of the Wellcome Trust Sanger Institute (WTSI), Hinxton, Cambridgeshire, UK and Prof. Brendan Wren of the London School of Hygiene and Tropical Medicine, London, UK. The work was carried out at the Wellcome Trust Sanger Institute.

All experiments and analysis were carried out by me apart from the following: Taxonomic classification of metagenomic sequence reads using Kraken, calculation of relative abundance of metagenomic sequence reads and generation of the sporulation signature gene list was carried out by Dr. Sam Forster (WTSI). David Goulding (WTSI) created the transmission electron microscopy images from the bacterial cultures provided. Library creation and DNA sequencing was performed by the DNA pipelines group (WTSI). Assembly and annotation of whole-genome sequences was performed by in-house pipelines developed by the Pathogen Informatics group (WTSI). Figures taken from the review article 'Transmission of the gut microbiota: spreading of health' published in Nature Reviews Microbiology, (Reference [1]), were drawn by a Nature Review Microbiology illustrator but were conceived and created by me in collaboration with Dr. Anne Neville, Dr. Sam Forster and Dr. Trevor Lawley.

This thesis does not exceed 100,000 words in length, including tables, figures and footnotes in accordance with the school regulations.

Abstract

The human gastrointestinal tract is colonised by a diverse range of health-associated bacteria, in addition to other microorganisms, termed the intestinal microbiota. Sequence-based, culture-independent approaches have revolutionised this field of study, however, due to the perception that these bacteria are largely unculturable, *in vitro* phenotypic analysis has been hindered. In this study, an anaerobic culturing workflow was developed which revealed that the majority of these bacteria can be cultured using one growth medium. In total, 137 characterised and novel bacterial species were isolated and whole-genome sequenced. Inter-host transmission of the intestinal microbiota may represent a means to maintain a diverse assortment of commensal bacteria within individuals, yet it remains a poorly understood process. Some anaerobic pathogens utilise resilient aero-tolerant spores to survive externally and to facilitate transmission to new hosts. To investigate if commensal spore-formers utilise similar mechanisms, a phenotypic screen was incorporated into the culturing workflow to target spore-forming bacteria. This resulted in the isolation of 66 phylogenetically diverse, spore-forming species which, through subsequent phenotypic characterisation are shown to be specialised for host-to-host transmission and intestinal colonisation. Further phylogenetic and evolutionary analysis revealed body-site associated, loss of sporulation has occurred in different taxa. Also, loss of sporulation is associated with features of host-adaptation that are not present in spore-formers such as a smaller genome size and loss of genetic redundancy. This suggests that the human intestinal microbiota is populated by commensal bacteria that have evolved to engage in opposing lifestyles, either orientated towards inter-host dispersal or within-host adaptation.

This study demonstrates the intestinal microbiota is not unculturable. In addition, commensal microbial transmission may be more prevalent than once thought as a significant proportion of these bacteria can survive outside of a host through the use of spores that are intrinsically resistant to environmental stresses.

Acknowledgements

I would like to thank my supervisors Trevor Lawley and Brendan Wren for guidance, encouragement and advice and for providing me with the opportunity to undertake this study. Thanks also to those at the Wellcome Trust Sanger Institute who I have been fortunate to work with and who have contributed through discussion, advice and training- members of the Host-Microbiota Interactions Laboratory, in particular Anne Neville, Sam Forster, Nitin Kumar, Mark Stares and Blessing Anonye and also, Sally Kay and Dave Goulding. Finally, thanks to Ann and Cillian for their support and motivation.

Contents

List of Figures

List of Tables

Glossary

ABC	ATP-binding cassette transporter
ANI	Average Nucleotide Identity
ATP	Adenosine triphosphate
BCCM	Belgian Co-ordinated Collection of Micro-organisms
c.f.u.	Colony-forming units
CCUG	Culture Collection, University of Gothenburg
CRAP	Culture, Re-streak, Archive and Phenotype
DDH	DNA-DNA hybridisation
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen Gmbh
EBI	European Bioinformatics Institute
FISH	Fluorescence *in sitiu* hybridisation
FMT	Faecal Microbiota Transplantation
GTR	generalized time-reversible
HMP	Human Microbiome Project
HPMCD	Human Pan-Microbes Community Database
IBD	Inflammatory Bowel Disease
JCM	Japan Collection of Microorganisms
LPS	Lipopolysaccharide

LSHTM	London School of Hygiene and Tropical Medicine
MALDI-TOF	matrix-assisted laser desorption ionisation time-of-flight
NGAL	neutrophil gelatinase-associated lipocalin
NOD-like	nucleotide-binding oligomerisation domain-like receptors
NSF	non-spore-forming species
OTU	Operational Taxonomic Unit
PaLoc	pathogenicity locus
PBS	Phosphate Buffer Saline solution
PCoA	Principle Coordinates Analysis
PTS	Phosphoenolpyruvate Carbohydrate Phosphotransferase
RDP	Ribosomal Database Project
RefSeq	NCBI reference sequence database
ROS	reactive oxygen species
s.d.	standard deviation
SCFA	Short Chain Fatty Acids
SF	spore-forming species
SPRI	solid phase reversible immobilisation
TEM	transmission electron microscopy
Tregs	regulatory T cells

VBNC	Viable but non-culturable
WTSI	Wellcome Trust Sanger Institute

Chapter 1 Introduction

1.1 The human intestinal microbiota

1.1.1 The microbiota of humans

Humans have evolved in a microbial world. Prokaryotes emerged nearly 3.5 billion years ago, the first human ancestors only appeared approximately 5 million years ago after diverging from chimpanzees [2-4]. During this time, co-evolution has occurred; humans and other animals are colonised by stable communities of beneficial or non-harmful prokaryotes termed a 'microbiota', which includes bacteria, fungi, viruses, archaea as well as microscopic eukaryotes [3-6]. Bacteria have received the most attention to date as they usually make up the largest proportion of biomass and have the largest metabolic activity [7, 8]. The genetic repertoire encoded by a microbiota is termed a 'microbiome'. Through their varied and diverse metabolic activity, colonising bacteria augment their hosts by provision of functions that the host is incapable of providing or cannot perform as efficiently as its microbiota [8-11]. In return for these benefits, the host provides nutrients for their microbes and a stable, safe environment to inhabit. In artificial conditions, some animals such as mice can survive without a microbiota but these are physiologically deficient compared to mice with a microbiota [12]. In other metazoans, such as insects, survival can be dependent on transfer of their bacterial symbiont to the offspring [13]. Hence, both host and microbes are reliant on each other to varying degrees in a relationship that has been described as a symbiosis [14].

Microbes colonise humans, both externally and internally [15, 16]. In some habitats such as the vagina, a simple community comprised principally of *Lactobacillus* species is indicative of health [15]. Similarly in the respiratory tract, a simple

community dominated by *Staphylococcus* and *Streptococcus* species represents health [17]. In other habitats, such as the gastrointestinal tract, a dense complex community is indicative of a healthy state [15]. Even within a particular habitat the density of bacteria varies. In the gastrointestinal tract, bacteria are present at low abundance (10^4 cells per ml) in the stomach reflecting the low pH, however in the large intestine, they are present at much higher levels (10^{11} cells per ml) (Figure 1.1) [7]. These communities vary in their temporal stability but some, such as the gut and oral microbiota are generally stable in composition over time [15]. While the species composition can vary between individuals, even within the same habitat, the functions provided are constant [15, 18]. A common feature of these communities is the intrinsic resistance provided to pathogen infection by competition for niches and nutrients, termed ‘colonisation resistance’ [7, 17, 19].

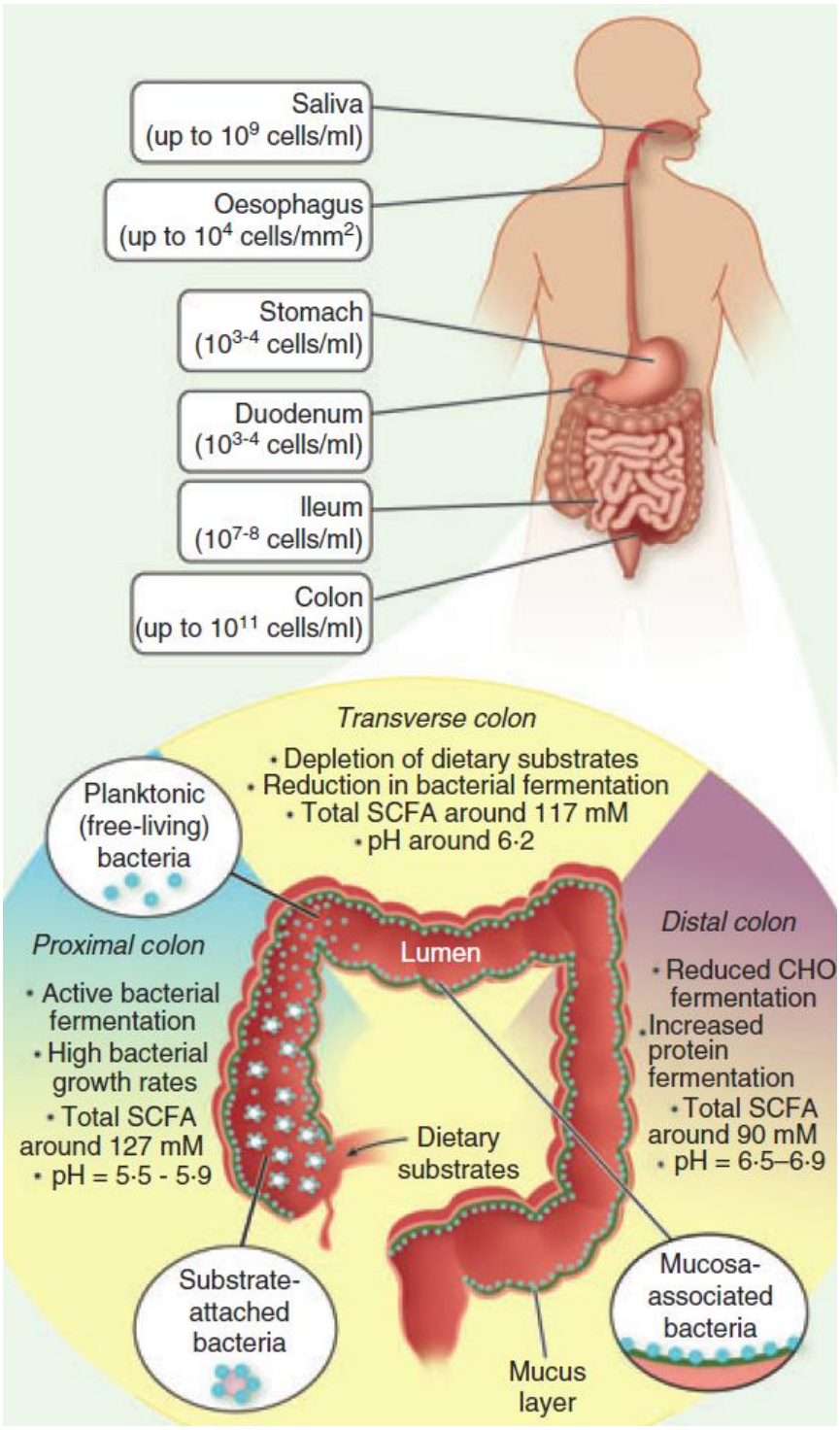


Figure 1.1 Distribution and abundance of bacteria in the human gastrointestinal tract

Bacteria within the human gastrointestinal tract differ in their abundance at different points with the highest concentration present in the colon (large intestine). Within the colon, active fermentation will be driven by substrate availability and pH. Figure taken from [7].

1.1.2 The human intestinal microbiota

The intestinal microbiota of humans is one of the most densely populated microbial communities known to exist (Figure 1.1) [20]. This community has important metabolic and protective roles in human health through metabolising indigestible carbohydrates, preventing infection by pathogenic bacteria and modulating host immune responses [6, 11, 21-23]. Accordingly, it is the best characterised microbial community in humans. Facultative anaerobic bacteria initially colonise the gastrointestinal tract at birth and during the first three years of life. However, these bacteria are gradually replaced by obligate anaerobes as the gastrointestinal tract becomes more anaerobic and the infant transitions to a solid food diet [24, 25]. The majority of bacteria in adults belong to two main phyla - the Bacteroidetes and the Firmicutes. These phyla, together with the Proteobacteria, Actinobacteria, Synergistetes and Fusobacteria, contain almost all of the bacterial species found in the human gastrointestinal tract [15, 26]. Most of these species are obligate anaerobes; however, the extent of aerotolerance varies among species in the Actinobacteria and Proteobacteria phyla [27-29]. Despite their abundance in the human gastrointestinal tract, these species represent only a small subset of all of the bacterial taxa on Earth [30]. Furthermore, many of these bacterial taxa are not found replicating outside of the intestinal environment, which reflects their adaptation to this specific niche [4, 31].

1.1.3 Defining a healthy intestinal microbiota

The factors that determine the optimal intestinal microbial community of an individual at any point in time are varied and include age, host genetics, diet and the local environment. Therefore, a core 'healthy' microbiota that is common to all individuals does not exist. Furthermore, the distinction between health-associated

commensal bacteria and harmful pathogenic bacteria is not always clear, as some bacterial species can promote health or cause disease depending on the specific strain or their location in the body. For example, *Bacteroides fragilis* produces immunomodulatory capsular polysaccharides that stimulate the production of anti-inflammatory cytokines. If this bacterium translocates from the intestine to the peritoneum, then the capsular polysaccharides can cause inflammation, which results in the formation of an abscess [32]. Abscesses can be considered as beneficial to the host by limiting the spread of disease; however, if left untreated they can cause obstructions and further bacterial dissemination if ruptured [33]. Depending on the strain and the virulence factors that are present, *Escherichia coli* is either considered to be a normal commensal of the intestinal microbiota or a pathogen [34]. Similarly, the gastric bacterium *Helicobacter pylori* is associated with an increase in the incidence of peptic ulcers and stomach cancer, but a decreased incidence of oesophageal cancer [35]. In general, a healthy state in the intestinal microbiota can be considered to consist of a diverse microbiota that is abundant in beneficial species, such as members of the *Bacteroidaceae*, *Ruminococcaceae* and *Lachnospiraceae* families in the Bacteroidetes and Firmicutes phyla, and with few pathobionts, such as many members of the Proteobacteria phylum [36].

1.1.4 Host selection of commensal intestinal bacteria

There is emerging evidence that hosts preferentially select communities of commensal bacteria through the modulation of the intestinal environment by a combination of host genetics and immune responses. Variation in genetic profiles between individuals is known to alter many aspects of health and disease, and it is now clear that it may also influence the composition of commensal bacterial communities. Despite initial studies that concluded that human genetics does not

substantially contribute to determining the bacterial species acquired [24, 37], recent studies have identified the presence of bacterial species that are associated with specific genetic polymorphisms, including abundant health-associated *Faecalibacterium* spp. [38, 39]. It has also been demonstrated that specific genes influence bacterial colonisation. For example, expression of the fucosyltransferase 2 (*FUT2*) gene results in the presentation of fucosylated substrates on intestinal epithelial cells, thus enhancing the recruitment of particular species of commensal bacteria to the epithelium and protecting against the translocation of pathogenic bacteria [40, 41]. The association between host genetics and the community composition of the microbiota remains poorly understood; however, it is now evident that host genetics may have an essential role in determining the optimal microbiota community for promoting health.

In addition to host genetics, the host immune system can distinguish between commensal and pathogenic bacteria to elicit different downstream signalling responses, through innate immune receptors, such as Toll-like receptors and nucleotide-binding oligomerisation domain-like receptors (NOD-like receptors). The recognition of commensal bacteria generally promotes intestinal homeostasis, whereas the recognition of pathogens results in a pro-inflammatory response [42]. Studies in genetically modified mouse models have shown that the absence of caspase 3 and caspase 4, which are involved in cell apoptosis and inflammatory responses, can also substantially alter the composition of the microbiota and disease susceptibility [43]. Host-derived antimicrobial peptides that are produced as part of a pro-inflammatory response to pathogens have been shown to specifically recognise pathogen lipopolysaccharide (LPS) structures and do not bind to LPS on commensal bacteria, owing to an altered charge on the commensal cell surface that prevents

binding [44]. These results demonstrate that host selection of the composition of the microbiota could be determined through host genetic background or induced in response to the presence of pathogens or commensals.

Host behaviours, such as dietary choices, may also determine the composition of the intestinal microbiota. This selection commences at birth. The presence of indigestible human milk oligosaccharides in breast milk promotes the expansion of commensal species, in particular *Bifidobacterium* spp., which have a wide range of glycoside hydrolases that can degrade these complex sugars before metabolising them [45, 46]. By importing human milk oligosaccharides into the bacterial cell before degrading them, *Bifidobacterium* spp. also limit nutrient availability to any pathogens that may be present in the intestinal environment [45, 46]. The composition of the microbiota in adults can also vary substantially with diet. For example, the prevalence of *Ruminococcus bromii* is known to increase in people who consume diets that are high in resistant starches [23, 47]. Taken together, it is clear that the combination of host genetics, responses to bacterial stimuli and environmental factors, such as diet, determines the current and optimal microbiota for an individual.

1.2 The host-microbiota interactions of other animals

Studying the microbiotas of other animals can inform research on the human microbiota. For example, by comparing host physiology and diet of humans and other animals, or by examining genomic features of host-adaptation present in the bacteria colonising other animals, we can gain a better understanding of how these processes have shaped the assembly and evolution of the human intestinal microbiota.

1.2.1 The rumen microbiota

In other animals, the actions of the microbiota in different parts of the gastrointestinal tract play a key role in health and metabolism. Ruminants such as cattle and sheep have a multi-chambered stomach that consists of the rumen where the majority of bacterial fermentation takes place, the reticulum which accumulates dense undigested material, the omasum which absorbs water and finally the abomasum or true stomach (Figure 1.2) [48]. These animals are distinct from hind-gut fermenter animals such as horses and rabbits which are mono-gastric but have an enlarged caecum which prolongs digestion of plant material. The rumen contains a dense community of prokaryotes with bacterial abundances comparable to the human large intestine (10^{11} cells per ml) [49]. These bacteria are specialised at degrading complex plant material to generate large amounts of Short Chain Fatty Acids (SCFA) which are a major energy source for ruminants, indeed, microbial fermentation in cattle can provide over 60% of their energy requirements [48]. This anaerobic community is, similar to the human large intestine, dominated by Firmicutes and Bacteroidetes but the composition of species within and their relative abundance is different [4, 49]. Major species within this community include *Ruminococcus flavefaciens* and *Prevotella* species [49]. As the rumen precedes the stomach and intestines, any bacteria that transit to the stomach are digested. Accordingly, the rumen microbiota is a major source of protein for ruminants [48]. The composition of the microbiota of fore- and hind-gut fermenters is driven by diet and physiology and is distinct from each other [4, 5]. Furthermore, the microbiota of the rumen and the small and large intestine within ruminants is also distinct from each other [4, 50].

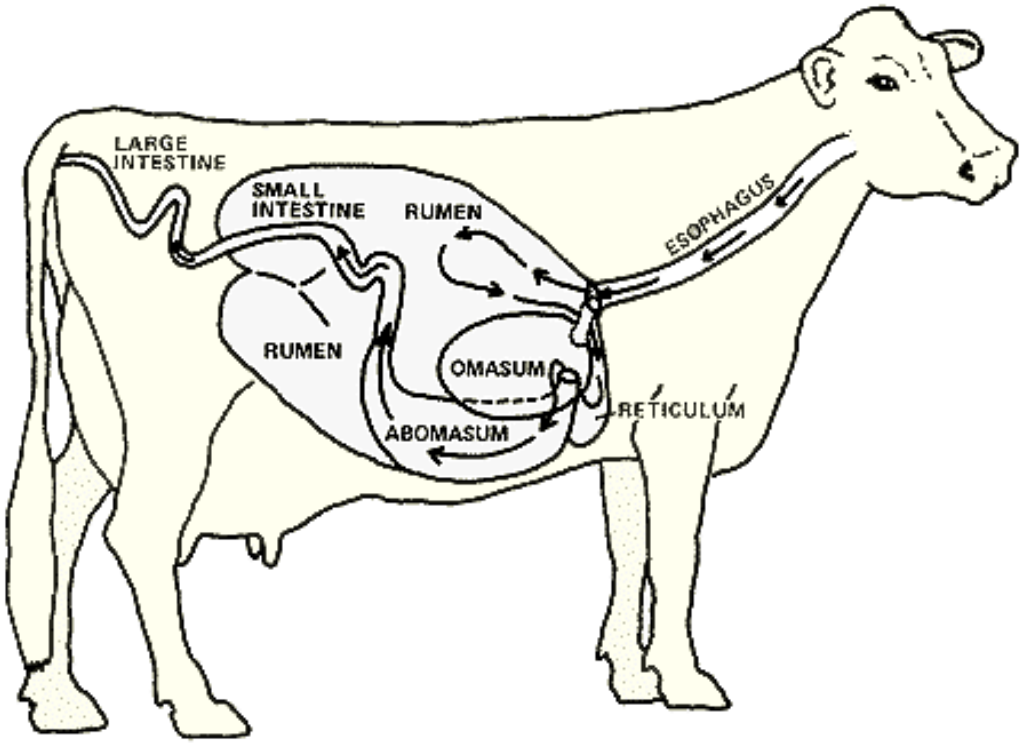


Figure 1.2 Rumen anatomy

Fermentation of plant material by indigenous bacteria in the rumen releases energy that can be utilised by the host. The omasum absorbs water and the reticulum separates dense particles from the main digesta. The abomasum is the true stomach and from this point the digestive anatomy resembles that of humans with a small and large intestine. Figure taken from [51].

1.2.2 Extreme host adaptation- the insect microbiota

The intestinal microbiota of humans and ruminants are characterised by complex microbial communities where the function provided is often more important than the species providing it. Many insects are colonised by simple communities of gut microbes and in some cases completely lack an intestinal microbiota, instead they are colonised by a single intracellular species of bacteria that are transmitted maternally [13, 52]. This is an extreme symbiosis, as in most cases neither party can survive without the other. *Buchnera aphidicola* which is a symbiont of pea aphids

(*Acyrthosiphon pisum*) resides in a specialised organ in the insect, termed a bacteriocyte (Figure 1.3). *B. aphidicola* is so host-adapted that it is completely reliant on the aphid for transmission and cannot colonise other insect species [13, 53, 54]. This extreme host-adaptation makes many of these bacteria difficult to study because the conditions necessary for survival cannot be replicated in the laboratory. In the case of *B. aphidicola,* the bacterium provides essential amino acids which the aphid cannot obtain from the nutrient poor plant phloem it feeds on. In return, the aphid provides nutrients and a stable environment to inhabit [55]. Other symbioses are not so benign. The *Wolbachia* genus is extremely abundant and has been estimated to colonise over 60% of insect species [56]. *Wolbachia* can influence the fecundity of their hosts to promote their own survival by infecting germ cells. The most common means of doing this is through cytoplasmic incompatibility where sperm from infected males can only successfully fertilise the eggs of infected females carrying the same *Wolbachia* strain [13]. Infected eggs can successfully be fertilised by infected or uninfected sperm. As *Wolbachia* is transmitted via the egg this means infected females will produce infected progeny regardless of the infection status of their male partner, while uninfected females can only reproduce with uninfected males (Figure 1.3). Although the mechanism behind this remains largely unknown, the end result is that the ratio of *Wolbachia* infected to uninfected progeny will increase in a population [57].

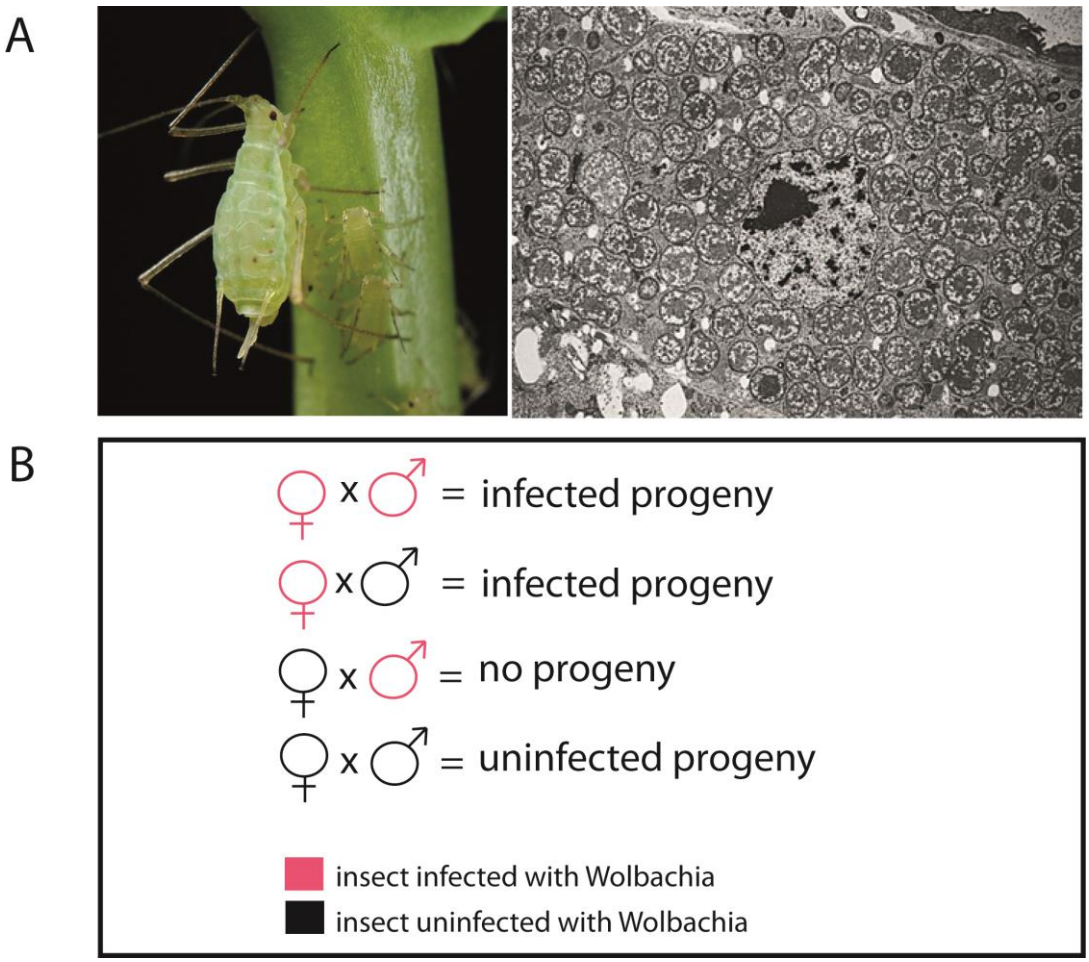


Figure 1.3 Bacterial symbionts of insects

(A) The pea aphid (left) is colonised by *Buchnera aphidicola* that resides within specialised cells called bacteriocytes (right). Both parties are completely reliant on each other for survival. In the image of the bacteriocyte the central object is the nucleus of the aphid bacteriocyte cell. The circular objects surrounding the nucleus in the cytoplasm are the *B. aphidicola* cells. (B) *Wolbachia* bacteria engage in cytoplasmic incompatibility to promote their propagation. Infected female insects can mate with infected and uninfected males to produce infected progeny. All other combinations do not produce infected progeny. Image of pea aphid taken from [58] and image of bacteriocyte taken from [59].

1.2.3 Genome reduction of extremely host-adapted bacteria

These insect symbionts all share similar genetic features that are a result of their long adaptation to the host environment, in some cases up to 200 million years of co-evolution [13] (Table 1.1). All have small genomes of less than 2 million base pairs and in some cases, smaller than 200,000 base pairs [60, 61]. They all encode a low level of genetic redundancy (reduction in copy number of genes that carry out the same function). Loss of regulatory genes and genes involved in metabolism is also common [54]. This process is accelerated by high mutation rates which favour a low G/C base content and a reduction in DNA repair genes [13]. Ultimately, these genetic changes reflect living in a constant, stable environment. Genome compaction has allowed these bacteria to discard unnecessary biosynthetic pathways and genes, the functions of which are either no longer required or are provided by their host [60, 62]. This level of genome reduction has not been observed in human associated bacteria but host-adapted *Lactobacilli* do have smaller genomes compared to environmental or plant-associated *Lactobacillus* species [63]. Some members of the *Prochlorococcus* genus, which are abundant, ocean dwelling Cyanobacteria, have also undergone genome reduction, in some cases up to 43% compared to non-reduced relatives [54] (Table 1.1). Analysis of a metagenomically assembled soil bacterium *Candidatus* Udaeobacter copiosus also reveals features of genome reduction [64]; hence while genome reduction is commonly associated with extremely adapted insect symbionts it is also found in free-living bacteria in different environments.

Genome characteristic	***Buchnera* spp. versus *Escherichia coli***	**Reduced *Prochlorococcus* spp. versus Non-reduced *Prochlorococcus* spp.**
Genome size	Reduced by up to 80 %	Reduced by up to 38 %
% G/C	Reduced to 26% (from 50 %)	Reduced to between 31 and 38 %
Gene number	Reduced by up to 80 %	Reduced by up to 43 %
Gene family size	Smaller	Smaller
Pseudogenes	Higher proportion	Possibly higher proportion
Recombination genes	Losses	Losses
DNA replication and repair genes	Losses	Losses
Regulation genes	Losses	Losses
Metabolic genes	Losses	Losses and Gains
Sequence evolution	Faster	Faster

Table 1.1 Shared features of genome reduction in host-associated symbionts and free-living bacteria

Some common features of genome reduction for an insect symbiont (*Buchnera* spp.) and free-living, ocean-dwelling *Prochlorococcus* spp. *Buchnera* are compared against their closest extant relative *Escherichia coli* and *Prochlorococcus* are compared against non-reduced *Prochlorococcus* species. Table adapted from and data taken from [54].

There are a number of theories to explain why bacterial genomes become reduced in size over long time scales. Genome streamlining is a feature of large population sizes

in free-living bacteria where unnecessary genes are deleted as they incur a fitness cost on the host (this has been proposed for *Prochlorococcus* species) [65]. In these large populations natural selection exerts a larger influence than genetic drift. This is accentuated in low-nutrient environments where a beneficial deletion becomes quickly fixed in the population [60, 65]. In much smaller population sizes that encounter regular population bottlenecks, Muller's ratchet can take effect where deleterious mutations result in an increase of non-functional genes which are subsequently deleted [60]. Muller's ratchet is proposed to be responsible for the genome reduction observed in many endosymbiotic bacteria. Other hypothesis include the Black Queen Hypothesis which states that microorganisms will lose the ability to carry out a function if it can rely on another member of the same community to supply the resource (and bear the cost of doing so) [66]. All of these theories are underpinned by an evolutionary trend in many studied bacterial genomes towards a smaller genome size [67, 68].

1.3 Studying the intestinal microbiota

1.3.1 Culturing and sequence-based approaches

Escherichia coli was the first bacterium isolated from the human gastrointestinal tract in 1885 by Theodor Escherich [26, 69]. Due to a lack of anaerobic culturing tools, many of the bacteria subsequently cultured were aero-tolerant. The development of anaerobic culturing techniques including those by Robert Hungate in the 1940's through to the 1960's led to the isolation of many of the dominant species of the intestinal microbiota [26, 70, 71]. The application of culture-independent methods to identify bacteria by sequencing of the 16S ribosomal RNA gene sequence initiated a new interest in the intestinal microbiota and led to the realisation that the microbial composition in the gut was more diverse than previously thought

[26, 29, 72]. This signalled a shift away from culture-based approaches to study the intestinal microbiota. With the development of Illumina-based sequencing technology at the start of this century, assessment of the functional capability within a microbial community could now be explored through shotgun metagenomic sequencing. During this time culturing continued with the isolation of important *Roseburia* and *Faecalibacterium* species [73-75], but sequence-based, culture-independent approaches were by now the preferred means to study the intestinal microbiota. The plethora of bacteria identified through sequence-based means was demonstrated to be far greater than the number isolated by culture-based approaches, which further reinforced the assumption that many of these bacteria were inherently unculturable [76, 77]. The Human Microbiome Project and the MetaHIT project, both launched in 2008, greatly expanded our knowledge of the microbiotas of the human body, through primarily sequence-based and computational approaches in addition to culturing and generation of whole genome sequences. These studies fuelled a further increase in predominately sequence-based analyses of the human intestinal microbiota [8, 15, 78].

One of the main findings from these sequence-based approaches was the observed discrepancy between bacteria identified through culture-independent means and those acquired through culturing. This discrepancy is well documented and is termed the ‘great plate count anomaly’. Despite this, several research groups have successfully cultured many novel species from the intestinal microbiota. These groups have utilised different approaches such as a wide variety of media and different culture conditions to isolate as many different species as possible, using only a few growth media and archiving isolates in a high throughput manner in 96-well plates or using a single media enriched with rumen fluid [79-81]. One barrier to

isolating bacteria in pure culture is that some bacteria are dependent on growth factors provided by neighbouring bacteria to grow. Examples of a growth factor are siderophores which are used by bacteria to scavenge available iron in the environment. This dependency can be overcome by addition of available forms of iron to the media which then allows isolation of pure cultures [76, 82]. In some situations, the preparation of culture media can be inhibitory to some bacteria. Autoclaving agar and phosphate buffer together results in an increase in inhibitory compounds such as hydrogen peroxide that prevent the growth of some bacteria. Autoclaving these media components separately removed the inhibition [83].

1.3.2 The benefits of bacterial culturing

Despite this progress, culture-independent methods are still utilised more to study the intestinal microbiota. Indeed, compared to culture-independent methods, culturing is laborious and requires specialised equipment such as anaerobic chambers to operate at scale. Despite this, bacterial culturing provides many important uses that complement sequence-based approaches. From a clinical perspective, pure cultures are necessary as a starting point to satisfy Koch's postulates and to determine if an isolate is causative for disease. It also allows determination of the antibiotic susceptibility of a disease causing isolate. Novel antibiotics produced by members of the human nasal microbiota have also been discovered through culture-based approaches [84]. While the intestinal microbiota is linked to many diseases, proper characterisation of the role of particular bacterial species cannot take place unless there are pure cultures of the implicated species. Culturing of pure isolates also permits whole genome sequencing. Not only is the genome sequence valuable in itself, but these sequences can also be used to improve the resolution of 16S rRNA gene amplicon and metagenomic approaches by adding them to reference databases

[77]. In recent years, as the cost of sequencing has decreased, the use of Average Nucleotide Identity (ANI) which compares nucleotide similarity based on whole genome sequences has being utilised as a means to differentiate species [85]. This is a less laborious means to distinguish species compared to the classical approach of the DNA-DNA hybridisation technique (DDH) and offers greater resolution and more information in comparison to 16S rRNA gene sequencing.

Culturing can also improve the taxonomy of bacteria through a combination of phylogenetic placement and phenotypic characterisation [86]. The taxonomy of bacteria within the intestinal microbiota was traditionally defined by broad morphological and phenotypic characteristics and did not always reflect correct phylogenetic placement. For example, the genus *Clostridium* was traditionally used as a repository for anaerobic spore-forming bacteria while the genus *Bacilli* was used as a repository for aerobic spore-forming bacteria. Over time *Clostridia* acquired more than 200 species [87]. The advent of 16S rRNA gene sequencing technologies combined with the culturing of novel bacteria led to the re-naming and movement of many species to new taxa to better reflect their phylogeny [88-90]. For example, *Clostridium orbiscindens* and *Eubacterium plautii* were combined and renamed as *Flavonifractor plautii* and moved to a new genus *Flavonifractor* [91]. Other name changes have not been so widely adapted. The intestinal pathogen *Clostridium difficile* has been renamed to *Peptoclostridium difficile* and then to *Clostridioides difficile* [90, 92], but due to the public awareness of the species as an important pathogen it is still commonly known by its original name. There are still many discrepancies in the systematics of the intestinal microbiota, for example, some species still have the same Latin prefix despite residing in different bacterial families e.g. *Ruminococcus gnavus* (*Lachnospiraceae* family) and *Ruminococcus bromii*

(*Ruminococcaceae* family). While some bacterial systematics will remain confusing, culturing can help alleviate this, as it allows phylogenetic placement through the generation of sequence data and facilitates the identification of distinguishing phenotypic characteristics.

1.4 Spore-forming bacteria within the intestinal microbiota

1.4.1 Sporulation and germination cycle and genetics

Sporulation is a form of bacterial dormancy that is utilised by certain members of the Firmicutes phylum which plays an important role in human health and disease. It is defined by the production of resilient structures called endospores (called spores hereafter) that maintain DNA integrity and ensure survival in unfavourable conditions such as paucity of nutrients, desiccation, and for anaerobic bacteria, oxygen exposure. Sporulation is believed to be an ancient bacterial phenotype and may have arisen when oxygen levels on Earth started to rise due to photosynthesis by Cyanobacteria, around the same time that Firmicutes diverged from their ancestors 2.7 billion years ago [93, 94]. Some thermophilic bacteria that emerged before spore-forming *Clostridia* and *Bacilli* classes within the Firmicutes also have the capability to form spores and thus sporulation may have evolved to allow bacteria to survive changing and hostile conditions in the early Earth's atmosphere [95]. Once a vegetative bacterial cell commits to sporulation (Figure 1.4, onset), the process proceeds through well-defined stages. Cell division with partitioning of DNA is followed by engulfment of the smaller forespore (Figure 1.4, asymmetric cell division and engulfment). The spore DNA is embedded in a dehydrated core abundant in dipicolonic acid and small acid-soluble proteins that protect the DNA. Surrounding the core is a peptidoglycan-rich cortex and multi-layered spore coat, both of which provide resistance to environmental insults (Figure 1.4, cortex

formation and coat formation). Sporulation concludes with lysing of the mother cell and release of the mature spore (Figure 1.4, mother cell lysis and spore release) [96, 97].

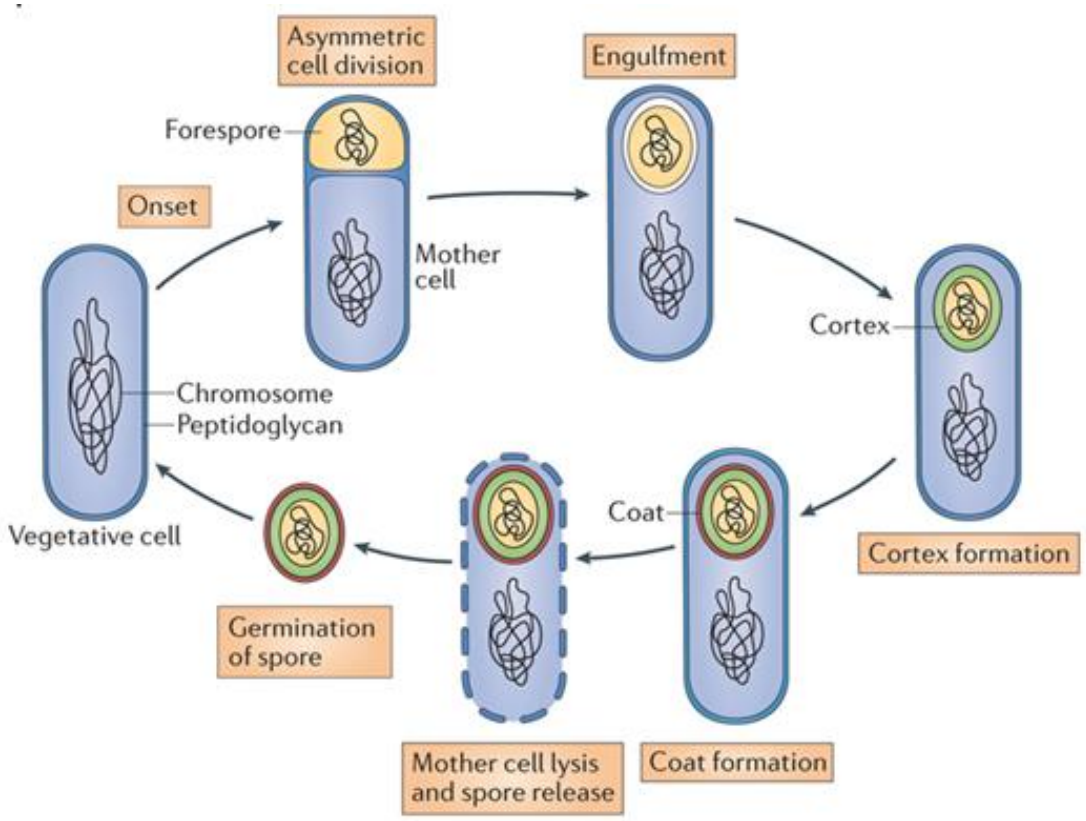


Figure 1.4 Sporulation and germination cycle

Sporulation typically commences when a stress in encountered by a vegetative cell capable of making spores. The process then proceeds through well-defined stages as indicated, that result in a resilient and stable spore that contains the DNA necessary to maintain fecundity. Sporulation concludes with the lysing of the mother cell and the release of the mature spore. The spore will typically germinate when favourable conditions return and a new vegetative cell is formed. Figure taken from [1].

Most spore-forming species produce one spore per vegetative cell, however some species can produce several such as *Metabacterium polyspora* which colonises the

guinea pig gastrointestinal tract. *M. polyspora* is believed to use spore-formation as its primary means of reproduction and as a means to transmit between hosts [98, 99]. Once formed, spores can remain dormant for long periods, a spore at least 25 million years old has been revived from amber, they are also extremely resilient, withstanding alcohol, disinfectants and even conditions outside of the Earth's atmosphere [100-102]. Their ability to remain dormant also facilitates transport over large distances via air currents, insects or human travel [103-105]. Spores can be released from dormancy upon sensing an external cue such as amino acids or nutrients that indicate favourable external conditions. Germination then ensues with a new vegetative cell emerging [96, 106]. Spore-forming bacteria are found in diverse environments such as soil, hot-springs, industrial waste and host-associated habitats of many animals [107].

As sporulation involves drastic changes in cell morphology and the destruction of the mother cell it is a tightly regulated process involving transcription of hundreds of genes [1, 108]. Spo0A is a transcription factor that is found in all spore-forming bacteria. However, presence of the *Spo0A* gene is not diagnostic for spore-formation, some bacteria possess *Spo0A*, but have not been demonstrated to form spores [109]. *Spo0A* is essential for sporulation as it coordinates hierarchical downstream activation of sporulation specific sigma factors, other sporulation specific genes and pleiotropic genes recruited for the sporulation process [108, 110, 111]. It is part of a two-component regulatory system- upon encountering an external signal, a membrane bound histidine kinase initiates a phosphorelay cascade which results in phosphorylation and activation of Spo0A. The Spo0A protein consists of a Che-Y-like signal receiver domain and a transcription activation domain (Figure 1.5). While *Spo0A* and associated sporulation-specific sigma factors are conserved amongst all

studied spore-formers, the genes under their control and their regulation are not shared amongst different taxonomic spore-formers [112]. For example, the main sigma factors F, E, G and K are activated in a sequential manner but in a different order in different species [97]. This flexibility has likely enabled spore-forming bacteria to adapt to different environments [109].

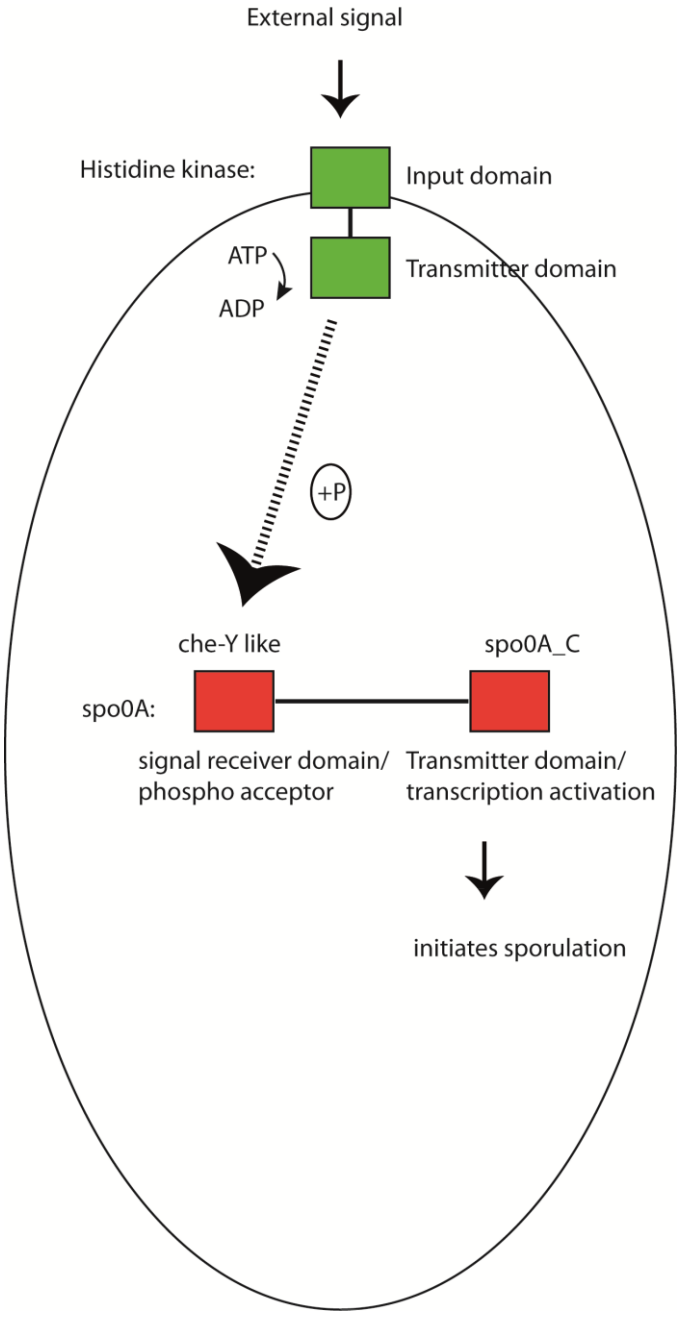


Figure 1.5 The two-component master regulator of sporulation *Spo0A*

The *Spo0A* gene is the master regulator of sporulation and is found in all spore-forming bacteria. A membrane bound histidine-kinase receives an external signal (e.g. lack of nutrients), this results in the phosphorylation of the signal-receiver domain of the Spo0A protein. This in turn activates Spo0A_C which initiates sporulation by activating transcription of sporulation-related genes in a DNA-binding mediated mechanism.

1.4.2 *Clostridium difficile*- the model enteric spore-former

Some spore-forming bacteria produce toxins which can cause disease in humans and other animals such as *Bacillus anthracis* (family *Bacillaceae*) and *Clostridium botulinum* (family *Clostridiaceae*) [113, 114]. Others are associated with food poisoning such as *Clostridium perfringens* (family *Clostridiaceae*) and *Bacillus cereus* (family *Bacillaceae*). *C. perfringens* has also been associated with necrotizing enterocolitis of new-born infants [115]. While the toxins of these bacteria are responsible for harming humans, it is the ability of these species to sporulate that makes them so resilient and difficult to eradicate. In recent years the anaerobic pathogen, *C. difficile* (family *Peptostreptococcaceae*) has become a major health burden in nosocomial settings.

Again, spores of *C. difficile* facilitate survival in external conditions and also make eradication in clinical settings difficult [102]. Once spores are ingested, they germinate in the small intestine in the presence of intestinal bile acids [97, 116, 117] (Figure 1.6). *C. difficile* has been detected in healthy adults that remain asymptomatic [118]. Its pathogenicity is linked to a depleted intestinal microbiota in the host, typically following antibiotic usage. People who are most at risk include immunocompromised or elderly individuals [119, 120]. The main virulence factors of *C. difficile* are toxins that target epithelial cells resulting in diarrhoea and in severe cases toxic megacolon and death [120]. The production of diarrhoea facilitates the dissemination of *C. difficile* through transmission of spores to new susceptible hosts (Figure 1.6). Due to its disease severity and the resilience of spores, *C. difficile* infection remains the most expensive infection to treat in the USA with over 29,000 fatalities occurring in 2011 [121]. It is thus, the most-studied and best-understood enteric spore-forming bacterium [105, 110, 119, 122].

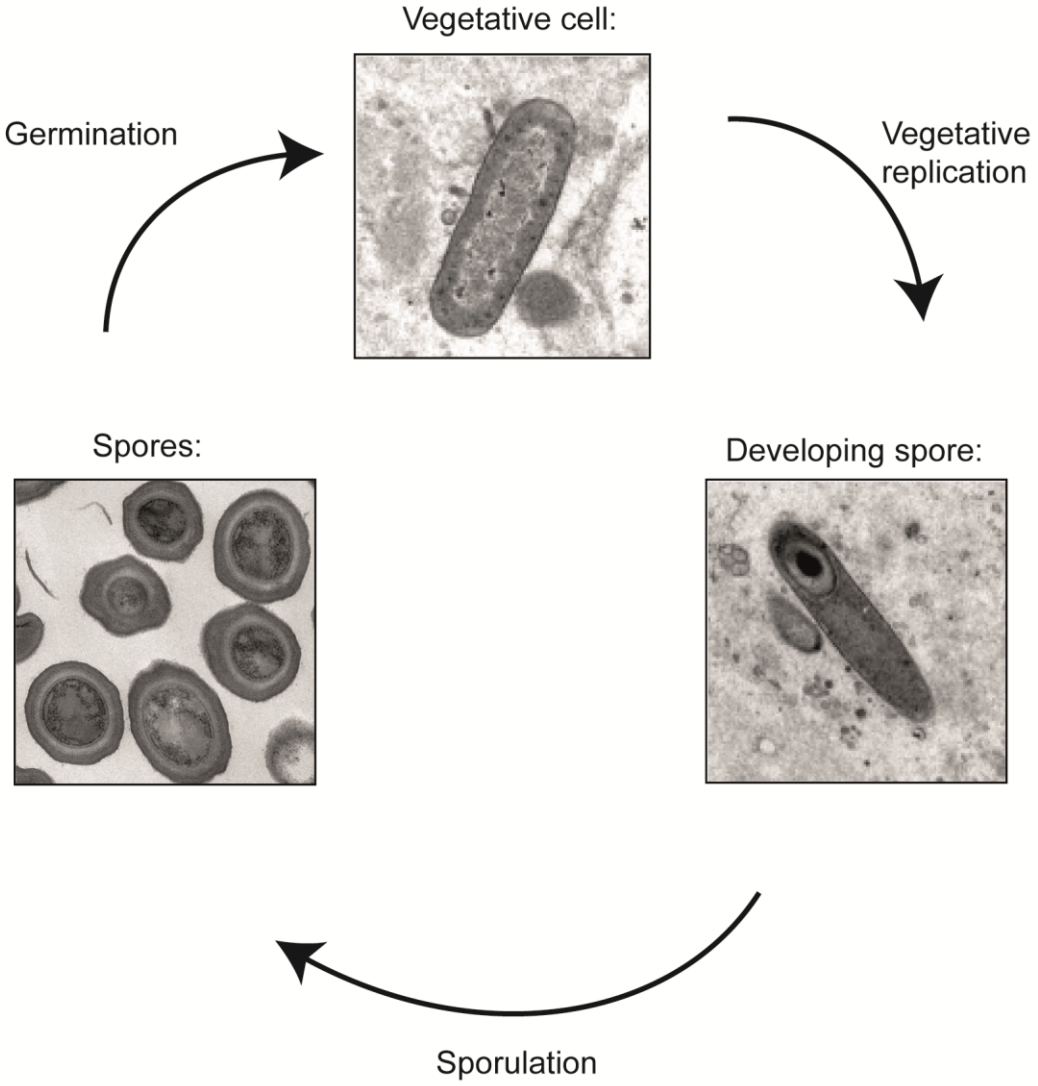


Figure 1.6 The transmission and infection life-cycle of *Clostridium difficile*

Clostridium difficile utilises resilient spores to survive externally and to transmit between hosts. Once ingested, the spores germinate in the small intestine forming vegetative cells which can produce harmful toxins in the large intestine, leading to disease. The vegetative cells produce new spores which are disseminated via toxin-induced diarrhoea, thus, allowing *C. difficile* to continue its life cycle. TEM images by David Goulding WTSI.

C. difficile is genetically diverse with different lineages or ribotypes being associated with different levels of morbidity. These include the 027 lineage which was responsible for several nosocomial-associated disease outbreaks globally and the 078 lineage which is associated with community acquisition and may also be zoonotically acquired [105, 123-125]. The *C. difficile* genome contains a large proportion of mobile genetic elements (approximately 11 %) which facilitates adaptation through the acquisition of new sequence [119]. Horizontal transfer of the toxin-encoding pathogenicity locus (PaLoc) from a toxin-producing to a non-toxin-producing lineage has been demonstrated with subsequent toxin production in the PaLoc recipient [126]. Antibiotic resistance genes are frequently encoded on mobile genetic elements and are believed to have contributed to the global spread of the pathogen [119, 127]. The spread of the 027 lineage from North America to Europe and South Korea is associated with the acquisition of resistance to fluoroquinolone which is widely used in North America [105]. Interestingly, this antibiotic is not widely used in Australia which may explain the low levels of the 027 lineage in this country [125].

C. difficile is adapted to the gastrointestinal tract through its ability to metabolise available carbohydrates and to compete with the resident microbiota [119, 128]. Furthermore, the ability of spores to recognise intestinal bile acids ensures germination occurs in a favourable environment. This bile acid recognition is not a feature of every spore-former and reflects the adaptation of *C. difficile* to the intestinal environment [97, 122]. Intestinal bile acids are synthesised in the liver from cholesterol and aid digestion by acting as emulsifying agents [129]. Cholate and its conjugated forms, taurocholate and glycocholate, act as germinants for *C. difficile* spores as does the secondary bile acid deoxycholate which is converted from

cholate to deoxycholate by 7-dehydroxylase encoding bacteria such as *Clostridium scindens*. Deoxycholate and other secondary bile acids such as lithocholate are toxic to vegetative cells of *C. difficile* and loss of colonisation resistance through antibiotic consumption may therefore occur through depletion of *C. scindens* and other 7-dehydroxylase encoding bacteria [116, 117, 130, 131].

1.5 Transmission of commensal bacteria

The presence of the intestinal microbiota in the human gut is the result of extensive immigration and competition that continues throughout life. The colonisation success of these health-associated symbiotic bacteria is attributable to their ability to spread and to be maintained in human populations [132]. Thus, transmission is an essential feature of the human microbiota that relies on the strategies used by bacteria to exit from one host (donor) and stably colonise another (recipient). The ubiquitous and sometimes exclusive presence of this select group of enteric bacteria in human populations demonstrates the existence of host-adapted colonisation processes and refined co-evolved transmission networks [5, 6, 31]. Most of our knowledge on the transmission mechanisms used by intestinal bacteria is derived from the study of pathogen transmission; this provides a conceptual framework to begin to understand commensal transmission [133]. Both commensal and pathogenic intestinal bacteria are primarily transmitted between hosts through the faecal–oral route. Commensal intestinal bacteria can also be transmitted through the vaginal–oral route at birth and through breast milk in early life. The transmission routes of commensal and pathogenic bacteria are distinguished largely by the colonisation strategy that is used once inside the host. Commensal bacteria provide health benefits to the host that are a result of their colonisation, whereas, depending on their virulence and infectious dose, pathogen colonisation can cause disease.

1.5.1 Shared transmission routes of commensal and pathogenic intestinal bacteria

The first step in a typical transmission route for an intestinal pathogen is the shedding of the bacterium from the host in faecal matter, which is followed by changes in bacterial metabolism or cellular architecture to maximise survival in the external environment. The pathogen must then persist in the external environment, possibly by using reservoirs, such as animals, the built environment, water sources or food chains, to increase the likelihood of entering a new susceptible host. Once the bacterium has successfully persisted in the external aerobic environment and has been ingested by a new host, it must colonise otherwise it will rapidly transit through the gut. Colonisation includes passage through the stomach, the establishment of a niche in the intestinal environment, the use of available nutrients, and replication to a level that will ensure stability and survival (Figure 1.7). A newly colonised host can then become a donor for the onward transmission of that bacterium. The colonising species will encounter competition from the resident microbiota, and this colonisation resistance has important roles in preventing invasion by pathogenic bacteria and in maintaining intestinal homeostasis [7]. The resident microbiota can compete directly through the use of available nutrients or by the secretion of toxins that target neighbouring bacteria, as has been demonstrated for *B. fragilis* [7, 134]. In addition to competition between bacterial species, the metabolism of available dietary substrates can facilitate cross-feeding between species, thus promoting cooperation and the colonisation of competing species [11, 135, 136].

It is likely that intestinal commensal bacteria use the same, or similar, strategies to those used by pathogenic bacteria to transmit between hosts (Figure 1.7). Recent evidence indicates that many of the survival mechanisms and environmental

reservoirs are also common between pathogenic and commensal bacteria. Moreover, colonisation factors, such as flagella and fimbriae, are also shared; these appendages are not unique to pathogens and are also a feature of commensal intestinal bacteria, including *Roseburia* spp. and *Bifidobacterium* spp. [137, 138]. Last, sequence-based studies of pathogen transmission networks have revealed that bacteria can disseminate both locally and globally through their human hosts, which indicates that the transmission of commensal bacteria is not spatially restricted [105, 139].

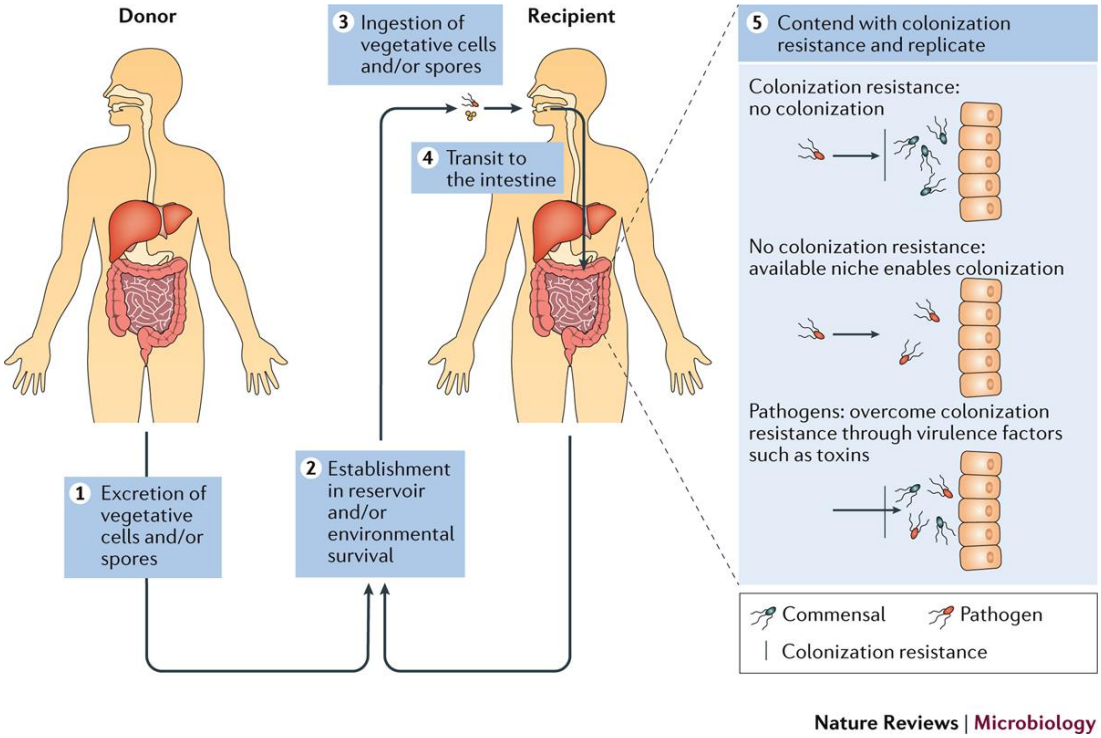


Figure 1.7 Transmission of pathogenic and commensal intestinal bacteria

Intestinal pathogens and commensal bacteria use similar mechanisms to transmit between host. Egestion from the host in faecal matter is the first stage in transmission (step 1). To promote dispersal and subsequent ingestion by a new host, pathogens may induce diarrhoea in the donor. Once in the external environment, survival mechanisms, such as aerotolerance, viable but non-culturable dormancy and sporulation, are used by these predominately anaerobic bacteria to survive and transmit. Environmental reservoirs, such as people, food, animals and the built environment, will function as a source or sink for transmission (step 2). Once ingested by a new host (step 3), the bacterium transits to the intestines (step 4). Competition from the resident microbiota can prevent colonisation (step 5, see colonisation resistance); however, bacteria can colonize if a niche is unoccupied (step 5, see no colonisation resistance). The restoration of bacterial species functions to maintain colonisation resistance and promote the diversity of health-associated bacteria in the gut. Pathogens can overcome colonisation resistance through the induction of the expression of virulence factors, such as toxins, which can lead to inflammation and perturb the resident microbiota (step 5, see pathogens). Metabolism of nutrients and replication promote persistence and support

further replication and subsequent onward transmission as the recipient now becomes a donor. Figure taken from [1].

1.5.2 Distinguishing the routes of transmission of commensal and pathogenic intestinal bacteria

Despite the similarities mentioned above, there are substantial differences between the mechanisms used by intestinal pathogens and commensal bacteria to transmit. Depending on the colonising dose, host susceptibility and environment, a pathogen can exist in a low-level asymptomatic state or can induce a high-level symptomatic super-shedding state in the host [140]. The low-level asymptomatic state is typically associated with relatively little perturbation of the intestinal microbiota and lower levels of transmission, thus rendering the host a silent carrier of potential pathogens. Bacteria such as enteropathogenic *E. coli*, *Vibrio cholerae* and *C. difficile* use virulence factors, such as toxins, during pathogenesis to maximise their colonisation, despite causing severe inflammatory symptoms and intestinal disease. The host can restrict pathogen colonisation through the secretion of antimicrobial peptides, such as neutrophil gelatinase-associated lipocalin (NGAL), which prevents microbial siderophores from binding to essential iron [141]. However, some pathogens can circumvent this response by producing modified siderophores, such as salmochelins, that are not bound by NGAL. Any resulting intestinal disease typically results in substantial perturbation and instability in the commensal microbiota, which often results in diarrhoea that may promote rapid pathogen dispersal and transmission at the expense of commensal colonisation and host health. Therefore, one distinction between pathogen and commensal colonisation in this context is that pathogenic

bacteria use a host-derived inflammation state to spread, whereas commensals do not, and therefore either decrease in number or are lost during dysbiosis [142, 143].

1.5.3 Survival in the environment

Once excreted from the body in faeces, intestinal bacteria must tolerate the local environment to enter and colonise a new host (Figure 1.7). As previously discussed spores are the transmission phenotype utilised by *C. difficile* but not all enteric spore-formers are pathogenic. There are also commensals some of which as demonstrated for *C. scindens*, exert protective effects by inhibiting *C. difficile* proliferation [144]. Other human-derived commensal bacteria from the Firmicutes have positive immunological effects by inducing regulatory T cells in mice [145]. These bacteria were isolated from chloroform-treated faecal samples and many of them are known spore-formers from the *Erysipelotrichaceae* and *Lachnospiraceae* families, including *C. scindens* again. Other intestinal spore-forming bacteria have been isolated from the *Erysipelotrichaceae* and *Lachnospiraceae* families in addition to other families that contain spore-forming species such as *Clostridiaceae*, *Ruminococcaceae* and *Peptostreptococcaceae* [26]. Spore-forming bacteria are therefore prevalent within the human intestinal microbiota but their extent remains unknown. For these commensals, spores could be utilised as a means to promote host-to-host transmission of anaerobic bacteria in a manner similar to *C. difficile* (Figure 1.6 and Figure 1.7).

Once shed by the host, intestinal bacteria in a vegetative state show varying levels of tolerance to atmospheric oxygen [27, 28, 73, 146-148] (Table 1.2). The damaging effects of oxygen in bacterial cells are due to the generation of reactive oxygen species (ROS), which damage DNA and proteins, and interfere with essential metabolic processes [149]. Aerobic bacteria and facultative anaerobic bacteria have

evolved mechanisms to avoid and repair the damage caused by ROS, including antioxidant enzymes such as catalases, peroxidases or superoxide dismutase [27, 149]. In a vegetative state, obligate anaerobic bacteria are typically sensitive to oxygen and may die within minutes of exposure [146]. Nevertheless, mechanisms to counter oxygen stress exist, even in these obligate anaerobes. *Faecalibacterium prausnitzii* relies on an extracellular flavin–thiol electron shuttle to grow in the presence of oxygen, which enables its survival in the oxygenated zone at the gut mucosa [150]. This oxygenated zone ensures that gut epithelial cells are protected from the majority of anaerobic bacteria in the lumen that could compromise the integrity of the epithelial cells [151, 152]. The extracellular flavin–thiol electron shuttle may also promote the survival of *F. prausnitzii* when it is exposed to atmospheric oxygen in the presence of the antioxidants inulin, cysteine and riboflavin [150, 153]. Other abundant intestinal bacteria, such as *Roseburia* spp., can only survive for a few minutes when exposed to atmospheric oxygen concentrations (Table 1.2). We hypothesise that *Roseburia* spp. either use a currently unknown survival mechanism or are extremely efficient at colonisation and can readily become established in new hosts to which they are in close proximity.

	Phylum	Family	Species	Survival time in ambient oxygen	Ref.
1	Proteobacteria	*Enterobacteriaceae*	*Escherichia coli*	Minimum 21 days	[146]
2	Firmicutes	*Lachnospiraceae*	*Clostridium aminovalericum*	45 minutes	[27]
3	Firmicutes	*Lachnospiraceae*	*Roseburia faecis*	2 minutes	[73]
4	Firmicutes	*Lachnospiraceae*	*Roseburia hominis*	2 minutes	[73]
5	Firmicutes	*Lachnospiraceae*	*Roseburia inulinivorans*	2 minutes	[73]
6	Firmicutes	*Lachnospiraceae*	*Roseburia intestinalis*	1 hour	[147]
7	Firmicutes	*Lachnospiraceae*	*Eubacterium hallii*	1 hour	[147]
8	Firmicutes	*Lachnospiraceae*	*Anaerostipes caccae*	Minimum 1 hour	[147]
9	Firmicutes	*Ruminococcaceae*	*Faecalibacterium prausnitzii*	1 hour	[147]
10	Firmicutes	*Veillonellaceae*	*Veillonella alcalescens*	72 hours	[148]
11	Firmicutes	*Clostridiales Incertae Sedis XI*	*Anaerococcus prevotii* ‡	72 hours	[148]
12	Firmicutes	*Lactobacillaceae*	*Lactobacillus plantarum*	Minimum 72 hours	[27]
13	Bacteroidetes	*Bacteroidaceae*	*Bacteroides uniformis*	144 hours	[146]
14	Bacteroidetes	*Bacteroidaceae*	*Bacteroides coprocola*	48 hours	[146]
15	Bacteroidetes	*Bacteroidaceae*	*Bacteroides vulgatus*	8-48 hours	[27, 146]
16	Bacteroidetes	*Bacteroidaceae*	*Bacteroides caccae*	72 hours	[146]

17	Bacteroidetes	*Rikenellaceae*	*Alistipes finegoldii*	72 hours	[146]
18	Bacteroidetes	*Bacteroidaceae*	*Bacteroides fragilis*	4 - 72 hours	[27, 148]
19	Actinobacteria	*Bifidobacteriaceae*	*Bifidobacterium adolescentis*	24-72 hours	[27, 28, 146]
20	Actinobacteria	*Bifidobacteriaceae*	*Bifidobacterium bifidum*	120 hours	[146]
21	Actinobacteria	*Bifidobacteriaceae*	*Bifidobacterium breve*	Minimum 48 hours	[28]
22	Actinobacteria	*Bifidobacteriaceae*	*Bifidobacterium longum*	Minimum 48 hours	[28]
23	Actinobacteria	*Bifidobacteriaceae*	*Bifidobacterium pseudocatenulatum*	Minimum 48 hours	[28]
24	Actinobacteria	*Coriobacteriaceae*	*Collinsella aerofaciens*	48 hours	[146]
25	Actinobacteria	*Coriobacteriaceae*	*Eggerthella lenta* †	45 minutes	[27]

‡ Previously known as *Peptococcus prevotii*. † Previously known as *Eubacterium lentum*

Table 1.2 Aerotolerance of non-spore forming intestinal bacteria.

A compilation of previously published experiments on the aerotolerance ability of a diverse selection of non-spore-forming intestinal bacteria. All assays involved *in vitro* inoculation of bacterial cultures on bacterial growth media that were exposed to ambient air for different time periods. This list is not exhaustive but is used to demonstrate broad patterns of varying aerotolerance amongst different taxa within the human intestinal microbiota. Reference numbers refer to the publication describing the original experiment.

There are other dormancy states that, unlike sporulation, are not reliant on the formation of specialised resistant structures. Viable but non-culturable (VBNC) state

is a form of bacterial dormancy that typically involves a decrease in metabolic activity and the generation of a strengthened cell wall that is achieved through modifications to its peptidoglycan structure. An increase or, more typically, a decrease in cell size has also been reported [154, 155]. These strategies all function to help bacteria withstand environmental stresses and preserve DNA integrity [154-156]. Similar to sporulation, the VBNC state is reversible through the removal of the inducing stress (for example, nutrient limitation or extreme temperature) or following exposure to growth stimulants, such as amino acids for *E. coli* or contact with intestinal cells for *V. cholerae* [154, 157]. Similar to spores, VBNC bacteria can remain dormant for long periods of time. For example, *Vibrio fluvialis* from marine sediment was successfully cultured after six years of dormancy following the addition of nutrients [158]. The majority of VBNC bacteria that have been identified thus far are human-associated pathogens, including *E. coli, Enterococcus faecium, V. cholerae, Salmonella enterica* subsp. enterica serovar Typhimurium and *Mycobacterium tuberculosis* [154, 159, 160]. Genetic and phenotypic characterisation of the VBNC state remains technically challenging, because the stimuli that are required to induce or culture bacteria from this dormant state are largely unknown or are difficult to simulate in a laboratory [161]. As VBNC dormancy is found in phylogenetically diverse bacterial species, it may be widespread in the intestinal microbiota and could be used as a strategy by non-spore-forming oxygen-sensitive commensal bacteria to survive in the environment until they are acquired by a new host. However, whether the VBNC state is induced in members of the commensal microbiota remains to be determined.

1.5.4 Reservoirs of commensal bacteria

Humans are the main reservoir of commensal intestinal bacteria, with transmission occurring readily between individuals. Childbirth is the first major life event in which the transmission of bacteria and colonisation occur. Depending on the mode of delivery, either the birth canal of the mother, or the hospital environment and the skin of the mother provide the initial inoculum of bacteria for the infant [162, 163]. Faecal–oral transmission could also occur during vaginal delivery, which would enable the immediate transmission of members of the intestinal microbiota to neonates at birth [162]. Compared with neonates that are born vaginally, it is thought that the composition of the microbiota of infants that are born by caesarean section may be more analogous to the skin microbiota than the vaginal microbiota in the early days of life [163]. Despite this, by six weeks, differences in the infant microbiota are determined by the body site and not the mode of delivery, which indicates that microbial convergence occurs early in life [164]. Evidently, there is no doubt that bacteria that can only be transmitted during vaginal delivery would be unable to colonise infants who are born by caesarean section. If no attempt is made to colonise infants who are born by caesarean section with these species, then, over generations, these species may be lost from the microbiota [165]. This decrease in diversity may have important health implications, as highlighted by reports that have associated immune disorders, such as asthma and allergies, in adult life with the abnormal development of the infant intestinal microbiota [165, 166]. After childbirth, inter-host transmission of intestinal bacteria continues, as shown by people who live in the same home sharing more species in common with each other than non-residents [24, 167-169]. The transmission rate of a bacterial species is affected by the number of hosts, their level of contact and their proximity to each

other, as well as by the inherent colonisation resistance of the microbiota in each individual, which is largely affected by age. A healthy adult has a broadly stable and resilient intestinal microbiota compared with an infant whose intestinal microbiota is still developing [167, 170].

Outside of family units, the effects of social interactions on the acquisition of the microbiota in large groups are best understood in non-human primates [171, 172]. Similar to humans, these are social animals that live and interact with each other in defined communities, and the composition of their intestinal microbiotas are influenced by the interactions of the social group. The higher the incidence of social interaction between individuals the more similar the composition of their intestinal microbiota, with species diversity increasing accordingly [171, 172]. The prevalence of anaerobic non-spore forming bacteria in baboons was associated with close social interactions between grooming pairs [171]. Although humans do not engage in social grooming, we physically interact through socially acceptable activities such as hand shaking, hugging and kissing, the frequency and intimacy of which increase as an individual interacts with a close family member or friend compared with a stranger. Thus, there is likely to be several social and cultural factors that contribute to the transmission of our intestinal microbiota.

Although the microbiota of an individual is largely structured and influenced by their diet, the microorganisms that are carried in food can also contribute to the intestinal microbiota [173]. From early life, infants acquire up to 8×10^6 bacteria daily, including intestinal-associated bacteria, through breast milk [174-176]. The mechanism by which intestinal bacteria translocate from the gut to the breast is unknown; however, an entero–mammary pathway that is facilitated by phagocytes that sample the gut lumen and subsequently translocate to the breast through the

bloodstream has been proposed [177-179]. Studies of various foods by culture-based methods have estimated that adults consume between 10^6 and 10^9 microorganisms daily [180]. Although most of the bacteria that are ingested do not survive transit through the stomach, those that do are not thought to colonise the gastrointestinal tract long-term [181]. The diversity of the microbiota that is acquired through food is dependent on diet [180, 182]; therefore, food provides a source of both exogenous bacterial species and genes for the resident microbiota to acquire through horizontal gene transfer [183].

Probiotics are defined as 'live microorganisms that, when administered in adequate amounts, confer a health benefit on the host' [184]. Probiotic bacteria, typically *Bifidobacterium* spp. and *Lactobacillus* spp., have been shown to alleviate the symptoms of several illnesses, including infectious diarrhoea, sepsis and atopic eczema [185-188]. The long-term colonisation efficiency of most probiotic bacterial species is variable [189, 190]; therefore, regular ingestion of probiotics is required to make a substantial long-term contribution to health [176, 191]. However, stable gut colonisation by *Bifidobacterium longum* six months after ingestion has been observed, which was attributed to the presence of an unoccupied niche that was vacated by a species that had similar carbohydrate-metabolising capabilities [192]. Overall, the variability in probiotic efficacy, coupled with host-specific responses to probiotics, means that the health benefits of ingesting these bacteria are not fully understood or predictable [193, 194].

Water is a major environmental reservoir for several intestinal bacterial pathogens, such as *Shigella flexneri, Shigella sonnei* and *V. cholerae*, which can cause debilitating gastrointestinal disease [195]. However, little is known about the fate or

the effect that commensal intestinal microorganisms that are found in water have on human health. The identification of bacteria in drinking water has primarily focused on pathogens, especially readily detectable indicator microorganisms, such as *E. coli*; however, the distinction between commensal and pathogenic strains of this species is not always made [196]. Despite an emphasis on pathogen detection, sequence-based culture-independent approaches have identified human-associated *Blautia* spp. in rivers [197]. Thus, it is clear that these species are transmitted through water; however, after the appropriate water treatment procedures, any strictly anaerobic non-spore-forming intestinal bacteria are likely to be killed, and these bacteria are therefore expected to have a low transmissibility and colonisation potential. Nevertheless, the full extent of the transmission of commensal intestinal bacteria through water is currently unknown.

Animals may also act as a reservoir for human microbiota with pets being the main source. The microbiotas of dogs and cats include taxa that are also found in the human microbiota; for example, genera such as *Roseburia*, *Faecalibacterium*, *Bacteroides*, *Prevotella* and *Ruminoccocus* [198, 199]. Farm animals are an additional source of bacteria. Analysis of the porcine intestinal microbiota has revealed similarities in taxonomic groups and functional capabilities with the human intestinal microbiota [5, 200]. Several human-associated pathogens, such as *Salmonella enterica* subsp. *enterica* serovar Enteritidis, *Campylobacter jejuni*, enteropathogenic *E. coli* and *C. difficile* [124, 201, 202], are transmitted between animals and humans; therefore, the potential for animals to transmit commensal species of bacteria is plausible. The treatment of animals that are to be used as food with antibiotics has also been linked to the acquisition of antibiotic-resistant strains of bacteria in humans [203, 204]. This highlights the need to recognise that human

health can be influenced through various diverse sources that are not directly connected to our own personal health decisions.

Interest in the microbes of the built environment has increased in recent years. Both buildings and transport systems adsorb our microbiota, which creates opportunities for microbial transmission across vast spatial areas and diverse human populations [205-208]. Humans are one of the main sources of indoor airborne bacteria that can spread through ventilation systems [209]. Outdoor air can also enter a building passively [205, 210]. Once bacteria become airborne (for example, through flushing a toilet or using a shower), viable bacteria can disperse around a room [211, 212]. In the built environment, the greatest density of human-associated bacteria will probably be found in bathrooms. Bacteria are abundant on surfaces that have been touched by human hands, on toilet seats or on floors [213, 214]. Skin-associated bacteria are the most dominant species on bathroom surfaces; however, a high proportion of intestinal-associated bacteria have also been found, such as members of the *Bacteroidaceae, Prevotellaceae, Ruminococcaceae* and *Lachnospiraceae* families. The presence of these intestinal-associated bacteria, together with poor hand-washing procedures, provides a reservoir for bacteria in the built environment that have the potential to transmit to humans [215]. A limitation of most of the studies on bacteria in the built environment is the lack of distinction between viable bacteria that have the potential to successfully colonise a new host and non-viable bacteria, which do not. It is estimated that only 1–10% of bacterial cells that are detected by culture-independent methods are viable [216]. Although culture-based methods can detect the viability of bacterial cells, the bacteria obtained will be an underrepresentation of the overall diversity in the sample.

1.6 Microbiota perturbation and restoration

1.6.1 Microbiota perturbation

The composition of an individual's microbiota remains largely stable once established after approximately three years [22, 25, 217]. Despite this, our microbiota experiences perturbations that can alter or damage its composition and functions [18, 181, 218-220]. Depending on the extent of the perturbation in the microbiota, and subsequent exposure to bacteria, the composition of the microbiota may be restored to a similar state or assume a new stable state that is composed of different bacterial species [221]. Therefore, a perturbation in the community will provide an opportunity for an externally derived bacterium to establish itself by reducing or eliminating competition from a resident species that occupies the same niche and requires the same nutrients [192, 222]. Factors that cause a microbiota perturbation are varied and range from antibiotic use, infection with a pathogen, a change in diet or travel [25, 181, 218, 223]. Changes in the composition of the microbiota have mostly been studied at the individual level; however, there is increasing evidence that suggests that changes in Western lifestyles and diet are altering the intestinal microbiota at larger population levels. Recently, it was observed that many traditional rural hunter–gatherer societies and agrarian groups that follow non-Western social behaviours and do not commonly use antibiotics or disinfectants have a more diverse intestinal microbiota that includes bacterial species that are now absent from the intestinal microbiota of developed world populations [224].

Any perturbation in an individual that eliminates certain bacterial species, or selects for some at the expense of other species, will prevent further onward transmission to other hosts [165, 225]. If the perturbation happens at the population level then the

effects may be compounded at a larger scale. For example, the consumption of a high-fat low-fibre diet, which is typical of Western populations, has been shown to cause the extinction of intestinal bacteria in mice if the diet is consumed over several generations [225]. Similarly, the use of antibiotics can negatively affect the diversity of intestinal bacteria, with repeated use preventing the restoration of the microbiota [219]. Although antibiotics and disinfection measures are essential for disease control and a high-calorie diet has greatly decreased undernutrition in Western societies; in this context, these changes may result in the indiscriminate elimination of commensal species, which could affect the diversity of the microbiota and microbial transmission [165, 226]. Indeed, a study in which the intestinal microbiotas of individuals who resided in either the United States or traditional agrarian societies in Papua New Guinea were compared attributed a lower α-diversity within, and higher β-diversity between, individuals in the United States cohort to decreased inter-host microbial transmission [227].

In addition to the observed decrease in the diversity of the intestinal microbiota in Western societies, an increase in autoimmune and allergic diseases in the developed world has been observed [165, 228, 229]. Originally termed the 'hygiene hypothesis', there is increasing evidence in humans and animal models that exposure to microorganisms early in life promotes the maturation of the immune system and decreases the incidence of autoimmune-related diseases [228, 230-232]. Consistent with this hypothesis, the use of antibiotics in childhood has been associated with an increased likelihood of developing paediatric inflammatory bowel disease and a pre-disposition to asthma and obesity in later life [219, 233]. These examples illustrate the importance of efficient microbial transmission networks and the potential effect on human health when they fail.

1.6.2 Microbiota restoration

Direct interventions currently provide the most immediate solution to establish or restore a diverse and beneficially functional microbiota across all age groups. Recent interventions in this area have included swabbing neonates born by caesarean section with gauze that has been pre-incubated in the vagina of the mother to mimic the natural transmission of the vaginal microbiota to the child [234]. In adults who are susceptible to recurrent infections with *C. difficile*, faecal microbiota transplantation (FMT) from a healthy donor has proven extremely effective at resolving such infections (see 'Therapeutics based on the intestinal microbota' section below) [235]. As the number of human gut commensal species that have been isolated and archived as pure cultures continues to increase [80, 236, 237], the development of live biotherapeutics for the treatment of disorders other than *C. difficile* infection will become feasible. Next-generation probiotics and functional foods that make use of the numerous diverse beneficial bacteria other than the widely used *Lactobacillus* spp. and *Bifidobacterium* spp. could potentially provide health benefits to individuals and to the wider interconnected human population. However, until there are improvements in culturing processes and an increase in the number of commensal bacteria isolated and archived for characterisation, their potential remains unfulfilled.

At a broader level, changes in living practices can promote the transmission of, and colonisation by, health-associated commensals at the expense of pathogen colonisation. A course of broad-spectrum antibiotics to treat a gastrointestinal infection also eliminates many beneficial commensals, thus rendering the microbiota-deficient host susceptible to infections with other pathogens [131, 223]. Indeed, narrow-spectrum antibiotics, or antimicrobials that have specific targets, such as bacteriocins or phage therapies, are desirable and under development as

alternatives to treatment with broad-spectrum antibiotics [238]. The effects of antibacterial hygiene products and hand sanitizers on the intestinal microbiota are unclear and require further study; however, efforts to use them more selectively may reduce the depletion of commensal bacteria [239]. Hygiene practices currently act to decrease the total number of bacteria on a surface or individual, whereas a more targeted approach that specifically removes pathogenic microorganisms should be given greater consideration [240].

1.6.3 Understanding microbiota perturbation through ecological theory

The human intestinal microbiota is an ecological community in terms of its individual members both cooperating and competing with each other to utilise available resources. Ecological theory provides a framework to predict and model the human microbiota [221]. This is important not only for our understanding of how the microbiota functions and responds to disturbances but also to predict its responses to perturbations. Metacommunity theory is defined by local communities within a habitat that are spatially separated from each other but are connected by the species migrating between them and can be used to understand dispersal and transmission of the intestinal microbiota between different human hosts [241]. Important processes in metacommunity theory include the dispersal of species between local host habitats, which maintains community diversity, the ability of species to adapt to a host environment and selection or filtering by the host to ensure only beneficial species colonise [221]. These processes are not independent of each other, hence, a low dispersal rate within a community will favour high adaptation and conversely a high dispersal rate will favour low adaptation [221].

A final important process in metacommunity theory is ecological drift which is caused by stochastic events that perturb the microbial community. Low-abundance

species are more susceptible to these events and are more likely to move towards extinction unless they can either adapt to ensure survival within the environment or they can disperse to a new host [221, 242, 243]. Examples of stochastic events could include a short-term change in diet, antibiotic usage or transient infection by a food-borne pathogen [181, 218]. Any species that becomes extinct will be replaced by another filling the newly vacant niche, hence dispersal is an important component of maintaining a diverse and functional microbiota [242]. Ecological drift will result in widespread variation of bacterial species amongst different individuals as species become extinct and are replaced. Indeed, this is a feature of the intestinal microbiota within individuals at lower taxonomic levels [244]. There is evidence to suggest that low abundance species possess dormancy mechanisms which either protect against ecological drift within a habitat or facilitate dispersal to a new environment if expelled [245]. In many microbial communities, a few species are disproportionally abundant and most species are rare, hence, ecological drift and dormancy may be prevalent in many microbial communities including the intestinal microbiota [245, 246].

1.7 Therapeutics based on the intestinal microbiota

The intestinal microbiota is associated with a considerable number of ailments including a wide range of intestinal-associated diseases, inflammation-associated disorders and neurological disorders [247]. The extent of the ailments the intestinal microbiota is associated with and the lack of substantive evidence demonstrating causality has led to calls for more mechanistic studies that demonstrate that the differences observed in a healthy verses disease state are not merely incidental [247, 248]. Until there are carefully designed studies that demonstrate that particular bacteria are causative or contributory, then any associative study should be treated

with caution. To avoid exaggeration of results, studies should be designed to identify bacterial species of interest, not merely examine changes in composition at high taxonomic levels which may not be informative. Admittedly, we currently lack genetic tools to engineer many species of the intestinal microbiota which would allow fulfilment of 'molecular Koch's postulates' necessary to implicate a gene in a disease, however this situation is changing [249, 250].

Despite the lack of mechanistic studies and genetic tools, there have been several important therapeutic breakthroughs, especially for treating *C. difficile* infection. Recurrence of *C. difficile* infection can occur in the same individual in 20% of cases with the chances of subsequent recurrence increasing thereafter. This is primarily due to antibiotic-mediated lack of commensal colonisation resistance and sub-optimal host immunity [251]. In these scenarios, an alternative treatment is faecal microbiota transplantation (FMT) from a healthy donor. While the exact mechanism remains unknown, it is believed that the microbiota in the donor faecal sample rapidly restores colonisation resistance in the form of a healthy functioning intestinal microbiota with subsequent displacement of *C. difficile*. The microbiota profile of the recipient post-FMT often resembles that of the donor, however the recipient can maintain some of their original strains which were present before *C. difficile* infection [252, 253]. This treatment is extremely effective for resolving *C. difficile* infection and clinical trials have reported success rates of over 90% [235]. As a result of its high efficacy, the demand for routine access to safe FMT material has stimulated the establishment of companies such as OpenBiome, that have created a bank of faecal samples from screened donors [254].

While effective, FMT is not an ideal therapeutic as it is not standardised, undetected pathogens may be present at low abundances and it will differ in microbial

composition from donor to donor. A more desirable treatment is a defined mix of bacteria that have the same efficacy at resolving *C. difficile* infection. Successful outcomes using a mix of six phylogenetically diverse bacteria in a mouse model and 33 phylogenetically diverse bacteria in humans have been reported [223, 255]. Defined mixes of bacteria are also used to target inflammatory conditions. These therapeutics are utilising bacteria that produce Short Chain Fatty Acids (SCFA) from dietary fibre which induce regulatory T cells (Tregs) that dampen down inflammation. The efficacy of this approach has been demonstrated using a mix of 17 human-derived bacteria in mice, and is now being developed as a commercial product [145, 254]. *C. difficile* infection and inflammatory-associated conditions are currently the most promising live therapeutic targets but there is also evidence that suggests that the intestinal microbiota may play a role in cancer, drug metabolism and neurological disorders [254, 256, 257]. Consequently, these 'live biotherapeutics' are now a commercial target for many microbiota-based companies.

1.8 Thesis aims

There were two broad aims in this study. The first aim was to develop a process to culture, isolate and archive a representative selection of bacteria from the intestinal microbiota of healthy humans. The second aim was to understand the relevance of spore-formation to the process of microbiota transmission, to explore the breadth of spore-formation in the human gut microbiota, and to evaluate the evolutionary dynamics of the phenotype.

The objectives relevant to the first aim were: (1) to determine the culturability of the human intestinal microbiota using a streamlined culturing process e.g. one growth medium, (2) to incorporate a phenotypic screen in the culturing process to culture

and isolate spore-forming bacteria and (3) to assess the phylogenetic distribution and novelty of the bacteria cultured.

The objectives relevant to the second aim were: (1) to examine the transmission mechanisms of spore-forming bacteria by assessing their ability to survive *ex vivo* in aerobic conditions and to colonise a new host by germination of spores once ingested, (2) to establish the extent of sporulation within the human intestinal microbiota using whole genome sequencing and metagenomic approaches, (3) to explore the phylogenetic and environmental distribution of spore-forming bacteria (from all environments, not just the gut) in the Firmicutes and (4) to compare genomes of spore-forming and non-spore-forming bacteria to explore genomic features of host-adaptation.

Chapter 2 Materials and Methods

2.1 Culturing of bacteria from the human intestinal microbiota

2.1.1 Acquisition of donor faecal samples

Ethical approval was obtained for collection of faeces and analysis of microbial components from appropriate ethical bodies at the Wellcome Trust Sanger Institute. Donors completed a questionnaire to assess their suitability for donation. To be included in the donation process, donors must not have experienced any gastrointestinal infections six months prior to donation, must not have used antibiotics six months prior to donation, must not have suffered from, or be recovering from chronic intestinal diseases such as Crohn's disease, Ulcerative colitis, Coeliac disease, Irritable bowel syndrome, stomach ulcers or Colorectal cancer. In addition, donors were required to be free of autoimmune diseases or allergies such as multiple sclerosis, asthma or psoriasis.

Metadata was also obtained from the donors including age-range, nationality, ethnicity, time resident within the UK and diet consumed (i.e. vegan, vegetarian or omnivore). To maintain donor anonymity, no identifying information such as date of birth, name or address was requested. Of the six donors who participated, all were Caucasian of different nationalities, four were male, two were female with an age range from 26 to 45 years old. Five were omnivores and one individual was vegetarian. All were resident in the UK for at least 2.5 years.

2.1.2 Anaerobic culturing

Fresh faecal samples were obtained from the donors and weighed (one faecal sample per donor: minimum 0.5 g per sample) before being placed in anaerobic conditions

within one hour of being passed to preserve the viability of anaerobic bacteria. All sample processing and culturing took place under anaerobic conditions (gas composition was 10 % carbon dioxide, 10 % hydrogen, 80% nitrogen) in a Whitley DG250 workstation at 37 °C. Culture media, Phosphate Buffer Saline solution (PBS) and all other materials that were used for culturing were placed in the anaerobic cabinet for at least 12 hours before use to reduce. The faecal samples were divided in two. One part was homogenized in reduced PBS (0.1 g stool per ml PBS) and was serially diluted down to 10^{-7} and plated directly onto YCFA agar supplemented with 0.002 g ml^{-1} each of glucose, maltose and cellobiose in large (13.5 cm diameter) Petri dishes (Table 2.1). The YCFA media was developed by the Rowett Institute in Aberdeen, Scotland [75]. Plating was performed by adding 250 µl of the faecal solution to the Petri dish and uniformly spreading it over the plate using a disposable hockey stick (plate spreader).

Ingredient	Amount		Components of solutions and mixes:	
Before Autoclaving			**Resazurin Solution**	
Agar (optional)	8 g		Resazurin	0.1 g
Tryptone	5.0 g		d. H_2O	100 ml
Yeast extract	1.25 g			
NaHCO3	2.0 g		**Mineral Solution I:**	
(D)+Glucose	1.0 g		K_2HPO_4	3 g
(D)+Maltose	1.0 g		d.H_2O	1 L
(D)+Cellobiose	1.0 g			
L-cysteine	0.5 g		**Mineral Solution II:**	
Mineral Solution I	75 ml		KH_2PO_4	3 g
Mineral Solution II	75 ml		$(NH_4)2SO_4$	6 g
Resazurin Solution	0.5 ml		NaCl	6 g
Haemin Solution	5 ml		$MgSO_4$	0.6 g
Vitamin solution I	0.5 ml		$CaCl_2$ (dry)	0.6 g
d.H2O	up to 500 ml		d.H_2O	1 L
VFA mix	3.1 ml			
NaOH	pH to 7.45		**VFA mix:**	
			Acetic acid	17 ml
After Autoclaving			Propionic acid	6 ml
Vitamin solution II	0.5 ml		n-Valeric acid	1 ml
			Isovaleric acid	1 ml
			Isobutyric acid	1 ml
			Haemin Solution:	
			KOH	0.28 g
			Ethanol 95 %	25 ml
			Haemin	0.1 g
			d.H_2O	up to 100 ml
			Vitamin Solution I:	
			Biotin	5 mg
			Cobalamin (Vitamin B12)	5 mg
			PABA (4-Aminobenzoic Acid)	15 mg
			Folic acid	25 mg
			Pyridoxine	75 mg
			d.H_2O	up to 500 ml
			Vitamin Solution II:	
			Thiamine hydrochloride	25 mg
			Riboflavin	25 mg
			d.H_2O	up to 500 ml

Table 2.1 YCFA media

Volumes are to make 500 ml of YCFA media (either agar or broth). The solutions and mixes are prepared separately and then added in the required volume when the media is being prepared.

The other part of the faecal sample was mixed with an equal volume of 70 % (v/v) ethanol and was vortexed approximately once every hour, over a four hour period at room temperature under ambient aerobic conditions to kill vegetative cells. Then, the mix was centrifuged, the ethanol was decanted and the solid material was then washed three times with PBS (an equal volume of PBS was added, the mix was vortexed, then centrifuged and the PBS was decanted). After washing it was resuspended in PBS, again at a concentration of 0.1g stool per ml PBS. Plating was performed as described earlier.

For the ethanol-treated samples, the medium was supplemented with 0.1 % sodium taurocholate to stimulate spore germination. Colonies were picked 72 hours after plating from a Petri dish of both ethanol-treated and non-ethanol-treated conditions harbouring non-confluent growth, (that is, plates on which the colonies were distinct and not touching). Approximately 100 colonies were picked from each plate, the aim was to pick colonies of different morphologies to ensure a diverse selection was isolated. The colonies that were picked were re-streaked onto new YCFA media on regular sized Petri dishes (9 cm diameter) to visually confirm purity before 16S rRNA gene sequencing profiling (below). Once colonies were identified, they were grown overnight in 10 ml YCFA broth and 500 ul of this inoculum was then frozen at -80 °C in a solution containing 50 % bacterial culture (500 ul) and 50 % glycerol (500 ul) (glycerol was 25 % final concentration).

2.1.3 Microbiota profiling by 16S rRNA gene sequencing

Identification of each picked isolate was performed by PCR amplification of the full-length 16S rRNA gene (using 7F (5′-AGAGTTTGATYMTGGCTCAG-3′) forward primer and 1510R (5′-ACGGYTACCTTGTTACGACTT-3′) reverse primer followed by capillary sequencing. Full-length 16S rRNA gene sequence reads were

aligned in the Ribosomal Database Project (RDP). RDP classifies sequences by comparison to a 16S rRNA gene sequence database that has taxonomic information assigned based upon Bergey's Taxonomic Outline of the Prokaryotes [258, 259]. These sequences were manually curated in ARB and mothur was then used to classify reads to operational taxonomic units (OTUs) using the cluster command [260, 261]. The R package seqinr version 3.1 was used to determine sequence similarity between OTUs and 98.7 % was used as a species-level cut-off [262, 263]. The full-length 16S rRNA gene sequence of each species-level OTU was compared to the RDP reference database to assign taxonomic designations to the genus level and a BLASTn search defined either a characterized or candidate novel species [258, 264].

Comparisons of 16S rRNA gene sequences with the Human Microbiome Project (HMP) were carried out using 97 % sequence similarity of the 16S rRNA gene sequence from the cultured bacteria to define a species because only partial 16S rRNA gene sequences were available. HMP data regarding the most wanted taxa and the completed sequencing projects were downloaded from http://hmpdacc.org/most_wanted/#data and http://hmpdacc.org/HMRGD/, respectively.

2.1.4. Submission of cultured bacteria to public repositories

To make the cultured isolates available to the wider research community, a representative species was sent to at least one of the following four public repositories- Deutsche Sammlung von Mikroorganismen und Zellkulturen Gmbh (DSMZ) in Braunschweig, Germany (http://www.dsmz.de), Japan Collection of Microorganisms (JCM) maintained by the Riken BioResource Center in Tsukuba, Japan (http://jcm.brc.riken.jp/en/), the Culture Collection, University of Gothenburg

(CCUG) maintained by the University of Gothenburg in Sweden (http://www.ccug.se/) and the Belgian Co-ordinated Collection of Micro-organisms (BCCM/LMG) hosted by the Laboratory of Microbiology at Ghent University, Belgium (http://bccm.belspo.be/). These were all sent as frozen glycerol stocks on dry ice except for isolates sent to the DSMZ which were sent as viable cultures on agar slants in 50 ml Falcon tubes. 110 of the 137 bacterial species cultured in this study were deposited in these public repositories. The accession numbers for the deposited isolates are listed in Appendix 2.

2.2 Generation of whole-genome, metagenomic and 16S rRNA gene amplicon sequence data

2.2.1 DNA extraction

Genomic DNA was extracted from at least one representative of each unique OTU using a phenol-chloroform-based DNA isolation procedure. The culture was grown in 10 ml YCFA media overnight before being washed three times by repeatedly centrifuging to a pellet and re-suspending in 10 ml PBS. The washed pellet was stored at -20 °C until DNA was ready to be extracted. DNA extraction involved release of DNA from the rest of the components of the bacterial cell using lysozyme, proteinase K and RNase A. DNA was then further isolated by repeated mixing with Phenol: Chloroform: Isoamyl Alcohol (ratio 25:24:1) (Sigma-Aldrich, Missouri, USA) in a phase lock tube (Qiagen, Hilden, Germany), followed by centrifugation to separate the aqueous phase (containing the DNA) on top and the denser organic phase on the bottom which contained the phenol, chloroform and organic bacterial components. Any phenol in the aqueous phase was then removed by phase separation using Chloroform: Isoamyl Alcohol (ratio 24:1) (Sigma-Aldrich, Missouri, USA), again using phase lock tubes. Cold 100 % ethanol was then added

to precipitate the DNA and DNA concentrations were quantified using a Qubit Fluorometer (ThermoFischer Scientific, Massachusetts, USA).

DNA was extracted directly from each faecal sample for whole-community metagenomic and 16S rRNA gene amplicon sequencing using the MP Biomedical FastDNA SPIN Kit for soil, the protocol was performed according to the manufacturer's recommendations, apart from the following modifications: Sodium phosphate buffer was added first to the Lysing Matrix E tube, followed by MT buffer, followed by 300 µl of 100 mg/ml of faeces homogenised in PBS. The Binding Matrix suspension was allowed to settle for 5 minutes, not 2 minutes and 700 µl of this mixture was added to SPIN Filters before centrifugation. Finally, before air drying the SPIN Filter containing DNA, it was centrifuged twice at 14,000 g for one minute each with the tube turned 180 degrees before the second spin. To enable comparisons with the complete community in the faecal samples, non-confluent cultures were removed from agar plates by adding sterile PBS as required and then scraping the cultures off using a disposable hockey stick into a sterile receptacle. DNA was again extracted using the MP Biomedical FastDNA SPIN Kit for soil.

2.2.2 Generation of whole-genome and metagenomic sequence data

All DNA samples were submitted to the DNA pipelines department at the WTSI for library creation and DNA sequencing. Library creation consisted of DNA fragmentation to the required size (200-300 bp) using the Covaris ultrasonicator (Covaris, Massachusetts, USA), followed by solid phase reversible immobilisation (SPRI) cleanup to remove unwanted, smaller fragment and to concentrate the DNA fragments of a desired size and then adapter ligation. These fragments were then PCR amplified and immobilised to the flow cell where cluster formation took place

according to Illumina's protocols. DNA was sequenced on the Illumina HiSeq platform generating paired-end fragments of between 200 and 300 bp and read lengths of 100 bp. Sequence data for whole-genome sequence and metagenomic sequence is stored at the ENA under accession number ERP012217.

2.2.3 Generation of 16S rRNA gene amplicon sequence data

16S rRNA gene amplicon libraries were made by PCR amplification of variable regions 1 and 2 of the 16S rRNA gene using the Q5 High-Fidelity Polymerase Kit supplied by New England Biolabs. Primers 27F AATGATACGGCGACCACCGAGATCTACAC (first part, Illumina adaptor) TATGGTAATT (second part, forward primer pad) CC (third part, forward primer linker) AGMGTTYGATYMTGGCTCAG (fourth part, forward primer) and 338R CAAGCAGAAGACGGCATACGAGAT (first part, reverse complement of 3′ Illumina adaptor) ACGAGACTGATT (second part, golay barcode) AGTCAGTCAG (third part, reverse primer pad) AA (fourth part, reverse primer linker) GCTGCCTCCCGTAGGAGT (fifth part, reverse primer) were used. Four PCR amplification reactions per sample were carried out; products were pooled and combined in equimolar amounts for sequencing using the Illumina MiSeq platform, generating 150 bp reads. ENA accession numbers are Donor 1 ERR671373, Donor 2 ERR671374, Donor 3 ERR671375, Donor 4 ERR671376, Donor 5 ERR671377 and Donor 6 ERR671378.

2.3 Analysis of sequence data generated through anaerobic culturing

2.3.1 Assembly and annotation of whole-genome sequence data

Assembly and annotation of whole genome sequence was carried out by the Pathogen Informatics group at the WTSI (this did not apply to metagenomic

sequence) using the methods described in reference [265]. VelvetOptimiser was used to assemble the sequence data with the resulting assemblies improved through sequence gap closure using GapFiller [266, 267]. The assemblies were annotated using the software tools contained in Prokka [268]. In brief, Prodigal was used to define coding-sequence boundaries and these were annotated using bacterial proteins in UniProt and the NCBI reference sequence (RefSeq) database, followed by protein domain annotation using Pfam and TIGRFAMS [269-273]. Non-protein coding annotation was also carried out and included rRNA genes, transfer RNA genes, signal leader peptides and non-coding RNA [274-277].

2.3.2 Analysis of metagenomic sequence data

Metagenomic sequence reads were analysed by Sam Forster of the WTSI. The Human Pan-Microbes Community Database (HPMCD) [278] is a manually curated database of metagenomic sequences sourced from the European Bioinformatics Institute (EBI) [279]. It also contains the metagenomic sequence generated in this study. Metagenomic reads contained in the HPMCD were taxonomically classified using Kraken which assigns a taxonomic classification using a custom database containing complete, high-quality reference bacterial, DNA viral and archaeal genomes in addition to the genomes sequenced in this research [280]. Kraken utilises small segments of query sequences (kmers) instead of the entire query sequence length to map to references sequences, which speeds up the process. Furthermore, it will map reads down a taxonomic classification as far as accurately possible (i.e. some reads may be mapped and classified at the phylum level and others may be mapped and classified at the species level). Resulting classified reads were log2 transformed and standardized by total abundance.

2.3.3 Analysis of 16S rRNA gene amplicon sequence data

Analysis of the partial 16S rRNA gene sequence generated from the 16S rRNA gene amplicon libraries was carried out using the mothur MiSeq SOP generating 7,549 OTUs across all samples [261, 281]. An online step by step description of the SOP is described here: https://www.mothur.org/wiki/MiSeq_SOP. This pipeline includes a screening step to remove ambiguous sequences, a filtering step to remove duplicated sequences (this reduces computational requirements), a mapping step in which the unique sequences are mapped to a database of taxonomically defined reference 16S rRNA gene sequences and a clustering step in which sequences are clustered together based on similarity. As these were partial 16S rRNA sequences of 150bp, a sequence similarity threshold of 97% was used to define an OTU. The Principle Coordinates Analysis (PCoA) plot displaying the separation of donor ethanol-treated and ethanol non-treated samples was constructed using the distance matrix (phylip format file) produced by this pipeline. The distance matrix captures the differences between the donors and culture conditions by comparing every sequence against every other sequence and converting the sequence differences to a value. A PCoA plot presents multi-dimensional data (in this scenario, the differences in sequences from different donors and different culture condition) in a two-dimensional format. The taxonomic information obtained by mapping the reads against the reference database was used to define the relative abundance of spore-formers at the genus and family level.

2.3.4 Comparison of culturing studies

The 1172 16S rRNA gene sequences from the Goodman *et al.* study [79] that were archived in 384-well plates were downloaded and clustered using mothur. As the reads from this study were ~200-300 bp in length, OTUs were clustered at 97%

sequence similarity generating 63 OTUs in total. The number of characterised and novel species from the two Lagier *et al.* studies was obtained directly from these papers [80, 282].

2.3.5 Gene sporulation signature

BLASTp was used to identify 21,342 conserved genes within the 694,300 genes annotated across the 234 sequenced genomes. A matrix of the conserved genes was then constructed populated by the BLASTp scores. Each isolate was assigned an ethanol-resistant or ethanol-sensitive status based on the culture condition of the isolate it was derived from (ethanol-resistant or ethanol-sensitive). A contrast-set machine learning based model was then used to identify a list of genes that best differentiated ethanol-resistant from ethanol-sensitive bacteria. This process was iterative in that different combinations of genes were tested until the final list of 66 genes was achieved that best distinguished the two groups. Detection of signature genes in a genome was performed using BLAST and the number of genes present was calculated as a percentage and presented as a score between 0 and 1 depending on the number of genes present. Scores greater than 0.5 were considered true spore-formers based on comparison to known spore-formers.

Using the sporulation signature Sam Forster calculated the abundance of spore-formers in 1351 publically available metagenomic data sets contained in the HPMCD. Kraken was used to assign taxonomic labels to the metagenomic sequences and the relative abundance of these sequences was then calculated. I calculated the abundance of spore-formers in the six donors and their associated culture plates using the taxonomically classified (using Kraken) relative abundance sequence data that was previously generated by Sam Forster. Genera were

considered spore-forming when all known species within that genus had a spore forming score greater than 0.5.

2.3.6 Comparison of abundance of bacteria in faecal samples and on culture plates

The relative abundance of taxonomically defined (by Kraken) metagenomic sequence reads generated in Section 2.3.2 from the original donor faecal samples and the isolates scraped off the culture plate were compared. There were 17 bacterial species detected on culture plates that were not detected in faecal samples, these were all less than -7.3(log10) in abundance, hence this was chosen as the limit of detection in the faecal sample and the culture plate. The average relative abundance across all six donors for a bacterial species detected in both samples above this cut-off value was plotted. In total there were 1079 species detected using these criteria.

2.3.7 Diversity analysis of spore-forming and non-spore-forming bacteria

Inverse Simpson's diversity index of spore-forming and non-spore-forming bacteria was calculated using the summary.single command in mothur. The input was the relative abundance of taxonomically defined (by Kraken) metagenomic sequence reads generated in Section 2.3.2 from the original donor faecal sample. This data included the ability of a bacterial species to make spores based on sporulation signature analysis. The Inverse Simpson diversity index is an alpha diversity measure that takes into account the number of species present (defined as the richness) and their abundance in a sample (defined as the evenness).

2.4 Phylogenetic analysis

2.4.1 16S rRNA gene sequence phylogeny of cultured bacteria

A maximum likelihood phylogeny of the culture-derived bacteria was generated from the RDP aligned full-length 16S rRNA gene sequence using FastTree version 2.1.3 with the following settings: a generalized time-reversible (GTR) model of nucleotide substitution and CAT approximation of the variation in rates across sites with 20 rate categories [283]. The ethanol-resistant phylogeny was derived directly from the entire culture phylogeny. Both phylogenetic trees were edited in ITOL [284].

2.4.2 Phylogeny of putative novel bacterial family

The full-length 16S rRNA gene sequence of *Clostridium thermocellum*_86% (renamed as *Falkowia sangerensis*) was aligned with 16S rRNA gene sequences from this study in addition to closely related sequences downloaded from NCBI. Closely-related sequences were determined based on a BLASTn search of the *Clostridium thermocellum*_86% sequence. The other sequences from this study provide phylogenetic context. MAFFT was used to align the sequences [285], this was visually examined using Seaview [286] and a maximum likelihood phylogeny with 1000 bootstrap replicates was generated using MEGA [287]. RDP and FastTree were not used for this phylogeny as alignment files and phylogenies produced using these methods were not as robust as MAFFT and MEGA and phylogenetic artifacts were present in the form of different families incorrectly clustering together. Note- regardless of method used the *Falkowiaceae* family always clustered separately to other families. These artifacts are most likely due to the smaller number of sequences used in this phylogeny (25) compared to the culture collection phylogeny (137).

2.4.3 Core gene phylogeny of Firmicutes

Using the FetchMG script [288], 40 universal, single-copy marker genes were extracted from Firmicutes whole genome sequences from the NCBI reference sequence (RefSeq) database (725 in total), the HMP (226 in total), an in-house collection of genomes from the Host-Microbiota Interactions Laboratory at the WTSI (506 in total) and from a study describing the first 1000 cultured species from the human gastrointestinal tract [26] (149 in total). These genomes are all from bacteria isolated from the gastrointestinal tract except for the NCBI RefSeq database which consists of curated and non-redundant genomes from a wide range of environments including host-associated and environmental habitats [289]. The resulting amino acid sequences were aligned using MAFFT, gaps representing poorly aligned sequence were removed using the Gblocks script and a maximum likelihood phylogeny was constructed using FastTree [283, 285, 290]. All support values (using the Shimodaira-Hasegawa test which estimates likelihood of local branches as opposed to resampling from the entire tree using a bootstrap approach) down to the family level of the phylogeny are greater than 0.8 (1 is maximum) except for the branch that divides *Staphlococcaceae* and *Bacillaceae* (0.39) and the branch that divides *Veillonellaceae* and *Erysipelotrichaceae* and the *Bacilli* (0.76). Actinobacteria, Fusobacteria and Proteobacteria sequences were used to root the tree.

For the comparison of the species phylogeny verses the Spo0A_C phylogeny, the amino acid sequences of the sporulation specific C-terminus domain sequence of the *Spo0A* gene (termed Spo0A_C) was extracted from the genomes, aligned using MAFFT, gaps representing poorly aligned sequence were removed using the Gblocks script and a maximum likelihood phylogeny was constructed using FastTree

[283, 285, 290]. Further details on acquisition of Spo0A_C sequence is described in Section 2.6.3. Both phylogenetic trees were edited in ITOL [284].

2.5 Phenotypic analysis of spore-forming bacteria

2.5.1 Oxygen sensitivity assay

Pure cultures were grown overnight in YCFA broth under anaerobic culture conditions as described above. The cultures were then spotted in a dilution series onto YCFA agar containing 0.1 % sodium taurocholate. Plates were incubated under ambient (aerobic) conditions at room temperature for specified time periods before being returned to the anaerobic cabinet. Colony-forming units (CFU) were counted 72 hours later. Cultures that were incubated anaerobically, and which were therefore not exposed to oxygen, acted as controls. Prior to the assay, all species were subjected to ethanol shock and were cultured anaerobically to determine their ability to sporulate. The viability of the oxygen-exposed cultures was expressed as a percentage of the viability of the anaerobic control cultures.

2.5.2 Germination response to intestinal bile acids assay

Pure cultures were grown overnight in YCFA broth under anaerobic conditions and were then washed three times by repeatedly centrifuging to a pellet and re-suspending in PBS. Vegetative cells were killed using an ethanol shock treatment as previously described and the cultures were then serially diluted and plated on YCFA agar with and without 0.1 % intestinal bile salts (taurocholate, cholate and glycocholate). CFU were counted 72 hours later and the fold change of the number of CFU present on plates in the presence of a particular germinant with respect to the number of CFU present on plates in the absence of a germinant was calculated. The limit of detection (200 $\mathrm{CFU\,ml^{-1}}$) was used for the number of CFU recovered

from *Clostridium hathewayi* plated without any germinants to allow a fold-change calculation. The experiment to determine the response of non-spore-formers to germinants was carried out similarly, except that vegetative cells were not treated with ethanol but rather were serially diluted and plated directly after washing.

2.5.3 Transmission electron microscopy

Spore images were generated using transmission electron microscopy (TEM) as previously described [291]. Bacterial isolates for imaging were prepared by streaking pure cultures from frozen glycerol stocks and confirming purity by full-length 16S rRNA gene sequencing after one round of sub-culture to obtain visible and isolated single colonies. TEM images were prepared from culture plates 72 hours after inoculation. Cultures were fixed using a solution containing 2 % paraformaldehyde and 2 % glutaraldehyde in 0.1 M sodium cacodylate buffer (pH 7.42). The cultures were then washed three times in sodium cacodylate buffer and were further fixed using 1 % osmium tetroxide, mordanted with 1 % tannic acid and rinsed in 1 % sodium sulphite. Repeated ethanol washes using increasing concentrations of ethanol were used to dehydrate the cultures, followed by staining with 2 % uranyl acetate before being embedded in Epon resin at 65 ˚C for 24 hours. Ultrathin sections of 40 nm thickness were cut on a Leica UC6 ultramicrotome, contrasted with uranyl acetate and lead citrate and viewed using a 120-kV FEI Spirit BioTWIN transmission electron microscope and a F415 Tietz charge-coupled device camera. The number of spore bodies visible in the TEM images was expressed as a percentage of the number of vegetative cells present and this ranged from 1 % for *Ruminococcus flavefaciens*_93% to 4 % for *Turicibacter sanguinis.*

2.6 Functional redundancy analysis

2.6.1 Ortholog analysis

A workflow outlining this process is described in Figure 5.7. Average Nucleotide Identity (ANI) was used to group genomes into species based on a whole genome sequence similarity cut-off of 95 % [292]. The associated environment of each ANI species was determined using the source of isolation of each species and a broader literature search. One genome from each representative ANI species was used for subsequent analysis. Using orthoMCL, the genes from the genomes were then clustered based on sequence similarity into homologous groups (all protein sequences were concatenated together into a single file and then all proteins were blasted back against this file. The output from the blast analysis was then used as the input for OrthoMCL- default settings: 1e-05 e-value for clustering, --mode3) [293]. The genes of each of the *Peptostreptococcaceae, Lachnospiraceae* and the *Erysipelotrichaceae* families were clustered separately. Details on the number of genomes and ANI species in each comparator group are contained in Table 2.2. This produced a list of clusters (homologous groups) comprised of genes with similar sequence. Each gene that comprised each homologous group was categorised according to the ANI species it was derived from, the associated environment of that species and the ability of that species to form spores. Genes in a homologous cluster can be from the same genome (paralogue) or from different genomes (orthologue). Spore-formation ability was determined by the presence of the Spo0A_C domain and separation of genomes based on sporulation signature score presented in Figure 5.4.

	Erysipelotrichaceae	*Peptostreptococcaceae*	*Lachnospiraceae*
Number of genomes	69	36	430
Number ANI species	34	24	197
Spore-forming gut species	21	12	139
Non-spore-forming gut species	9	1*	1*
Spore-forming oral	0	0	1*
Non-spore-forming oral species	2*	5	9
Non-spore-forming rumen species	2*	0	44
Spore-forming environmental	0	4*	2*
Spore-forming rumen	0	0	1*
Non-spore-forming environmental species	0	2*	0*
Total number genes	87599	59357	655862
Total number homologous groups	7299	5125	24994

Table 2.2 Dataset for orthoMCL analysis

These genomes (and the genes within) were used to compare spore-forming species against non-spore-forming species within the same bacterial family. Species marked with an asterisk were not included as there were insufficient numbers to compare.

2.6.2 Gene per genome analysis:

For each homologous group produced using orthoMCL, the gene per genome ratio was calculated according to the number of genes in that homologous group and the number of genomes the genes were derived from. Genomes were categorised according to their ability to form spores resulting in a gene per genome ratio for spore-forming and non-spore-forming genomes. Only homologous groups that contained genes from at least 50 % of spore-forming and/or 50 % of non-spore-forming ANI species from each taxonomic family were included. To compare within both spore-forming and non-spore-forming groups genes from greater than 50 % of ANI species of both groups had to be present in a homologous group for it to be included in the analysis. Only ratios equal to or greater than two were selected (i.e. the ratio of gene per genome value of a spore-forming group compared to the gene per genome value of a non-spore-forming group within the same homologous group was equal to or greater than two or vice versa when comparing non-spore-formers to spore-formers).

2.6.3 Presence and absence of sporulation associated genes

The sporulation specific C-terminus domain sequence of the *C. difficile* CD630 *Spo0A* gene (Pfam entry: PF08769), termed Spo0A_C, was used to search the genomes for the presence of the Spo0A_C sequence. tBLASTn performed the search- the amino acid sequence of the Spo0A_C domain was searched against the nucleotide sequence of the genomes. The presence of the sporulation signature genes in the genomes was determined by using tBLASTn- the amino acid sequences of the sporulation signature genes were searched against the nucleotide sequences of the genomes. An e-value cut-off of 1e-40 was used in both cases. The heatmap

displaying presence or absence of sporulation signature genes was made using the package ‘gplots’ in R.

2.6.4 Functional annotation of unique genes and genes in gene per genome analysis

Gene sequences were extracted from a homologous group, aligned using muscle [294] and a consensus sequence was created using the cons script described in EMBOSS [295]. Each of these homologous groups was then annotated by using the annotation already present in the genes that comprised the homologous group, by comparison against the well annotated *Clostridium difficile* CD630 genome and by searches against the Pfam and KEGG databases [110, 119, 272, 296]. To compare genes across different groups for similarity BLASTp using an e-value cutoff of 1e-20 was used [264]. Functional classes were manually assigned using the functional classification scheme developed for *E. coli* [297].

Chapter 3 Culturing of the human intestinal microbiota

3.1 Introduction

Our knowledge of the human intestinal microbiota and the role it plays in health and disease has greatly improved due to culture-independent, sequence-based approaches such as 16S rRNA gene amplicon sequencing and whole genome shotgun metagenomic sequencing [8, 15, 29]. While 16S rRNA gene amplicon sequencing is primarily a quantitative approach, metagenomic sequencing provides additional functional context. Despite the ability of these approaches to quickly generate large amounts of data and to reveal the composition of a microbial community, they are not without their drawbacks. 16S rRNA gene amplicon sequencing can reveal the taxa present to the genus level, but lacks the resolution needed to delineate species or strains. Inferences of functional capability can be made from the taxonomic information but this will be associated with a reference species and not with the strain in question [298]. Metagenomic sequencing is capable of sequencing the entire genetic repertoire of a sample and thus can provide both detailed compositional and functional information. However, while annotated reference databases continue to improve as the volume of sequence data increases, many sequenced genes are still of unknown function (hypothetical) [282, 299].

Many culture-independent studies of the intestinal microbiota are based upon comparison of sequenced data-sets from samples obtained from healthy and diseased individuals. From this, inferences about the role played by various microorganisms in health and disease can be made based upon the enrichment or depletion of a bacterial species in an individual. This approach is very informative and can identify

bacterial taxa that may be associated with the disease in question (enriched in disease-state compared to health-state) or that may protect against the disease (enriched in health-state compared to health-state). However, these approaches do not confirm causality and do not lead to a deeper mechanistic understanding of the underlying biology [247, 248]. Having the implicated bacterial species or strain as a pure culture allows progression from identification of a health or disease-associated species to confirming causality using animal models or laboratory studies. A recent study in adult mice focused on identifying bacteria responsible for conferring colonisation resistance against intestinal pathogens. Through a 16S rRNA gene amplicon sequencing approach, chloroform-resistant spore-forming members of the *Lachnospiraceae* and *Ruminococcaceae* families were identified [300]. These bacteria are not abundant in neonates and their absence may contribute to the high rate of pathogen infections which are not observed in adults. It is reasonable to assume a subset of these spore-forming bacteria are sufficient to restore colonisation resistance, however until the individual species are isolated and characterised, the identity of the bacteria in question and the underlying mechanism will remain unknown.

Culturing of bacteria therefore enables phenotypic characterisation to be carried out to validate results generated purely through genomics. Despite these benefits, culturing has inherent drawbacks compared to sequence-based approaches. It is a laborious process and ill-suited to studying the dynamics of an entire microbial community. The culturing of anaerobic bacteria also requires specialised equipment to maintain anaerobic atmospheric conditions, ideally using anaerobic cabinets if culturing is to be done at scale. However, when used in combination, culturing and genomics offers a powerful approach that can proceed from genomic analysis of an

entire community to characterisation of important individual species [77]. Furthermore, isolation of pure cultures also improves the resolution of reference genome datasets used by metagenomic approaches to assign taxonomic and functional information to a sequence [278].

While a large number of species have been cultured from the human gut since *Escherichia coli* was isolated over 130 years ago [26], microscopic analyses and more recently culture-independent, sequence-based approaches have revealed the complexity of the intestinal microbiota. This has contributed to the prevailing belief that the majority of these bacteria are unculturable. As a result of this, much of the underlying biology and phenotypes of the intestinal microbiota remains poorly understood. For example, enteric spore-forming bacteria can be pathogenic (such as *C. difficile*) or commensal (such as the colonisation resistance restoring bacteria discussed above) and the ability to form spores could aid in the transmission of oxygen-sensitive bacteria between human hosts [95, 131, 301]. Despite this, the extent of sporulation within the gut and the role it plays in the commensal intestinal microbiota remains unknown. In recent years, there have been a number of culturing studies that have isolated many novel taxa [79, 282]. As such, the issue is not that the majority of the intestinal microbiota is inherently unculturable but rather that perhaps, they have not yet been cultured [76, 77].

I therefore sought to establish a methodology to culture a representative selection of the human intestinal microbiota and to gain insights into the extent of sporulation (Figure 3.1). The workflow incorporates several steps, which are Culture, Re-streak, Archive and Phenotype (CRAP). Fresh faecal samples are left untreated or are treated to select for bacteria with a desired phenotype, in this case, sporulation. The stool sample is homogenised and serially diluted and then aliquots of the

homogenate are inoculated on YCFA agar to culture bacteria. Isolates are then identified by selecting single colonies that are streaked to purity and full-length 16S rRNA genes are amplified and sequenced. Each unique, novel and desired isolate is archived frozen in a culture collection and a whole-genome sequence is generated for each. Phenotypic characterisation of the isolates can then be carried out.

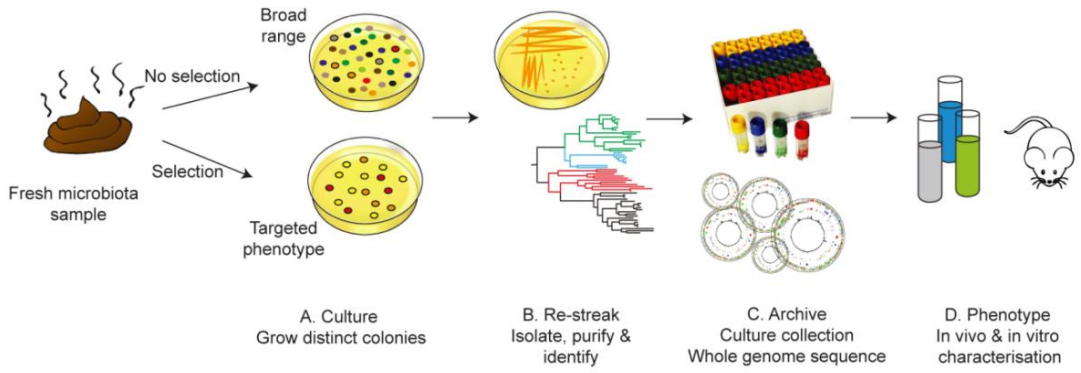


Figure 3.1 Schematic diagram of the anaerobic culturing workflow.

Schematic diagram of the culturing workflow, encompassing bacterial culturing and genomics to isolate and characterize bacterial species from the human intestinal microbiota.

3.2 Results

3.2.1 Establishment and assessment of culture process

Faecal samples were obtained from six healthy human donors. The aim was to culture health-associated intestinal bacteria, therefore, the donors initially completed a questionnaire which assessed their suitability and gave informed consent to participate in the study. To participate, donors should not have taken antibiotics in the six months prior to donation and should have no history of gastrointestinal

disorders. All six donors were adults and resident in the UK. Faecal samples were immediately processed upon receipt to ensure viability of anaerobic bacteria (see Materials and Methods). Faecal samples were either plated onto YCFA media (broad range culturing) or were immersed in ethanol to kill vegetative cells leaving bacterial spores (targeted phenotype culturing) (Figure 3.1) [302]. To enhance recovery, the spores were plated on YCFA media with sodium taurocholate, a known germinant for *C. difficile* [303].

The YCFA media was developed by the Rowett Institute in Aberdeen and was chosen as it is a broad-range media containing carbohydrates and fatty acids that are utilised by the intestinal microbiota (see Chapter 2- Materials and Methods for constituents) [75]. We first wished to assess how suitable YCFA was as a medium for culturing intestinal anaerobic bacteria. Accordingly, we compared the bacterial species in the original faecal sample to the bacterial species growing on the culture plates without selection.

DNA was extracted from both the faecal samples and the bacteria growing on the culture plate for all six donors. Metagenomic sequencing was carried out by the DNA pipelines department (core facility of the Wellcome Trust Sanger Institute). Sam Forster of the Host-Microbiota Interactions Laboratory added the sequence generated to the Human Pan-Microbes Community Database (HPMCD) and used Kraken to assign taxonomic information to the sequence. Following this, the relative abundances of the sequences were calculated (see Chapter 2- Materials and Methods for a full description). I then used the taxonomic and abundance information to perform the subsequent analysis described here. There was a strong correlation between the relative abundances of the taxa in the two samples indicating that bacteria growing on the culture plate were representative of the original faecal

sample (Figure 3.2). Therefore, the majority of intestinal bacteria, present in faecal samples can be isolated in culture if picked from the culture plate.

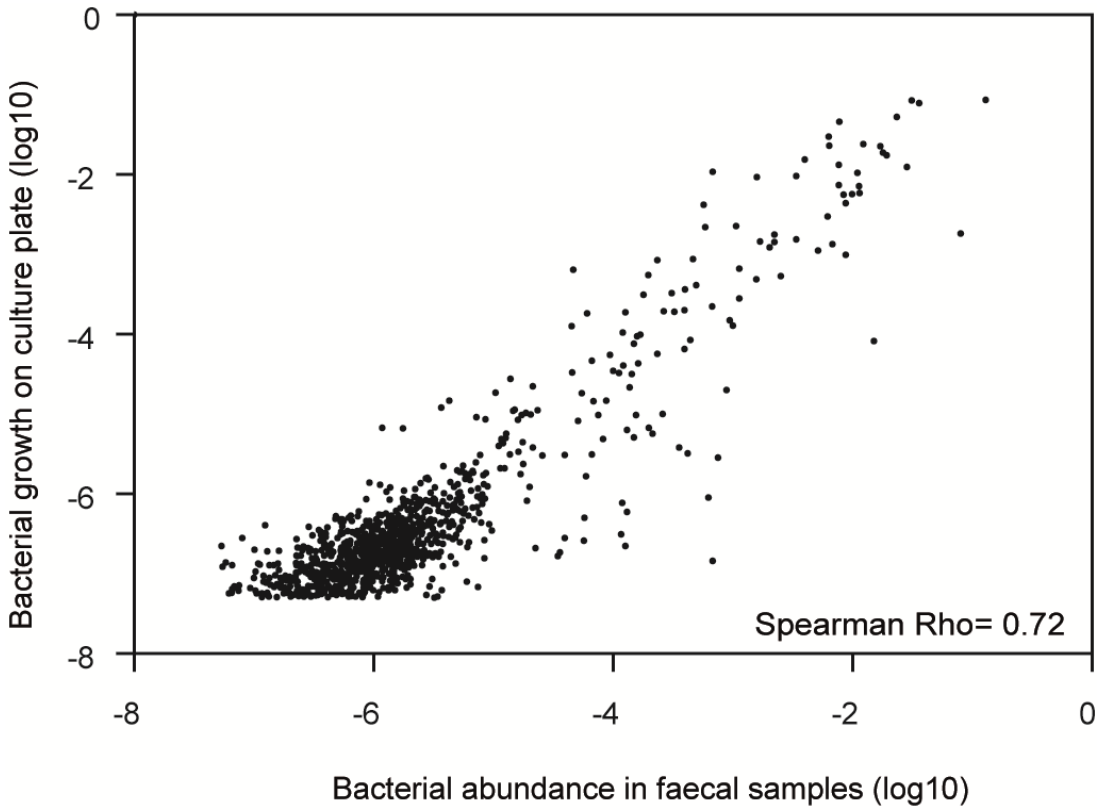


Figure 3.2 Bacteria grown on YCFA agar are representative of the complete faecal samples.

Relative abundance of bacteria in faecal samples (x axis) compared with relative abundance of bacteria growing on YCFA agar plates (y axis) as determined by metagenomic sequencing. Bacteria grown on YCFA agar are representative of the complete faecal samples as indicated by Spearman $\rho = 0.72$ (n = 6 donors).

To isolate spore-forming bacteria, faecal samples were immersed in 70 % ethanol for four hours to kill vegetative cells. During this time period, the immersed faecal samples were regularly vortexed. To assess the suitability of ethanol to isolate spore-forming bacteria, 16S rRNA gene amplicon sequencing of the original faecal sample,

the bacterial cultures recovered from the untreated culture plates (no ethanol treatment) and the ethanol treated plates was carried out. Principal Coordinates analysis (PCoA) of the 16S rRNA gene amplicon sequences showed a clear separation between sequences derived from ethanol treated samples and those of untreated samples and the original faecal samples (Figure 3.3).

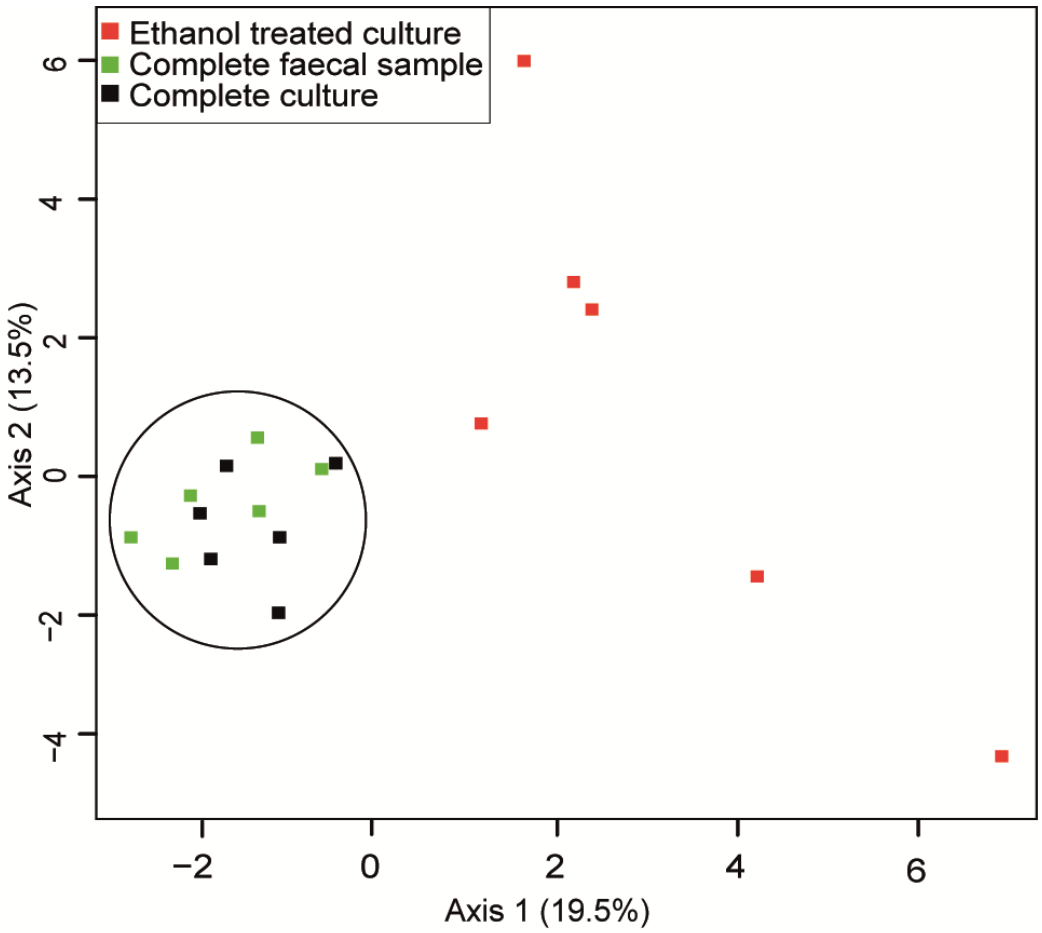


Figure 3.3 Ethanol selection separates spore-forming bacteria from non-spore-forming bacteria allowing their subsequent isolation.

Principal coordinates analysis plot of 16S rRNA gene sequences detected from six donor faecal samples (n = 6), representing bacteria in complete faecal samples (green), faecal bacterial colonies recovered from YCFA agar plates without ethanol pre-treatment (black) or with ethanol pre-treatment to select for ethanol-resistant spore-forming bacteria (red). Culturing without ethanol selection is representative of the complete faecal sample, ethanol treatment shifts the profile, enriching for ethanol-resistant spore-forming bacteria and allowing their subsequent isolation. The percentage of variation displayed by each axis is shown in brackets.

3.2.2 Extensive culturing and isolation of characterised and novel bacteria

Knowing that the culturing process using YCFA as the growth media would allow the growth of a representative proportion of the intestinal microbiota and that ethanol selection facilitated isolation of spore-forming bacteria, the samples were processed through the remainder of the CRAP protocol. Approximately 2000 colonies were picked from plates containing distinct, non-confluent colonies from the six donors. Colonies were picked from the original plate and streaked onto new plates (six colonies per petri dish) and received a unique identifier. PCR amplification of the full-length 16S rRNA gene was carried out, followed by capillary sequencing. The resulting sequences were then aligned using the Ribosomal Database Project (RDP) [304], edited in ARB [260] and clustered into Operational Taxonomic Units (OTUs) using mothur [261]. An isolate was considered a unique species using a 98.7% cut-off sequence identity over the full length of the 16S rRNA gene sequence [262, 263]. For each unique OTU identified, a representative isolate was selected and re-streaked on a full petri-dish to ensure purity. The RDP classifier was used to assign taxonomic information to the genus level and a BLASTn search defined an isolate as either a novel or previously characterised species [264]. All unique isolates were then archived as frozen stocks for future whole genome sequencing and phenotypic analysis.

In total, 137 unique species were isolated and archived from the six donors (Figure 3.4 and Appendix 1). 110 of these species were deposited in public culture collections (Appendix 2). 66 species were isolated from the ethanol-resistant culture condition and 71 from the untreated condition. 68 of the 137 species are novel and 69 are previously characterised. Thus, over 40 % of the species isolated from both

conditions had not been previously identified further undermining the notion that the intestinal microbiota is unculturable.

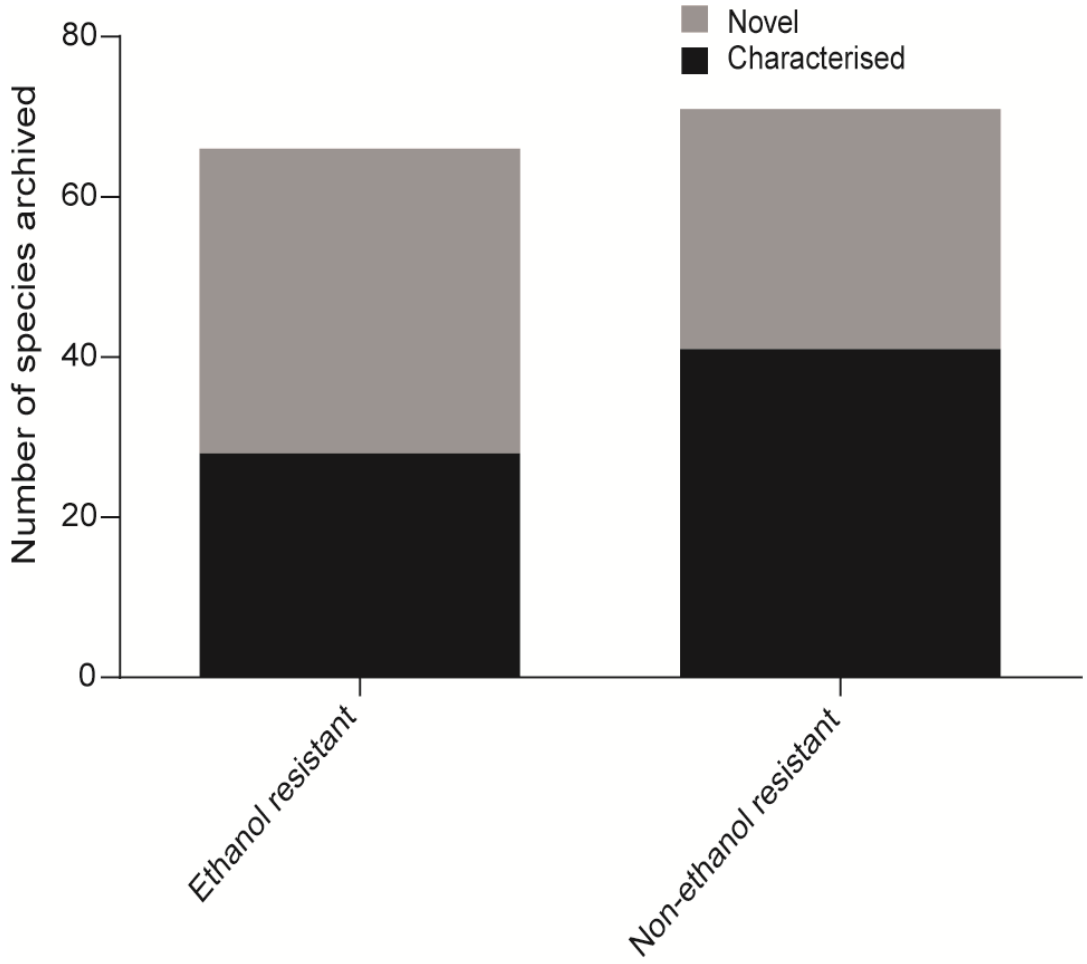


Figure 3.4 The number of bacterial species cultured in this study.

In total, approximately 2000 isolates were picked and 137 species were cultured. Over 40 % of bacteria isolated from both ethanol-resistant (38 out of 66) and non-ethanol-resistant (ethanol-sensitive) (30 out of 71) conditions were novel.

These species belong to the main phyla of the intestinal microbiota, the Firmicutes, Bacteroidetes and Actinobacteria (Figure 3.5 and Appendix 1). Proteobacteria were not isolated probably as they are not abundant in the intestinal microbiota and may

have been outcompeted on the culture plate by obligate anaerobes. Of the 68 novel species isolated, 45 are characterised as members of novel genera and a further three are classified as members of novel families by RDP (two species belong to the same novel family). The novelty captured is agnostic of taxonomy as there are novel isolates present in all of the families indicated on the phylogeny which includes the major families of the intestinal microbiota, the *Bacteroidaceae, Ruminococcaceae* and the *Lachnospiraceae.* A major aim of the HMP was to create a catalogue of reference genomes from human-associated bacteria. As part of this process they defined a list of 'most wanted' bacteria which represent novel taxa or taxa with few cultured representatives based upon 16S rRNA gene amplicon sequencing studies [78, 305]. Of the 137 species cultured here, 90 are part of the HMP's most wanted list (Appendix 1).

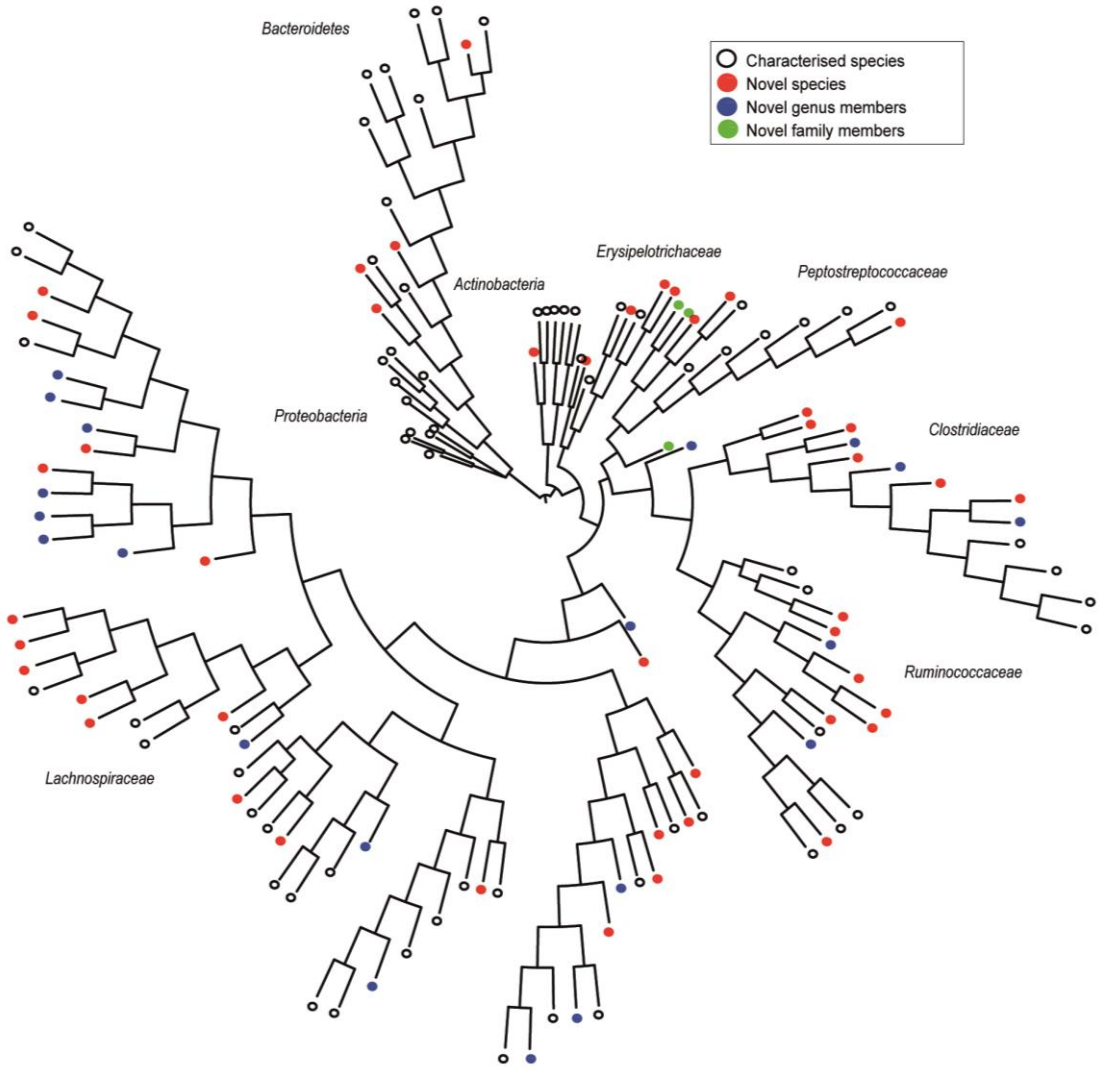


Figure 3.5 Phylogenetic tree of bacteria cultured from the six donors constructed from full-length 16S rRNA gene sequences.

137 bacterial species were cultured in total. Novel candidate species (red), genera (blue) and families (green) are shown by dot colours. Major phyla and family names are indicated. Proteobacteria were not cultured, but are included for context.

Each cultured and isolated species was whole genome sequenced, assembled and annotated by the DNA pipelines department and by bioinformatics pipelines

developed by the Pathogen Informatics team at the WTSI. The assembled whole genome sequences were added to the previously described Kraken database created by Sam Forster and this information was used to taxonomically classify the faecal-derived metagenomic sequences. The relative proportions of the taxonomically classified metagenomic sequence were again calculated by Sam Forster [280]. I then used this database to assess the proportions of bacterial genera and species in each of the six donors. Based on the average relative abundance across the six donors 96 % of the bacterial abundance at the genus level was cultured (Figure 3.6) and 90 % of the bacterial abundance at the species level was cultured (Figure 3.7).

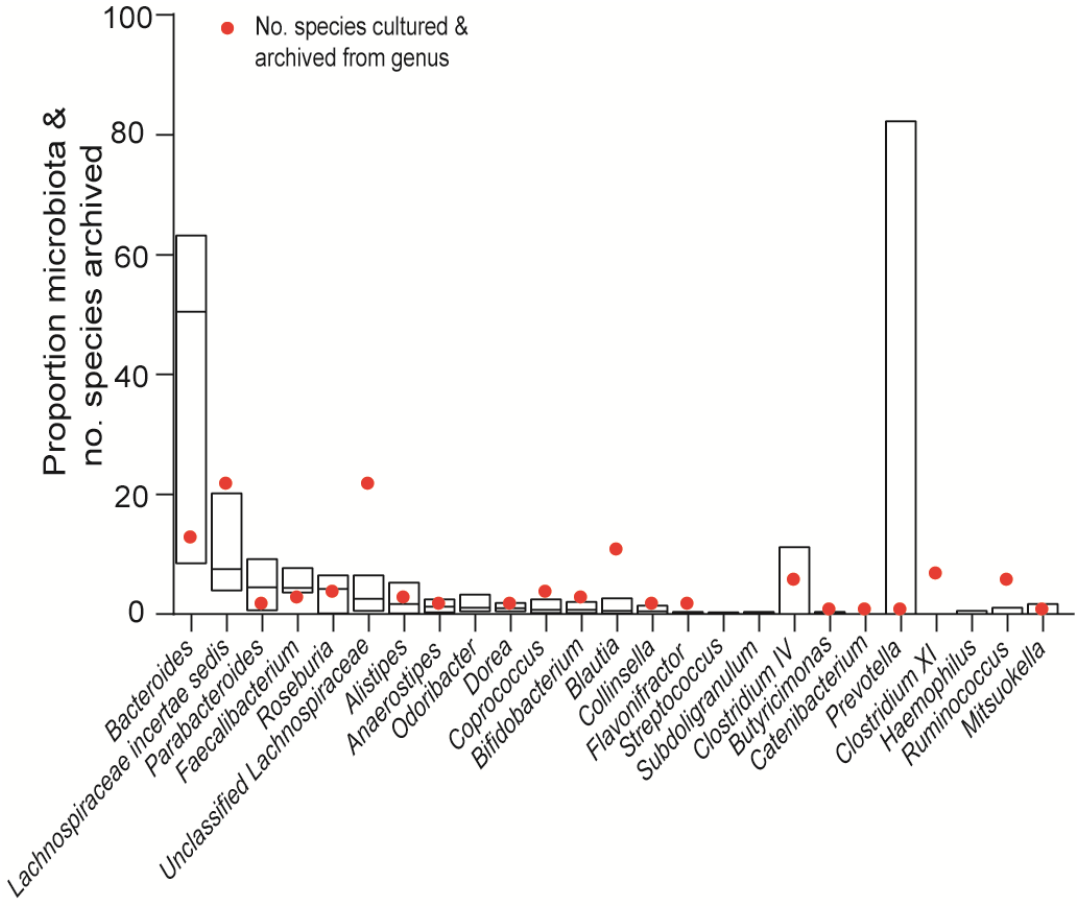


Figure 3.6 Archiving of bacterial diversity and novelty through anaerobic culturing- most abundant genera

Representative species from 21 of the 25 most abundant bacterial genera were isolated and archived (abundance was determined by metagenomic sequencing and based on average relative abundance across the six donors (n = 6)). This represents 96 % of the average relative abundance at the genus level across the six donors. A red dot indicates the number of species archived from each genus. *Lachnospiraceae incertae sedis*, unclassified *Lachnospiraceae*, *Clostridium* IV and *Clostridium* XI are not strict genera and represent currently unclassified species. Median and range is presented for the above with taxa ranked by median.

The most abundant genera in the six donors represent common genera typically found in other cohorts derived from Western populations, including *Bacteroides*,

Faecalibacterium, *Roseburia* and *Bifidobacterium* [15, 306]. *Prevotella* was highly abundant in one donor, hence the wide range present in Figure 3.6 and Figure 3.7. The most abundant species across the six donors are also representative of larger populations with species such as *Bacteroides vulgatus*, *Faecalibacterium prasunitzii* and *Roseburia faecis* prevalent and abundant in the gut microbiota of Western individuals [77].

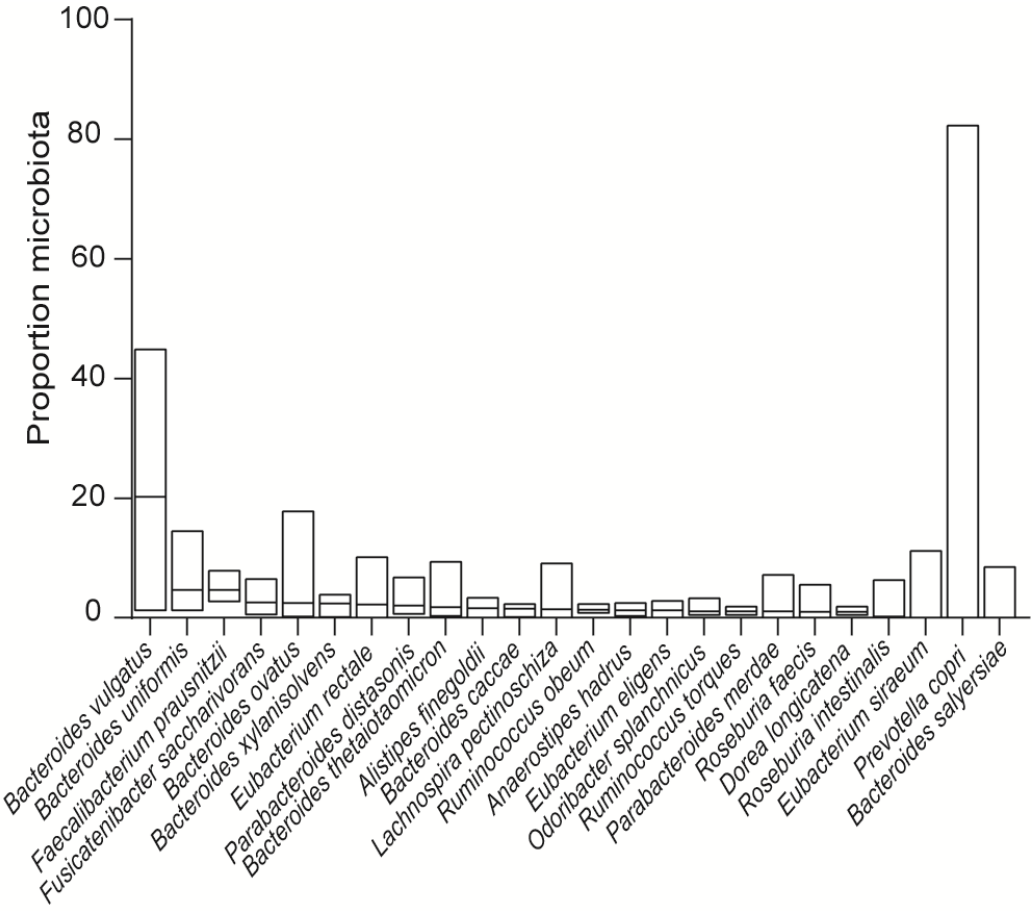


Figure 3.7 Archiving of bacterial diversity and novelty through anaerobic culturing- most abundant species

Representative species from 23 of the 24 most abundant species were isolated and archived (abundance was determined by metagenomic sequencing and based on average relative abundance across the six donors (n = 6)). This represents 90 % of the average relative abundance at the species level across the six donors. *Odoribacter splanchnicus* was the only species not archived. Median and range is presented for the above with taxa ranked by median value.

Genera that were present at low average relative abundance (<0.1 %) were also isolated (Figure 3.8). These are not restricted to any readily-culturable taxonomic groups but are distributed across different bacterial classes. For example, *Turicibacter* and *Erysipelotrichaceae incertae sedis* belong to the *Erysipelotrichia*

class, *Clostridium sensu stricto*, *Oscollibacter* and *Sarcina* belong to the *Clostridia* class, while *Megasphaera* belongs to the *Negativicutes* class.

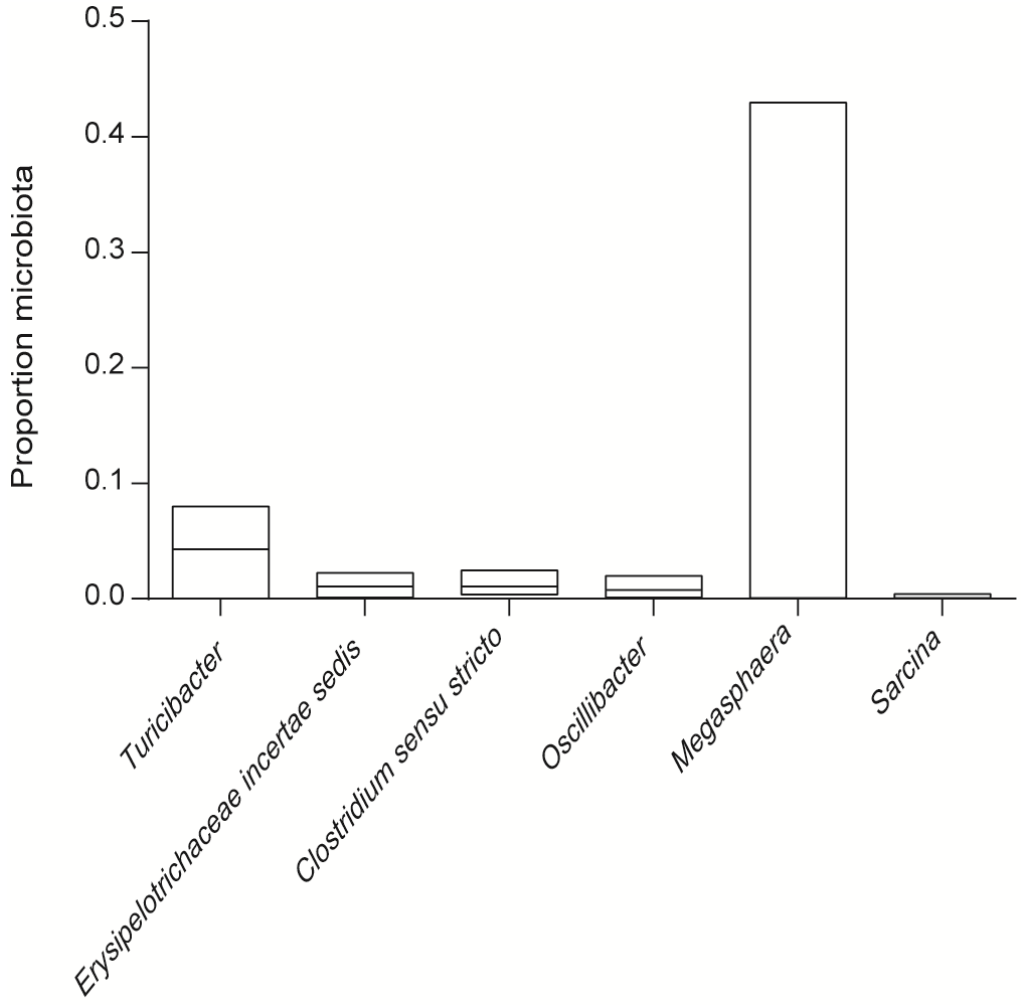


Figure 3.8 Archiving of lowly represented members of the intestinal microbiota

Lowly represented intestinal microbiota members were also cultured from the six donors. At least one representative species from each of the genera presented were cultured. Median and range is presented for the above with taxa ranked by median value.

3.2.3 Phylogenetic analysis of an isolate from a putative novel family

Next, one of the isolates belonging to a novel taxonomic family, *Clostridium thermocellum*_86% was examined in more detail (86 % denotes the 16S rRNA gene nucleotide sequence similarity to the nearest characterised species, in this case, *Clostridium thermocellum*). This isolate is putatively named here as *Falkowia sangerensis*, in honour of Professor Stanley Falkow of Stanford University for his pioneering work on molecular microbial pathogenesis and the Wellcome Trust Sanger Institute where the isolate was first cultured and described. A BLASTn search of the full length 16S rRNA gene sequence revealed the nearest characterised species to *F. sangerensis* was *Christensenella minuta* strain YIT (accession no. NR_112900) with 88 % nucleotide sequence identity over 97 % query coverage. The next nearest characterised species was *Catabacter hongkongensis* strain JCM 17853 (accession no. AB671763) with 88 % sequence identity over 97 % sequence coverage). *C. hongkongenesis* has been implicated in sepsis, however like *C. minuta*, its natural environment is believed to be the human gut [307]. Both *C. hongkongenesis* and *C. minuta* are characterised as species belonging to two different families, however, previous reports suggest that they may belong to the same family [26]. Using full-length 16S rRNA gene sequence, Yarza *et al.* defined the cut-off for defining a novel family as 86.5 % by examination of over 200 taxa [308]. This places *F. sangerensis* at the threshold for defining a new family based on 16S rRNA gene sequence comparison which is putatively name here *Falkowiaceae*. This is the most stringent cut-off with other reports placing the cut-off for novel families at 95 % [309].

Phylogenetic analysis based on full length 16S rRNA gene sequence established that this isolate does form a distinct clade, separate from other characterised families, that

resides within the *Clostridiales* order of the Firmicutes phylum (Figure 3.9). The *Christensenellaceae* and *Catabacteriaceae* families do cluster together and may indeed be part of the same family but they are distinct from *Falkowiaceae*. Also distinct from *Falkowiaceae* is the *Gracilibacteraceae* family which contains isolates of environmental origin [310]. Similar 16S rRNA gene sequences from uncultured bacteria were identified which cluster with *F. sangerensis* which are gastrointestinal in origin. While further phenotypic analysis would be required to properly characterise this isolate, comparison of the 16S rRNA gene sequence and phylogenetic analysis does indicate *F. sangerensis* is a member of a novel family.

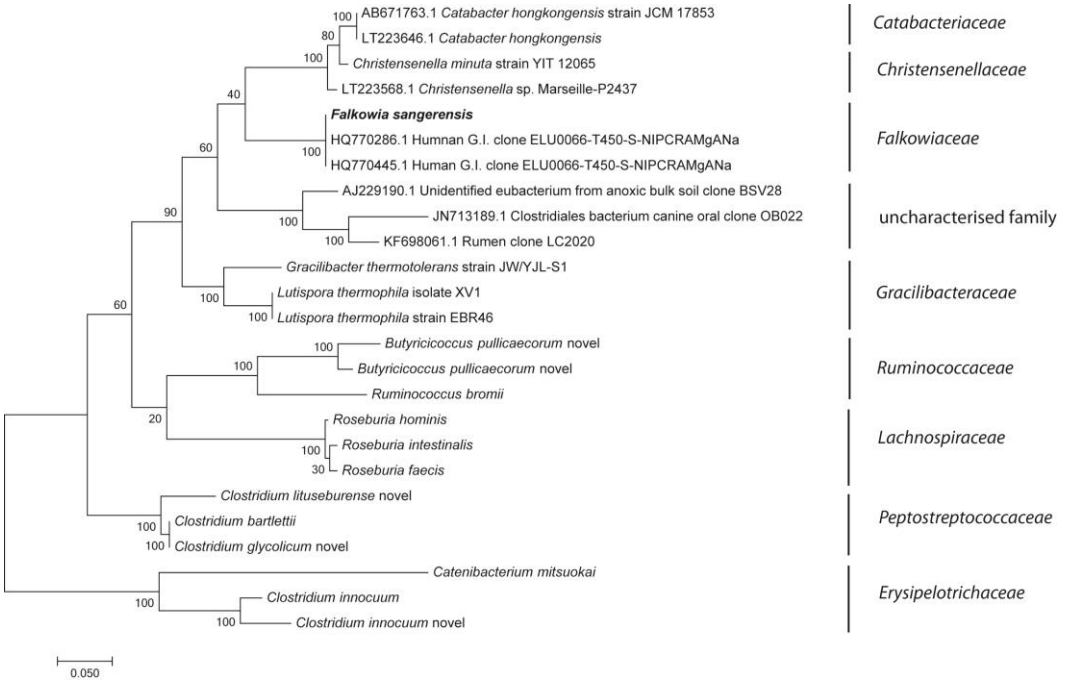


Figure 3.9 Phylogenetic tree of putatively named novel *Falkowia sangerensis* isolate and closely related species based on 16S rRNA gene sequence

To provide context, other species from human intestinal microbiota families were included (*Lachnospiraceae*, *Ruminococcaceae*, *Peptostreptococcaceae* and *Erysipelotrichaceae*) in addition to the isolates identified as the closest relatives based on a BLASTn search. Taxonomic families are indicated on right hand side of the phylogenetic tree. Bootstrap values are present where indicated.

3.2.4 Comparison with other culturing studies

To put the extent of the culturing into context I compared my results with other recent studies that have cultured from the human intestinal microbiota. In this comparison, I considered the number of whole genome sequences generated. Ease of reproduction is also important, while culturing of enteric anaerobic bacteria does require anaerobic culturing facilities; a streamlined approach as chosen here by using

one growth medium will alleviate the work required. The isolates cultured in this study were compared with other large-scale culturing studies [79, 80, 282] (Table 3.1). The two Lagier *et al.* studies cultured more characterised species than this study but they used 70-200 different culture conditions including different growth media, growth temperatures and filtration steps. In comparison, I used one growth medium (YCFA) combined with and without an ethanol treatment step to enrich for spore-forming bacteria. Despite the range of culture conditions used in the Lagier *et al* 2012 study they did not culture more novel species than obtained here. They used matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) to identify known species, and any species unidentified by the MALDI-TOF were subjected to 16S rRNA gene sequencing using the same species level cut-off of 98.7%. Selection for spore-forming bacteria was included in their culturing protocol so the reason for the low number of novel bacteria cultured is not clear. Two of the three stool samples were frozen at -20 ˚C and then -80 ˚C before culturing commenced, hence some loss of bacterial viability may have occurred during this period. In summary, the culturing results reported here compare favourably with other studies, especially when considering the number of species cultured per culture conditions. Using one growth medium and two culture conditions can isolate a large number of bacterial species, including many that are novel.

Culturing study	No. characterised species cultured	No. novel species cultured	No. samples	No. culture conditions	No. whole genome sequences generated	No. isolates deposited in public culture collections	No. species isolated per culture condition
This study	69	68	6	2	137	110	68.5
Goodman *et al.* 2011 [79]	48	15	1	1	0	0	63
Lagier *et al.* 2012 [282]	309	31	2	212	31	0	1.6
Lagier *et al.* 2016 [80]	860	197	973	70	197	197	15.1

Table 3.1 Comparison of culturing studies

Culturing results of this study was compared to other recent culturing studies (references are in culturing study column). The number of species isolated per culture condition column reflects how streamlined the culturing was in terms of number of culture conditions used. For this study ethanol treated and non-ethanol treated samples count as two culture conditions.

3.3 Discussion

In this study I utilised a streamlined approach to culture, isolate, archive and whole genome sequence 137 characterised and novel bacteria from the human gastrointestinal tract. The bacteria cultured are representative of the samples they were derived from and consist of species present at high and low abundances in the gut. Incorporated in the culturing workflow was a "targeted phenotypic" screen

designed to select for ethanol-resistant spore-forming bacteria, from which 66 characterised and novel bacterial species were isolated. Using one of the novel species isolated, a phylogenetic approach determined that this isolate is a member of a novel bacterial family. This demonstrates the ability of this approach to culture and isolate highly novel members of the human intestinal microbiota.

While other studies have utilised a large number of culture conditions and media to isolate as many species as possible [80, 282], the focus here was to design a streamlined workflow that was effective (Figure 3.1). Hence, one broad culture medium was used that would support the growth of bacteria with different nutritional requirements. The combination of this broad range culture media with the phenotypic selection of spore-forming bacteria contributed to the large number of novel species isolated. There were more novel bacterial species isolated from the ethanol-resistant condition compared to the non-ethanol-resistant condition, highlighting the value of a targeted approach to enrich for bacteria that may be less abundant and therefore less likely to be picked from a plate under normal circumstances. The isolation of bacteria displaying an ethanol-resistant phenotype allows for subsequent *in vitro* characterisation and genotypic analysis which will be explored in Chapter 4.

The novelty captured here extended to novel bacteria genera and even novel bacterial families. In this study, the RDP classifier was chosen to classify sequences to the genus level and comparison of full-length 16S rRNA sequence was used to classify to the species level. Taxonomic classification of bacteria has always been inherently problematic. For example the traditional assignment of anaerobic spore-forming bacteria to the genus *Clostridium* has resulted in a large number of bacterial species with the *Clostridium* prefix that now reside in different genera or families [88].

Correcting the nomenclature by renaming these species is difficult as new names are often not adopted by the scientific community [89]. The isolation, whole genome sequencing and archival of novel bacterial species can help improve taxonomic assignments.

First, it helps to fill in the taxonomic gaps between species that have been already characterised. As more species are added to the *Christensellaceae* and *Catabacteriaceae* families, it may become clear that they should merged into a single family. *C. minuta* was only characterised in 2012 as the type species of a novel family isolated from the human gut [311]. Only three years later its abundance in people was shown to be influenced by human genetics, highlighting the value of culturing and characterising novel bacteria [38]. Second, it clarifies the boundaries of a particular taxon. As more species are added to the *Falkowenciae* family, a clearer picture will emerge of the extent of this family and the genera contained within. Third, whole genome sequencing and archival of cultured bacteria allows analysis to be carried out at the genome and phenotype level, this adds further granularity to the defining characteristics of a given taxon.

The majority of the isolates cultured were deposited in public culture collections. The deposition of isolates from this and other large-scale culturing studies [80] (Table 3.1) provides a valuable resource for other members of the scientific community to access. In addition, the whole genome sequences generated from these isolates improved the resolution of the Kraken database allowing detailed identification of metagenomic sequences. Ultimately, the novelty archived here provides a means to move from a sequence-based, quantitative approach to proceed to a more phenotypic and mechanistic understanding of this microbial community

[312]. In the next chapter, using the bacteria cultured here, the sporulation phenotype will be examined in the context of inter-host transmission and colonisation.

Chapter 4 Characterisation of intestinal spore-forming bacteria in the context of inter-host transmission

4.1 Introduction

Assembly of the intestinal microbiota begins at birth and develops rapidly during the first three years of life, at which point a stable microbial community is attained [22, 24, 25, 163]. Individuals living in close contact have a more similar microbiota than more removed individuals which indicates transmission and acquisition of new species occurs beyond childhood throughout life [167, 168, 170]. As the majority of the intestinal microbiota cannot tolerate extended periods in aerobic conditions, transmission between individuals in close contact is more likely [73, 148, 170]. A colonising bacterium in this scenario is more likely to be quickly ingested by a new host in close proximity. Regardless of proximity to new hosts and the length of time exposed to aerobic conditions, any means to prolong viability outside of a host will be advantageous. As discussed in Chapter 1 (Introduction) some of these survival mechanisms could include aerotolerance, spore-formation and other dormancy mechanisms. Transmission of bacteria between human hosts is best understood during birth, however the extent of transmission between older individuals is poorly understood as are the mechanisms of how this could occur.

C. difficile utilises resilient spores that can tolerate disinfectants and aerobic conditions to survive outside of a host for extended time periods [291]. In addition to promoting external survival, *C. difficile* spores can recognise intestinal bile acids once ingested by a new host [122]. This acts as a signal that the spores are back in

the intestinal environment which triggers germination and subsequent colonisation [116]. In the previous chapter I cultured and isolated 66 spore-forming species which constituted nearly half of the bacterial species cultured (66 out of 137). I hypothesise that the enteric commensal spore-formers isolated may exhibit similar transmission and colonisation patterns to *C. difficile.* If true, this would provide a means for oxygen-sensitive bacteria to survive outside of a host and to colonise a new host.

In this chapter I place the large number of spore-formers cultured in a phylogenetic context to ascertain the extent of the sporulation phenotype in the gut. Then, using the well-studied *C. difficile* as a comparator, I examine the ability of these enteric spore-formers to survive in external aerobic conditions and to recognise the presence of intestinal bile-acids, both of which are necessary to ensure transmission and colonisation. Examination of publicly available metagenomic data-sets is then used to ascertain the extent of intestinal spore-forming bacteria within individuals.

4.2 Results

4.2.1 Phylogenetic analysis of cultured ethanol-resistant spore-forming bacteria

Ethanol-resistant bacteria were isolated from five characterised taxonomic families that belong to two taxonomic classes. These are the *Erysipelotrichaceae* family within the *Erysipelotrichia* class, and the *Clostridiaceae, Peptostreptococcaceae, Ruminococcaceae* and *Lachnospiraceae* families all within the *Clostridia* class. Ethanol-resistant bacteria were also isolated from two putative novel families (Figure 4.1). Some of the bacteria displaying an ethanol-resistant phenotype include species previously classified as non-spore-formers such as *Turicibacter sanguinis* within the *Erysipelotrichaceae* family [313]. Other ethanol-resistant bacteria are closely related to non-spore-forming bacteria. For example, novel bacteria closely related to non-

spore-forming members of the *Roseburia* genus were isolated [11, 74]. The *Roseburia* genus contains species abundant within the intestinal microbiota that are known butyrate producers. Butyrate is a SCFA that acts not only as an energy source for epithelial cells, but also has anti-inflammatory properties [11]. Finally, other ethanol-resistant bacteria have been suspected of making spores but until now have never been demonstrated to do so. These include *Eubacterium eligens, Eubacterium rectale* and *Coprococcus comes* [106]. Traditionally, spore-forming bacteria are associated with a rod-like bacillus morphology such as *Bacillus subtilis* or *C. difficile.* The TEM images, which were generated by David Goulding of the WTSI, show ethanol-resistant bacteria with the typical rod morphology in addition to cocci or variants of cocci such as *Coprococcus eutactus.* Hence, the sporulation phenotype is phylogenetically diverse within the Firmicutes and is morphologically heterogeneous.

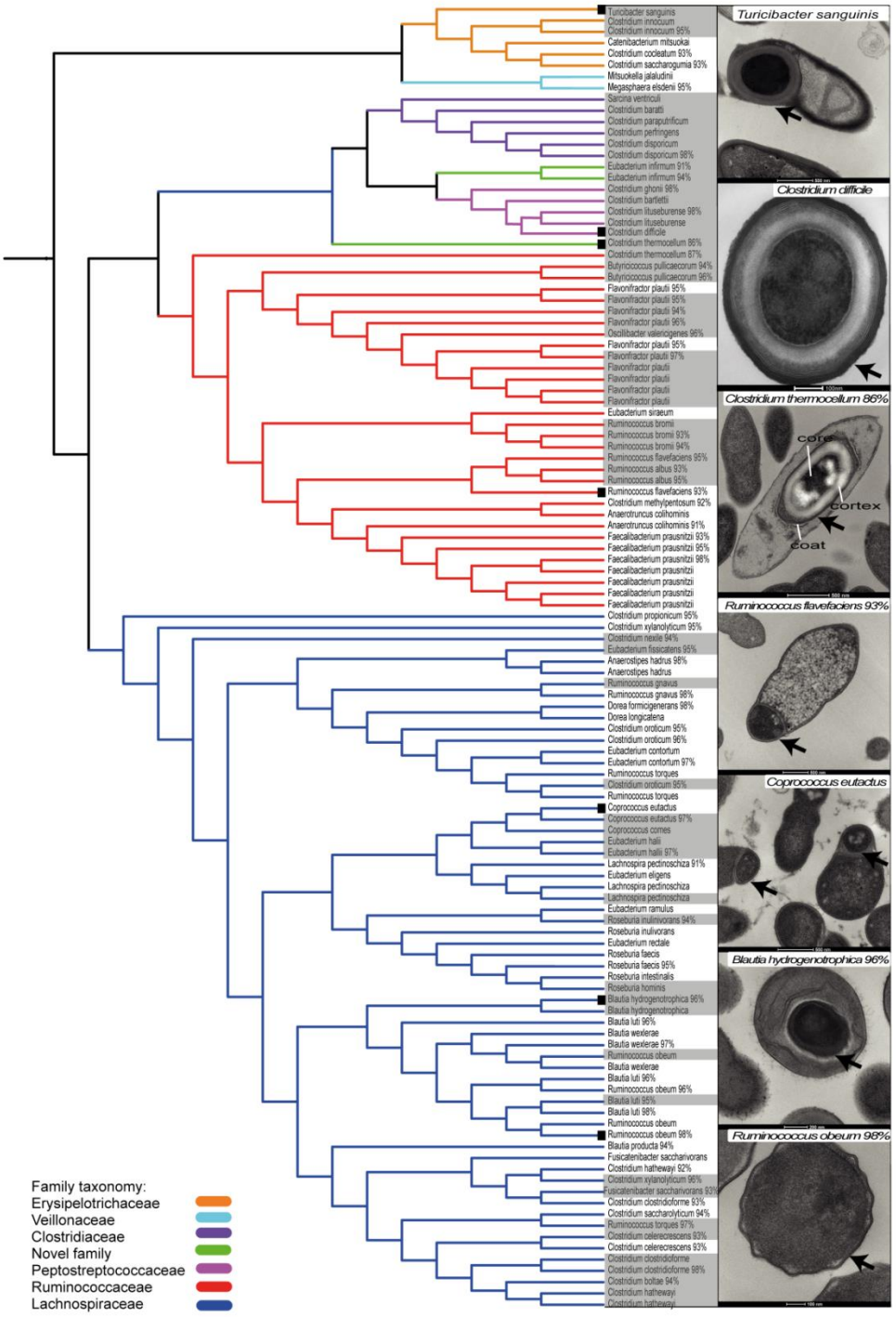


Figure 4.1 Phylogeny of intestinal spore-forming bacteria

Full length 16S rRNA gene sequence phylogeny illustrating the taxonomic relationship of ethanol-resistant bacteria within the Firmicutes cultured from the donor faecal samples. Branch colours indicate distinct families. Shaded text indicates species cultured from an ethanol-treated faecal sample and unshaded

text indicates species cultured from a non-ethanol-treated faecal sample. Percentage values represent closest identity to a characterized species. Transmission electron micrographs (TEMs) of spore ultrastructures for a phylogenetically diverse selection of cultured bacteria are shown with an arrow in images and include a candidate novel family with 86 % identity to the 16S rRNA gene sequence from *Clostridium thermocellum* which is the isolate putatively named as *Falkowia sangerensis* in Chapter 3. Typical spore structures are defined and illustrated in the same image. TEMs are ordered according to boxes next to the species name. Scale bars are shown at the bottom of each image. *C. difficile* is included for context and was not cultured in this study. TEM Images were generated by David Goulding (WTSI).

4.2.2 Phenotypic characterisation of intestinal spore-forming bacteria

I next wanted to assess if the ability to produce ethanol-resistant spores would provide an advantage for transmission. To successfully transmit between hosts, a bacterial species must exit from a host in faecal matter, and then survive adverse environmental conditions which will potentially include disinfectants used in cleaning regimes and the presence of oxygen before being ingested by a new host [1]. Accordingly, the number of plated spores that germinated following exposure to a disinfectant was measured (colony forming units quantified on culture plates after ethanol exposure was used as a proxy for germinated spores). Ethanol was chosen as a disinfectant as it is readily available and widely used in hand sanitisers. A taxonomically diverse range of spore-forming and non-spore-forming bacteria cultured in the previous chapter (in addition to *C. difficile* and *E. coli*) were exposed to a 70 % ethanol solution for four hours before being washed and plated under anaerobic conditions. This concentration of ethanol is within the range considered effective and recommended by the Food and Drug Administration (FDA) for hand sanitisers [314]. Only spore-forming bacteria survived prolonged exposure to ethanol

(Figure 4.2). The non-spore-forming *Bacteroides uniformis*, *Bifidobacterium bifidum* and *E. coli* all failed to grow post exposure.

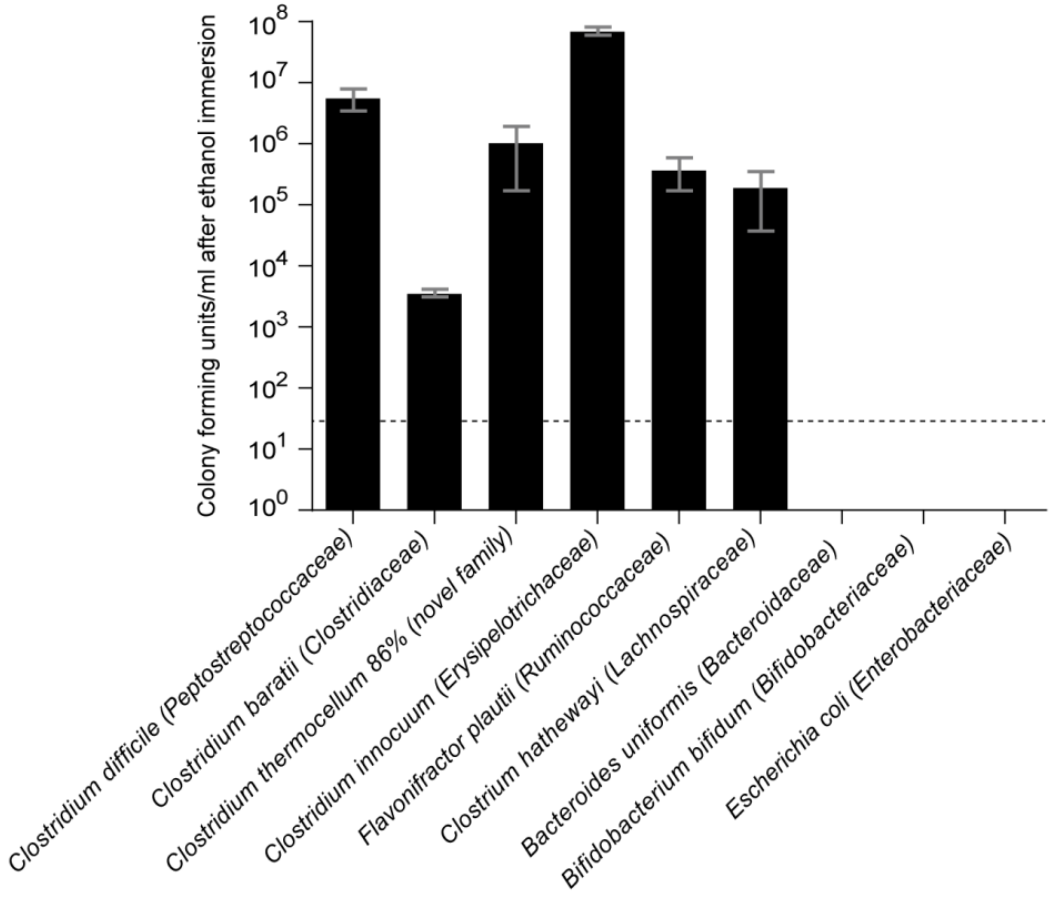


Figure 4.2 Spore-forming bacteria are more resilient than non-spore-forming bacteria to environmental stresses such as disinfectants.

Pure bacterial cultures were immersed in ethanol for 4 hours before being washed and inoculated onto YCFA growth medium with sodium taurocholate as a germinant. Only spore-forming bacteria survived. Taxonomic family names are shown in brackets. The dashed line indicates the culture detection limit of 50 CFU ml^{-1}. Mean ± s.d., n = 3 biological replicates for each species tested.

Next, the ability of spore-forming bacteria to survive exposure to aerobic conditions was tested. Again, a taxonomically diverse range of spore-forming and non-spore-forming bacteria cultured in the previous chapter, in addition to *C. difficile* and *E. coli* were utilised. These bacteria were inoculated onto media plates in different concentrations as spot dilutions and were then exposed to ambient aerobic and temperature conditions on the laboratory bench for varying time periods. They were then returned to anaerobic conditions. The number of colonies recovered for each time point were counted and compared as a percentage to a control group which was not exposed to aerobic conditions. The spore-forming bacteria survived to the end of the experiment which was 504 hours (21 days) (Figure 4.3). The non-spore-forming bacteria survived a maximum of 144 hours (6 days) apart from *E. coli* which is a facultative anaerobe. Hence, anaerobic enteric spore-forming bacteria are able to survive for longer in aerobic conditions than anaerobic enteric non-spore-forming bacteria. Interestingly, for many of the spore-forming bacteria, a sudden decrease in the percentage of viable cells recovered can be observed between 24 and 96 hours, after which the percentage of viable cells recovered remains constant. This could represent the point at which the vegetative cells of these spore-formers died. Any spores present would remain dormant in aerobic conditions. Once returned to anaerobic conditions, the spores would germinate, forming vegetative cells and allowing a colony count to be made.

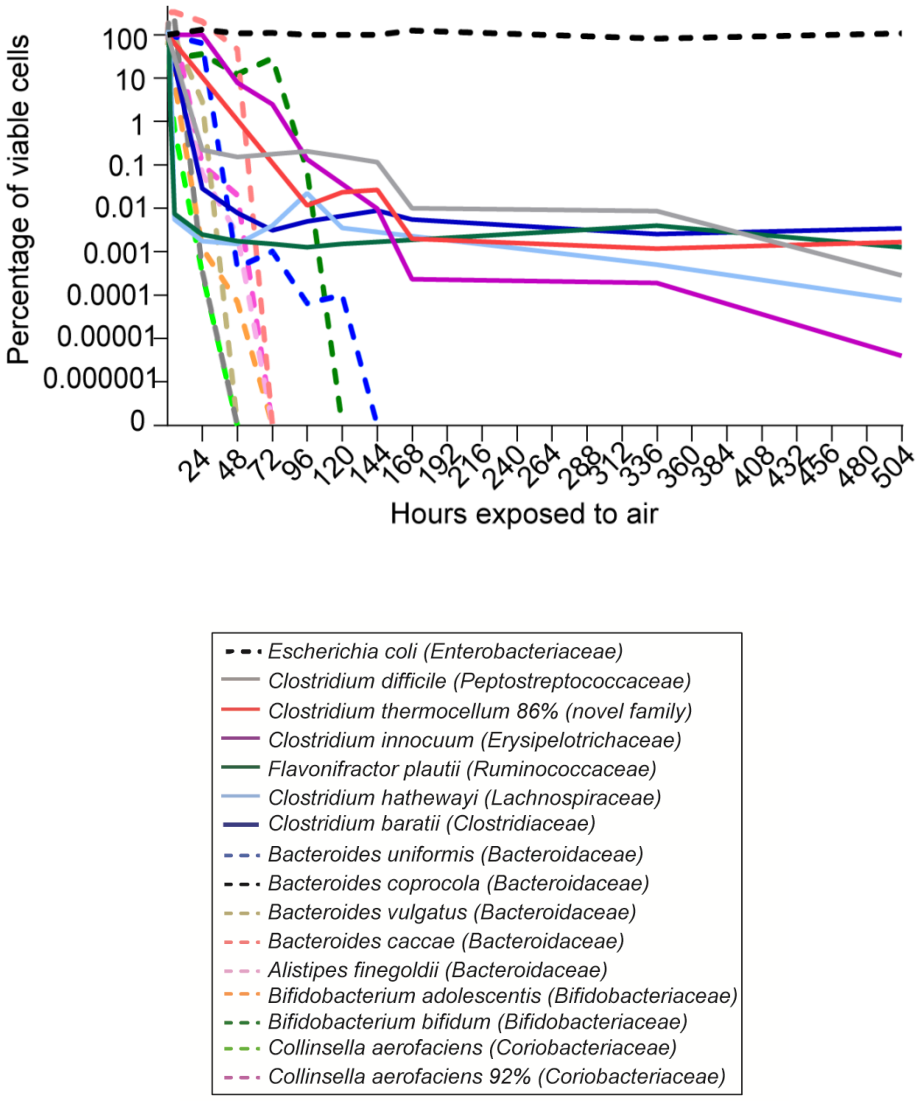


Figure 4.3 Oxygen tolerance of phylogenetically diverse intestinal spore-forming bacteria

Once exposed to oxygen, only 1 % of the original inoculum of non-spore-forming bacteria (dashed lines) were viable after 96 hours (4 days) and none were viable after 144 hours (6 days). Spore-forming bacteria (solid lines) persist owing to spore formation. The experiment was stopped after 504 hours (21 days). Taxonomic families of each species tested are shown in brackets (n = 3 biological replicates for each strain).

The ability to survive oxygen exposure and disinfectants such as ethanol enables enteric spores to survive in the external environment. Once ingested by a new host the spores need to recognise the intestinal environment in order to germinate and commence colonisation. I next tested if the spores utilise intestinal bile-acids as a germinant in a manner similar to that used by *C. difficile*. A selection of spore-forming and non-spore-forming bacteria were subjected to an ethanol shock treatment and were then plated on media with the common bile acids taurocholate, glycocholate and cholate. Taurocholate was extremely effective at increasing the number of colonies recovered following plating, termed 'germination potency' in Figure 4.4. Taurocholate in the media increased the number of colonies recovered by between 8 and 70,000 fold. Glycocholate and cholate were not as effective as taurocholate but did significantly increase the germination potency for *Clostridium innocuum.* Non-spore-forming bacteria did not survive the ethanol shock treatment and are not presented in Figure 4.4.

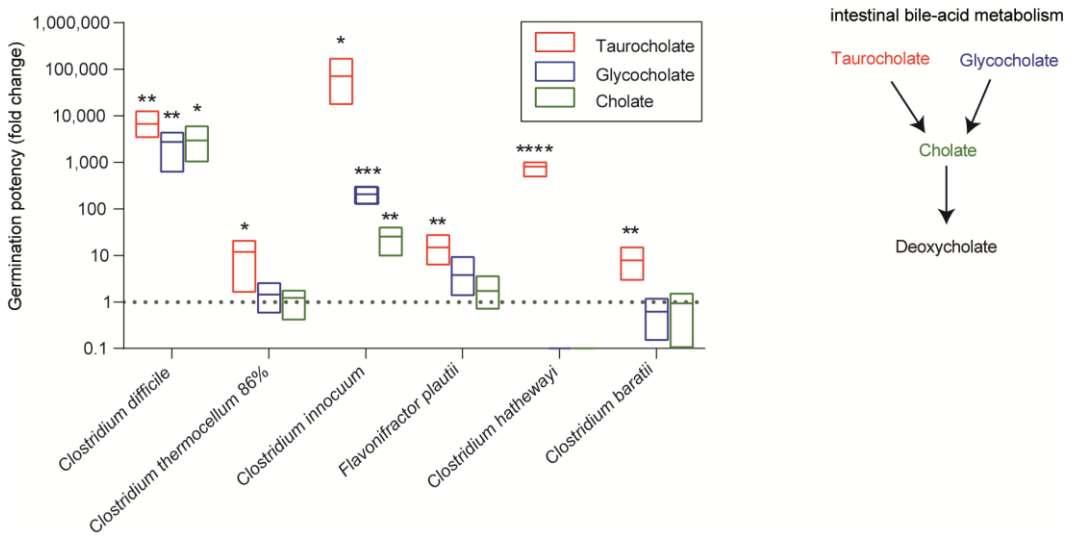


Figure 4.4 Intestinal spore-formers respond to bile-acid germinants.

The number of colony-forming units (CFU) (representing germinated spores) present on plates in the presence of a particular germinant is expressed as a fold change with respect to the number of CFU recovered on plates in the absence of a germinant. Spore-formers and non-spore-formers were subjected to ethanol shock before being plated (n = 6 biological replicates for each strain). A fold change of one (dashed line) would indicate that a germinant had no effect on the number of CFU recovered. Schematic summarises the cholate-derived bile acid metabolism in the mammalian intestine. Mean and range, Welch's unpaired two-tailed t-test (*P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001).

Next, the non-spore-forming bacteria *Collinsella aerofaciens* and *Bacteroides uniformis* were plated on media containing the same intestinal bile acids but were not subjected to ethanol shock treatment beforehand. No significant difference in the germination potency was observed for any of the three bile-acids (Figure 4.5). Hence, the response to intestinal bile-acids is unique to spore-forming bacteria and is based upon germination of spores following bile-acid recognition.

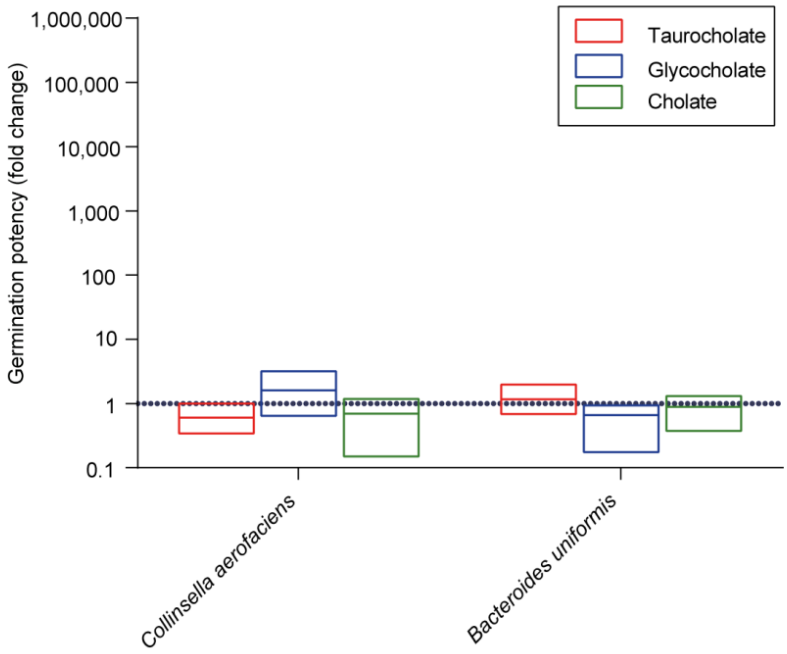


Figure 4.5 Intestinal non-spore-formers do not respond to bile-acid germinants.

The number of CFU present on plates in the presence of a particular germinant expressed as a fold change with respect to the number of CFU present on plates in the absence of a germinant. No ethanol shock treatment was performed beforehand. A fold change of one (dashed line) would indicate that a germinant had no effect on the number of CFU recovered from the bacteria. Mean and range, n = 3 biological replicates for both species.

4.2.3 Genomic analysis of intestinal spore-forming bacteria and sporulation dynamics

Next, the genotypic pathways that underlie these resistant phenotypes were examined. In total 234 ethanol-resistant and ethanol-sensitive bacteria were cultured. These were whole genome sequenced using the DNA pipeline department of the WTSI and assembled and annotated using bioinformatics pipelines developed by the

Pathogen Informatics team at the WTSI. Sam Forster of the Host-Microbiota Interactions Laboratory then used a machine learning approach to identify from the 694,300 genes in this dataset, a list of 66 genes that were enriched in the ethanol-resistant bacteria compared to the ethanol-sensitive bacteria (see Chapter 2 Material and Methods for details). I then used this sporulation signature for the subsequent analysis described here. Unlike previous studies that relied on prior biological assumptions, this gene list was unbiased in its approach [95, 112, 315]. This sporulation signature is enriched with known sporulation-associated genes from stages I–V of the spore formation and germination cycles (significant at $q < 3.0 \times 10^{-37}$, Fisher's exact test) (Figure 4.6). Genes associated with regulation are present with at least 10 genes coding for regulatory or DNA-binding roles ($q < 1.4 \times 10^{-5}$, Fisher's exact test). Genes not previously associated with sporulation are also present and these have putative roles as heat shock, membrane-associated proteins and DNA-polymerase-associated proteins.

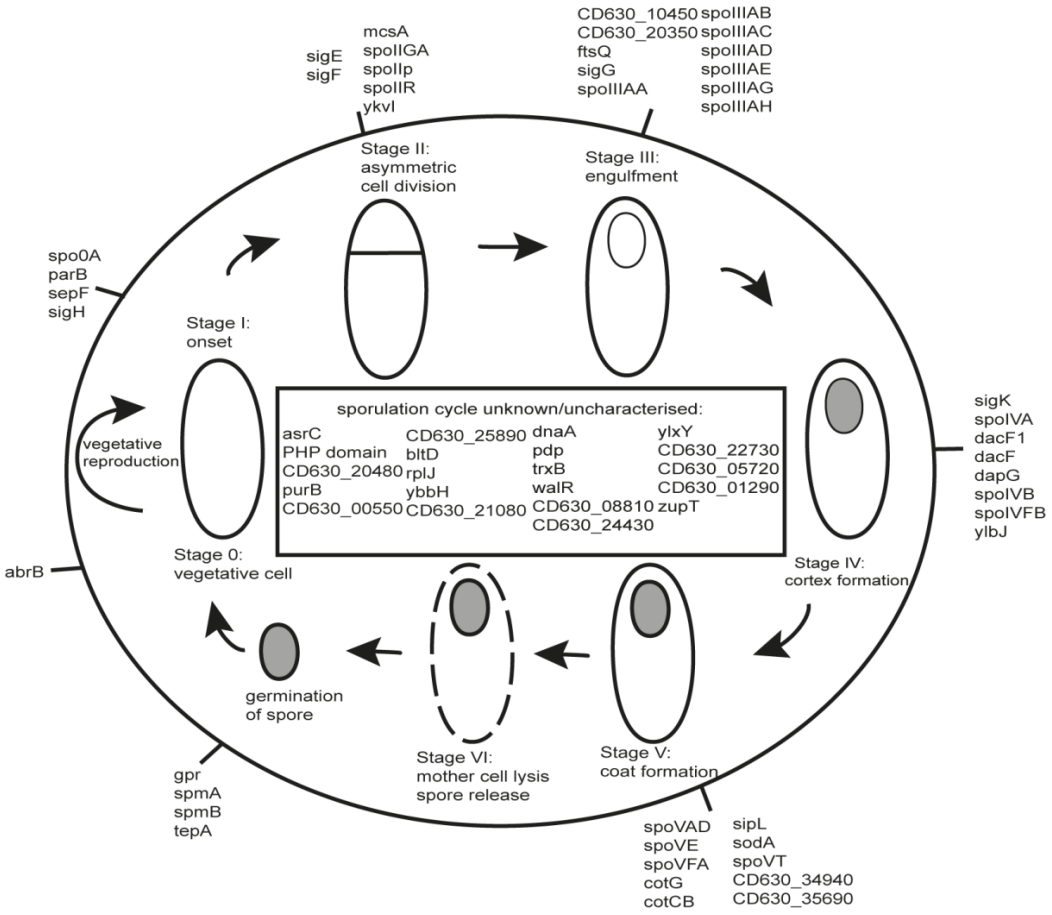


Figure 4.6 A genomic signature for identifying spore-forming bacterial species

The sporulation signature contains 66 sporulation- and germination-associated genes and genes not previously associated with sporulation. Characterised sporulation genes are on the outer circle, genes not associated with a specific sporulation cycle or uncharacterised genes are in the inside rectangle. *C. difficile* strain 630 gene names are used when possible, otherwise locus tag identifiers are shown. *Bacillus subtilis* gene names are used when no *C. difficile* homologue is available.

The presence of these genes in a genome can now be used to predict the likelihood of a species being a spore-former. Interrogating the genomes of known spore formers,

known non-spore-formers and the bacteria cultured in this study with the signature reveals a clear separation between spore-formers and non-spore-formers (Figure 4.7 and Appendix 3). The signature also accurately identifies spore-formers from different environments. For example, included in the 'known spore formers' group are bacteria normally resident in the soil such as *Bacillus pumilus* and *Thermosediminibacter oceani, a* thermophilic species isolated from sediment on the seafloor [316]. This suggests that the genetic machinery of spore-formers is broadly conserved amongst extant species regardless of habitat.

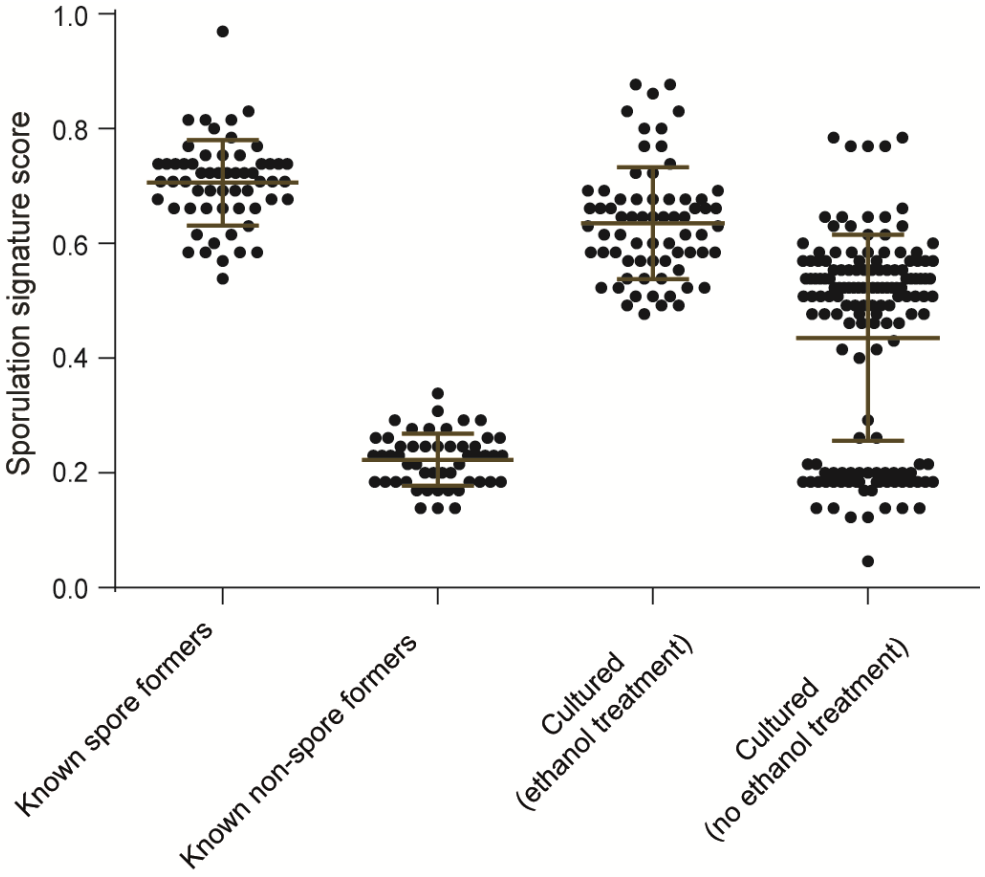


Figure 4.7 The sporulation signature accurately distinguishes spore-forming and non-spore-forming bacteria from this study and from different environments.

The higher the score the more likely a species is to be a spore-former (known spore-formers $n = 57$, known non-spore-formers $n = 50$, cultured after ethanol treatment $n = 69$, cultured after no ethanol treatment $n = 149$). Mean ± s.d.

The sporulation signature was next used to assess the abundance of spore-forming bacteria in metagenomic datasets. Taxonomic classification of the metagenomic sequences contained in the HPMCD was performed using Kraken [280]. The relative abundance of the taxonomically identified metagenomic reads was determined and the sporulation signature was then applied to the whole genomes of the species identified by Kraken. Sam Forster performed the analysis of the metagenomic data-

sets in Figure 4.8. I performed the analysis of the donor faecal samples and the culture plate samples using the relative abundance data generated by Sam Forster.

On average, across the metagenomic datasets, spore-forming bacteria comprised approximately 60 % of the genera (Figure 4.8a). When these genera were weighted by their abundance, spore-forming bacteria comprise 30 % of the microbial abundance in the intestinal microbiota (Figure 4.8b). These proportions are consistent in both the metagenomic sequence derived from the donor faecal samples in our study and the 1351 public samples.

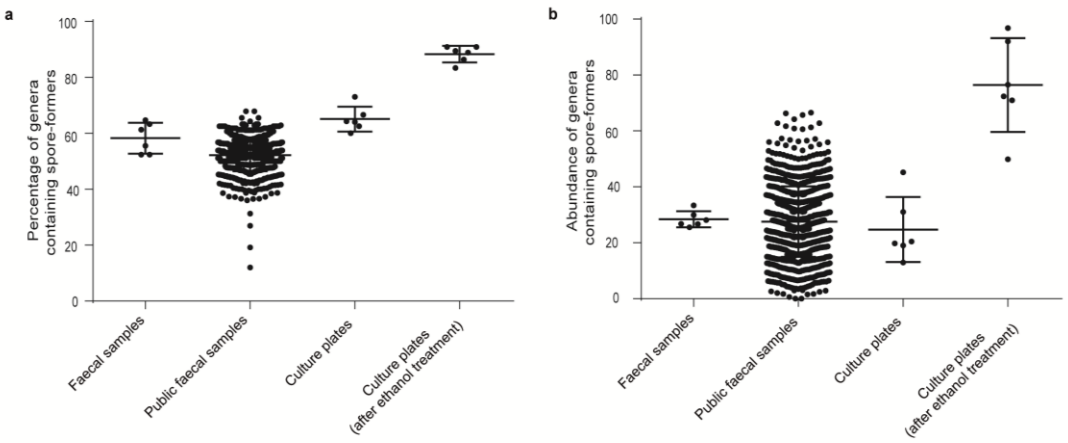


Figure 4.8 Abundance of spore-forming bacteria within the human intestinal microbiota

Using the genomic signature to interrogate public (n = 1,351) and complete faecal sample metagenomic data sets from this study (n = 6) reveals the proportion of spore-formers as a count of the total number of genera (a) and as total microbial abundance (b). Mean ± standard deviation (s.d.).

To validate the abundance of enteric spore-formers as derived by metagenomic sequencing, 16S rRNA gene amplicon sequencing of the six donor faecal samples

was carried out (Figure 4.9). The spore-formers cultured were assigned to their respective taxonomic genera and families. The abundance of these genera and families as estimated by 16S rRNA gene amplicon sequencing was then used as the lower and upper limit for calculating the proportion of spore-formers at a taxonomic level. Specific genera and families were only included if they contained a species that was cultured after ethanol shock treatment. Based on these parameters the relative abundance of spore-formers closely matches the metagenomics based estimate of 30 %.

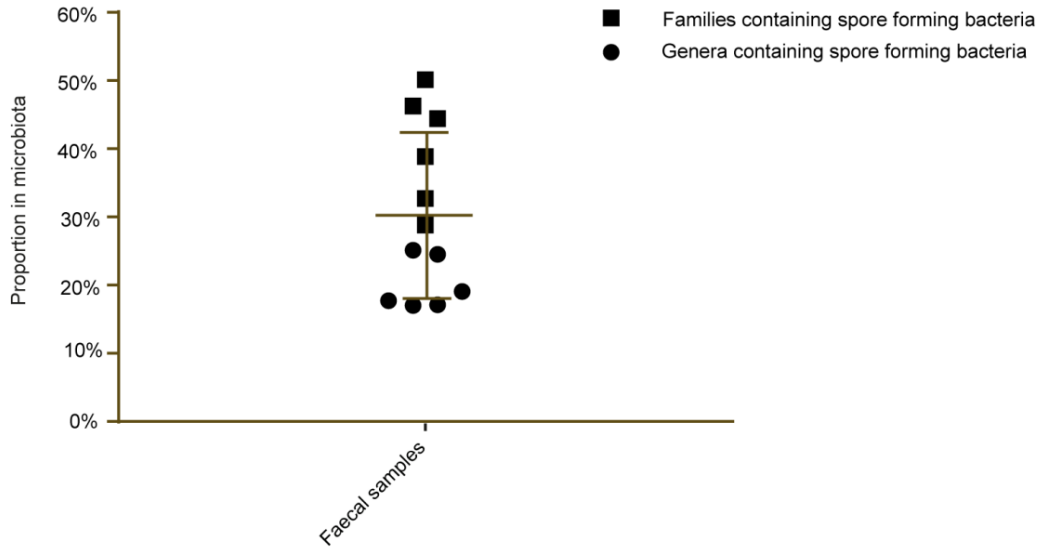


Figure 4.9 Validation of the estimation of the proportion of spore-formers in the intestinal microbiota by 16S rRNA gene amplicon sequencing

The genus (circle) and family (square) taxonomic ranks were designated as the lower and upper limits of intestinal spore-formers. Mean ± s.d.

The higher than expected abundance of spore-forming bacteria in the intestinal microbiota indicates that spore-formation may provide an advantage for transmission of aero-sensitive bacteria. If spore-formers can transmit more readily then they could be more diverse and dynamic within individuals compared to non-spore forming bacteria. Indeed, spore-forming bacteria were more diverse than non-spore forming bacteria in our dataset (Figure 4.10a). To assess if spore-formers were more dynamic over time I sampled from the six donors a year after their initial donation. Comparison of the metagenomic sequence generated a year later with the original sequence revealed that spore-forming bacteria change more in abundance than non-spore forming bacteria over time (Figure 4.10b). These results indicate that spore-forming and non-spore-forming bacteria in the intestinal microbiota display different transmission and colonisation dynamics that influence the overall composition of the intestinal microbiota.

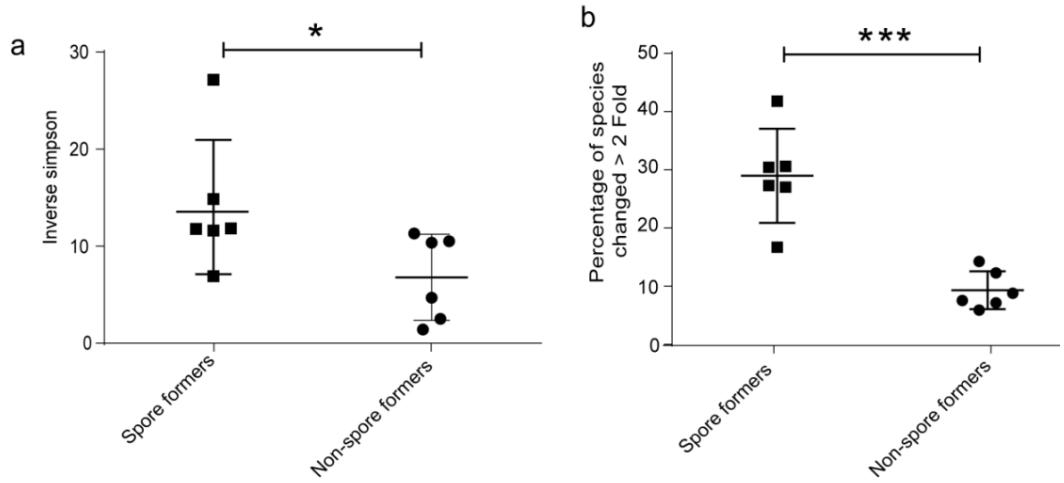


Figure 4.10 Dynamic sporulation capacity within the human intestinal microbiota

Metagenomic sequencing of donor faecal samples (n = 6) one year later demonstrates that spore-forming bacteria are more diverse than non-spore-forming bacteria (a) and that a significantly increased proportion of species show two-fold or greater change over the same time period (b). Mean ± standard deviation (s.d.), two-tailed paired t-test (*P < 0.05, ***P < 0.001).

4.3 Discussion

In this study, I characterised the enteric spore-forming bacteria cultured in Chapter 3. Using *in vitro* phenotypic characterisation and analysis of metagenomic data-sets the ability of spore-forming bacteria to readily transmit via resilient spores and how this impacts the composition of the intestinal microbiota in individuals was explored. Based on these results a hypothetical model can be formed that predicts the different transmission dynamics of spore-forming and non-spore-forming bacteria (Figure 4.11). Owing to their resistance to environmental stresses and aerotolerance, spore-forming bacteria are not as spatially and temporally restricted during transmission as

non-spore-forming bacteria. For individuals who are in regular contact with, and close proximity to, each other (for example, co-residents) both spore-forming bacteria and non-spore-forming bacteria can transmit with the same efficiency. However, as spatial and temporal distances increase, non-spore-forming oxygen-sensitive bacteria will become restricted in their ability to transmit until eventually transmission will not be possible. As spore-forming bacteria can remain viable for extended periods of time in external aerobic environments, they are not reliant on close contact between individuals to transmit. For example, spores that are shed by an individual can potentially be acquired by other individuals several weeks later.

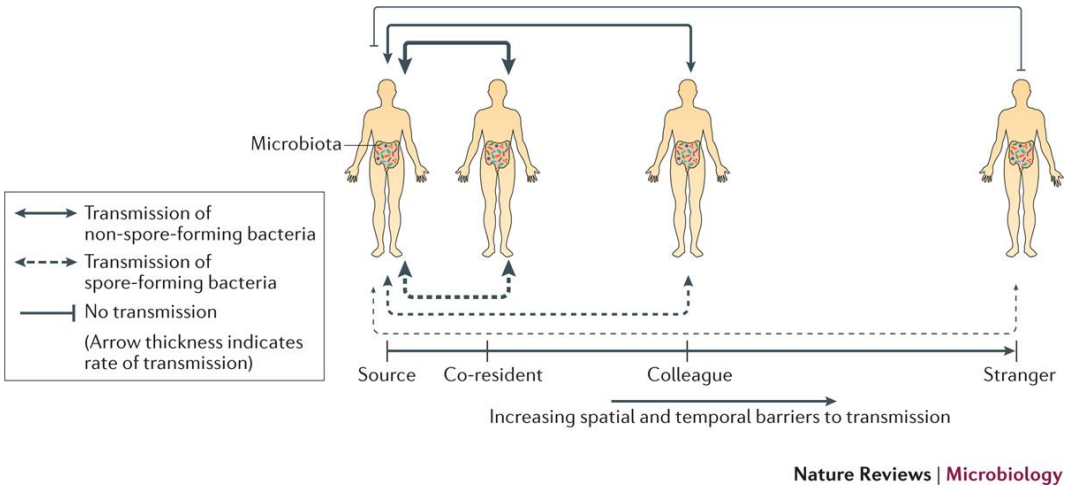


Figure 4.11 Inter-host transmission dynamics of spore-forming and non-spore-forming intestinal bacteria.

A hypothetical model exploring the different transmission dynamics of spore-forming and non-spore-forming intestinal bacteria. See main text for details. Figure taken from [1].

Here, tolerance to ethanol exposure was used to test resistance to disinfectants. Other studies have shown that, compared to non-spore-forming bacteria, *C. difficile* spores are more resistant to other disinfectants too [102]. We now spend up to 90 % of our

time indoors; therefore our homes, work places and other built environments and the cleaning regimes we use within them, play an important role in how we interact with the microbes that surround us [209, 317, 318]. This is not the environment our microbiota initially evolved with. For example, increasing urbanisation has led to cleaner, more sterile living spaces with more rooms for individuals which affords greater privacy and correspondingly less contact time with co-residents [319]. In this scenario spore-forming bacteria could prevail to the detriment of extremely aero-sensitive bacteria that may rely more on closer contact between individuals to transmit and are more sensitive to disinfectants. Other studies have highlighted the high microbiota diversity of people from traditional rural-gatherer societies and agrarian groups compared to developed Western populations [224, 320]. While diet may drive a large part of this diversity the differences in the built environment utilised by both groups may also play a role. As such, the full influence of the built environment on transmission of the intestinal microbiota is currently unknown, however as research in this discipline increases, its importance will become more apparent [209].

All of the anaerobic non-spore-forming bacteria tested remained viable for at least 48 hours. While this experimental scenario probably does not reflect real life conditions it does indicate that anaerobic non-spore-forming bacteria may survive long enough in order to successfully transmit. While not quantified, the oxygen exposure experiment will also have incorporated other environmental stresses such as being exposed to temperatures other than body temperature and ultraviolet radiation. Thus, while spore-forming bacteria have an advantage over non-spore-formers in terms of transmission, non-spore forming anaerobic bacteria may be able to survive long enough in order to successfully colonise a new host in close proximity. As discussed

in Chapter 1 (Introduction Chapter), other survival mechanisms are utilised by bacteria which could be a feature of non-spore-forming anaerobic bacteria.

There was a big difference in the response of spores to different bile acids with some responding to sodium taurocholate, cholate and glycocholate while others only responded to sodium taurocholate, such as *Clostridium hathewayi* (Figure 4.4). This suggests that the germination response of intestinal commensal spores is not uniform and reinforces the fact that much remains to be learned about intestinal spore-forming bacteria. The spores that were cultured and germinated in Chapter 3 were all from healthy individuals with no recent antibiotic exposure. Hence, they were probably produced under normal homeostatic conditions in the intestinal environment and not as a result of encountering a stress. Intestinal spore-formation may therefore be a normal feature of the life cycle of these bacteria that occurs independently of external perturbations.

Despite their abundance at the genus level, spore-formers are less abundant than non-spore-formers and change more in composition than non-spore-formers. Perhaps, as spore-formers are less abundant they are more at risk of encountering extinction or expulsion from the host due to ecological drift. In this scenario spore-formation provides a reliable escape mechanism and ensures survival [245]. More abundant or more persistent species are not as dependent on dispersal to survive, indeed a trade-off can be envisaged where adaptation to the gut environment ensures sufficient abundancy to negate the need to concentrate resources on dispersal [221]. The ability to form spores could be a means to disperse to the outside environment and survive in aerobic conditions before being ingested by a new host. In the next chapter I will explore the hypothesis that if adept at dispersal, spore-forming bacteria are not as reliant on host-adaptation.

Chapter 5 Host-adaptation of commensal bacteria through loss of sporulation and reductive genome evolution

5.1 Introduction

Human intestinal spore-formers present a dichotomy for bacteria within the intestinal microbiota. On the one hand they are extremely adapted and evolved to the host environment. As demonstrated in the previous chapter, they utilise intestinal bile acids to trigger germination and initiate colonisation. Furthermore, they interact with the host through their immunomodulatory properties [131, 145]. On the other hand, they are independent of the host because they are capable of surviving *ex vivo* as spores for extended periods of time. Speciation of non-spore-forming *Bacteroidaceae* and *Bifidobacteriaceae* align with the speciation events of our early *Hominidae* ancestors over 15 million years ago to modern humans today [3]. However, spore-forming *Lachnospiraceae* do not display the same patterns of co-speciation [3]. This suggests that host-associated spore-forming bacteria may have a different evolutionary history that is not as connected to their host as the evolutionary history of non-spore-forming bacteria. In addition, spore-formation is a metabolically expensive process requiring the orchestration of hundreds of genes and resulting in the destruction of the original vegetative cell [108, 110]. As such, little is known of the ecological forces that maintain sporulation within host-associated bacteria.

Here, I explore the differences between host-associated spore-forming and non-spore-forming bacteria in the context of host adaptation. I demonstrate that loss of

sporulation within commensal host-associated bacterial families is habitat dependent, with the sporulation phenotype broadly maintained in the gut and lost in other body habitats. Loss of sporulation within these bacteria leads to genetic features associated with host adaptation. These include a reduction in genome size and a reduction in genetic redundancy defined by loss of paralogous genes within a genome. Conversely, spore-forming bacteria within the same family maintain larger genomes and a greater genetic redundancy which could provide the flexibility to respond to different environmental conditions *in vivo* and *ex vivo*.

5.2 Results

5.2.1 Large and small scale loss of sporulation within the Firmicutes

Currently, there are five defined taxonomic classes within the Firmicutes, the *Clostridia, Bacilli, Erysipelotrichia, Negativicutes* and *Thermolithobacteria* [321-324]. All contain spore-forming species, except for *Thermolithobacteria*, however, only two species have been isolated from this class and no whole genome sequences are currently available [107]. I first sought to establish a robust phylogeny of the Firmicutes to investigate the distribution of the sporulation phenotype. Thus, 715 Firmicute genomes from the NCBI curated RefSeq database, in addition to whole genome sequences from intestinal isolates from the HMP, whole genome sequences from a comprehensive study describing the first 1000 intestinal cultured species and an in-house collection of 506 sequences derived from our bacterial culture collection were used to construct a phylogeny based on 40 universal genes present in each genome [26, 146, 271, 288, 325].

In total, 1658 whole genome sequences from over 45 bacterial families within the Firmicutes were included, from environments as diverse as salt mines, hot-springs,

fermented foods, soil and host-associated environments such as the intestine and mouth. Based on its proximity to the root, this phylogeny (Figure 5.1), places the order *Halanaerobiales* (family *Halobacteroidaceae*) as an early emerging member of the Firmicutes. These halophilic bacteria are currently placed within the *Clostridia* class, but based on this phylogenetic placement, which is also supported by previous studies, supports movement to a distinct class of their own [26, 326]. Also, based on this phylogeny, the *Negativicutes* and the *Thermolithobacteria* are not true phylogenetic classes, but are clustered within the *Clostridia*, and again, this is supported by other studies [26, 326]. To assess sporulation ability, the 66 genes that comprise the sporulation signature identified in Chapter 4 were mapped onto the phylogeny [146]. Using the previously established parameters, a species is considered capable of spore-formation if its genome has a sporulation signature score greater than 50 %. The three *Halanaerobiales* within the phylogeny, *Halonatronum saccharophilum*, *Halanaerobium saccharolyticum* and *Orenia marismortui* are spore-formers based on these criteria and have been characterised as spore-forming suggesting that the early ancestor of the Firmicutes was a spore-former [327-329].

The sporulation phenotype is not distributed throughout the Firmicutes but is absent within certain taxa including some contained within the *Erysipelotrichaceae, Bacilli* and *Clostridia* classes. This indicates subsequent loss of the phenotype as the Firmicutes evolved. The *Bacilli* class has large-scale absences of sporulation within the *Staphylococcaceae* family and within the *Lactobacillales* order which contains the *Lactobacillaceae* and *Streptococcaceae* families. The presence of the sporulation specific C-terminal domain of the *Spo0A* gene (Spo0A_C) was also mapped onto the phylogeny (Figure 5.1). There is a strong concordance with the presence of

Spo0A_C and a high sporulation signature score. Conversely, an absence of Spo0A_C is associated with a low sporulation score. There are a few exceptions, some species have maintained Spo0A_C but have a low sporulation score (<40 %). These include some *Staphylococcus* species and some *Exiguobacterium* species in the *Bacillales*. However, the low sporulation score combined with no reports in the literature of spore-formation suggests these species do not make spores [330, 331]. In summary, this phylogeny demonstrates that sporulation evolved in an early ancestor of the Firmicutes with subsequent large and small scale loss of the phenotype occurring.

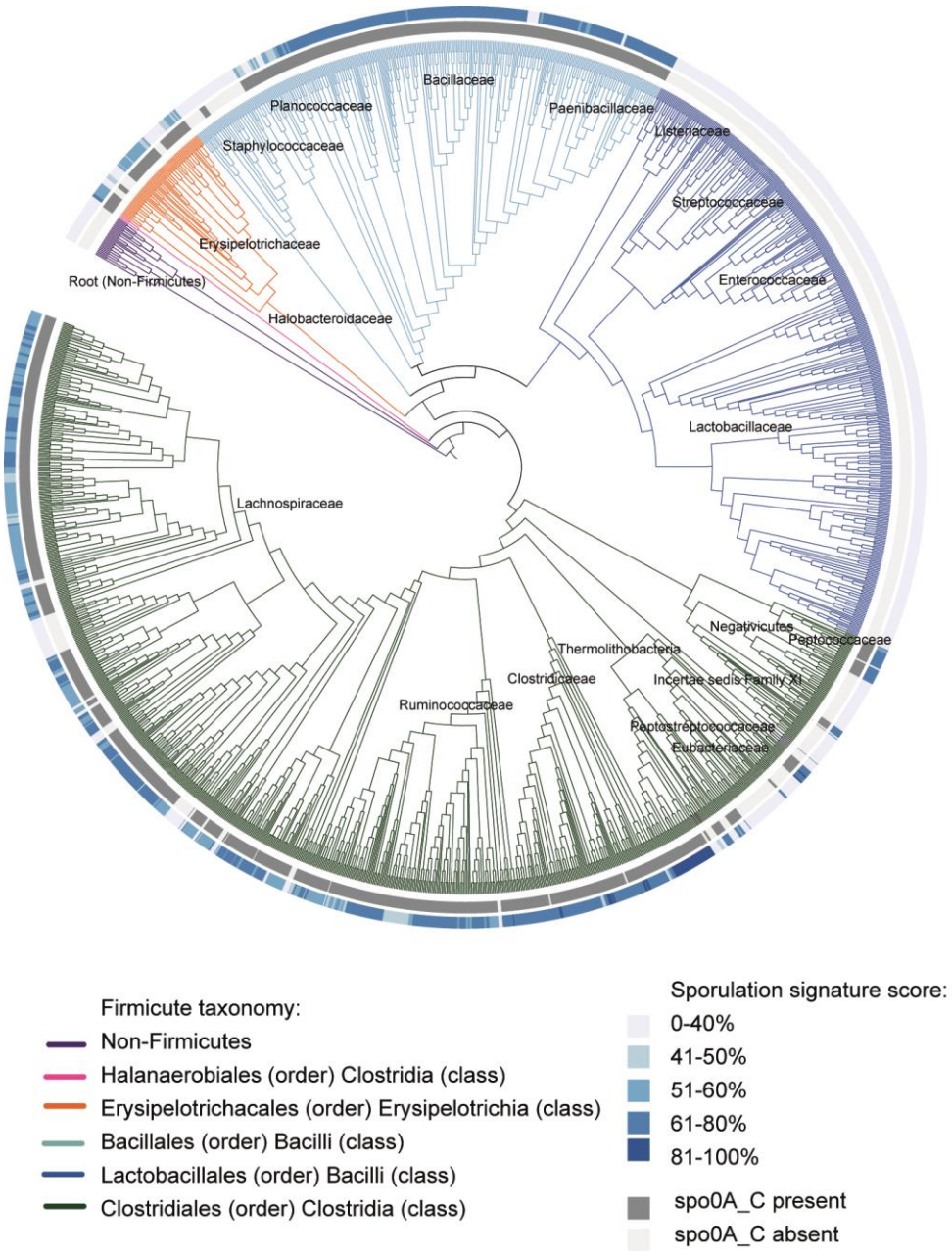


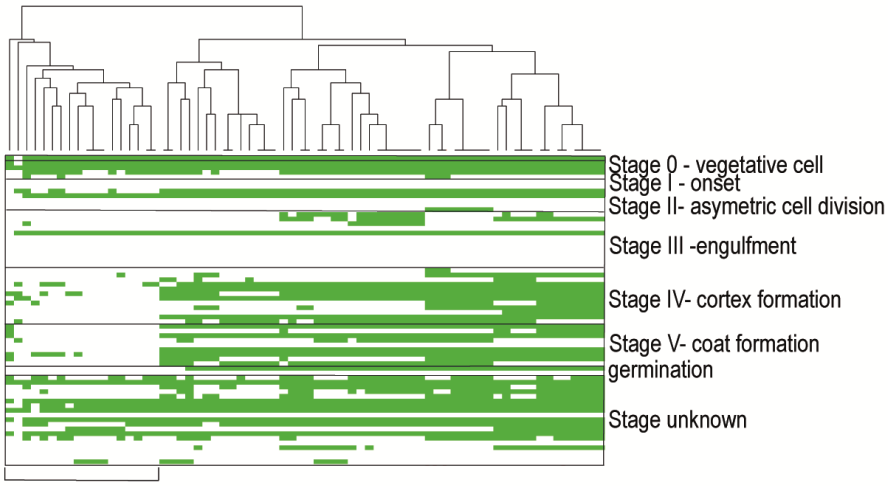
Figure 5.1 Distribution of the sporulation phenotype within the Firmicutes

Sporulation has been lost at large and small scales within the Firmicutes. Phylogeny of the Firmicutes constructed from 40 universal protein coding genes extracted from 1658 whole genome sequences. Sporulation has been lost at large (*Lactobacillales*) and small taxonomic scales (within the *Erysipelotrichaceae*, *Peptostreptococcaceae* and *Lachnospiraceae* families). Major taxonomic orders are indicated by branch colours within the phylogeny and major families within these orders are indicated by text. The

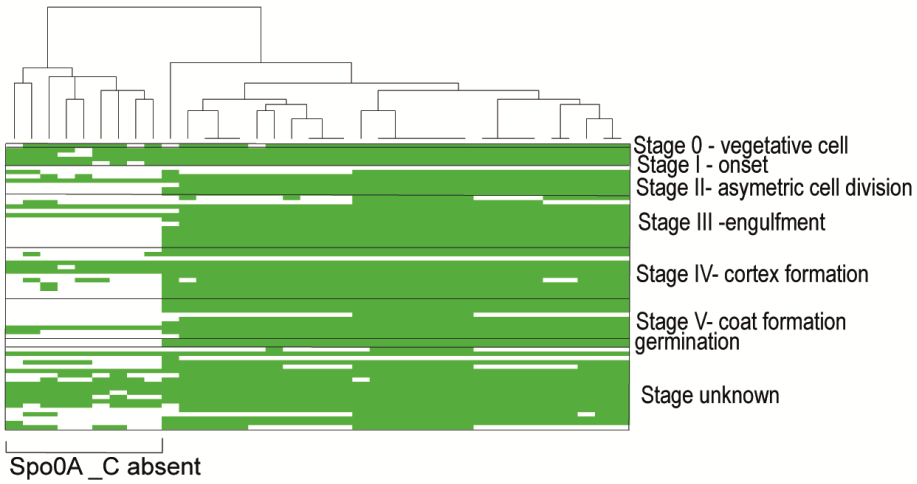
Thermolithobacteria are classified as a distinct class but cluster here within the *Clostridiales* in addition to the *Negativicutes*, both of which are annotated on the tree. Non-Firmicutes bacteria were used to root the tree. Sporulation ability is defined a sporulation signature score of greater than 50 %. The presence of the Spo0A_C protein domain in a genome is also indicated.

The association of a low sporulation score combined with an absence of Spo0A_C suggests that once the sporulation phenotype is lost, sporulation-associated genes are not retained, but are globally degraded as there is no advantage in maintaining them. Also of note is that sporulation-specific genes in the sporulation signature belonging to Stages II (asymmetric cell division), III (engulfment), IV (cortex-formation), V (coat-formation) and germination are largely absent from species within host-adapted families that also lack Spo0A_C (Figure 5.2) (*Ruminococcaceae* and *Clostridiaceae* genomes were not included in this analysis due to the low number of non-spore-forming genomes present).

Erysipelotrichaceae



Peptostreptococcaceae



Lachnospiraceae

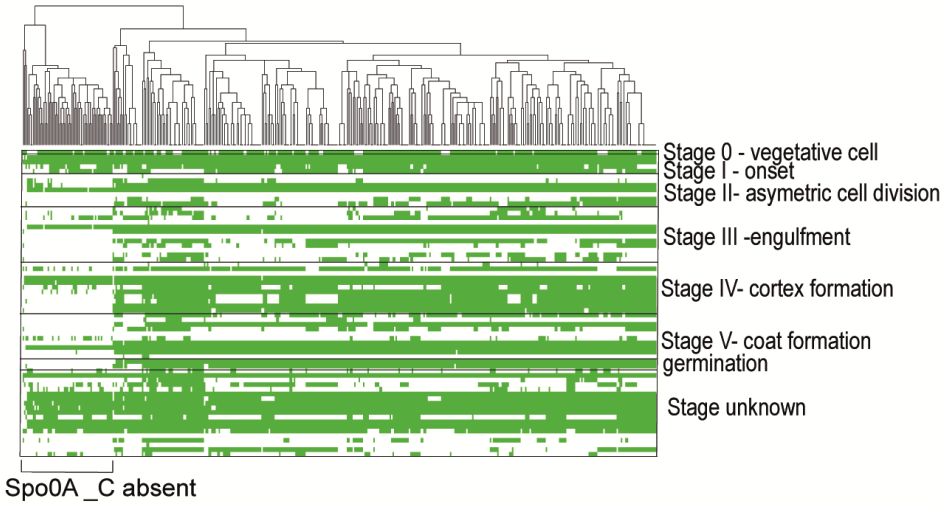


Figure 5.2 Loss of sporulation is accompanied with widespread loss of sporulation-associated genes.

Heatmaps displaying presence (green) or absence (white) of 66 genes in sporulation signature within genomes of species from gut associated bacterial families. Genomes are clustered (cladogram on top) by presence of the 66 genes (grouped according to sporulation stage on vertical plane). Bacterial species that lack the Spo0A_C domain and are therefore non-spore-formers also lack many of the sporulation signature genes.

As sporulation is a complex process requiring synchronisation of many genes acting in a hierarchical fashion it is unlikely the phenotype has been transferred horizontally. However, to investigate this possibility a phylogeny using the Spo0A_C domain sequence was constructed. The architecture of this phylogeny is broadly congruent with the species tree (Figure 5.3). This supports the hypothesis that sporulation has evolved in tandem with the Firmicutes and has not been transferred horizontally amongst taxa [95].

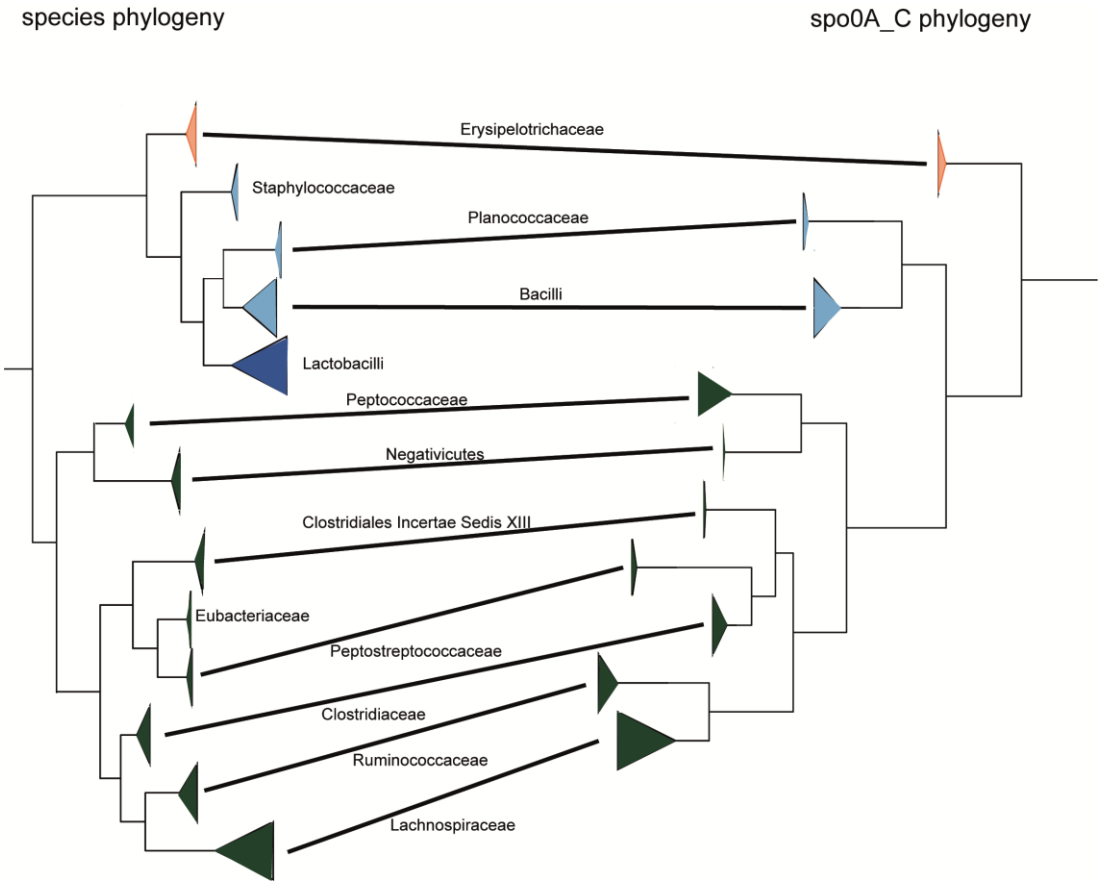


Figure 5.3 The sporulation phenotype has not been horizontally transferred.

Species phylogeny from Figure 5.1 and phylogeny of the Spo0A_C sequence extracted from the species where present. The two phylogenies are broadly congruent as indicated by the line linking taxa in both phylogenies which indicates sporulation has not been horizontally transferred amongst the Firmicutes. For conciseness, the phylogeny is compressed to only display major families.

I next sought to explore the phylogenetic small scale loss of sporulation within host-associated Firmicutes families. As these bacteria colonise multiple body sites, genomes from spore-forming (SF) and non-spore-forming (NSF) bacteria were assigned to their host habitat (Figure 5.4). There is a clear boundary between genomes with and without Spo0A_C and for *Peptostreptococcaceae* and

Lachnospiraceae, the habitat the bacteria reside in. This boundary does not fall on the 50 % sporulation signature score previously used. As the presence of Spo0A_C is a good proxy for sporulation ability (Figure 5.1) and the absence of Spo0A_C corresponds to an inability to make spores, I chose to use this boundary to define spore-forming and non-spore-forming species within each family. Hence, in Figure 5.4 genomes above the dotted line represent a spore-forming species while those below the dotted line represent non-spore-forming species.

Sporulation is broadly maintained in the gut amongst all families examined (Figure 5.4). *Erysipelotrichaceae* are the exception with bacteria containing gut-associated SF and NSF species. Unlike the *Erysipelotrichaceae,* gut-associated *Lactobacillaceae* are exclusively NSF. While some species are host-associated, this family is principally found in a wide range of environmental habitats and is not abundant within the human gut. Based on this analysis it appears sporulation provides a selective advantage in the large intestine, which is absent in other host-associated environments such as the rumen or mouth, resulting in the loss of the phenotype. Interestingly, only one of the top ten most abundant species in the intestinal microbiota of the six donors presented in Figure 3.7 were classified as spore-forming. Therefore, as discussed in the Introduction chapter, spore-formation may act as a means to negate elimination through ecological drift for these species that are present at lower abundances.

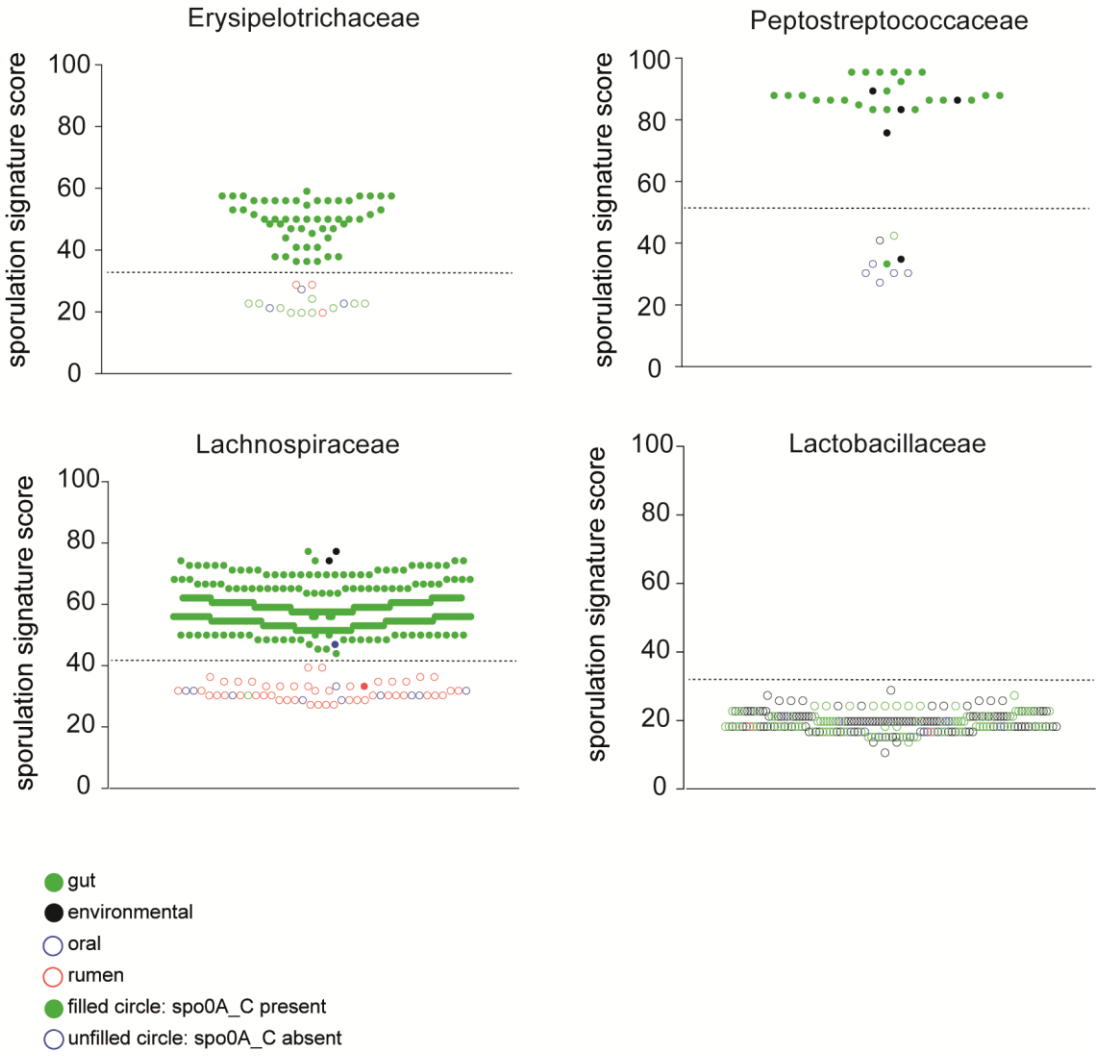


Figure 5.4 Sporulation is maintained in the gut but has been lost from other host-associated environments.

The habitat of Firmicutes families that contain spore-forming and non-spore-forming host-associated species is presented. Each dot represents a genome (*Erysipelotrichaceae* n=69, *Peptostreptococcaceae* n=36, *Lachnospiraceae* n=430, *Lactobacillaceae* n=230). Genomes above the dotted line represent spore-forming bacteria, genomes below the dotted line represent non-spore-forming bacteria. There is a clear delineation between spore-formers and non-spore-formers and their associated host habitat based on the presence or absence of Spo0A_C. *Lactobacillaceae* are exclusively non-spore-forming.

5.2.2 Genetic features of host-adaptation in non-spore-forming bacteria

Bacterial genomes reflect the environmental and evolutionary selective forces acting on them. Bacteria that inhabit dynamic, unstable or multiple environments often have large genomes which encode the functional capabilities required to survive and respond to different scenarios. A reduction in genome size is often associated with bacteria that are host-adapted and rely on their host for nutrients and survival [68]. Comparisons of genome sizes between closely related bacterial taxa can therefore provide insights into the functional capabilities of these groups and the evolutionary forces driving these differences. Within the same host-associated Firmicutes family, NSF genomes are significantly smaller than SF genomes regardless of the habitat they reside in (Figure 5.5). *Lachnospiraceae* NSF rumen genomes are reduced on average by 12 % in size, *Lachnospiraceae* NSF oral genomes by 39 %, *Peptostreptococcaceae* oral genomes by 39 % and *Erysipelotrichaceae* NSF gut genomes by 37 %. Notably, ocean dwelling *Prochlorococcus* bacteria have undergone similar levels of genome reduction of 38 % compared to closely related bacteria that have not undergone genome reduction [54]. Therefore, while best understood in endosymbiotic bacteria, genome reduction also features in free-living bacteria. Other studies have reported the genomes of oral and rumen-associated bacteria are smaller than gut-associated bacteria, regardless of sporulation status. The oral and rumen-associated genome sizes in this dataset are comparable with other datasets [332] suggesting there may be an optimum genome size for bacteria that is influenced by the habitat they reside in.

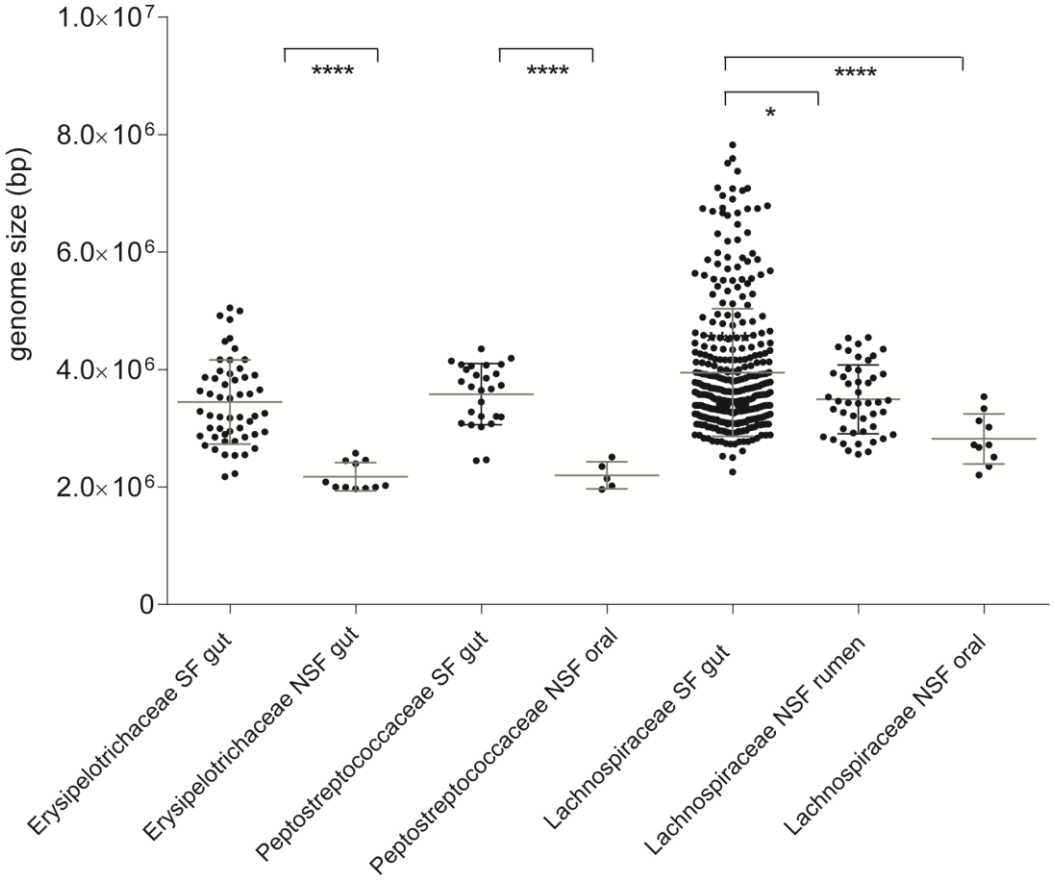


Figure 5.5 Non-spore-forming bacteria have smaller genomes than spore-forming bacteria within the same host-associated family

Absence of sporulation is associated with a reduced genome size within host-associated Firmicutes families. The genome sizes of non-spore-forming (NSF) *Lachnospiraceae, Peptostreptococcaceae* and *Erysipelotrichaceae* bacteria are smaller than spore-forming (SF) bacteria from the same family (Mann-Whitney test to compare genome size of spore-forming and non-spore-forming bacteria within the same family, * = P< 0.05 & ****= P < 0.0001, Mean with SD).

Coping with a varied or challenging environments requires the ability to maintain functionality despite perturbations faced [333]. A feature of a robust genome is genetic redundancy where more than one gene encodes the same or related function in a genome (paralogue). This ensures that even if one gene is inactivated the function is maintained. Conversely, if robustness is no longer required due to a stable

and constant environment, then loss of genetic redundancy can be expected where the function is maintained in the genome but there is no longer a back-up gene encoding the same function [60]. There is a strong correlation between genome size and gene number in the Firmicutes dataset as has been previously reported [67] (Spearman's rho= 0.9842) (Figure 5.6). This correlation applies regardless of the sporulation capabilities of the bacterial species. Hence, I next sought to investigate if there was a difference in genetic redundancy at the family taxonomic level, between the genomes of spore-forming and non-spore-forming bacteria which could account for some of the difference in genome size reported in Figure 5.5.

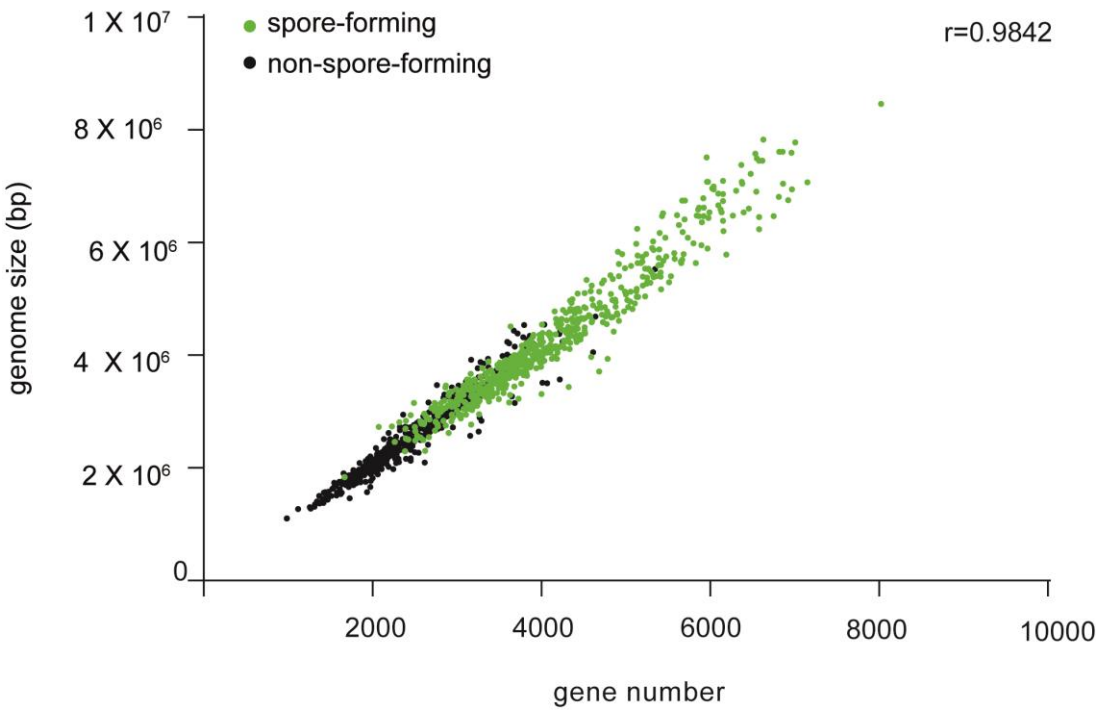


Figure 5.6 Genome size and gene number are correlated in the Firmicutes.

Genome size and gene number are strongly correlated. Each dot represents a genome from the entire Firmicutes dataset, hence, this correlation applies regardless of habitat or sporulation ability, Spearman rho correlation coefficient = 0.9842.

Figure 5.7 below outlines the workflow used for this analysis. The genomes were first grouped into species using Average Nucleotide Identity (ANI). ANI groups genomes based on nucleotide similarity, a threshold of 95 % similarity is used to denote genomes of the same species [292]. A representative genome from each ANI species was then utilised for further analysis. For each of the *Erysipelotrichaceae, Peptostreptococcaceae* and *Lachnospiraceae* families, orthoMCL was used to cluster genes from the representative genomes into homologous groups based on their sequence similarity [293]. The genes in a homologous group can be from the same spore-forming or non-spore-forming genome (paralogue) or from different spore-forming or non-spore-forming genomes (orthologue). Regardless of which genome the genes are derived from, each gene in a homologous group will encode the same function as determined by their sequence similarity. I compared genomes from *Erysipelotrichaceae* SF gut (n=21 ANI species) against *Erysipelotrichaceae* NSF gut (n=9 ANI species), *Peptostreptococcaceae* SF gut (n=12 ANI species) against *Peptostreptocacceae* NSF oral (n=5 ANI species) and *Lachnospiraceae* SF gut (n= ANI 139 species) against *Lachnospiraceae* NSF rumen (n= ANI 44 species) and *Lachnospiraceae* NSF oral (n= ANI 9 species). Other genomes from bacteria in different environments were not included due to their low numbers. Further details on the number of genes and homologous groups are described in Table 2.2.

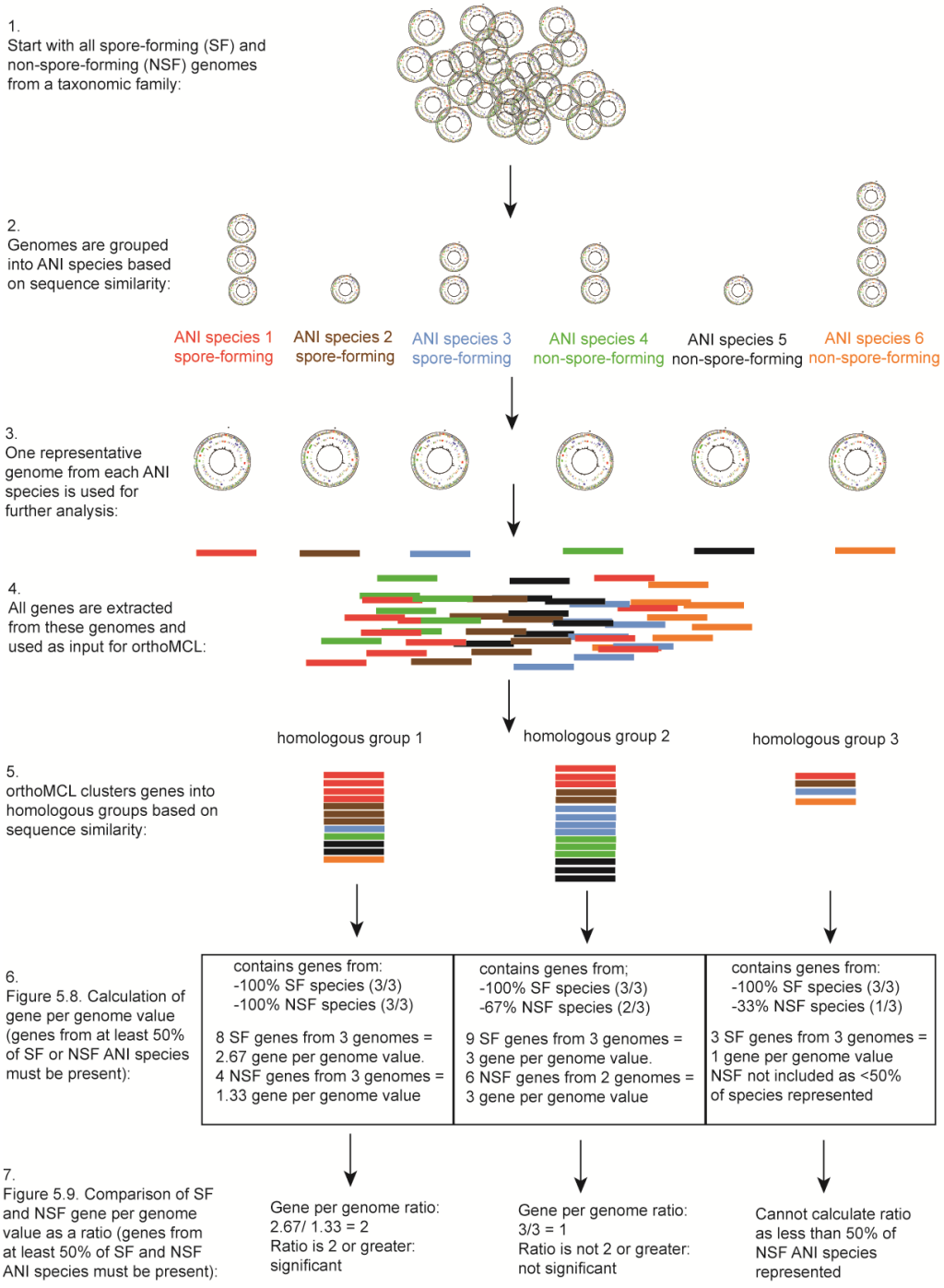


Figure 5.7 Workflow for genetic redundancy analysis

A simplified version of the workflow used to create Figures 5.8 and 5.9. For brevity, in this scenario only three homologous groups were identified and no distinction is made between genes from bacteria found in different environments. This workflow was carried out separately on genomes from *Erysipelotrichaceae*, *Peptostreptococcaceae* and *Lachnospiraceae* families.

To examine the genetic redundancy within SF and NSF bacteria in a particular family, the number of genes per genome for each homologous group was calculated. Homologous groups that did not contain genes from at least 50 % of the SF or NSF ANI species in that family were excluded (for example, if a homologous group contained genes from four or less of the nine *Erysipelotrichaceae* NSF gut ANI species, it was excluded). Next, the gene per genome value for each remaining homologous group was calculated, (i.e. the number of genes from SF or NSF bacteria in a homologous group compared to the number of SF or NSF genomes those genes were derived from). SF bacteria have a higher average gene per genome value compared to the NSF bacteria within the same family (i.e. greater genetic redundancy) (Figure 5.8). This applies to all the families examined.

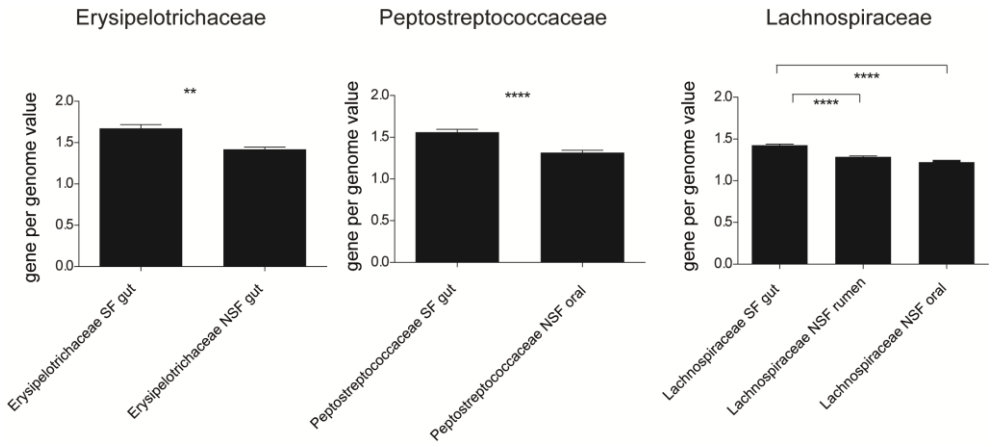


Figure 5.8 Genetic redundancy is greater in spore-forming host-associated Firmicutes compared to non-spore-forming bacteria within the same family.

Spore-forming (SF) host-associated bacteria have more genetic redundancy compared to non-spore-forming bacteria (NSF) from the same family as determined by the gene per genome value of a homologous group. The mean of all the genes per genome values for each homologous group is presented. (error bars- SEM. Paired t-test comparing gene per genome value of spore-formers within a homologous group against the gene per genome value of non-spore-formers within the same homologous group. $P<0.01$ for *Erysipelotrichaceae* and $P<0.0001$ for *Peptostreptococcaceae* and *Lachnospiraceae*).

Next, the functions of homologous groups that are shared by both spore-forming and non-spore-forming bacteria but which have different levels of genetic redundancy were determined (Figure 5.9). A homologous group was included in the analysis if it contained genes present in greater than 50 % of species from both SF and NSF bacteria within the same family, i.e. the homologous group is present in the majority of the species in that family. For each homologous group that passed this criterion, the ratio of the gene per genome value for both SF and NSF bacteria was calculated. A ratio of 2 or greater was considered significant, i.e. in a homologous group, the

gene per genome value of genes derived from SF (or NSF) bacteria compared to the gene per genome value of the other group is two or greater. SF bacteria are present in a higher number of homologous groups that contain more genetic redundancy compared to NSF bacteria (Figure 5.9). There are a wide range of functions encoded in the different homologous groups but regulatory functions and transport and binding functions are predominant in all four comparisons.

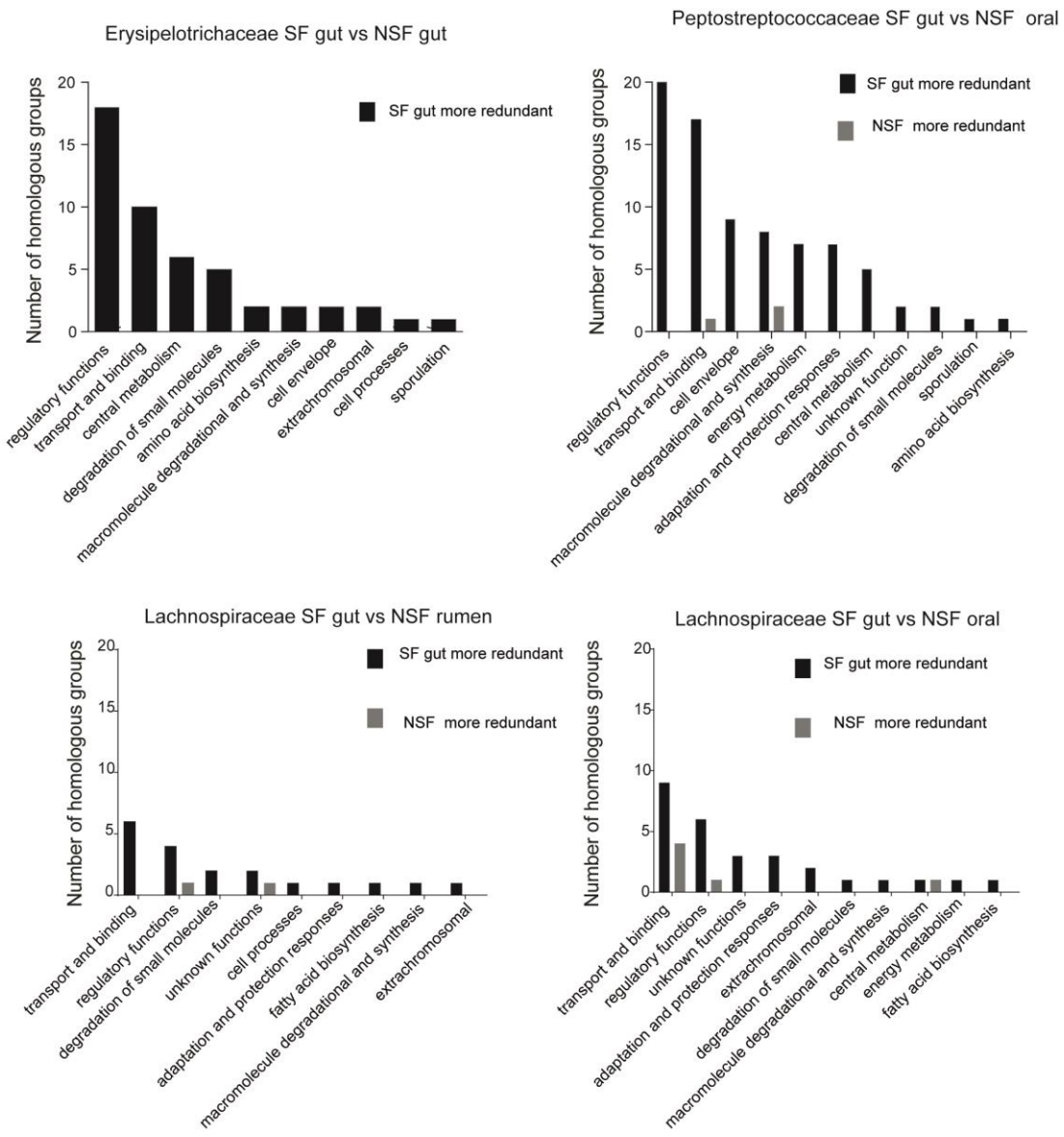


Figure 5.9 Genetic redundancy is associated with shared functions in spore-forming bacteria in different families.

Spore-forming bacteria (SF) have a greater level of genetic redundancy compared to non-spore-forming bacteria (NSF) within shared homologous groups. These homologous groups encode different functions but transport and binding and regulatory functions are predominant.

Within the *Erysipelotrichaceae* SF bacteria there are ten homologous groups with a high level of genetic redundancy with transport and binding functions. Of these,

seven are associated with the phosphoenolpyruvate carbohydrate phosphotransferase (PTS) system which is involved in transport of carbohydrates into the bacterial cell. This is a multi-enzyme system that involves phosphorylation of the imported carbohydrate which prevents diffusion back out of the cell [334]. The PTS system consists of carbohydrate-general cytoplasmic proteins and membrane bound carbohydrate-specific proteins. The seven redundant homologous groups are all annotated as carbohydrate-specific and are specific for lactose, cellobiose, mannose, fructose and sorbose. Of the 18 homologous groups with a high level of genetic redundancy annotated with regulatory roles, five are annotated as part of two-component regulatory systems. Similar to the PTS system, this is also a sensory system used by bacteria to sense and respond to external stimuli [335]. The Spo0A protein is the response regulator of a two-component system.

Within the *Peptostreptococcaceae* SF group, redundant homologous groups of note include the pleiotropic transcriptional regulator *ccpA* which plays a role in colonisation, virulence, biofilm formation and plays a major role in glucose metabolism mediating transcription of several hundred genes in *C. difficile* [336]. Of the 17 homologous groups annotated as involved in transport and binding roles, there are nine homologous groups annotated as ATP-binding cassette (ABC) transporters. These are ubiquitous proteins found in eukaryotes and prokaryotes that transport solutes either in or out of the cell. They function by binding solutes to a transmembrane protein, this is followed by an ATP-driven conformational change in the protein which imports or exports the solute [337]. The annotation of the homologous groups suggest they are involved in the importation of solutes, the function of which could include carbohydrates, vitamins or metals [337]. Of these ABC annotated homologous groups, two are siderophores with homology to *fhuB*

and *fhuC*. Similar to the *Erysipelotrichaceae* SF genomes, two-component regulatory system genes are prevalent (7 out of 20 redundant regulatory homologous groups).

The redundant *Lachnospiraceae* homologous groups follow a similar pattern. Most are redundant in SF bacteria, and PTS system associated and ABC transporters comprise the majority of the redundant transport and binding homologous groups. Again, there are redundant two-component system homologous groups present in the SF bacteria when compared to both rumen and oral genomes. Finally, there were no redundant NSF *Erysipelotrichaceae* homologous groups and only three redundant *Peptostreptoccocaceae* NSF oral and three *Lachnospiraceae* NSF rumen homologous groups. Interestingly, of the six redundant *Lachnospiraceae* NSF oral homologous groups, four are ABC-type genes, two of which are annotated as involved with cobalt transport.

5.3 Discussion

In this study, I used a robust phylogeny based on 40 universal genes found in over 1600 Firmicutes genomes to demonstrate that the sporulation phenotype most likely evolved once in an early ancestor. Since then, sporulation has been lost multiple times, in entire families and also within families. Examination of genomes where sporulation has been lost within families demonstrates that within two families (the *Peptostreptococcaceae* and the *Lachnospiraceae*) this loss is habitat dependent. Sporulation is maintained in gut-associated bacteria within these families but lost in oral and rumen-associated bacteria. Sporulation has also been lost amongst members of the *Erysipelotrichaceae* family but these bacteria still reside in the gut.

Examination of the genomes of bacteria that are non-spore-forming reveals that these bacteria display genetic features of host adaptation such as smaller genome size and lower genetic redundancy.

The complete absence of sporulation within *Lactobacillaceae, Staphylococcaceae* and *Streptococcaceae* families has been proposed to occur as a result of adaptation to nutrient-rich environments. In this scenario, sporulation is lost as it is no longer required [107, 338, 339]. Some studies have reported the lack of sporulation within *Lachnospiraceae* residing in the mouth and rumen [31], but to the best of my knowledge this is the first report of loss of sporulation occurring within some, but not all, taxa within multiple host-associated families. There was a remarkable divide, determined by habitat, within the *Lachnospiraceae* and the *Peptostreptococcaceae* spore-forming and non-spore-forming bacteria. The digestive physiology of ruminants and mono-gastric animals such as humans differs in terms of diet and the site of primary microbial fermentation. These differences likely drive differences in the microbial composition within [5]. Microbial fermentation in the rumen precedes passaging of food material to the stomach and large intestine whereas the majority of microbial fermentation in humans takes place in the large intestine. The longer retention time of digesta in the gut of ruminants allows degradation of recalcitrant plant material by the rumen microbiota. Perhaps, rumen-associated bacteria are not exposed to aerobic conditions as often as intestinal-associated bacteria due to their location in the digestive tract, as such, sporulation is under reduced selection pressure in these conditions compared to mono-gastric associated animals where regular expulsion of microbes in faeces occurs.

Sporulation is advantageous in unstable dynamic environments as it provides a mechanism to ensure survival if prevailing conditions become deleterious. The oral

and intestinal microbiotas are stable in composition both in abundance and over time compared to other body sites, suggesting another factor is important for maintaining the phenotype in the gut but not in the mouth [217, 340, 341]. The oral cavity, similar to the gut, is initially colonised at birth with a mature and stable community developing during childhood [342]. Horizontal transfer between individuals also occurs [342, 343], and as transmission between oral habitats can occur directly and immediately through kissing or through hand to mouth contact, then exposure to aerobic conditions may be limited for colonising anaerobic bacteria. In this scenario, as opposed to the gut environment, sporulation is not required for transmission and may be selected against. The reason behind the absence of sporulation within *Erysipelotrichaceae* bacteria that reside in the gut remains unknown. Little is known of the role of this bacterial family within the intestinal microbiota, and most reports focus on their association with disease, as such, their role as commensals and how they differ in functionality from other host-adapted bacterial families (and classes) is unclear [344, 345].

Loss of genetic redundancy is associated with host adaptation. Features of loss of genetic redundancy include a reduced genome size, a reduction in paralogues within genomes, and a loss of metabolic and regulatory genes [54, 60, 62, 68]. The absence of these features in spore-forming bacteria reflects a lifestyle that incorporates regular exposure to adverse environmental conditions, perhaps due to their lower abundance. Spore-formation provides a means to survive these adverse conditions, regulatory genes such as two-component regulatory system genes provide a means to assess and respond to the changing conditions. In a similar fashion, the greater redundancy of PTS and ABC transporter genes within spore-forming genomes may provide these bacteria with the flexibility to transport carbohydrates and nutrients in

different environmental conditions or different hosts. For non-spore-forming bacteria this may not be as relevant if their environment is constant, hence they can afford to lose extra copies of these genes. Interestingly, the *ccpA* gene which is redundant in *Peptostreptococcaceae* spore-forming bacteria is involved in carbon catabolite repression and regulates toxin production in *C. difficile* depending on PTS system mediated carbohydrate availability [336]. In summary, this study reveals that the human intestinal microbiota is populated by bacteria, that depending on their ability to form resilient spores, have evolved to engage in opposing lifestyles either orientated towards inter-host dispersal or within-host adaptation.

Chapter 6 Summary and future directions

The factors that influence our intestinal microbiota are becoming more apparent, partly owing to technological advances in microbiology, genomics and bioinformatics, and partly owing to the realisation that assembling and maintaining a healthy intestinal microbiota may depend not only on our diet, lifestyle choices and general health, but through inter-host transmission, may also depend on the microbiota and the health of others. The health status of the donors that we acquire our microbiota from may affect the composition of our own intestinal microbiota. In theory, donors that have the greatest diversity of commensal bacteria in the highest numbers are most likely to replenish the depleted microbiota of potential recipients (Figure 6.1). Suboptimal donors are people that were once healthy donors, but through antibiotic exposure or other disease conditions have lost intestinal microbiota diversity. The microbiota of suboptimal donors may potentially include higher levels of pathogens, which may be transmitted at a higher frequency than in healthy donors [1]. The culturing, computational analysis and *in vitro* analysis described in this thesis provides a platform to validate a model for transmission of the intestinal microbiota as described in Figure 6.1.

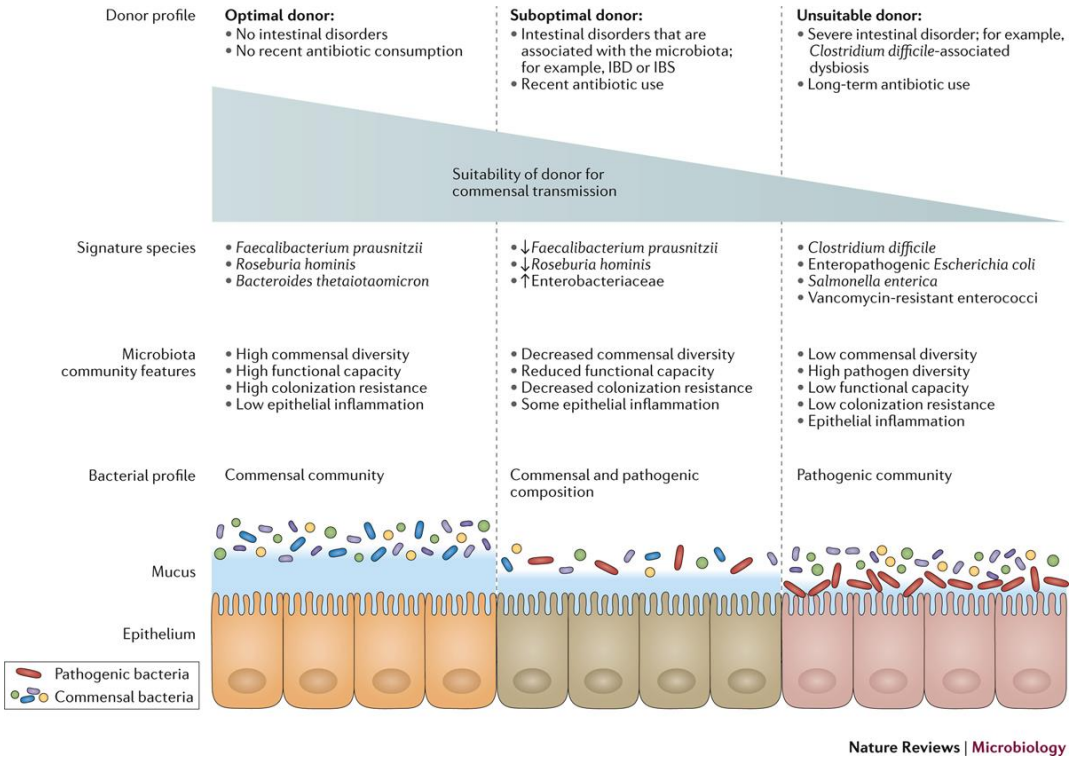


Figure 6.1 A model for transmission of commensal intestinal bacteria and the influence of donor health status.

Healthy donors who have no history of intestinal disorders or recent antibiotic treatment will typically have a diverse intestinal microbiota that exhibits high colonisation resistance. Healthy donors are optimal donors of commensal microorganisms because they will regularly contribute health-associated bacteria to their environment. Conversely, donors who have lower levels of commensal diversity, decreased colonisation resistance and a higher proportion of pathogenic bacteria are not considered optimal donors. These suboptimal and unsuitable donors would be more likely to shed pathogenic bacteria into the external environment that are not beneficial to human health. The signature species that categorise donors in this model are not comprehensive and are included on the basis of current research in the field. IBD, inflammatory bowel disease; IBS, irritable bowel syndrome. Figure taken from [1].

During my PhD thesis, an anaerobic culturing workflow was developed to culture and isolate a broad range of bacteria from the human intestinal microbiota. Using

metagenomic sequencing of the donor's faecal samples it was established that the bacteria growing on the culture plate are representative of the same bacteria in faecal samples with a strong concordance in abundance. This resulted in the isolation of 137 distinct bacterial species [146]. Just under half of these species (68 in total) were novel and included bacteria from novel genera and novel families. Incorporated in the workflow was a targeted phenotypic screen to isolate ethanol-resistant spore-forming bacteria. The phenotypic screen was successful and resulted in the isolation of 66 ethanol-resistant spore-forming bacteria. Analysis of metagenomic sequence demonstrated that intestinal spore-forming bacteria are more abundant than previously thought, comprising up to 30% of the microbial abundance. *In vitro* analysis demonstrated that these bacteria are resistant to aerobic conditions and germinate in the presence of bile-acids commonly found in the human gastrointestinal tract, thus explaining how a significant proportion of the intestinal microbiota can exploit spore-formation to transmit between individuals.

Phylogenetic and evolutionary analysis of host-associated bacteria from the Firmicutes revealed that sporulation is highly maintained in gut bacteria but has been lost in bacteria from other host-associated environments such as the oral cavity of humans and the rumen of ruminants. Genomic analysis of spore-forming and non-spore-forming bacteria from the same host-associated taxonomic family showed that non-spore-forming bacteria have undergone genome reduction which could reflect adaptation to the host. Conversely spore-forming bacteria may not be as reliant on host adaptation as they can readily survive *ex vivo.* This raises interesting questions as to the role of spore-forming bacteria within the gut. The host selects for bacteria that are beneficial [1, 40, 42], in addition, while an individual's microbiota profile is unique, the same species (both spore-forming and non-spore-forming) are

consistently observed in different people [8, 15, 16]. While normally considered a response to stress, under normal conditions, *C. difficile* continuously produces spores at a low rate, which are able to persist by adhering to intestinal epithelial cells and mucin [346, 347]. Furthermore, spores of *B. subtilis* have been demonstrated to germinate stochastically without an external cue. While risky for the germinating spore in question, if environmental conditions are favourable it allows the population to quickly expand and occupy available niches [348]. Hence, the presence of intestinal bile acids may not always be necessary to stimulate germination, especially considering bile acids are present at lower concentrations in the large intestine where the majority of the intestinal microbiota reside [117, 129]. Therefore, spore-formation may, in addition to facilitating transmission, promote colonisation resistance by ensuring that the diversity of the community is maintained through spore germination and expansion into vacant niches, especially following a perturbation [349].

Mixtures of spore-formers from intestinal bacteria are being developed as a commercial therapeutic to treat *C. difficile* infection. In theory, spores provide a stable delivery system that should require a low level of processing to maintain viability. While initial studies looked promising, the recent failure of clinical trials raises questions regarding their suitability [350, 351]. Other studies suggest intestinal spore-forming bacteria are not as proficient at colonising as non-spore-forming bacteria, both in the scenario of FMT to treat *C. difficile* infection and in early colonisation of infants with bacteria derived from their mothers [169, 352]. This highlights the need for continued study to better understand the role of spore-forming bacteria in the intestinal microbiota.

The large number of different bacterial species isolated in this study using a single growth medium and carried out by one individual, demonstrates that the human intestinal microbiota is readily culturable. The limiting factor in culturing bacteria from the intestinal microbiota is the number of colonies picked and not any intrinsic media requirements or technological limitations. While other studies have cultured novel bacteria [79-81, 282], this is the first study to culture a large number using a streamlined workflow with a single growth medium, to quantify the bacteria cultured in terms of their abundance in the intestinal microbiota, to deposit the bacteria in public repositories and to then characterise a phenotype that is shared by a large number of these cultured bacteria. As such, this study unlocks the intestinal microbiota for future phenotypic analysis and facilitates more mechanistic experiments which can make the connection between a change in abundance in a disease state and an actual causative role for the implicated bacteria.

Another means to untangle the relationship between the intestinal microbiota and a disease condition is by using gnotobiotic mice which are germ-free mice colonised with known bacteria, in some cases, originally derived from humans. These gnotobiotic mice allow many variables to be controlled including diet, genetics, the external environment and most importantly, the intestinal microbiota. Gnotobiotic mice stably maintain their microbial communities across generations, hence they provide a powerful tool to understand the functioning of the intestinal microbiota and their response to perturbations *in vivo* [353, 354]. Using these mouse models, investigators have examined the effect of pathogen infection on a defined community of bacteria [355], the response of individual bacterial species to a change in diet [353] or the production of metabolites following introduction of a single bacterial species [356]. The spatial distribution of a defined bacterial community

along the gastrointestinal tract of mice has also been examined using Fluorescence *in sitiu* hybridisation (FISH) [357]. Other studies have colonised germ-free mice with bacteria engineered with fluorescent markers [249].

The ability to manipulate bacterial genomes through inactivation of genes or increased transcription of genes allows important insights to be gained on the functional capabilities of a bacterial species and how it responds to the surrounding environment [110, 250]. Many of the studies in gnotobiotic mice described above have utilised members of the *Bacteroides* genus, which are abundant in the human gut and are tractable to genetic engineering [249, 358]. While methods for engineering *C. difficile* are available [359, 360], there are few existing genetic tools that can engineer a broad selection of commensal species of the Firmicutes [361]. However, the availability of a wide range of commensal Firmicutes cultured here along with their whole genome sequences is a starting point to alleviate this bottleneck. Furthermore, the large number of Firmicutes cultured in this study provides a valuable resource to use for gnotobiotic mice experiments.

The bacteria cultured in this study also provide a valuable resource to develop bacterial based therapeutics. FMT has been extremely effective at resolving *C. difficile* infection [235], but a more likely therapeutic in the future and a more acceptable alternative will be one composed of a defined mix of bacteria [223, 362]. While *C. difficile* infection may transpire to be the most amenable to a bacterial-based therapeutic, other promising applications include the treatment of Inflammatory Bowel Disease (IBD) through the use of regulatory T-cell inducing bacteria which dampen down intestinal inflammation [145, 362].

The whole-genome sequences generated from these isolates will also complement culture-independent, sequence-based studies. There were 694,300 genes sequenced from 234 genomes of cultured bacteria in this study. Elucidating the functions of sequenced genes has always been problematic [77], however linking genes back to a cultured isolate allows a more detailed exploration of functionality to be made, using RNA-seq for example. These cultured isolates will also improve reference databases that are used to classify sequences in metagenomic studies. There were 90 genomes generated here that are part of the Human Microbiome Project's (HMP) 'most wanted' list of underrepresented taxa [305]. Included in this this list and cultured in this study is *Eubacterium rectale* which was noted in the recent HMP paper as an isolate requiring reference genomes [16]. Finally, while computational assemblies of genomes directly from metagenomic sequences can now assemble hundreds of genomes from different species and represents a significant computational advance, it is still not as accurate or complete as generating a whole genome sequence from a cultured isolate [77, 363].

Resistance to ethanol exposure was used as the phenotypic screen in the culturing process to isolate spore-forming bacteria. This approach was based on starting with a phenotype (ethanol-resistance) and working back to the genotype (sporulation signature and subsequent analysis of metagenomic sequence). When combined, inferences were made on the transmission dynamics of intestinal spore-forming bacteria. This culturing workflow is modular as the phenotypic screen can be altered to examine different phenotypes important in the intestinal microbiota. For example, plating faecal samples on growth media with different antibiotics would allow isolation of bacteria resistant to those antibiotics. Increasing levels of antibiotics in the growth media could be utilised to examine the level of resistance. Utilisation of

carbon sources is an important function of intestinal bacteria and plays a key role in the assembly and stability of the microbial community [23, 181, 364]. The host exploits preferential utilisation of carbons sources such as fucose or sialic acid by commensal bacteria at the epithelial layer as a means to restrict pathogen growth [40, 358]. Hence, plating faecal samples in minimal media with different carbon sources could be a means to isolate commensal bacteria with these protective effects. Furthermore, in both of these examples, the generation of genomes of bacteria that are isolated through the phenotypic screen would allow exploration of the genetic mechanisms underlying the phenotype.

Based on the topics discussed above, studies to build on the work described in this thesis should include:

1. **Transmission of the intestinal microbiota**

In Chapter 4 it was established that intestinal spores promote inter-host transmission as a result of their resistant nature and response to intestinal bile acids. Related avenues of research include:

A. Experimental examination of the transmission dynamics of spore-formers and non-spore-formers:

To further explore the transmission dynamics of spore-formers and non-spore-formers, mice could be left in cages for a few days and then removed. The uncleaned cages are then sealed for defined periods of time at which point, gnotobiotic mice are introduced. Longitudinal faecal sampling and sequencing from the gnotobiotic mice would provide information on the colonisation patterns- which spore-formers

colonise first and which non-spore-forming bacteria can colonise and upto what time point.

B. Aerotolerance of vegetative cells:

Are vegetative cells of non-spore-formers more aero-tolerant than vegetative cells of spore-formers? This would be expected as they cannot rely on a spore phenotype to aid transmission. Experiments to elucidate this would be designed on the oxygen tolerance test in Figure 4.3 and would include a larger selection of intestinal bacteria. For spore-formers in the experiment it will be necessary to distinguish between spores and vegetative cells. The rapid initial drop in viability will be due to vegetative cells dying. As the curve levels off only spores will be viable and this provides a means to differentiate between a mixed population of spores and vegetative cells from the same species.

C. The effect of the donor health-status on microbiota transmission:

Individuals living in the same house have a more similar microbiota than individuals living separately. Therefore, as presented above in Figure 6.1 the health status of the donor could influence whether beneficial or potentially pathogenic bacteria are transmitted. Experiments to investigate this would utilise mice colonised with a microbiota that would make them optimal, suboptimal or unsuitable donors. Placement of gnotobiotic or germ-free mice in the same cage and longitudinal sequencing of faecal samples would determine if the recipient mice also acquire a ‘healthy’ or ‘unhealthy’ microbiota.

2. Characterisation of a novel or understudied taxa

The culturing of a large number of novel taxa provides an opportunity to better characterise members of the intestinal microbiota.

A. Further characterisation of *Falkowia sangerensi*s:

Continued characterisation of this isolate would include the generation of phenotypic data. To investigate substrate utilisation, the Biolog System could be employed. This platform involves 96 well plates each containing a different growth substrate (such as different carbon sources) and a reducing dye for high-throughput phenotypic screening. Respiration by the inoculated bacterium in the presence of the substrate reduces the dye causing a measurable colour change which indicates a positive result. Additional characterisation could also include the use of Etest strips to determine antibiotic resistance profiles.

B. Characterisation of a taxonomic group

Similar to the characterisation of *F. sangernesis*, taxa could be selected for further characterisation. Two candidates are the *Erysipelotrichaceae* family or the *Blautia* genus within the *Lachnospiraceae* family. Six *Erysipelotrichaceae* were cultured in this study, of which three were novel. Twelve *Blautia* were cultured, of which eight were novel. Both of these taxonomic groups remain understudied. Characterisation would start with the establishment of a phylogeny using the whole genome sequences generated in this study and publicly available ones. Using the genomes, a core and accessory genome for the taxa could be established. This could then be followed by phenotypic characterisation similar to *F. sangerensis* such as utilisation

of the Biolog system. This would allow determination of which substrates are utilised by all members of the taxon and which are unique to certain species.

3. Germination and colonisation dynamics of intestinal spore formers

Intestinal spores were shown to germinate in response to different bile-acids and at different levels (Figure 4.4). This demonstrates much remains to be learnt about the germination and colonisation patterns of intestinal spore-formers. Experiments to increase knowledge in this area would include:

A. Bile acid response experiments:

A larger selection of intestinal spore-formers could be tested using the same experimental design presented in Figure 4.4. In addition, a larger selection of bile acids could be tested to assess their germination potency. A literature search on the abundance and presence of different bile acids in different parts of the gastrointestinal tract could then be used to link the response of different bacterial spores with their intestinal ecology.

B. Genetics of intestinal spore-formers:

Many spore-formers utilise *Ger*-type genes to recognise an environmental germinant (for example *GerA* in *B. subtilis*), however, *C. difficile* utilises *Cspc* to recognise bile acids [122, 365]. Besides the sporulation signature used to identify spore-formers in this study, little is known of the genetics of intestinal spore-formers. BLAST searches of these and other sporulation genes could be used to build a picture of the shared and unique sporulation genes and to link this to a phylogenetic signal within the intestinal microbiota.

Given the importance of the intestinal microbiota and the role transmission may play in maintaining microbial diversity within individuals, a greater understanding of the transmission of commensal and symbiotic microbiota is required. Traditionally, the study of bacterial transmission networks has focused on pathogens because restricting pathogen transmission is important for preventing the spread of disease. The bacterial isolates cultured in this study and the knowledge gained on the transmission capabilities of spore-forming bacteria can be applied in the future to the study of intestinal commensal transmission. The challenge moving forward will be to use these resources to validate the hypothesis that commensal bacteria spread health [1].

References

1. Browne, H.P., et al., *Transmission of the gut microbiota: spreading of health.* Nature Reviews Microbiology, 2017. **15**: p. 531.
2. Zimmer, C., *On the Origin of Eukaryotes.* Science, 2009. **325**(5941): p. 666-668.
3. Moeller, A.H., et al., *Cospeciation of gut microbiota with hominids.* Science, 2016. **353**(6297): p. 380-382.
4. Ley, R.E., et al., *Worlds within worlds: evolution of the vertebrate gut microbiota.* Nat Rev Microbiol, 2008. **6**(10): p. 776-88.
5. Ley, R.E., et al., *Evolution of mammals and their gut microbes.* Science, 2008. **320**(5883): p. 1647-51.
6. Sekirov, I., et al., *Gut microbiota in health and disease.* Physiol Rev, 2010. **90**(3): p. 859-904.
7. Lawley, T.D. and A.W. Walker, *Intestinal colonization resistance.* Immunology, 2013. **138**(1): p. 1-11.
8. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing.* Nature, 2010. **464**(7285): p. 59-65.
9. El Kaoutari, A., et al., *The abundance and variety of carbohydrate-active enzymes in the human gut microbiota.* Nat Rev Microbiol, 2013. **11**(7): p. 497-504.
10. Hooper, L.V., D.R. Littman, and A.J. Macpherson, *Interactions between the microbiota and the immune system.* Science, 2012. **336**(6086): p. 1268-73.
11. Louis, P., G.L. Hold, and H.J. Flint, *The gut microbiota, bacterial metabolites and colorectal cancer.* Nat Rev Microbiol, 2014. **12**(10): p. 661-72.
12. Sommer, F. and F. Backhed, *The gut microbiota--masters of host development and physiology.* Nat Rev Microbiol, 2013. **11**(4): p. 227-38.
13. Moran, N.A., J.P. McCutcheon, and A. Nakabachi, *Genomics and Evolution of Heritable Bacterial Symbionts.* Annual Review of Genetics, 2008. **42**(1): p. 165-190.
14. Bordenstein, S.R. and K.R. Theis, *Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes.* PLoS Biol, 2015. **13**(8): p. e1002226.
15. Human Microbiome Project, C., *Structure, function and diversity of the healthy human microbiome.* Nature, 2012. **486**(7402): p. 207-14.
16. Lloyd-Price, J., et al., *Strains, functions and dynamics in the expanded Human Microbiome Project.* Nature, 2017. **advance online publication**.
17. Man, W.H., W.A.A. de Steenhuijsen Piters, and D. Bogaert, *The microbiota of the respiratory tract: gatekeeper to respiratory health.* Nat Rev Micro, 2017. **15**(5): p. 259-270.
18. Lozupone, C.A., et al., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. **489**(7415): p. 220-30.
19. Britton, R.A. and V.B. Young, *Role of the intestinal microbiota in resistance to colonization by Clostridium difficile.* Gastroenterology, 2014. **146**(6): p. 1547-53.
20. Whitman, W.B., D.C. Coleman, and W.J. Wiebe, *Prokaryotes: the unseen majority.* Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6578-83.

21. O. Sheridan, P., et al., *Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes.* Microbial Genomics, 2016. **2**(2): p. e000043.
22. Ottman, N., et al., *The function of our microbiota: who is out there and what do they do?* Front Cell Infect Microbiol, 2012. **2**: p. 104.
23. Flint, H.J., et al., *The role of the gut microbiota in nutrition and health.* Nature Reviews Gastroenterology & Hepatology, 2012. **9**(10): p. 577-589.
24. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography.* Nature, 2012. **486**(7402): p. 222-7.
25. Koenig, J.E., et al., *Succession of microbial consortia in the developing infant gut microbiome.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4578-85.
26. Rajilic-Stojanovic, M. and W.M. de Vos, *The first 1000 cultured species of the human gastrointestinal microbiota.* FEMS Microbiol Rev, 2014. **38**(5): p. 996-1047.
27. Rolfe, R.D., et al., *Factors related to the oxygen tolerance of anaerobic bacteria.* Appl Environ Microbiol, 1978. **36**(2): p. 306-13.
28. Andriantsoanirina, V., et al., *Tolerance of Bifidobacterium human isolates to bile, acid and oxygen.* Anaerobe, 2013. **21**: p. 39-42.
29. Eckburg, P.B., et al., *Diversity of the Human Intestinal Microbial Flora.* Science, 2005. **308**(5728): p. 1635-1638.
30. Hug, L.A., et al., *A new view of the tree of life.* Nature Microbiology, 2016. **1**: p. 16048.
31. Meehan, C.J. and R.G. Beiko, *A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria.* Genome Biol Evol, 2014. **6**(3): p. 703-13.
32. Surana, N.K. and D.L. Kasper, *The yin yang of bacterial polysaccharides: lessons learned from B. fragilis PSA.* Immunol Rev, 2012. **245**(1): p. 13-26.
33. Wexler, H.M., *Bacteroides: the Good, the Bad, and the Nitty-Gritty.* Clinical Microbiology Reviews, 2007. **20**(4): p. 593-621.
34. Kamada, N., et al., *Control of Pathogens and Pathobionts by the Gut Microbiota.* Nature immunology, 2013. **14**(7): p. 685-690.
35. Chow, W.-H., et al., *An Inverse Relation between cagA+ Strains of Helicobacter pylori Infection and Risk of Esophageal and Gastric Cardia Adenocarcinoma.* Cancer Research, 1998. **58**(4): p. 588-590.
36. Petersen, C. and J.L. Round, *Defining dysbiosis and its influence on host immunity and disease.* Cellular Microbiology, 2014. **16**(7): p. 1024-1033.
37. Turnbaugh, P.J., et al., *A core gut microbiome in obese and lean twins.* Nature, 2008. **457**(7228): p. 480-484.
38. Goodrich, J.K., et al., *Human genetics shape the gut microbiome.* Cell, 2014. **159**(4): p. 789-99.
39. Turpin, W., et al., *Association of host genome with intestinal microbial composition in a large healthy cohort.* Nat Genet, 2016. **advance online publication**.
40. Pham, T.A., et al., *Epithelial IL-22RA1-Mediated Fucosylation Promotes Intestinal Colonization Resistance to an Opportunistic Pathogen.* Cell Host Microbe, 2014. **16**(4): p. 504-16.
41. Bry, L., et al., *A Model of Host-Microbial Interactions in an Open Mammalian Ecosystem.* Science, 1996. **273**(5280): p. 1380-1383.

42. Rakoff-Nahoum, S., et al., *Recognition of Commensal Microflora by Toll-Like Receptors Is Required for Intestinal Homeostasis.* Cell, 2004. **118**(2): p. 229-241.
43. Brinkman, B.M., et al., *Gut Microbiota Affects Sensitivity to Acute DSS-induced Colitis Independently of Host Genotype.* Inflammatory Bowel Diseases, 2013. **19**(12): p. 2560-2567.
44. Cullen, T.W., et al., *Gut microbiota. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation.* Science, 2015. **347**(6218): p. 170-5.
45. Sela, D.A., et al., *The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome.* Proc Natl Acad Sci U S A, 2008. **105**(48): p. 18964-9.
46. Charbonneau, M.R., et al., *A microbial perspective of human developmental biology.* Nature, 2016. **535**(7610): p. 48-55.
47. Walker, A.W., et al., *Dominant and diet-responsive groups of bacteria within the human colonic microbiota.* The ISME Journal, 2010. **5**(2): p. 220-230.
48. STEVENS, C.E. and I.D. HUME, *Contributions of Microbes in Vertebrate Gastrointestinal Tract to Production and Conservation of Nutrients.* Physiological Reviews, 1998. **78**(2): p. 393-427.
49. White, B.A., et al., *Biomass Utilization by Gut Microbiomes.* Annual Review of Microbiology, 2014. **68**(1): p. 279-296.
50. Mao, S., et al., *Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function.* 2015. **5**: p. 16116.
51. Michigan, U.o. *Rumen Anatomy*. Available from: https://www.extension.umn.edu/agriculture/dairy/feed-and-nutrition/feeding-the-dairy-herd/ruminant-anatomy-and-physiology.html.
52. Engel, P. and N.A. Moran, *The gut microbiota of insects – diversity in structure and function.* FEMS Microbiology Reviews, 2013. **37**(5): p. 699-735.
53. Shigenobu, S., et al., *Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS.* Nature, 2000. **407**(6800): p. 81-86.
54. Batut, B., et al., *Reductive genome evolution at both ends of the bacterial population size spectrum.* Nat Rev Micro, 2014. **12**(12): p. 841-850.
55. Douglas, A.E., *Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera.* Annual Review of Entomology, 1998. **43**(1): p. 17-37.
56. Hilgenboecker, K., et al., *How many species are infected with Wolbachia? – a statistical analysis of current data.* FEMS Microbiology Letters, 2008. **281**(2): p. 215-220.
57. Werren, J.H., L. Baldo, and M.E. Clark, *Wolbachia: master manipulators of invertebrate biology.* Nat Rev Micro, 2008. **6**(10): p. 741-751.
58. Wu, S. *pea aphid image*. pea aphid image]. Available from: http://journals.plos.org/plosbiology/issue?id=10.1371/issue.pbio.v08.i02.
59. Hoff, M., *When bacteria lose a single DNA base, aphids suffer.* PLoS Biol, 2007. **5**(5): p. e126.
60. Martínez-Cano, D.J., et al., *Evolution of small prokaryotic genomes.* Frontiers in Microbiology, 2015. **5**(742).
61. McCutcheon, J.P. and N.A. Moran, *Extreme genome reduction in symbiotic bacteria.* Nat Rev Micro, 2012. **10**(1): p. 13-26.

62. Mendonça, A.G., R.J. Alves, and J.B. Pereira-Leal, *Loss of Genetic Redundancy in Reductive Genome Evolution.* PLOS Computational Biology, 2011. **7**(2): p. e1001082.
63. Sun, Z., et al., *Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera.* Nat Commun, 2015. **6**: p. 8322.
64. Brewer, T.E., et al., *Genome reduction in an abundant and ubiquitous soil bacterium 'Candidatus Udaeobacter copiosus'.* 2016. **2**: p. 16198.
65. Giovannoni, S.J., J. Cameron Thrash, and B. Temperton, *Implications of streamlining theory for microbial ecology.* ISME J, 2014. **8**(8): p. 1553-1565.
66. Morris, J.J., R.E. Lenski, and E.R. Zinser, *The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss.* mBio, 2012. **3**(2).
67. Mira, A., H. Ochman, and N.A. Moran, *Deletional bias and the evolution of bacterial genomes.* Trends in Genetics, 2001. **17**(10): p. 589-596.
68. Ochman, H., *Genomes on the shrink.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(34): p. 11959-11960.
69. Shulman, S.T., H.C. Friedmann, and R.H. Sims, *Theodor Escherich: The First Pediatric Infectious Diseases Physician?* Clinical Infectious Diseases, 2007. **45**(8): p. 1025-1029.
70. Moore, W.E.C. and L.V. Holdeman, *Human Fecal Flora: The Normal Flora of 20 Japanese-Hawaiians.* Applied Microbiology, 1974. **27**(5): p. 961-979.
71. Hungate, R.E., *Evolution of a Microbial Ecologist.* Annual Review of Microbiology, 1979. **33**(1): p. 1-21.
72. Zoetendal, E.G., M. Rajilić-Stojanović, and W.M. de Vos, *High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota.* Gut, 2008. **57**(11): p. 1605-1615.
73. Duncan, S.H., et al., *Proposal of Roseburia faecis sp. nov., Roseburia hominis sp. nov. and Roseburia inulinivorans sp. nov., based on isolates from human faeces.* Int J Syst Evol Microbiol, 2006. **56**(Pt 10): p. 2437-41.
74. Duncan, S.H., et al., *Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces.* Int J Syst Evol Microbiol, 2002. **52**(Pt 5): p. 1615-20.
75. Duncan, S.H., et al., *Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov.* Int J Syst Evol Microbiol, 2002. **52**(Pt 6): p. 2141-6.
76. Stewart, E.J., *Growing unculturable bacteria.* J Bacteriol, 2012. **194**(16): p. 4151-60.
77. Walker, A.W., et al., *Phylogeny, culturing, and metagenomics of the human gut microbiota.* Trends in Microbiology, 2014. **22**(5): p. 267-274.
78. Human Microbiome Jumpstart Reference Strains, C., et al., *A catalog of reference genomes from the human microbiome.* Science, 2010. **328**(5981): p. 994-9.
79. Goodman, A.L., et al., *Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice.* Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6252-7.
80. Lagier, J.-C., et al., *Culture of previously uncultured members of the human gut microbiota by culturomics.* Nature Microbiology, 2016. **1**: p. 16203.

81. Barcenilla, A., et al., *Phylogenetic Relationships of Butyrate-Producing Bacteria from the Human Gut.* Applied and Environmental Microbiology, 2000. **66**(4): p. 1654-1661.
82. D'Onofrio, A., et al., *Siderophores from Neighboring Organisms Promote the Growth of Uncultured Bacteria.* Chemistry & biology, 2010. **17**(3): p. 254-264.
83. Tanaka, T., et al., *A Hidden Pitfall in the Preparation of Agar Media Undermines Microorganism Cultivability.* Applied and Environmental Microbiology, 2014. **80**(24): p. 7659-7666.
84. Zipperer, A., et al., *Human commensals producing a novel antibiotic impair pathogen colonization.* Nature, 2016. **535**(7613): p. 511-516.
85. Kim, M., et al., *Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes.* International Journal of Systematic and Evolutionary Microbiology, 2014. **64**(2): p. 346-351.
86. Rosselló-Mora, R. and R. Amann, *The species concept for prokaryotes.* FEMS Microbiology Reviews, 2001. **25**(1): p. 39-67.
87. Parte, A.C., *LPSN—list of prokaryotic names with standing in nomenclature.* Nucleic Acids Research, 2014. **42**(D1): p. D613-D616.
88. Collins, M.D., et al., *The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations.* Int J Syst Bacteriol, 1994. **44**(4): p. 812-26.
89. Beiko, R.G., *Microbial Malaise: How Can We Classify the Microbiome?* Trends Microbiol, 2015.
90. Yutin, N. and M.Y. Galperin, *A genomic update on clostridial phylogeny: Gram-negative spore-formers and other misplaced clostridia.* Environmental microbiology, 2013. **15**(10): p. 2631-2641.
91. Carlier, J.-P., et al., *Proposal to unify Clostridium orbiscindens Winter et al. 1991 and Eubacterium plautii (Séguin 1928) Hofstad and Aasjord 1982, with description of Flavonifractor plautii gen. nov., comb. nov., and reassignment of Bacteroides capillosus to Pseudoflavonifractor capillosus gen. nov., comb. nov.* International Journal of Systematic and Evolutionary Microbiology, 2010. **60**(3): p. 585-590.
92. Lawson, P.A., et al., *Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prévot 1938.* Anaerobe, 2016. **40**(Supplement C): p. 95-99.
93. Battistuzzi, F.U., A. Feijao, and S.B. Hedges, *A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land.* BMC Evolutionary Biology, 2004. **4**(1): p. 44.
94. Bekker, A., et al., *Dating the rise of atmospheric oxygen.* Nature, 2004. **427**(6970): p. 117-120.
95. Abecasis, A.B., et al., *A genomic signature and the identification of new sporulation genes.* J Bacteriol, 2013. **195**(9): p. 2101-15.
96. Errington, J., *Regulation of endospore formation in Bacillus subtilis.* Nat Rev Microbiol, 2003. **1**(2): p. 117-26.
97. Paredes-Sabja, D., A. Shen, and J.A. Sorg, *Clostridium difficile spore biology: sporulation, germination, and spore structural proteins.* Trends Microbiol, 2014. **22**(7): p. 406-16.

98. Angert, E.R. and R.M. Losick, *Propagation by sporulation in the guinea pig symbiont Metabacterium polyspora.* Proc Natl Acad Sci U S A, 1998. **95**(17): p. 10218-23.
99. Hutchison, E.A., D.A. Miller, and E.R. Angert, *Sporulation in Bacteria: Beyond the Standard Model.* Microbiol Spectr, 2014. **2**(5).
100. Cano, R. and M. Borucki, *Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber.* Science, 1995. **268**(5213): p. 1060-1064.
101. Horneck, G., H. Bucker, and G. Reitz, *Long-term survival of bacterial spores in space.* Adv Space Res, 1994. **14**(10): p. 41-5.
102. Lawley, T.D., et al., *Use of Purified Clostridium difficile Spores To Facilitate Evaluation of Health Care Disinfection Regimens.* Applied and Environmental Microbiology, 2010. **76**(20): p. 6895-6900.
103. Gorbushina, A.A., et al., *Life in Darwin's dust: intercontinental transport and survival of microbes in the nineteenth century.* Environmental Microbiology, 2007. **9**(12): p. 2911-2922.
104. Davies, M.P., M. Anderson, and A.C. Hilton, *The housefly Musca domestica as a mechanical vector of Clostridium difficile.* Journal of Hospital Infection, 2016. **94**(3): p. 263-267.
105. He, M., et al., *Emergence and global spread of epidemic healthcare-associated Clostridium difficile.* Nat Genet, 2013. **45**(1): p. 109-13.
106. Paredes-Sabja, D., P. Setlow, and M.R. Sarker, *Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved.* Trends Microbiol, 2011. **19**(2): p. 85-94.
107. Galperin, M.Y., *Genome Diversity of Spore-Forming Firmicutes.* Microbiology spectrum, 2013. **1**(2): p. TBS-0015-2012.
108. Fimlaid, K.A., et al., *Global analysis of the sporulation pathway of Clostridium difficile.* PLoS Genet, 2013. **9**(8): p. e1003660.
109. Galperin, M.Y., et al., *Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes.* Environ Microbiol, 2012. **14**(11): p. 2870-90.
110. Pettit, L.J., et al., *Functional genomics reveals that Clostridium difficile Spo0A coordinates sporulation, virulence and metabolism.* BMC Genomics, 2014. **15**: p. 160.
111. Molle, V., et al., *The Spo0A regulon of Bacillus subtilis.* Molecular Microbiology, 2003. **50**(5): p. 1683-1701.
112. de Hoon, M.J., P. Eichenberger, and D. Vitkup, *Hierarchical evolution of the bacterial sporulation network.* Curr Biol, 2010. **20**(17): p. R735-45.
113. Spencer, R.C., *Bacillus anthracis.* Journal of Clinical Pathology, 2003. **56**(3): p. 182-187.
114. Rossetto, O., M. Pirazzini, and C. Montecucco, *Botulinum neurotoxins: genetic, structural and mechanistic insights.* Nat Rev Micro, 2014. **12**(8): p. 535-549.
115. Dittmar, E., et al., *Necrotizing enterocolitis of the neonate with Clostridium perfringens: diagnosis, clinical course, and role of alpha toxin.* European Journal of Pediatrics, 2008. **167**(8): p. 891-895.
116. Sorg, J.A. and A.L. Sonenshein, *Bile salts and glycine as cogerminants for Clostridium difficile spores.* J Bacteriol, 2008. **190**(7): p. 2505-12.

117. Shen, A., *A Gut Odyssey: The Impact of the Microbiota on Clostridium difficile Spore Formation and Germination.* PLoS Pathog, 2015. **11**(10): p. e1005157.
118. Viscidi, R., S. Willey, and J.G. Bartlett, *Isolation rates and toxigenic potential of Clostridium difficile isolates from various patient populations.* Gastroenterology. **81**(1): p. 5-9.
119. Sebaihia, M., et al., *The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome.* Nat Genet, 2006. **38**.
120. Abt, M.C., P.T. McKenney, and E.G. Pamer, *Clostridium difficile colitis: pathogenesis and host defence.* Nat Rev Micro, 2016. **14**(10): p. 609-620.
121. Lessa, F.C., et al., *Burden of Clostridium difficile Infection in the United States.* New England Journal of Medicine, 2015. **372**(9): p. 825-834.
122. Francis, M.B., et al., *Bile acid recognition by the Clostridium difficile germinant receptor, CspC, is important for establishing infection.* PLoS Pathog, 2013. **9**(5): p. e1003356.
123. Goorhuis, A., et al., *Emergence of Clostridium difficile Infection Due to a New Hypervirulent Strain, Polymerase Chain Reaction Ribotype 078.* Clinical Infectious Diseases, 2008. **47**(9): p. 1162-1170.
124. Knetsch, C.W., et al., *Whole genome sequencing reveals potential spread of Clostridium difficile between humans and farm animals in the Netherlands, 2002 to 2011.* Euro Surveill, 2014. **19**(45): p. 20954.
125. Martin, J.S.H., T.M. Monaghan, and M.H. Wilcox, *Clostridium difficile infection: epidemiology, diagnosis and understanding transmission.* Nature Reviews Gastroenterology &Amp; Hepatology, 2016. **13**: p. 206.
126. Brouwer, M.S.M., et al., *Horizontal gene transfer converts non-toxigenic Clostridium difficile strains into toxin producers.* Nature Communications, 2013. **4**: p. 2601.
127. He, M., et al., *Evolutionary dynamics of Clostridium difficile over short and long time scales.* Proc Natl Acad Sci U S A, 2010. **107**(16): p. 7527-32.
128. Dawson, L.F., et al., *The analysis of para-cresol production and tolerance in Clostridium difficile 027 and 012 strains.* BMC Microbiology, 2011. **11**(1): p. 86.
129. Ridlon, J.M., D.-J. Kang, and P.B. Hylemon, *Bile salt biotransformations by human intestinal bacteria.* Journal of Lipid Research, 2006. **47**(2): p. 241-259.
130. Sorg, J.A. and A.L. Sonenshein, *Inhibiting the Initiation of Clostridium difficile Spore Germination using Analogs of Chenodeoxycholic Acid, a Bile Acid.* Journal of Bacteriology, 2010. **192**(19): p. 4983-4990.
131. Buffie, C.G., et al., *Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile.* Nature, 2014.
132. Falkow, S., *Who Speaks for the Microbes?* Emerging Infectious Diseases, 1998. **4**(3): p. 495-497.
133. Falkow, S., *What is a pathogen?* ASM News, 1997. **63**: p. 359-365.
134. Wexler, A.G., et al., *Human symbionts inject and neutralize antibacterial toxins to persist in the gut.* Proc Natl Acad Sci U S A, 2016. **113**(13): p. 3639-44.
135. Duncan, S.H., P. Louis, and H.J. Flint, *Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product.* Appl Environ Microbiol, 2004. **70**(10): p. 5810-7.

136. Rakoff-Nahoum, S., K.R. Foster, and L.E. Comstock, *The evolution of cooperation within the gut microbiota.* Nature, 2016. **533**(7602): p. 255-9.
137. Turroni, F., et al., *Expression of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in response to environmental gut conditions.* FEMS Microbiol Lett, 2014. **357**(1): p. 23-33.
138. Neville, B.A., et al., *Pro-Inflammatory Flagellin Proteins of Prevalent Motile Commensal Bacteria Are Variably Abundant in the Intestinal Microbiome of Elderly Humans.* PLOS ONE, 2013. **8**(7): p. e68919.
139. Harris, S.R., et al., *Evolution of MRSA During Hospital Transmission and Intercontinental Spread.* Science, 2010. **327**(5964): p. 469-474.
140. Chase-Topping, M., et al., *Super-shedding and the link between human infection and livestock carriage of Escherichia coli O157.* Nat Rev Micro, 2008. **6**(12): p. 904-912.
141. Flo, T.H., et al., *Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron.* Nature, 2004. **432**(7019): p. 917-921.
142. Stecher, B., et al., *Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota.* PLoS Biol, 2007. **5**(10): p. 2177-89.
143. Rivera-Chávez, F., et al., *Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella.* Cell Host & Microbe, 2016. **19**(4): p. 443-454.
144. Buffie, C.G., et al., *Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile.* Nature, 2015. **517**(7533): p. 205-8.
145. Atarashi, K., et al., *Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota.* Nature, 2013. **500**(7461): p. 232-6.
146. Browne, H.P., et al., *Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation.* Nature, 2016. **533**(7604): p. 543-546.
147. Flint, H.J., et al., *Interactions and competition within the microbial community of the human colon: links between diet and health.* Environ Microbiol, 2007. **9**(5): p. 1101-11.
148. Tally, F.P., et al., *Oxygen tolerance of fresh clinical anaerobic bacteria.* J Clin Microbiol, 1975. **1**(2): p. 161-4.
149. Miller, R.A. and B.E. Britigan, *Role of oxidants in microbial pathophysiology.* Clin Microbiol Rev, 1997. **10**(1): p. 1-18.
150. Khan, M.T., et al., *The gut anaerobe Faecalibacterium prausnitzii uses an extracellular electron shuttle to grow at oxic-anoxic interphases.* ISME J, 2012. **6**(8): p. 1578-1585.
151. Donaldson, G.P., S.M. Lee, and S.K. Mazmanian, *Gut biogeography of the bacterial microbiota.* Nature Reviews Microbiology, 2015. **14**(1): p. 20-32.
152. Albenberg, L., et al., *Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota.* Gastroenterology, 2014. **147**(5): p. 1055-63 e8.
153. Khan, M.T., J.M. van Dijl, and H.J.M. Harmsen, *Antioxidants Keep the Potentially Probiotic but Highly Oxygen-Sensitive Human Gut Bacterium Faecalibacterium prausnitzii Alive at Ambient Air.* PLOS ONE, 2014. **9**(5): p. e96097.
154. Li, L., et al., *The importance of the viable but non-culturable state in human bacterial pathogens.* Front Microbiol, 2014. **5**: p. 258.

155. Signoretto, C., et al., *Cell wall chemical composition of Enterococcus faecalis in the viable but nonculturable state.* Appl Environ Microbiol, 2000. **66**(5): p. 1953-9.
156. Rittershaus, Emily S.C., S.-H. Baek, and Christopher M. Sassetti, *The Normalcy of Dormancy: Common Themes in Microbial Quiescence.* Cell Host & Microbe, 2013. **13**(6): p. 643-651.
157. Senoh, M., et al., *Conversion of viable but nonculturable Vibrio cholerae to the culturable state by co-culture with eukaryotic cells.* Microbiol Immunol, 2010. **54**(9): p. 502-7.
158. Amel, B.K., B. Amine, and B. Amina, *Survival of Vibrio fluvialis in seawater under starvation conditions.* Microbiol Res, 2008. **163**(3): p. 323-8.
159. Xu, H.S., et al., *Survival and viability of nonculturableEscherichia coli andVibrio cholerae in the estuarine and marine environment.* Microb Ecol, 1982. **8**(4): p. 313-23.
160. Gupte, A.R., C.L. De Rezende, and S.W. Joseph, *Induction and resuscitation of viable but nonculturable Salmonella enterica serovar typhimurium DT104.* Appl Environ Microbiol, 2003. **69**(11): p. 6669-75.
161. Dworkin, J. and I.M. Shah, *Exit from dormancy in microbial organisms.* Nat Rev Microbiol, 2010. **8**(12): p. 890-6.
162. Bäckhed, F., et al., *Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life.* Cell Host & Microbe, 2015. **17**(5): p. 690-703.
163. Dominguez-Bello, M.G., et al., *Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.* Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11971-5.
164. Chu, D.M., et al., *Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery.* Nature Medicine, 2017. **23**(3): p. 314-326.
165. Blaser, M.J. and S. Falkow, *What are the consequences of the disappearing human microbiota?* Nat Rev Microbiol, 2009. **7**(12): p. 887-94.
166. Arrieta, M.-C., et al., *The Intestinal Microbiome in Early Life: Health and Disease.* Frontiers in Immunology, 2014. **5**.
167. Schloss, P.D., et al., *The dynamics of a family's gut microbiota reveal variations on a theme.* Microbiome, 2014. **2**(1): p. 25.
168. Song, S.J., et al., *Cohabiting family members share microbiota with one another and with their dogs.* Elife, 2013. **2**: p. e00458.
169. Nayfach, S., et al., *An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography.* Genome Res, 2016. **26**(11): p. 1612-1625.
170. Faith, J.J., J.-F. Colombel, and J.I. Gordon, *Identifying strains that contribute to complex diseases through the study of microbial inheritance.* Proceedings of the National Academy of Sciences, 2015. **112**(3): p. 633-640.
171. Tung, J., et al., *Social networks predict gut microbiome composition in wild baboons.* Elife, 2015. **4**.
172. Moeller, A.H., et al., *Social behavior shapes the chimpanzee pan-microbiome.* Science Advances, 2016. **2**(1): p. e1500997-e1500997.
173. De Filippo, C., et al., *Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa.* Proceedings of the National Academy of Sciences, 2010. **107**(33): p. 14691-14696.

174. Heikkila, M.P. and P.E.J. Saris, *Inhibition of Staphylococcus aureus by the commensal bacteria of human milk.* Journal of Applied Microbiology, 2003. **95**(3): p. 471-478.
175. Fernández, L., et al., *The human milk microbiota: Origin and potential roles in health and disease.* Pharmacological Research, 2013. **69**(1): p. 1-10.
176. Derrien, M. and J.E. van Hylckama Vlieg, *Fate, activity, and impact of ingested bacteria within the human gut microbiota.* Trends Microbiol, 2015. **23**(6): p. 354-66.
177. Jost, T., et al., *Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding.* Environ Microbiol, 2014. **16**(9): p. 2891-904.
178. Vazquez-Torres, A., et al., *Extraintestinal dissemination of Salmonella by CD18-expressing phagocytes.* Nature, 1999. **401**(6755): p. 804-8.
179. Martin, R., et al., *The commensal microflora of human milk: new perspectives for food bacteriotherapy and probiotics.* Trends in Food Science & Technology, 2004. **15**(3-4): p. 121-127.
180. Lang, J.M., J.A. Eisen, and A.M. Zivkovic, *The microbes we eat: abundance and taxonomy of microbes consumed in a day's worth of meals for three diet types.* PeerJ, 2014. **2**: p. e659.
181. David, L.A., et al., *Diet rapidly and reproducibly alters the human gut microbiome.* Nature, 2014. **505**(7484): p. 559-63.
182. Leff, J.W. and N. Fierer, *Bacterial communities associated with the surfaces of fresh fruits and vegetables.* PLoS One, 2013. **8**(3): p. e59310.
183. Brito, I.L., et al., *Mobile genes in the human microbiome are structured from global to individual scales.* Nature, 2016. **535**(7612): p. 435-439.
184. Hill, C., et al., *The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic.* Nature Reviews Gastroenterology &Amp; Hepatology, 2014. **11**: p. 506.
185. Kalliomaki, M., et al., *Probiotics in primary prevention of atopic disease: a randomised placebo-controlled trial.* Lancet, 2001. **357**(9262): p. 1076-9.
186. Allen, S.J., et al., *Probiotics for treating acute infectious diarrhoea*, in *Cochrane Database of Systematic Reviews*. 2010, Wiley-Blackwell.
187. Hill, C., et al., *Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic.* Nat Rev Gastroenterol Hepatol, 2014. **11**(8): p. 506-14.
188. Panigrahi, P., et al., *A randomized synbiotic trial to prevent sepsis among infants in rural India.* Nature, 2017. **548**(7668): p. 407-412.
189. Malinen, E., et al., *PCR-ELISA II: Analysis of Bifidobacterium populations in human faecal samples from a consumption trial with Bifidobacterium lactis Bb-12 and a galacto-oligosaccharide preparation.* Syst Appl Microbiol, 2002. **25**(2): p. 249-58.
190. Charbonneau, D., R.D. Gibb, and E.M. Quigley, *Fecal excretion of Bifidobacterium infantis 35624 and changes in fecal microbiota after eight weeks of oral supplementation with encapsulated probiotic.* Gut Microbes, 2013. **4**(3): p. 201-11.
191. McNulty, N.P., et al., *The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins.* Sci Transl Med, 2011. **3**(106): p. 106ra106.

192. Maldonado-Gómez, María X., et al., *Stable Engraftment of Bifidobacterium longum AH1206 in the Human Gut Depends on Individualized Features of the Resident Microbiome.* Cell Host & Microbe, 2016. **20**(4): p. 515-526.
193. Reid, G., et al., *Responders and non-responders to probiotic interventions: how can we improve the odds?* Gut Microbes, 2010. **1**(3): p. 200-4.
194. Sanders, M.E., et al., *An update on the use and investigation of probiotics in health and disease.* Gut, 2013. **62**(5): p. 787-96.
195. Natchu, U.c.m. and S. Bhatnagar, *Diarrhoea in children: identifying the cause and burden.* The Lancet, 2013. **382**(9888): p. 184-186.
196. Tallon, P., et al., *Microbial Indicators of Faecal Contamination in Water: A Current Perspective.* Water, Air, and Soil Pollution, 2005. **166**(1): p. 139-166.
197. Koskey, A.M., et al., *Blautia and Prevotella sequences distinguish human and animal fecal pollution in Brazil surface waters.* Environmental Microbiology Reports, 2014. **6**(6): p. 696-704.
198. Handl, S., et al., *Massive parallel 16S rRNA gene pyrosequencing reveals highly diverse fecal bacterial and fungal communities in healthy dogs and cats.* FEMS Microbiol Ecol, 2011. **76**(2): p. 301-10.
199. Hand, D., et al., *Pyrosequencing the canine faecal microbiota: breadth and depth of biodiversity.* PLoS One, 2013. **8**(1): p. e53115.
200. Lamendella, R., et al., *Comparative fecal metagenomics unveils unique functional capacity of the swine gut.* BMC Microbiol, 2011. **11**: p. 103.
201. Moller-Stray, J., et al., *Two outbreaks of diarrhoea in nurseries in Norway after farm visits, April to May 2009.* Euro Surveill, 2012. **17**(47).
202. Toro, M., et al., *Whole-Genome Sequencing Analysis of Salmonella enterica Serovar Enteritidis Isolates in Chile Provides Insights into Possible Transmission between Gulls, Poultry, and Humans.* Appl Environ Microbiol, 2016. **82**(20): p. 6223-6232.
203. Iovine, N.M. and M.J. Blaser, *Antibiotics in Animal Feed and Spread of Resistant Campylobacter from Poultry to Humans.* Emerging Infectious Diseases, 2004. **10**(6): p. 1158-1189.
204. Gupta, A., et al., *Antimicrobial Resistance among Campylobacter Strains, United States, 1997–2001.* Emerging Infectious Diseases, 2004. **10**(6): p. 1102-1109.
205. Dunn, R.R., et al., *Home life: factors structuring the bacterial diversity found within and between homes.* PLoS One, 2013. **8**(5): p. e64133.
206. Lax, S., C.R. Nagler, and J.A. Gilbert, *Our interface with the built environment: immunity and the indoor microbiota.* Trends Immunol, 2015. **36**(3): p. 121-3.
207. Meadow, J.F., et al., *Humans differ in their personal microbial cloud.* PeerJ, 2015. **3**: p. e1258.
208. Hsu, T., et al., *Urban Transit System Microbial Communities Differ by Surface Type and Interaction with Humans and the Environment.* mSystems, 2016. **1**(3).
209. Kelley, S.T. and J.A. Gilbert, *Studying the microbiology of the indoor environment.* Genome Biol, 2013. **14**(2): p. 202.
210. Kembel, S.W., et al., *Architectural design influences the diversity and structure of the built environment microbiome.* ISME J, 2012. **6**(8): p. 1469-79.

211. Barker, J. and M.V. Jones, *The potential spread of infection caused by aerosol contamination of surfaces after flushing a domestic toilet.* Journal of Applied Microbiology, 2005. **99**(2): p. 339-347.
212. Perkins, S.D., et al., *Potentially pathogenic bacteria in shower water and air of a stem cell transplant unit.* Appl Environ Microbiol, 2009. **75**(16): p. 5363-72.
213. Flores, G.E., et al., *Microbial biogeography of public restroom surfaces.* PLoS One, 2011. **6**(11): p. e28132.
214. Gibbons, S.M., et al., *Ecological succession and viability of human-associated microbiota on restroom surfaces.* Appl Environ Microbiol, 2015. **81**(2): p. 765-73.
215. Snelling, A.M., et al., *Comparative evaluation of the hygienic efficacy of an ultra-rapid hand dryer vs conventional warm air hand dryers.* Journal of Applied Microbiology, 2010. **110**(1): p. 19-26.
216. Vaishampayan, P., et al., *New perspectives on viable microbial communities in low-biomass cleanroom environments.* ISME J, 2013. **7**(2): p. 312-24.
217. Faith, J.J., et al., *The Long-Term Stability of the Human Gut Microbiota.* Science, 2013. **341**(6141).
218. David, L.A., et al., *Host lifestyle affects human microbiota on daily timescales.* Genome Biol, 2014. **15**(7): p. R89.
219. Korpela, K., et al., *Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children.* Nature Communications, 2016. **7**: p. 10410.
220. Claesson, M.J., et al., *Gut microbiota composition correlates with diet and health in the elderly.* Nature, 2012. **488**(7410): p. 178-184.
221. Costello, E.K., et al., *The application of ecological theory toward an understanding of the human microbiome.* Science, 2012. **336**(6086): p. 1255-62.
222. Lee, S.M., et al., *Bacterial colonization factors control specificity and stability of the gut microbiota.* Nature, 2013. **501**(7467): p. 426-9.
223. Lawley, T.D., et al., *Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice.* PLoS Pathog, 2012. **8**(10): p. e1002995.
224. Clemente, J.C., et al., *The microbiome of uncontacted Amerindians.* Sci Adv, 2015. **1**(3).
225. Sonnenburg, E.D., et al., *Diet-induced extinctions in the gut microbiota compound over generations.* Nature, 2016. **529**(7585): p. 212-215.
226. Blaser, M.J., *Antibiotic use and its consequences for the normal microbiome.* Science, 2016. **352**(6285): p. 544-5.
227. Martinez, I., et al., *The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes.* Cell Rep, 2015. **11**(4): p. 527-38.
228. Strachan, D.P., *Hay fever, hygiene, and household size.* BMJ : British Medical Journal, 1989. **299**(6710): p. 1259-1260.
229. Beasley, R., *Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC.* The Lancet, 1998. **351**(9111): p. 1225-1232.
230. Stiemsma, L., et al., *The hygiene hypothesis: current perspectives and future therapies.* ImmunoTargets and Therapy, 2015: p. 143.
231. Gomez de Aguero, M., et al., *The maternal microbiota drives early postnatal innate immune development.* Science, 2016. **351**(6279): p. 1296-302.

232. Blaser, M.J., *Who are we? Indigenous microbes and the ecology of human diseases.* EMBO Rep, 2006. **7**(10): p. 956-60.
233. Shaw, S.Y., J.F. Blanchard, and C.N. Bernstein, *Association Between the Use of Antibiotics in the First Year of Life and Pediatric Inflammatory Bowel Disease.* The American Journal of Gastroenterology, 2010. **105**(12): p. 2687-2692.
234. Dominguez-Bello, M.G., et al., *Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer.* Nat Med, 2016. **22**(3): p. 250-3.
235. van Nood, E., et al., *Duodenal infusion of donor feces for recurrent Clostridium difficile.* N Engl J Med, 2013. **368**(5): p. 407-15.
236. Rettedal, E.A., H. Gumpert, and M.O. Sommer, *Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria.* Nat Commun, 2014. **5**: p. 4714.
237. Lau, J.T., et al., *Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling.* Genome Medicine, 2016. **8**(1): p. 72.
238. Czaplewski, L., et al., *Alternatives to antibiotics—a pipeline portfolio review.* The Lancet Infectious Diseases, 2016. **16**(2): p. 239-251.
239. Yee, A.L. and J.A. Gilbert, *MICROBIOME. Is triclosan harming your microbiome?* Science, 2016. **353**(6297): p. 348-9.
240. Vandegrift, R., et al., *Cleanliness in context: reconciling hygiene with a modern microbial perspective*. 2016, Cold Spring Harbor Laboratory Press.
241. Leibold, M.A., et al., *The metacommunity concept: a framework for multi-scale community ecology.* Ecology Letters, 2004. **7**(7): p. 601-613.
242. Shade, A., et al., *Fundamentals of microbial community resistance and resilience.* Front Microbiol, 2012. **3**: p. 417.
243. Locey, K.J., *Synthesizing traditional biogeography with microbial ecology: the importance of dormancy.* Journal of Biogeography, 2010. **37**(10): p. 1835-1841.
244. Shafquat, A., et al., *Functional and phylogenetic assembly of microbial communities in the human microbiome.* Trends Microbiol, 2014. **22**(5): p. 261-6.
245. Nemergut, D.R., et al., *Patterns and processes of microbial community assembly.* Microbiol Mol Biol Rev, 2013. **77**(3): p. 342-56.
246. Nemergut, D.R., et al., *Global patterns in the biogeography of bacterial taxa.* Environmental Microbiology, 2011. **13**(1): p. 135-144.
247. Cani, P.D., *Gut microbiota [mdash] at the intersection of everything?* Nat Rev Gastroenterol Hepatol, 2017. **14**(6): p. 321-322.
248. Hanage, W.P., *Microbiology: Microbiome science needs a healthy dose of scepticism.* Nature, 2014. **512**(7514): p. 247-8.
249. Whitaker, W.R., E.S. Shepherd, and J.L. Sonnenburg, *Tunable Expression Tools Enable Single-Cell Strain Distinction in the Gut Microbiome.* Cell, 2017. **169**(3): p. 538-546.e12.
250. Falkow, S., *Molecular Koch's postulates applied to bacterial pathogenicity [mdash] a personal recollection 15 years later.* Nat Rev Micro, 2004. **2**(1): p. 67-72.
251. Kelly , C.P. and J.T. LaMont *Clostridium difficile — More Difficult Than Ever.* New England Journal of Medicine, 2008. **359**(18): p. 1932-1940.

252. Jalanka, J., et al., *Long-term effects on luminal and mucosal microbiota and commonly acquired taxa in faecal microbiota transplantation for recurrent Clostridium difficile infection.* BMC Medicine, 2016. **14**(1).
253. Li, S.S., et al., *Durable coexistence of donor and recipient strains after fecal microbiota transplantation.* Science, 2016. **352**(6285): p. 586-589.
254. Garber, K., *Drugging the gut microbiome.* Nat Biotech, 2015. **33**(3): p. 228-231.
255. Petrof, E.O., et al., *Stool substitute transplant therapy for the eradication of Clostridium difficile infection: 'RePOOPulating' the gut.* Microbiome, 2013. **1**(1): p. 3.
256. Spanogiannopoulos, P., et al., *The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism.* Nat Rev Micro, 2016. **14**(5): p. 273-287.
257. Cohen, L.J., et al., *Commensal bacteria make GPCR ligands that mimic human signalling molecules.* Nature, 2017. **549**(7670): p. 48-53.
258. Wang, Q., et al., *Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.* Appl Environ Microbiol, 2007. **73**(16): p. 5261-7.
259. Garrity, G.M., J.A. Bell, and T.G. Lilburn, *Taxonomic outline of the prokaryotes. Bergey's manual of systematic bacteriology.* Springer, New York, Berlin, Heidelberg, 2004.
260. Ludwig, W., et al., *ARB: a software environment for sequence data.* Nucleic Acids Res, 2004. **32**(4): p. 1363-71.
261. Schloss, P.D., et al., *Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.* Appl Environ Microbiol, 2009. **75**(23): p. 7537-41.
262. Bosshard, P.P., et al., *Ribosomal DNA sequencing for identification of aerobic gram-positive rods in the clinical laboratory (an 18-month evaluation).* J Clin Microbiol, 2003. **41**(9): p. 4134-40.
263. Clarridge, J.E., 3rd, *Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases.* Clin Microbiol Rev, 2004. **17**(4): p. 840-62, table of contents.
264. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
265. Page, A.J., et al., *Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data.* Microbial Genomics, 2016. **2**(8).
266. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs.* Genome Res, 2008. **18**(5): p. 821-9.
267. Boetzer, M. and W. Pirovano, *Toward almost closed genomes with GapFiller.* Genome Biol, 2012. **13**(6): p. R56.
268. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9.
269. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification.* BMC Bioinformatics, 2010. **11**: p. 119.
270. Dimmer, E.C., et al., *The UniProt-GO Annotation database in 2011.* Nucleic Acids Res, 2012. **40**(Database issue): p. D565-70.
271. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Research, 2016. **44**(D1): p. D733-D745.

272. Punta, M., et al., *The Pfam protein families database.* Nucleic Acids Res, 2012. **40**(Database issue): p. D290-301.
273. Haft, D.H., et al., *TIGRFAMs and Genome Properties in 2013.* Nucleic Acids Research, 2013. **41**(D1): p. D387-D395.
274. Lagesen, K., et al., *RNAmmer: consistent and rapid annotation of ribosomal RNA genes.* Nucleic Acids Res, 2007. **35**(9): p. 3100-8.
275. Laslett, D. and B. Canback, *ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences.* Nucleic Acids Res, 2004. **32**(1): p. 11-6.
276. Petersen, T.N., et al., *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nat Methods, 2011. **8**(10): p. 785-6.
277. Kolbe, D.L. and S.R. Eddy, *Fast filtering for RNA homology search.* Bioinformatics, 2011. **27**(22): p. 3102-3109.
278. Forster, S.C., et al., *HPMCD: the database of human microbial communities from metagenomic datasets and microbial reference genomes.* Nucleic Acids Res, 2015.
279. Mitchell, A., et al., *EBI metagenomics in 2016 - an expanding and evolving resource for the analysis and archiving of metagenomic data.* Nucleic Acids Research, 2016. **44**(D1): p. D595-D603.
280. Wood, D.E. and S.L. Salzberg, *Kraken: ultrafast metagenomic sequence classification using exact alignments.* Genome Biol, 2014. **15**(3): p. R46.
281. Kozich, J.J., et al., *Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform.* Appl Environ Microbiol, 2013. **79**(17): p. 5112-20.
282. Lagier, J.C., et al., *Microbial culturomics: paradigm shift in the human gut microbiome study.* Clin Microbiol Infect, 2012. **18**(12): p. 1185-93.
283. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree 2--approximately maximum-likelihood trees for large alignments.* PLoS One, 2010. **5**(3): p. e9490.
284. Letunic, I. and P. Bork, *Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W475-8.
285. Katoh, K., et al., *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.* Nucleic Acids Research, 2002. **30**(14): p. 3059-3066.
286. Galtier, N., M. Gouy, and C. Gautier, *SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny.* Comput Appl Biosci, 1996. **12**(6): p. 543-8.
287. Kumar, S., et al., *MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences.* Briefings in Bioinformatics, 2008. **9**(4): p. 299-306.
288. Sunagawa, S., et al., *Metagenomic species profiling using universal phylogenetic marker genes.* Nat Meth, 2013. **10**(12): p. 1196-1199.
289. Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.* Nucleic Acids Research, 2007. **35**(Database issue): p. D61-D65.
290. Castresana, J., *Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.* Mol Biol Evol, 2000. **17**(4): p. 540-52.

291. Lawley, T.D., et al., *Proteomic and genomic characterization of highly infectious Clostridium difficile 630 spores.* J Bacteriol, 2009. **191**(17): p. 5377-86.
292. Richter, M. and R. Rosselló-Móra, *Shifting the genomic gold standard for the prokaryotic species definition.* Proceedings of the National Academy of Sciences, 2009. **106**(45): p. 19126-19131.
293. Li, L., C.J. Stoeckert, and D.S. Roos, *OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes.* Genome Research, 2003. **13**(9): p. 2178-2189.
294. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Research, 2004. **32**(5): p. 1792-1797.
295. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: The European Molecular Biology Open Software Suite.* Trends in Genetics, 2000. **16**(6): p. 276-277.
296. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
297. Riley, M., *Functions of the gene products of Escherichia coli.* Microbiol Rev, 1993. **57**(4): p. 862-952.
298. Langille, M.G.I., et al., *Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.* Nature biotechnology, 2013. **31**(9): p. 814-821.
299. Galperin, M.Y. and E.V. Koonin, *From complete genome sequence to "complete" understanding?* Trends in biotechnology, 2010. **28**(8): p. 398-406.
300. Kim, Y.-G., et al., *Neonatal acquisition of Clostridia species protects against colonization by bacterial pathogens.* Science, 2017. **356**(6335): p. 315-319.
301. Lawley, T.D., et al., *Antibiotic treatment of clostridium difficile carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts.* Infect Immun, 2009. **77**(9): p. 3661-9.
302. Riley, T.V., et al., *Comparison of alcohol shock enrichment and selective enrichment for the isolation of Clostridium difficile.* Epidemiol Infect, 1987. **99**(2): p. 355-9.
303. Wilson, K.H., M.J. Kennedy, and F.R. Fekety, *Use of sodium taurocholate to enhance spore recovery on a medium selective for Clostridium difficile.* Journal of Clinical Microbiology, 1982. **15**(3): p. 443-446.
304. Cole, J.R., et al., *Ribosomal Database Project: data and tools for high throughput rRNA analysis.* Nucleic Acids Research, 2014. **42**(Database issue): p. D633-D642.
305. Fodor, A.A., et al., *The "most wanted" taxa from the human microbiome for whole genome sequencing.* PLoS One, 2012. **7**(7): p. e41294.
306. Tap, J., et al., *Towards the human intestinal microbiota phylogenetic core.* Environmental Microbiology, 2009. **11**(10): p. 2574-2584.
307. Lau, S.K.P., et al., *Catabacter hongkongensis gen. nov., sp. nov., Isolated from Blood Cultures of Patients from Hong Kong and Canada.* Journal of Clinical Microbiology, 2007. **45**(2): p. 395-401.
308. Yarza, P., et al., *Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences.* Nat Rev Micro, 2014. **12**(9): p. 635-645.

309. Bosshard, P.P., et al., *Ribosomal DNA Sequencing for Identification of Aerobic Gram-Positive Rods in the Clinical Laboratory (an 18-Month Evaluation).* Journal of Clinical Microbiology, 2003. **41**(9): p. 4134-4140.
310. Lee, Y.-J., et al., *Gracilibacter thermotolerans gen. nov., sp. nov., an anaerobic, thermotolerant bacterium from a constructed wetland receiving acid sulfate water.* International Journal of Systematic and Evolutionary Microbiology, 2006. **56**(9): p. 2089-2093.
311. Morotomi, M., F. Nagai, and Y. Watanabe, *Description of Christensenella minuta gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order Clostridiales, and proposal of Christensenellaceae fam. nov.* Int J Syst Evol Microbiol, 2012. **62**(Pt 1): p. 144-9.
312. Frick, J.S. and D. Haller, *Intestinal microbiota: From sequencing to function.* Int J Med Microbiol, 2016.
313. Bosshard, P.P., R. Zbinden, and M. Altwegg, *Turicibacter sanguinis gen. nov., sp. nov., a novel anaerobic, Gram-positive bacterium.* Int J Syst Evol Microbiol, 2002. **52**(Pt 4): p. 1263-6.
314. Reynolds, S.A., F. Levy, and E.S. Walker, *Hand Sanitizer Alert.* Emerging Infectious Diseases, 2006. **12**(3): p. 527-529.
315. Onyenwoke, R.U., et al., *Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes).* Arch Microbiol, 2004. **182**(2-3): p. 182-92.
316. Lee, Y.-J., et al., *Thermosediminibacter oceani gen. nov., sp. nov. and Thermosediminibacter litoriperuensis sp. nov., new anaerobic thermophilic bacteria isolated from Peru Margin.* Extremophiles, 2005. **9**(5): p. 375-383.
317. Custovic, A., S.C. Taggart, and A. Woodcock, *House dust mite and cat allergen in different indoor environments.* Clin Exp Allergy, 1994. **24**(12): p. 1164-8.
318. Leung, M.H. and P.K. Lee, *The roles of the outdoors and occupants in contributing to a potential pan-microbiome of the built environment: a review.* Microbiome, 2016. **4**(1): p. 21.
319. Ruiz-Calderon, J.F., et al., *Walls talk: Microbial biogeography of homes spanning urbanization.* Science Advances, 2016. **2**(2): p. e1501061.
320. Obregon-Tito, A.J., et al., *Subsistence strategies in traditional societies distinguish gut microbiomes.* Nat Commun, 2015. **6**: p. 6505.
321. Ludwig, W., K.-H. Schleifer, and W.B. Whitman, *Revised road map to the phylum Firmicutes*, in *Bergey's Manual® of Systematic Bacteriology: Volume Three The Firmicutes*, P. De Vos, et al., Editors. 2009, Springer New York: New York, NY. p. 1-13.
322. Marchandin, H., et al., *Negativicoccus succinicivorans gen. nov., sp. nov., isolated from human clinical samples, emended description of the family Veillonellaceae and description of Negativicutes classis nov., Selenomonadales ord. nov. and Acidaminococcaceae fam. nov. in the bacterial phylum Firmicutes.* International Journal of Systematic and Evolutionary Microbiology, 2010. **60**(6): p. 1271-1279.
323. Sokolova, T., et al., *Novel chemolithotrophic, thermophilic, anaerobic bacteria Thermolithobacter ferrireducens gen. nov., sp. nov. and Thermolithobacter carboxydivorans sp. nov.* Extremophiles, 2007. **11**(1): p. 145-157.
324. Ogawa, Y., et al., *The Genome of Erysipelothrix rhusiopathiae, the Causative Agent of Swine Erysipelas, Reveals New Insights into the Evolution of*

Firmicutes and the Organism's Intracellular Adaptations. Journal of Bacteriology, 2011. **193**(12): p. 2959-2971.
325. Ciccarelli, F.D., et al., *Toward Automatic Reconstruction of a Highly Resolved Tree of Life.* Science, 2006. **311**(5765): p. 1283-1287.
326. Antunes, L.C.S., et al., *Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the Firmicutes.* eLife, 2016. **5**: p. e14589.
327. Rainey, F.A., et al., *The taxonomic status of the fermentative halophilic anaerobic bacteria: description of Haloanaerobiales ord. nov., Halobacteroidaceae fam. nov., Orenia gen. nov. and further taxonomic rearrangements at the genus and species level.* Anaerobe, 1995. **1**(4): p. 185-99.
328. Oren, A., H. Pohla, and E. Stackebrandt, *Transfer of Clostridium lortetii to a New Genus Sporohalobacter gen. nov. as Sporohalobacter lortetii comb. nov., and Description of Sporohalobacter marismortui sp. nov.* Systematic and Applied Microbiology, 1987. **9**(3): p. 239-246.
329. Zhilina, T.N., et al., *[Halonatronum saccharophilum gen. nov. sp. nov--a new haloalkalophilic bacteria from the order Haloanaerobiales from Lake Magadi]*. Mikrobiologiia, 2001. **70**(1): p. 77-85.
330. Crapart, S., et al., *Exiguobacterium profundum sp. nov., a moderately thermophilic, lactic acid-producing bacterium isolated from a deep-sea hydrothermal vent.* International Journal of Systematic and Evolutionary Microbiology, 2007. **57**(2): p. 287-292.
331. FARROW, J.A.E., S. WALLBANKS, and M.D. COLLINS, *Phylogenetic Interrelationships of Round-Spore-Forming Bacilli Containing Cell Walls Based on Lysine and the Non-Spore-Forming Genera Caryophanon, Exiguobacterium, Kurthia, and Planococcus.* International Journal of Systematic and Evolutionary Microbiology, 1994. **44**(1): p. 74-82.
332. Nayfach, S. and K.S. Pollard, *Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome.* Genome Biology, 2015. **16**(1): p. 51.
333. Kitano, H., *Biological robustness.* Nat Rev Genet, 2004. **5**(11): p. 826-837.
334. Postma, P.W., J.W. Lengeler, and G.R. Jacobson, *Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria.* Microbiological Reviews, 1993. **57**(3): p. 543-594.
335. Capra, E.J. and M.T. Laub, *The Evolution of Two-Component Signal Transduction Systems.* Annual review of microbiology, 2012. **66**: p. 325-347.
336. Antunes, A., et al., *Global transcriptional control by glucose and carbon regulator CcpA in Clostridium difficile.* Nucleic Acids Research, 2012. **40**(21): p. 10701-10718.
337. Davidson, A.L. and J. Chen, *ATP-Binding Cassette Transporters in Bacteria.* Annual Review of Biochemistry, 2004. **73**(1): p. 241-268.
338. Maughan, H., C.W. Birky, Jr., and W.L. Nicholson, *Transcriptome divergence and the loss of plasticity in Bacillus subtilis after 6,000 generations of evolution under relaxed selection for sporulation.* J Bacteriol, 2009. **191**(1): p. 428-33.
339. Makarova, K., et al., *Comparative genomics of the lactic acid bacteria.* Proceedings of the National Academy of Sciences, 2006. **103**(42): p. 15611-15616.

340. Zhou, Y., et al., *Biogeography of the ecosystems of the healthy human body.* Genome Biology, 2013. **14**(1): p. R1.
341. Costello, E.K., et al., *Bacterial Community Variation in Human Body Habitats Across Space and Time.* Science, 2009. **326**(5960): p. 1694-1697.
342. Sampaio-Maia, B. and F. Monteiro-Silva, *Acquisition and maturation of oral microbiome throughout childhood: An update.* Dental Research Journal, 2014. **11**(3): p. 291-301.
343. Kort, R., et al., *Shaping the oral microbiota through intimate kissing.* Microbiome, 2014. **2**(1): p. 41.
344. Kaakoush, N.O., *Insights into the Role of Erysipelotrichaceae in the Human Host.* Frontiers in Cellular and Infection Microbiology, 2015. **5**: p. 84.
345. Alexander, C.J., et al., *Identification and antimicrobial resistance patterns of clinical isolates of Clostridium clostridioforme, Clostridium innocuum, and Clostridium ramosum compared with those of clinical isolates of Clostridium perfringens.* Journal of Clinical Microbiology, 1995. **33**(12): p. 3209-15.
346. Janoir, C., et al., *Adaptive strategies and pathogenesis of Clostridium difficile from in vivo transcriptomics.* Infect Immun, 2013. **81**(10): p. 3757-69.
347. Mora-Uribe, P., et al., *Characterization of the Adherence of Clostridium difficile Spores: The Integrity of the Outermost Layer Affects Adherence Properties of Spores of the Epidemic Strain R20291 to Components of the Intestinal Mucosa.* Frontiers in Cellular and Infection Microbiology, 2016. **6**(99).
348. Sturm, A. and J. Dworkin, *Phenotypic Diversity as a Mechanism to Exit Cellular Dormancy.* Current Biology, 2015. **25**(17): p. 2272-2277.
349. Lennon, J.T. and S.E. Jones, *Microbial seed banks: the ecological and evolutionary implications of dormancy.* 2011. **9**: p. 119.
350. Khanna, S., et al., *A Novel Microbiome Therapeutic Increases Gut Microbial Diversity and Prevents Recurrent Clostridium difficile Infection.* J Infect Dis, 2016. **214**(2): p. 173-81.
351. Ratner, M., *Seres's pioneering microbiome drug fails mid-stage trial.* Nat Biotech, 2016. **34**(10): p. 1004-1005.
352. Lee, S.T.M., et al., *Tracking microbial colonization in fecal microbiota transplantation experiments via genome-resolved metagenomics.* Microbiome, 2017. **5**(1): p. 50.
353. Turnbaugh, P.J., et al., *The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice.* Science translational medicine, 2009. **1**(6): p. 6ra14-6ra14.
354. Collins, J., et al., *Humanized microbiota mice as a model of recurrent Clostridium difficile disease.* Microbiome, 2015. **3**(1): p. 35.
355. Wagner, V.E., et al., *Effects of a gut pathobiont in a gnotobiotic mouse model of childhood undernutrition.* Science Translational Medicine, 2016. **8**(366): p. 366ra164-366ra164.
356. Marcobal, A., et al., *A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice.* ISME J, 2013. **7**(10): p. 1933-1943.
357. Mark Welch, J.L., et al., *Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice.* Proceedings of the National Academy of Sciences, 2017.
358. Ng, K.M., et al., *Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens.* Nature, 2013. **502**(7469): p. 96-99.

359. Mullany, P., et al., *Gene Cloning in Clostridium difficile Using Tn916 as a Shuttle Conjugative Transposon.* Plasmid, 1994. **31**(3): p. 320-323.
360. Heap, J.T., et al., *The ClosTron: A universal gene knock-out system for the genus Clostridium.* Journal of Microbiological Methods, 2007. **70**(3): p. 452-464.
361. Cuív, P.Ó., et al., *Isolation of Genetically Tractable Most-Wanted Bacteria by Metaparental Mating.* Scientific Reports, 2015. **5**: p. 13282.
362. Olle, B., *Medicines from microbiota.* Nat Biotechnol, 2013. **31**(4): p. 309-15.
363. Nielsen, H.B., et al., *Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes.* Nat Biotech, 2014. **32**(8): p. 822-828.
364. Scott, K.P., et al., *The influence of diet on the gut microbiota.* Pharmacological Research, 2013. **69**(1): p. 52-60.
365. Paredes-Sabja, D., P. Setlow, and M.R. Sarker, *Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved.* Trends in Microbiology. **19**(2): p. 85-94.

Appendix 1 Cultured Isolates

The table summarises details of the isolates cultured in this study. It includes their taxonomic placement, novelty based on 16S rRNA gene sequence similarity, ethanol resistance, sporulation signature score and if they are included in the Human Microbiome Projects most wanted list of underrepresented taxa in cultured isolates. The isolates are listed according to their order on Figure 3.5 and therefore include three Proteobacteria species which were not cultured but were included to provide phylogenetic context. Hence, the total number listed here is 140 (three Proteobacteria and 137 cultured in this study).

order on Fig. 3.5	Phylum	Class	Family†	Genus†	closest 16S BLAST match to named bacterium- ≥99% unless otherwise stated	novelty	etoh resistance	spore sig. score	HMP most wanted
1	Proteobacteria	Betaproteobacteria	Oxalobacteraceae	Oxalobacter	Oxalobacter formigenes				
2	Proteobacteria	Gammaproteobacteria	Enterobacteriaceae	Proteus	Proteus mirabilis				
3	Proteobacteria	Gammaproteobacteria	Enterobacteriaceae	Escherichia	Escherichia coli				
4	Bacteroidetes	Bacteroidia	Rikenellaceae	Alistipes	Alistipes finegoldii	characterised	ethanol-sensitive	0.185	low
5	Bacteroidetes	Bacteroidia	Rikenellaceae	Alistipes	Alistipes finegoldii	characterised	ethanol-sensitive	0.1846	medium
6	Bacteroidetes	Bacteroidia	Porphyromonadaceae	Parabacteroides	Parabacteroides merdae	characterised	ethanol-sensitive	0.215	low
7	Bacteroidetes	Bacteroidia	Porphyromonadaceae	Parabacteroides	Parabacteroides distasonis	characterised	ethanol-sensitive	0.2	low
8	Bacteroidetes	Bacteroidia	Porphyromonadaceae	Parabacteroides	Parabacteroides distasonis	characterised	ethanol-	0.215	low

							sensitive		
9	Bacteroidetes	Bacteroidia	Porphyromonadaceae	Parabacteroides	Parabacteroides distasonis	characterised	ethanol-sensitive	0.2	low
10	Bacteroidetes	Bacteroidia	Prevotellaceae	Prevotella	Prevotella copri	characterised	ethanol-sensitive	0.046	low
11	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides coprocola_94%	novel species	ethanol-sensitive	0.185	no
12	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides plebius_95%	novel species	ethanol-sensitive	0.169	no
13	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides vulgatus	characterised	ethanol-sensitive	0.185	medium
14	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides uniformis	characterised	ethanol-sensitive	0.185	low
15	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides intestinalis_98%	novel species	ethanol-sensitive	0.185	low
16	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides salyersiae	characterised	ethanol-sensitive	0.185	low
17	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides caccae	characterised	ethanol-sensitive	0.2	low

18	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides xylanisolvens	characterised	ethanol-sensitive	0.2	low
19	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides ovatus	characterised	ethanol-sensitive	0.185	low
20	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides finegoldi	characterised	ethanol-sensitive	0.185	low
21	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides thetaiotaomicron	characterised	ethanol-sensitive	0.2	medium
22	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides thetaiotaomicron	characterised	ethanol-sensitive	0.2	medium
23	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides finegoldii_98%	novel species	ethanol-sensitive	0.185	low
24	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides	Bacteroides thetaiotaomicron	characterised	ethanol-sensitive	0.2	medium
25	Actinobacteria	Actinobacteria	Coriobacteriaceae	Collinsella	Collinsella aerofaciens_92%	novel species	ethanol-sensitive	0.215	no
26	Actinobacteria	Actinobacteria	Coriobacteriaceae	Collinsella	Collinsella aerofaciens	characterised	ethanol-sensitive	0.185	low
27	Actinobacteria	Actinobacteria	Coriobacteriaceae	Collinsella	Collinsella aerofaciens	characterised	ethanol-	0.2	low

							sensitive		
28	Actinobacteria	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	Bifidobacterium adolescentis	characterised	ethanol-sensitive	0.138	low
29	Actinobacteria	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	Bifidobacterium adolescentis	characterised	ethanol-sensitive	0.138	low
30	Actinobacteria	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	Bifidobacterium bifidum	characterised	ethanol-sensitive	0.138	no
31	Actinobacteria	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	Bifidobacterium pseudocatenulatum	characterised	ethanol-sensitive	0.123	low
32	Firmicutes	Negativicutes	Veillonellaceae	Mitsuokella	Mitsuokella jalaludinii	characterised	ethanol-sensitive	0.292	no
33	Firmicutes	Negativicutes	Veillonellaceae	Megasphaera	Megasphaera elsdenii_95%	novel species	ethanol-sensitive	0.262	no
34	Firmicutes	Erysipelotrichia	Erysipelotrichaceae	Turicibacter	Turicibacter sanguinis	characterised	ethanol-resistant	0.569	low
35	Firmicutes	Erysipelotrichia	Erysipelotrichaceae	Erysipelotrichaceae_incertae_sedis	Clostridium innocuum	characterised	ethanol-resistant	0.492	no
36	Firmicutes	Erysipelotrichia	Erysipelotrichaceae	Erysipelotrichaceae_incertae_sedis	Clostridium innocuum_95%	novel species	ethanol-resistant	0.477	low

37	Firmicutes	Erysipelotrichia	Erysipelotrichaceae	Catenibacterium	Catenibacterium mitsuokai	characterised	ethanol-sensitive	0.262	no
38	Firmicutes	Erysipelotrichia	Erysipelotrichaceae	Clostridium XVIII	Clostridium cocleatum_93%	novel species	ethanol-sensitive	0.415	no
39	Firmicutes	Erysipelotrichia	Erysipelotrichaceae	Clostridium XVIII	Clostridium saccharogumia_93%	novel species	ethanol-sensitive	0.415	low
40	Firmicutes	Clostridia	unclassified		Eubacterium infirmum_91%	novel family 1	ethanol-resistant	0.538	no
41	Firmicutes	Clostridia	unclassified		Eubacterium infirmum_94%	novel family 1	ethanol-resistant	0.662	no
42	Firmicutes	Clostridia	Peptostreptococcaceae	Clostridium XI	Clostridium ghonii_98%	novel species	ethanol-resistant	0.831	no
43	Firmicutes	Clostridia	Peptostreptococcaceae	Clostridium XI	Clostridium lituseburense_98%	novel species	ethanol-resistant	0.862	medium
44	Firmicutes	Clostridia	Peptostreptococcaceae	Clostridium XI	Clostridium bartlettii	characterised	ethanol-resistant	0.877	low
45	Firmicutes	Clostridia	Clostridiaceae	Sarcina	Sarcina ventriculi	characterised	ethanol-resistant	0.692	no
46	Firmicutes	Clostridia	Clostridiaceae	Clostridium sensu	Clostridium baratti	characterised	ethanol-	0.769	no

				stricto			resistant		
47	Firmicutes	Clostridia	Clostridiaceae	Clostridium sensu stricto	Clostridium paraputrificum	characterised	ethanol-resistant	0.769	no
48	Firmicutes	Clostridia	Clostridiaceae	Clostridium sensu stricto	Clostridium disporicum	characterised	ethanol-resistant	0.723	no
49	Firmicutes	Clostridia	Clostridiaceae	Clostridium sensu stricto	Clostridium disporicum	characterised	ethanol-resistant	0.785	medium
50	Firmicutes	Clostridia	Clostridiaceae	Clostridium sensu stricto	Clostridium perfringens	characterised	ethanol-resistant	0.8	medium
51	Firmicutes	Clostridia	Clostridiaceae	Clostridium sensu stricto	Clostridium disporicum_98%	novel species	ethanol-resistant	0.8	medium
52	Firmicutes	Clostridia	unclassified		Clostridium thermocellum_86%	novel family 2	ethanol-resistant	0.646	no
53	Firmicutes	Clostridia	Ruminococcaceae	Acetivibrio	Clostridium thermocellum_87%	novel genus	ethanol-resistant	0.662	high
54	Firmicutes	Clostridia	Ruminococcaceae	Butyricicoccus	Butyricicoccus pullicaecorum_94%	novel species	ethanol-resistant	0.569	no
55	Firmicutes	Clostridia	Ruminococcaceae	Butyricicoccus	Butyricicoccus pullicaecorum_94%	novel species	ethanol-resistant	0.569	high

56	Firmicutes	Clostridia	Ruminococcaceae	Flavonifractor	Flavonifractor plautii_95%	novel species	ethanol-sensitive	0.662	medium
57	Firmicutes	Clostridia	Ruminococcaceae	unclassified	Flavonifractor plautii_95%	novel genus	ethanol-resistant	0.585	low
58	Firmicutes	Clostridia	Ruminococcaceae	unclassified	Flavonifractor plautii_94%	novel species	ethanol-resistant	0.646	medium
59	Firmicutes	Clostridia	Ruminococcaceae	unclassified	Flavonifractor plautii_96%	novel genus	ethanol-resistant	0.677	no
60	Firmicutes	Clostridia	Ruminococcaceae	Oscillibacter	Oscillibacter valericigenes_96%	novel species	ethanol-resistant	0.646	low
61	Firmicutes	Clostridia	Ruminococcaceae	Pseudoflavonifractor	Flavonifractor plautii_95%	novel species	ethanol-sensitive	0.646	no
62	Firmicutes	Clostridia	Ruminococcaceae	unclassified	Flavonifractor plautii_97%	novel genus	ethanol-resistant	0.62	medium
63	Firmicutes	Clostridia	Ruminococcaceae	Flavonifractor	Flavonifractor plautii	characterised	ethanol-resistant	0.723	low
64	Firmicutes	Clostridia	Ruminococcaceae	Flavonifractor	Flavonifractor plautii	characterised	ethanol-resistant	0.692	low
65	Firmicutes	Clostridia	Ruminococcaceae	Flavonifractor	Flavonifractor plautii	characterised	ethanol-	0.692	low

							resistant		
66	Firmicutes	Clostridia	Ruminococcaceae	Flavonifractor	Flavonifractor plautii	characterised	ethanol-resistant	0.508	low
67	Firmicutes	Clostridia	Ruminococcaceae	Clostridium IV	Eubacterium siraeum	characterised	ethanol-sensitive	0.477	low
68	Firmicutes	Clostridia	Ruminococcaceae	Clostridium IV	Ruminococcus bromii	characterised	ethanol-sensitive	0.569	low
69	Firmicutes	Clostridia	Ruminococcaceae	Clostridium IV	Ruminococcus bromii_93%	novel species	ethanol-resistant	0.585	low
70	Firmicutes	Clostridia	Ruminococcaceae	Clostridium IV	Ruminococcus bromii_94%	novel species	ethanol-resistant	0.662	low
71	Firmicutes	Clostridia	Ruminococcaceae	unclassified	Ruminococcus flavefaciens_93%	novel genus	ethanol-sensitive	0.538	low
72	Firmicutes	Clostridia	Ruminococcaceae	Ruminococcus	Ruminococcus flavefaciens_95%	novel species	ethanol-resistant	0.477	no
73	Firmicutes	Clostridia	Ruminococcaceae	Ruminococcus	Ruminococcus albus_98%	novel species	ethanol-resistant	0.585	no
74	Firmicutes	Clostridia	Ruminococcaceae	Ruminococcus	Ruminococcus albus_95%	novel species	ethanol-resistant	0.508	medium

75	Firmicutes	Clostridia	Ruminococcaceae	Clostridium XIV	Clostridium methylpentosum_92%	novel species	ethanol-resistant	0.646	no
76	Firmicutes	Clostridia	Ruminococcaceae	unclassified	Anaerotruncus colihominis	characterised	ethanol-resistant	0.662	no
77	Firmicutes	Clostridia	Ruminococcaceae	unclassified	Anaerotruncus colihominis_91%	novel genus	ethanol-resistant	0.662	no
78	Firmicutes	Clostridia	Ruminococcaceae	Faecalibacterium	Faecalibacterium prausnitzii	characterised	ethanol-sensitive	0.138	medium
79	Firmicutes	Clostridia	Ruminococcaceae	Faecalibacterium	Faecalibacterium prausnitzii	characterised	ethanol-sensitive	0.462	medium
80	Firmicutes	Clostridia	Ruminococcaceae	Faecalibacterium	Faecalibacterium prausnitzii_98%	novel species	ethanol-sensitive	0.462	medium
81	Firmicutes	Clostridia	Ruminococcaceae	Faecalibacterium	Faecalibacterium prausnitzii	characterised	ethanol-sensitive	0.462	medium
82	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Clostridium xylanolyticum_95%	novel genus	ethanol-sensitive	0.569	low
83	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_ incertae_sedis	Clostridium nexile_94%	novel species	ethanol-resistant	0.677	no
84	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_	Eubacterium fissicatens_95%	novel species	ethanol-	0.646	no

				incertae_sedis			resistant		
85	Firmicutes	Clostridia	Lachnospiraceae	Anaerostipes	Anaerostipes hadrus	characterised	ethanol-sensitive	0.554	no
86	Firmicutes	Clostridia	Lachnospiraceae	Anaerostipes	Anaerostipes hadrus_98%	novel species	ethanol-sensitive	0.569	no
87	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_ incertae_sedis	Ruminococcus gnavus	characterised	ethanol-resistant	0.585	no
88	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_ incertae_sedis	Ruminococcus gnavus_98%	novel species	ethanol-sensitive	0.615	low
89	Firmicutes	Clostridia	Lachnospiraceae	Dorea	Dorea formicigenerans_98%	novel species	ethanol-sensitive	0.554	no
90	Firmicutes	Clostridia	Lachnospiraceae	Dorea	Dorea longicatena	characterised	ethanol-sensitive	0.523	no
91	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Clostridium oroticum_95%	novel genus	ethanol-resistant	0.569	no
92	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_ incertae_sedis	Clostridium oroticum_96%	novel species	ethanol-resistant	0.523	no
93	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Eubacterium contortum	characterised	ethanol-resistant	0.631	no

94	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Eubacterium contortum_97%	novel genus	ethanol-sensitive	0.585	no
95	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Ruminococcus torques	characterised	ethanol-sensitive	0.585	low
96	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Clostridium oroticum_95%	novel genus	ethanol-resistant	0.615	no
97	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Ruminococcus torques	characterised	ethanol-sensitive	0.554	low
98	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Eubacterium ramulus	characterised	ethanol-sensitive	0.523	low
99	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Roseburia inulinivorans_94%	novel species	ethanol-resistant	0.615	no
100	Firmicutes	Clostridia	Lachnospiraceae	Roseburia	Roseburia inulinvorans	characterised	ethanol-sensitive	0.631	low
101	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Eubacterium rectale	characterised	ethanol-resistant	0.477	low
102	Firmicutes	Clostridia	Lachnospiraceae	Roseburia	Roseburia faecis	characterised	ethanol-sensitive	0.554	low
103	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Roseburia faecis_95%	novel genus	ethanol-	0.631	medium

							sensitive		
104	Firmicutes	Clostridia	Lachnospiraceae	Roseburia	Roseburia intestinalis	characterised	ethanol-sensitive	0.615	low
105	Firmicutes	Clostridia	Lachnospiraceae	Roseburia	Roseburia hominis	characterised	ethanol-resistant	0.6	low
106	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Lachnospira pectinoschiza_91%	novel genus	ethanol-sensitive	0.508	high
107	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Eubacterium eligens	characterised	ethanol-resistant	0.554	medium
108	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Lachnospira pectinoschiza	characterised	ethanol-sensitive	0.523	low
109	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Lachnospira pectinoschiza	characterised	ethanol-resistant	0.523	low
110	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Eubacterium hallii_97%	novel species	ethanol-sensitive	0.554	low
111	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Eubacterium hallii	characterised	ethanol-resistant	0.585	low
112	Firmicutes	Clostridia	Lachnospiraceae	Coprococcus	Coprococcus comes	characterised	ethanol-resistant	0.523	low

113	Firmicutes	Clostridia	Lachnospiraceae	Coprococcus	Coprococcus eutactus_97%	novel species	ethanol-resistant	0.538	low
114	Firmicutes	Clostridia	Lachnospiraceae	Coprococcus	Coprococcus eutactus	characterised	ethanol-resistant	0.523	low
115	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Blautia hydrogenotrophica_96%	novel genus	ethanol-resistant	0.615	no
116	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Blautia hydrogenotrophica	characterised	ethanol-resistant	0.6	no
117	Firmicutes	Clostridia	Lachnospiraceae	Blautia	Balutia luti_96%	novel species	ethanol-sensitive	0.462	no
118	Firmicutes	Clostridia	Lachnospiraceae	Blautia	Blautia wexlerae	characterised	ethanol-sensitive	0.492	low
119	Firmicutes	Clostridia	Lachnospiraceae	Blautia	Blautia obeum	characterised	ethanol-resistant	0.508	low
120	Firmicutes	Clostridia	Lachnospiraceae	Blautia	Blautia luti_96%	novel species	ethanol-sensitive	0.554	medium
121	Firmicutes	Clostridia	Lachnospiraceae	Blautia	Ruminococcus obeum_96%	novel species	ethanol-sensitive	0.492	medium
122	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_	Ruminococcus obeum	characterised	ethanol-	0.492	low

				incertae_sedis			sensitive		
123	Firmicutes	Clostridia	Lachnospiraceae	Blautia	Ruminococcus obeum_98%	novel species	ethanol-sensitive	0.538	low
124	Firmicutes	Clostridia	Lachnospiraceae	Blautia	Blautia luti_95%	novel species	ethanol-resistant	0.538	medium
125	Firmicutes	Clostridia	Lachnospiraceae	Blautia	Blautia luti_98%	novel species	ethanol-sensitive	0.538	medium
126	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Blautia producta_94%	novel species	ethanol-sensitive	0.585	no
127	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Fusicatenibacter saccharivorans	novel genus	ethanol-resistant	0.538	low
128	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Fusicatenibacter saccharivorans_93%	novel genus	ethanol-resistant	0.6	medium
129	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Clostridium clostridioforme_93%	novel genus	ethanol-sensitive	0.523	medium
130	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Clostridium hathewayi_92%	novel genus	ethanol-sensitive	0.538	no
131	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_incertae_sedis	Clostridium xylanolyticum_96%	novel species	ethanol-resistant	0.585	medium

132	Firmicutes	Clostridia	Lachnospiraceae	Lachnospiraceae_ incertae_sedis	Clostridium saccharolyticum_94%	novel species	ethanol-sensitive	0.538	no
133	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Ruminococcus torques_97%	novel genus	ethanol-resistant	0.538	medium
134	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Clostridium celerecrescens_93%	novel genus	ethanol-resistant	0.569	medium
135	Firmicutes	Clostridia	Lachnospiraceae	unclassified	Clostridium celerescens_93%	novel genus	ethanol-sensitive	0.569	medium
136	Firmicutes	Clostridia	Lachnospiraceae	Clostridium XIVa	Clostridium clostridioforme	characterised	ethanol-resistant	0.646	low
137	Firmicutes	Clostridia	Lachnospiraceae	Clostridium XIVa	Clostridium clostridioforme_98%	novel species	ethanol-resistant	0.677	low
138	Firmicutes	Clostridia	Lachnospiraceae	Clostridium XIVa	Clostridium boltae_94%	novel species	ethanol-resistant	0.677	no
139	Firmicutes	Clostridia	Lachnospiraceae	Clostridium XIVa	Clostridium hathewayi	characterised	ethanol-resistant	0.646	no
140	Firmicutes	Clostridia	Lachnospiraceae	Clostridium XIVa	Clostridium hathewayi	characterised	ethanol-resistant	0.677	no

†Incerate sedis refers to an uncertain taxonomic placement within a taxonomic class. Unclassified – isolate is novel at this taxonomic level.

Appendix 2 Isolates deposited in public culture collections

110 of 137 isolates were deposited in the following public culture collections: DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen Gmbh (DSMZ) in Braunschweig, Germany, JCM: Japan Collection of Microorganisms (JCM) maintained by the Riken BioResource Center in Tsukuba, Japan, CCUG: the Culture Collection, University of Gothenburg (CCUG) maintained by the University of Gothenburg in Sweden, BCCM: Belgian Co-ordinated Collection of Micro-organisms (BCCM/LMG) hosted by the Laboratory of Microbiology at Ghent University, Belgium. The isolates are listed according to their order on Figure 3.5 and therefore include three Proteobacteria species which were not cultured but were included to provide phylogenetic context. Hence, the total number listed here is 140 (three Proteobacteria and 137 cultured in this study).

order on tree Fig. 3.5	closest 16S BLAST match to named bacterium-99 or 100% match unless otherwise stated	Culture collection	Public strain designation	Culture collection accession number
1	Oxalobacter formigenes	n/a		n/a
2	Proteus mirabilis	n/a		n/a
3	Escherichia coli	n/a		n/a
4	Alistipes finegoldii			
5	Alistipes finegoldii	CCUG	Sanger_38	CCUG 68735
6	Parabacteroides merdae	CCUG	Sanger_39	CCUG 68661
7	Parabacteroides distasonis	CCUG	Sanger_40	CCUG 68616
8	Parabacteroides distasonis	CCUG	Sanger_41	CCUG 68699
9	Parabacteroides distasonis	CCUG	Sanger_42	CCUG 68700
10	Prevotella copri	CCUG	Sanger_43	CCUG 68549
11	Bacteroides coprocola_94%	DSMZ	Sanger_22	DSM 102145
12	Bacteroides plebius_95%	DSMZ	Sanger_21	DSM 102146
13	Bacteroides vulgatus	CCUG	Sanger_44	CCUG 68662
14	Bacteroides uniformis	CCUG	Sanger_45	CCUG 68683
15	Bacteroides intestinalis_98%	JCM	Sanger_46	JCM 31249

16	Bacteroides salyersiae	BCCM	Sanger_47	LMG 29389
17	Bacteroides caccae	BCCM	Sanger_48	LMG 29390
18	Bacteroides xylanisolvens	CCUG	Sanger_49	CCUG 68584
19	Bacteroides ovatus	CCUG	Sanger_50	CCUG 68701
20	Bacteroides finegoldi	CCUG	Sanger_52	CCUG 68636
21	Bacteroides thetaiotaomicron	CCUG	Sanger_53	CCUG 68702
22	Bacteroides thetaiotaomicron	CCUG	Sanger_54	CCUG 68684
23	Bacteroides finegoldii_98%	BCCM	Sanger_51	LMG 29391
24	Bacteroides thetaiotaomicron	CCUG	Sanger_55	CCUG 68746
25	Collinsella aerofaciens_92%	BCCM	Sanger_56	LMG 29392
26	Collinsella aerofaciens	BCCM & CCUG	Sanger_57	LMG 29393 & CCUG 68712
27	Collinsella aerofaciens			
28	Bifidobacterium adolescentis	BCCM	Sanger_59	LMG 29394
29	Bifidobacterium adolescentis	BCCM	Sanger_60	LMG 29395
30	Bifidobacterium bifidum	BCCM	Sanger_61	LMG 29396
31	Bifidobacterium pseudocatenulatum	BCCM	Sanger_62	LMG 29397
32	Mitsuokella jalaludinii	BCCM & CCUG	Sanger_63	LMG 29398 & CCUG 68585

33	Megasphaera elsdenii_95%	DSMZ	Sanger_24	DSM 102144
34	Turibacter sanguinis	CCUG	Sanger_64	CCUG 68586
35	Clostridium innocuum	CCUG	Sanger_65	CCUG 68747
36	Clostridium innocuum_95%	BCCM	Sanger_66	LMG 29399
37	Catenibacterium mitsuokai	JCM	Sanger_67	JCM 31250
38	Clostridium cocleatum_93	BCCM	Sanger_68	LMG 29400
39	Clostridium saccharogumia_93%	BCCM	Sanger_69	LMG 29401
40	Eubacterium infirmum_91%			
41	Eubacterium infirmum_94%			
42	Clostridium ghonii_98%	BCCM	Sanger_70	LMG 29402
43	Clostridium lituseburense_98%	CCUG	Sanger_73	CCUG 68538
44	Clostridium bartlettii	BCCM	Sanger_71	LMG 29403
45	Sarcina ventriculi	JCM	Sanger_74	JCM 31252
46	Clostridium baratti	JCM	Sanger_75	JCM 31253
47	Clostridium paraputrificum	JCM	Sanger_76	JCM 31254
48	Clostridium disporicum	JCM	Sanger_72	JCM 31251
49	Clostridium disporicum			

50	Clostridium perfringens			
51	Clostridium disporicum_98%	CCUG	Sanger_79	CCUG 68587
52	Clostridium thermocellum_86%			
53	Clostridium thermocellum_87%			
54	Butyricicoccus pullicaecorum_94%	DSMZ	Sanger_34	DSM 102882
55	Butyricicoccus pullicaecorum_94%	BCCM & CCUG	Sanger_80	LMG 29404 & CCUG 68588
56	Flavonifractor plautii_95%			
57	Flavonifractor plautii_95%	DSMZ	Sanger_09	DSM 102137
58	Flavonifractor plautii_94%	DSMZ	Sanger_35	DSM 102175
59	Flavonifractor plautii_96%	DSMZ	Sanger_08	DSM 102116
60	Oscillibacter valericigenes_96%	DSMZ	Sanger_26	DSM 102152
61	Flavonifractor plautii_95%	CCUG	Sanger_82	CCUG 68613
62	Flavonifractor plautii_97%			
63	Flavonifractor plautii	BCCM	Sanger_83	LMG 29405
64	Flavonifractor plautii	BCCM	Sanger_84	LMG 29406
65	Flavonifractor plautii	BCCM & CCUG	Sanger_85	CCUG 68710 & LMG 29407
66	Flavonifractor plautii	BCCM & CCUG	Sanger_86	LMG 29408 & CCUG 68637

67	Eubacterium siraeum	BCCM	Sanger_87	LMG 29409
68	Ruminococcus bromii			
69	Ruminococcus bromii_93%	DSMZ	Sanger_36	DSM 102803
70	Ruminococcus bromii_94%	CCUG	Sanger_89	CCUG 68614
71	Ruminococcus flavefaciens_93%	DSMZ	Sanger_06	DSM 102115
72	Ruminococcus flavefaciens_95%	BCCM & DSMZ	Sanger_90	LMG 29410 & DSM 102167
73	Ruminococcus albus_98%	DSMZ	Sanger_91	DSM 102227
74	Ruminococcus albus_95%	DSMZ	Sanger_31	DSM 102216
75	Clostridium methylpentosum_92%	DSMZ	Sanger_27	DSM 102153
76	Anaerotruncus colihominis	JCM	Sanger_92	JCM 31255
77	Anaerotruncus colihominis_91%	DSMZ	Sanger_05	DSM 102114
78	Faecalibacterium prausnitzii	CCUG	Sanger_93	CCUG 68711
79	Faecalibacterium prausnitzii	CCUG	Sanger_94	CCUG 68745
80	Faecalibacterium prausnitzii_98%			
81	Faecalibacterium prausnitzii			
82	Clostridium xylanolyticum_95%	DSMZ	Sanger_04	DSM 102317
83	Clostridium nexile_94%	DSMZ	Sanger_33	DSM 102154

84	Eubacterium fissicatens_95%	CCUG & DSMZ	Sanger_97	CCUG 68796 & DSM 102166
85	Anaerostipes hadrus			
86	Anaerostipes hadrus_98%	CCUG	Sanger_99	CCUG 68539
87	Ruminococcus gnavus	CCUG	Sanger_100	CCUG 68638
88	Ruminococcus gnavus_98%			
89	Dorea formicigenerans_98%	CCUG & JCM	Sanger_102	CCUG 68540 & JCM 31256
90	Dorea longicatena			
91	Clostridium oroticum_95%	DSMZ	Sanger_03	DSM 102260
92	Clostridium oroticum_96%	JCM	Sanger_104	JCM 31257
93	Eubacterium contortum	JCM	Sanger_105	JCM 31258
94	Eubacterium contortum_97%	DSMZ	Sanger_02	DSM 102136
95	Ruminococcus torques			
96	Clostridium oroticum_95%	DSMZ	Sanger_01	DSM 102316
97	Ruminococcus torques	JCM	Sanger_107	JCM 31259
98	Eubacterium ramulus			
99	Roseburia inulinivorans_94%	DSMZ	Sanger_109	DSM 102148
100	Roseburia inulinvorans	JCM	Sanger_110	JCM 31260

101	Eubacterium rectale			
102	Roseburia faecis	JCM	Sanger_112	JCM 31261
103	Roseburia faecis_95%	DSMZ	Sanger_19	DSMZ 102150
104	Roseburia intestinalis	JCM	Sanger_113	JCM 31262
105	Roseburia hominis			
106	Lachnospira pectinoschiza_91%	DSMZ	Sanger_20	DSM 102349
107	Eubacterium eligens			
108	Lachnospira pectinoschiza			
109	Lachnospira pectinoschiza	CCUG	Sanger_117	CCUG 68639
110	Eubacterium hallii_97%			
111	Eubacterium hallii	JCM	Sanger_119	JCM 31263
112	Coprococcus comes	JCM	Sanger_120	JCM 31264
113	Coprococcus eutactus_97%	CCUG	Sanger_121	CCUG 68541
114	Coprococcus eutactus	JCM	Sanger_122	JCM 31265
115	Blautia hydrogenotrophica_96%			
116	Blautia hydrogenotrophica	JCM	Sanger_124	JCM 31266
117	Balutia luti_96%	DSMZ	Sanger_23	DSMZ 102163

118	Blautia wexlerae	JCM	Sanger_125	JCM 31267
119	Blautia wexlerae			
120	Ruminococcus obeum_96%	DSMZ	Sanger_28	DSM 102165
121	Ruminococcus obeum_96%	CCUG	Sanger_127	CCUG 68542
122	Ruminococcus obeum			
123	Ruminococcus obeum_98%	DSMZ	Sanger_25	DSM 102164
124	Blautia luti_95%	CCUG	Sanger_129	CCUG 68550
125	Blautia luti_98%	CCUG	Sanger_130	CCUG 68551
126	Blautia producta_94%	DSMZ	Sanger_32	DSM 102174
127	Fusicatenibacter saccharivorans	CCUG & JCM	Sanger_131	CCUG 68552 & JCM 31268
128	Fusicatenibacter saccharivorans_93%	DSMZ	Sanger_17	DSM 102348
129	Clostridium clostridioforme_93%	DSMZ	Sanger_16	DSM 102825
130	Clostridium hathewayi_92%	DSMZ	Sanger_18	DSM 102261
131	Clostridium xylanolyticum_96%	DSMZ	Sanger_132	DSM 102147
132	Clostridium saccharolyticum_94%	DSMZ	Sanger_29	DSM 102151
133	Ruminococcus torques_97%	DSMZ	Sanger_15	DSM 102149
134	Clostridium celerecrescens_93%			

135	Clostridium celerescens_93%	DSMZ	Sanger_13	DSM 102317
136	Clostridium clostridioforme	CCUG	Sanger_133	CCUG 68660
137	Clostridium clostridioforme_98%	CCUG	Sanger_134	CCUG 68553
138	Clostridium boltae_94%	CCUG	Sanger_135	CCUG 68615
139	Clostridium hathewayi	CCUG	Sanger_136	CCUG 68736
140	Clostridium hathewayi	CCUG	Sanger_137	CCUG 68640

Appendix 3 Validation of the sporulation signature

Sporulation signature scores of ethanol-resistant and ethanol-sensitive isolates from this study are presented in addition to known spore-formers and known non-spore-formers from different environments. This data was used to populate Figure 4.7. Note: Genomes were sequenced for multiple cultures from the same species in the culture collection. Subsequently multiple copies of the same species are included in the analysis here.

	Species Name (Closest BLAST Match, ≥98.7% match unless otherwise stated)	**Public identifier for species not cultured in this study**	**Category**	**signature score**
1	Akkermansia muciniphila	GCA_000020225.1	known non-spore formers	0.25
2	Alicyclobacillus acidocaldarius	GCA_000024285.1	known spore formers	0.66
3	Alistipes finegoldii	cultured in this study	Cultures from untreated plates	0.16
4	Alistipes finegoldii	cultured in this study	Cultures from untreated plates	0.17
5	Alkaliphilus oremlandii	GCA_000018325.1	known spore formers	0.75
6	Ammonifex degensii KC4	GCA_000024605.1	known spore formers	0.66
7	Anaeromyxobacter dehalogenans	GCA_000013385.1	known non-spore formers	0.17
8	Anaerostipes hadrus	cultured in this study	Cultures from untreated plates	0.58
9	Anaerostipes hadrus	cultured in this study	Cultures from untreated plates	0.56
10	Anaerostipes hadrus	cultured in this study	Cultures from untreated plates	0.55
11	Anaerostipes hadrus	cultured in this study	Cultures from untreated plates	0.55
12	Anaerostipes hadrus	cultured in this study	Cultures from untreated plates	0.55

13	Anaerotruncus colihominis	cultured in this study	Cultures from ethanol treated plates	0.66
14	Anaerotruncus colihominis_91%	cultured in this study	Cultures from ethanol treated plates	0.66
15	Aquifex aeolicus	GCA_000008625.1	known non-spore formers	0.17
16	Bacillus amyloliquefaciens	GCA_000015785.1	known spore formers	0.73
17	Bacillus anthracis	GCA_000007845.1	known spore formers	0.73
18	Bacillus cereus	GCA_000007825.1	known spore formers	0.73
19	Bacillus clausii	GCA_000009825.1	known spore formers	0.69
20	Bacillus halodurans	GCA_000011145.1	known spore formers	0.70
21	Bacillus licheniformis	GCA_000008425.1	known spore formers	0.72
22	Bacillus pumilus	GCA_000017885.1	known spore formers	0.72
23	Bacillus subtilis	GCA_000009045.1	known spore formers	0.73
24	Bacillus thuringiensis serovar konkukian	GCA_000008505.1	known spore formers	0.73
25	Bacillus weihenstephanensis	GCA_000018825.1	known spore formers	0.73
26	Bacteroides caccae	cultured in this study	Cultures from untreated plates	0.19
27	Bacteroides caccae	cultured in this study	Cultures from untreated plates	0.17
28	Bacteroides finegoldi	cultured in this study	Cultures from untreated plates	0.17
29	Bacteroides ovatus	cultured in this study	Cultures from untreated plates	0.17

30	Bacteroides salyersiae	cultured in this study	Cultures from untreated plates	0.17
31	Bacteroides finegoldii_98%	cultured in this study	Cultures from untreated plates	0.17
32	Bacteroides coprocola_94%	cultured in this study	Cultures from untreated plates	0.17
33	Bacteroides plebius_95%	cultured in this study	Cultures from untreated plates	0.16
34	Bacteroides intestinalis_98%	cultured in this study	Cultures from untreated plates	0.17
35	Bacteroides thetaiotaomicron	cultured in this study	Cultures from untreated plates	0.19
36	Bacteroides thetaiotaomicron	cultured in this study	Cultures from untreated plates	0.19
37	Bacteroides thetaiotaomicron	cultured in this study	Cultures from untreated plates	0.19
38	Bacteroides thetaiotaomicron	cultured in this study	Cultures from untreated plates	0.19
39	Bacteroides uniformis	cultured in this study	Cultures from untreated plates	0.19
40	Bacteroides uniformis	cultured in this study	Cultures from untreated plates	0.17
41	Bacteroides uniformis	cultured in this study	Cultures from untreated plates	0.17
42	Bacteroides uniformis	cultured in this study	Cultures from untreated plates	0.17
43	Bacteroides uniformis	cultured in this study	Cultures from untreated plates	0.17
44	Bacteroides uniformis	cultured in this study	Cultures from untreated plates	0.19
45	Bacteroides vulgatus	cultured in this study	Cultures from untreated plates	0.17
46	Bacteroides vulgatus	cultured in this study	Cultures from untreated plates	0.17

47	Bacteroides vulgatus	cultured in this study	Cultures from untreated plates	0.17
48	Bacteroides xylanisolvens	cultured in this study	Cultures from untreated plates	0.19
49	Bacteroides fragilis	GCA_000009925.1	known non-spore formers	0.17
50	Bifidobacterium adolscentis	cultured in this study	Cultures from untreated plates	0.11
51	Bifidobacterium adolescentis	cultured in this study	Cultures from untreated plates	0.13
52	Bifidobacterium adolescentis	cultured in this study	Cultures from untreated plates	0.13
53	Bifidobacterium bifidum	cultured in this study	Cultures from untreated plates	0.13
54	Bifidobacterium pseudocatenulatum	cultured in this study	Cultures from untreated plates	0.11
55	Bifidobacterium bifidum	GCA_000165905.1	known non-spore formers	0.13
56	Blautia hydrogenotrophica	cultured in this study	Cultures from ethanol treated plates	0.59
57	Blautia luti_97%	cultured in this study	Cultures from ethanol treated plates	0.50
58	Ruminococcus obeum_96%	cultured in this study	Cultures from untreated plates	0.55
59	Blautia luti_96%	cultured in this study	Cultures from ethanol treated plates	0.50
60	Blautia luti_95%	cultured in this study	Cultures from untreated plates	0.52
61	Blautia hydrogenotrophica_96%	cultured in this study	Cultures from ethanol treated plates	0.61
62	Blautia luti_96%	cultured in this study	Cultures from untreated plates	0.53
63	Blautia luti_95%	cultured in this study	Cultures from untreated plates	0.50

64	Balutia luti_96%	cultured in this study	Cultures from untreated plates	0.45
65	Blautia luti_98%	cultured in this study	Cultures from untreated plates	0.53
66	Blautia producta_94%	cultured in this study	Cultures from untreated plates	0.58
67	Blautia luti_96%	cultured in this study	Cultures from ethanol treated plates	0.48
68	Blautia wexlerae	cultured in this study	Cultures from untreated plates	0.48
69	Blautia wexlerae	cultured in this study	Cultures from untreated plates	0.48
70	Bordetella parapertussis	GCA_000317935.1	known non-spore formers	0.30
71	Brevibacillus brevis	GCA_000010165.1	known spore formers	0.70
72	Brucella melitensis	GCA_000022625.1	known non-spore formers	0.17
73	Brucella suis	GCA_000018905.1	known non-spore formers	0.19
74	Butyricicoccus pullicaecorum_96%	cultured in this study	Cultures from ethanol treated plates	0.56
75	Butyricicoccus pullicaecorum_94%	cultured in this study	Cultures from ethanol treated plates	0.56
76	Caldanaerobacter subterraneus subsp tengcongensis	GCA_000007085.1	known spore formers	0.81
77	Caldicellulosiruptor bescii DSM6725	GCA_000022325.1	known spore formers	0.63
78	Caldicellulosiruptor hydrothermalis 108	GCA_000166355.1	known spore formers	0.61
79	Caldicellulosiruptor kristjanssonii I77R1B	GCA_000166695.1	known spore formers	0.58
80	Caldicellulosiruptor kronotskyensis 2002	GCA_000166775.1	known spore formers	0.61

82	Caldicellulosiruptor obsidiansis OB47	GCA_000145215.1	known spore formers	0.58
83	Caldicellulosiruptor owensensis OL	GCA_000166335.1	known spore formers	0.58
84	Caldicellulosiruptor saccharolyticus DSM8903	GCA_000016545.1	known spore formers	0.59
85	Campylobacter jejuni	GCA_000009085.1	known non-spore formers	0.16
86	Candidatus Desulforudis audaxviator MP104C	GCA_000018425.1	known spore formers	0.67
87	Carboxydothermus hydrogenoformans	GCA_000012865.1	known spore formers	0.67
88	Catenibacterium mitsuokai	cultured in this study	Cultures from untreated plates	0.25
89	Catenibacterium mitsuokai	cultured in this study	Cultures from untreated plates	0.20
90	Clostridium baratii	cultured in this study	Cultures from ethanol treated plates	0.77
91	Clostridium baratii	cultured in this study	Cultures from untreated plates	0.77
92	Clostridium bartlettii	cultured in this study	Cultures from ethanol treated plates	0.88
93	Clostridium clostridioforme	cultured in this study	Cultures from ethanol treated plates	0.64
94	Clostridium disporicum	cultured in this study	Cultures from untreated plates	0.77
95	Clostridium disporicum	cultured in this study	Cultures from untreated plates	0.78
96	Clostridium disporicum	cultured in this study	Cultures from ethanol treated plates	0.80
97	Clostridium hathewayi	cultured in this study	Cultures from ethanol treated plates	0.67
98	Clostridium hathewayi	cultured in this study	Cultures from ethanol treated plates	0.64

99	Clostridium innocuum	cultured in this study	Cultures from ethanol treated plates	0.48
100	Clostridium orbiscindens	cultured in this study	Cultures from ethanol treated plates	0.72
101	Clostridium paraputrificum	cultured in this study	Cultures from ethanol treated plates	0.77
102	Clostridium paraputrificum	cultured in this study	Cultures from untreated plates	0.77
103	Clostridium perfringens	cultured in this study	Cultures from ethanol treated plates	0.72
104	Clostridium celerecrescens_94%	cultured in this study	Cultures from untreated plates	0.55
105	Clostridium celerecrescens_94%	cultured in this study	Cultures from untreated plates	0.58
106	Clostridium saccharolyticum_94%	cultured in this study	Cultures from untreated plates	0.53
107	Clostridium methylpentosum_92%	cultured in this study	Cultures from ethanol treated plates	0.64
108	Clostridium oroticum_95%	cultured in this study	Cultures from ethanol treated plates	0.61
109	Clostridium clostridioforme_98%	cultured in this study	Cultures from ethanol treated plates	0.67
110	Clostridium straminisolvens_89%	cultured in this study	Cultures from untreated plates	0.64
111	Clostridium beijerinckii_96%	cultured in this study	Cultures from ethanol treated plates	0.73
112	Clostridium cocleatum_93%	cultured in this study	Cultures from untreated plates	0.41
113	Clostridium spiroforme_93%	cultured in this study	Cultures from untreated plates	0.42
114	Clostridium orbiscindens_95%	cultured in this study	Cultures from ethanol treated plates	0.58
115	Clostridium xylanolyticum_93%	cultured in this study	Cultures from untreated plates	0.63

116	Clostridium hathewayi_92%	cultured in this study	Cultures from untreated plates	0.53
117	Clostridium oroticum_95%	cultured in this study	Cultures from untreated plates	0.58
118	Clostridium saccharogumia_93%	cultured in this study	Cultures from untreated plates	0.41
119	Clostridium lituseburense_98%	cultured in this study	Cultures from ethanol treated plates	0.86
120	Clostridium boltae_93%	cultured in this study	Cultures from untreated plates	0.56
121	Clostridium celerescens_93%	cultured in this study	Cultures from untreated plates	0.56
122	Clostridium oroticum_96%	cultured in this study	Cultures from ethanol treated plates	0.52
123	Clostridium xylanolyticum_92%	cultured in this study	Cultures from ethanol treated plates	0.58
124	Clostridium xylanolyticum_95%	cultured in this study	Cultures from untreated plates	0.56
125	Clostridium innocuum_95%	cultured in this study	Cultures from ethanol treated plates	0.47
126	Clostridium glycolicum_97%	cultured in this study	Cultures from ethanol treated plates	0.88
127	Clostridium nexile_94%	cultured in this study	Cultures from ethanol treated plates	0.67
128	Clostridium oroticum_95%	cultured in this study	Cultures from untreated plates	0.56
129	Clostridium clostridioforme_93%	cultured in this study	Cultures from untreated plates	0.52
130	Clostridium celerecrescens_93%	cultured in this study	Cultures from ethanol treated plates	0.56
131	Clostridium boltae_94%	cultured in this study	Cultures from ethanol treated plates	0.67
132	Clostridium thermocellum_87%	cultured in this study	Cultures from ethanol treated plates	0.66

133	Clostridium orbiscindens_94%	cultured in this study	Cultures from ethanol treated plates	0.64
134	Clostridium thermocellum_86%	cultured in this study	Cultures from ethanol treated plates	0.64
135	Clostridium spiroforme_93%	cultured in this study	Cultures from untreated plates	0.39
136	Clostridium ghonii_98%	cultured in this study	Cultures from ethanol treated plates	0.83
137	Clostridium ghonii_98%	cultured in this study	Cultures from ethanol treated plates	0.83
138	Clostridium disporicum_98%	cultured in this study	Cultures from ethanol treated plates	0.80
139	Clostridium disporicum_98%	cultured in this study	Cultures from untreated plates	0.78
140	Clostridium symbiosum	cultured in this study	Cultures from ethanol treated plates	0.67
141	Clostridium acetobutylicum	GCA_000008765.1	known spore formers	0.73
142	Clostridium beijerinckii	GCA_000016965.1	known spore formers	0.73
143	Clostridium botulinum	GCA_000017045.1	known spore formers	0.83
144	Clostridium kluyveri	GCA_000016505.1	known spore formers	0.78
145	Clostridium novyi	GCA_000014125.1	known spore formers	0.75
146	Clostridium perfringens	GCA_000009685.1	known spore formers	0.72
147	Clostridium tetani E88	GCA_000007625.1	known spore formers	0.77
148	Collinsella aerofaciens	cultured in this study	Cultures from untreated plates	0.17
149	Collinsella aerofaciens	cultured in this study	Cultures from untreated plates	0.19

150	Collinsella aerofaciens	cultured in this study	Cultures from untreated plates	0.17
151	Collinsella aerofaciens_92%	cultured in this study	Cultures from untreated plates	0.20
152	Coprococcus comes	cultured in this study	Cultures from ethanol treated plates	0.52
153	Coprococcus comes	cultured in this study	Cultures from untreated plates	0.52
154	Coprococcus comes	cultured in this study	Cultures from untreated plates	0.50
155	Coprococcus comes	cultured in this study	Cultures from untreated plates	0.50
156	Coprococcus eutactus	cultured in this study	Cultures from untreated plates	0.52
157	Coprococcus eutactus	cultured in this study	Cultures from ethanol treated plates	0.52
158	Coprococcus eutactus	cultured in this study	Cultures from untreated plates	0.50
159	Coprococcus eutactus	cultured in this study	Cultures from untreated plates	0.52
160	Coprococcus eutactus_97%	cultured in this study	Cultures from ethanol treated plates	0.53
161	Coprococcus nexile_95%	cultured in this study	Cultures from ethanol treated plates	0.50
162	Desulfitobacterium hafniense	GCA_000010045.1	known spore formers	0.75
163	Desulfotomaculum acetoxidans	GCA_000024205.1	known spore formers	0.72
164	Desulfotomaculum kuznetsovii	GCA_000214705.1	known spore formers	0.67
165	Desulfotomaculum reducens	GCA_000016165.1	known spore formers	0.72
166	Desulfovibrio vulgaris	GCA_000195755.1	known non-spore formers	0.27

167	Dorea longicatena	cultured in this study	Cultures from untreated plates	0.52
168	Dorea longicatena	cultured in this study	Cultures from untreated plates	0.52
169	Dorea longicatena	cultured in this study	Cultures from untreated plates	0.48
170	Dorea longicatena	cultured in this study	Cultures from untreated plates	0.48
171	Dorea longicatena	cultured in this study	Cultures from untreated plates	0.52
172	Dorea formicigenerans_98%	cultured in this study	Cultures from untreated plates	0.55
173	Eggerthella lenta	GCA_000024265.1	known non-spore formers	0.17
174	Enterococcus faecalis V583	GCA_000007785.1	known non-spore formers	0.23
175	Escherichia coli	GCA_000005845.2	known non-spore formers	0.22
176	Ethanoligenens harbinense YUAN-3	GCA_000178115.2	known spore formers	0.66
177	Eubacterium contortum	cultured in this study	Cultures from ethanol treated plates	0.63
178	Eubacterium eligens	cultured in this study	Cultures from untreated plates	0.53
179	Eubacterium eligens	cultured in this study	Cultures from untreated plates	0.55
180	Eubacterium halii	cultured in this study	Cultures from ethanol treated plates	0.58
181	Eubacterium hallii	cultured in this study	Cultures from untreated plates	0.58
182	Eubacterium ramulus	cultured in this study	Cultures from untreated plates	0.52
183	Eubacterium rectale	cultured in this study	Cultures from untreated plates	0.47

184	Eubacterium rectale	cultured in this study	Cultures from untreated plates	0.47
185	Eubacterium rectale	cultured in this study	Cultures from untreated plates	0.47
186	Eubacterium siraeum	cultured in this study	Cultures from untreated plates	0.47
187	Eubacterium infirmum_91%	cultured in this study	Cultures from untreated plates	0.53
188	Eubacterium fissicatens_95%	cultured in this study	Cultures from ethanol treated plates	0.64
189	Eubacterium hallii_98%	cultured in this study	Cultures from ethanol treated plates	0.58
190	Eubacterium contortum_97%	cultured in this study	Cultures from untreated plates	0.58
191	Eubacterium hallii_98%	cultured in this study	Cultures from untreated plates	0.56
192	Eubacterium infirmum_94%	cultured in this study	Cultures from ethanol treated plates	0.66
193	Anaerostipes hadrum_98%	cultured in this study	Cultures from untreated plates	0.56
194	Eubacterium hallii_97%	cultured in this study	Cultures from untreated plates	0.55
196	Faecalibacterium prausnitzii	cultured in this study	Cultures from untreated plates	0.45
197	Faecalibacterium prausnitzii	cultured in this study	Cultures from untreated plates	0.13
198	Faecalibacterium prausnitzii	cultured in this study	Cultures from untreated plates	0.45
199	Faecalibacterium prausnitzii_98%	cultured in this study	Cultures from untreated plates	0.45
200	Faecalibacterium prausnitzii_98%	cultured in this study	Cultures from untreated plates	0.45
201	Flavonifractor plautii	cultured in this study	Cultures from untreated plates	0.50

202	Flavonifractor plautii	cultured in this study	Cultures from ethanol treated plates	0.69
203	Flavonifractor plautii	cultured in this study	Cultures from ethanol treated plates	0.69
204	Flavonifractor plautii_95%	cultured in this study	Cultures from ethanol treated plates	0.58
205	Flavonifractor plautii_95%	cultured in this study	Cultures from untreated plates	0.66
206	Flavonifractor plautii_95%	cultured in this study	Cultures from untreated plates	0.64
207	Flavonifractor plautii_95%	cultured in this study	Cultures from ethanol treated plates	0.61
208	Flavonifractor plautii_96%	cultured in this study	Cultures from ethanol treated plates	0.67
209	Fusicatenibacter saccharivorans	cultured in this study	Cultures from ethanol treated plates	0.55
210	Fusicatenibacter saccharivorans	cultured in this study	Cultures from untreated plates	0.52
211	Fusicatenibacter saccharivorans	cultured in this study	Cultures from untreated plates	0.53
212	Fusicatenibacter saccharivorans_93%	cultured in this study	Cultures from ethanol treated plates	0.59
213	Fusobacterium necrophorum	GCA_000242215.1	known non-spore formers	0.22
214	Fusobacterium nucleatum GCF 000007325.1 ASM732v1	GCA_000007325.1	known non-spore formers	0.20
215	Geobacillus kaustophilus	GCA_000009785.1	known spore formers	0.70
216	Geobacillus thermodenitrificans	GCA_000015745.1	known spore formers	0.73
217	Geobacter sulfurreducens	GCA_000007985.2	known non-spore formers	0.22
218	Haemophilus influenzae	GCA_000027305.1	known non-spore formers	0.17

219	Halothermothrix orenii H168	GCA_000020485.1	known spore formers	0.69
220	Helicobacter pylori	GCA_000008525.1	known non-spore formers	0.13
221	Heliobacterium modesticaldum	GCA_000019165.1	known spore formers	0.70
222	Kyrpidia tusciae	GCA_000092905.1	known spore formers	0.69
223	Lachnoclostridium phytofermentans	GCA_000018685.1	known spore formers	0.69
224	Lachnospira pectinoschiza	cultured in this study	Cultures from untreated plates	0.52
225	Lachnospira pectinoschiza	cultured in this study	Cultures from untreated plates	0.52
226	Lachnospira pectinoshiza_91%	cultured in this study	Cultures from untreated plates	0.50
227	Lactobacillus acidophilus	GCA_000786395.1	known non-spore formers	0.22
228	Leptospira interrogans	GCA_000092565.1	known non-spore formers	0.16
229	Listeria innocua	GCA_000195795.1	known non-spore formers	0.33
230	Lysinibacillus sphaericus	GCA_000568835.1	known spore formers	0.58
231	Megasphaera elsdenii_95%	cultured in this study	Cultures from untreated plates	0.25
232	Mitsuokella jalaludinii	cultured in this study	Cultures from untreated plates	0.28
233	Moorella thermoacetica	GCA_000013105.1	known spore formers	0.72
234	Mycobacterium avium	GCA_000007865.1	known non-spore formers	0.23
235	Mycobacterium bovis	GCA_000195835.1	known non-spore formers	0.23

236	Mycobacterium leprae	GCA_000195855.1	known non-spore formers	0.22
237	Mycobacterium marinum	GCA_000018345.1	known non-spore formers	0.23
238	Mycobacterium smegmatis	GCA_000015005.1	known non-spore formers	0.25
239	Mycobacterium tuberculosis	GCA_000195955.2	known non-spore formers	0.23
240	Mycobacterium ulcerans	GCA_000013925.1	known non-spore formers	0.22
241	Myxococcus xanthus	GCA_000012685.1	known non-spore formers	0.22
242	Natranaerobius thermophilus JWNM-WN-LF	GCA_000020005.1	known spore formers	0.66
243	Neisseria meningtidis	GCA_000008805.1	known non-spore formers	0.20
244	Nitrosomonas europaea	GCA_000009145.1	known non-spore formers	0.19
245	Oceanobacillus iheyensis	GCA_000011245.1	known spore formers	0.70
246	Oscillibacter valericigenes_96%	cultured in this study	Cultures from untreated plates	0.64
247	Oscillibacter valericigenes_95%	cultured in this study	Cultures from ethanol treated plates	0.66
248	Paenibacillus polymyxa	GCA_000146875.1	known spore formers	0.72
249	Parabacteroides distasonis	cultured in this study	Cultures from untreated plates	0.17
250	Parabacteroides distasonis	cultured in this study	Cultures from untreated plates	0.19
251	Parabacteroides distasonis	cultured in this study	Cultures from untreated plates	0.20
252	Parabacteroides distasonis	cultured in this study	Cultures from untreated plates	0.19

253	Parabacteroides merdae	cultured in this study	Cultures from untreated plates	0.20
254	Pelotomaculum thermopropionicum	GCA_000010565.1	known spore formers	0.69
255	Peptoclostridium difficile	GCA_000009205.1	known spore formers	0.97
256	Photobacterium profundum	GCA_000196255.1	known non-spore formers	0.25
257	Prevotella copri	cultured in this study	Cultures from untreated plates	0.03
258	Prevotella stercorea	GCA_000235885.1	known non-spore formers	0.16
259	Prochlorococcus marinus	GCA_000007925.1	known non-spore formers	0.13
260	Propionibacterium acnes	GCA_000008345.1	known non-spore formers	0.19
261	Proteus mirabilis	GCA_000069965.1	known non-spore formers	0.25
262	Pseudomonas fluorescens	GCA_000006765.1	known non-spore formers	0.28
263	Pseudomonas syringae	GCA_000012245.1	known non-spore formers	0.28
264	Roseburia faecis	cultured in this study	Cultures from untreated plates	0.55
265	Roseburia hominis	cultured in this study	Cultures from untreated plates	0.59
266	Roseburia intestinalis	cultured in this study	Cultures from untreated plates	0.61
267	Roseburia inulinivorans	cultured in this study	Cultures from untreated plates	0.64
268	Roseburia inulinvorans	cultured in this study	Cultures from untreated plates	0.63
269	Roseburia faecis_95%	cultured in this study	Cultures from untreated plates	0.63

270	Roseburia inulinivorans_94%	cultured in this study	Cultures from ethanol treated plates	0.61
271	Ruminiclostridium thermocellum	GCA_000015865.1	known spore formers	0.81
272	Ruminococcus bromii	cultured in this study	Cultures from untreated plates	0.56
273	Ruminococcus gnavus	cultured in this study	Cultures from ethanol treated plates	0.58
274	Ruminococcus obeum	cultured in this study	Cultures from untreated plates	0.47
275	Ruminococcus obeum	cultured in this study	Cultures from untreated plates	0.50
276	Ruminococcus obeum	cultured in this study	Cultures from untreated plates	0.50
277	Ruminococcus obeum	cultured in this study	Cultures from untreated plates	0.48
278	Ruminococcus obeum	cultured in this study	Cultures from untreated plates	0.50
279	Ruminococcus bromii_94%	cultured in this study	Cultures from untreated plates	0.13
280	Ruminococcus torques_96%	cultured in this study	Cultures from untreated plates	0.53
281	Ruminococcus torques_97%	cultured in this study	Cultures from untreated plates	0.59
282	Ruminococcus torques_97%	cultured in this study	Cultures from untreated plates	0.56
283	Ruminococcus albus_95%	cultured in this study	Cultures from ethanol treated plates	0.64
284	Ruminococcus albus_95%	cultured in this study	Cultures from ethanol treated plates	0.52
285	Ruminococcus bromii_92%	cultured in this study	Cultures from ethanol treated plates	0.59
286	Ruminococcus obeum_98%	cultured in this study	Cultures from untreated plates	0.47

287	Ruminococcus obeum_98%	cultured in this study	Cultures from untreated plates	0.53
288	Ruminococcus gnavus_98%	cultured in this study	Cultures from untreated plates	0.61
289	Ruminococcus torques_96%	cultured in this study	Cultures from untreated plates	0.53
290	Ruminococcus albus_95%	cultured in this study	Cultures from untreated plates	0.50
291	Ruminococcus obeum_96%	cultured in this study	Cultures from untreated plates	0.48
292	Ruminococcus bromii_94%	cultured in this study	Cultures from ethanol treated plates	0.66
293	Ruminococcus flavefaciens_95%	cultured in this study	Cultures from untreated plates	0.47
294	Ruminococcus flavefaciens_93%	cultured in this study	Cultures from untreated plates	0.53
295	Ruminococcus bromii_93%	cultured in this study	Cultures from ethanol treated plates	0.58
296	Ruminococcus albus_95%	cultured in this study	Cultures from ethanol treated plates	0.48
297	Ruminococcus albus_98%	cultured in this study	Cultures from ethanol treated plates	0.63
298	Ruminococcus torques	cultured in this study	Cultures from untreated plates	0.52
299	Ruminococcus torques	cultured in this study	Cultures from untreated plates	0.55
300	Ruminococcus torques	cultured in this study	Cultures from untreated plates	0.53
301	Ruminococcus torques	cultured in this study	Cultures from untreated plates	0.52
302	Ruminococcus torques	cultured in this study	Cultures from untreated plates	0.58
303	Eubacterium rectale	cultured in this study	Cultures from untreated plates	0.56

304	Salmonella enterica	GCA_000006945.1	known non-spore formers	0.28
305	Sarcina ventriculi	cultured in this study	Cultures from ethanol treated plates	0.69
306	Shigella flexneri	GCA_000006925.2	known non-spore formers	0.22
307	Staphylococcus aureus	GCA_000512505.1	known non-spore formers	0.23
308	Streptococcus mutans UA159	GCA_000007465.2	known non-spore formers	0.20
309	Streptococcus pneumoniae	GCA_000007045.1	known non-spore formers	0.23
310	Streptococcus pyogenes	GCA_000006785.2	known non-spore formers	0.17
311	Streptococcus thermophilus	GCA_000011845.1	known non-spore formers	0.23
312	Streptomyces avermitilis	GCA_000009765.1	known non-spore formers	0.27
313	Streptomyces coelicolor	GCA_000203835.1	known non-spore formers	0.27
314	Sulfobacillus acidophilus	GCA_000219855.1	known spore formers	0.53
315	Sutterella parvirubra	GCA_000250875.1	known non-spore formers	0.16
316	Symbiobacterium thermophilum	GCA_000009905.1	known spore formers	0.66
317	Synergistes jonesii	GCA_000712295.1	known non-spore formers	0.22
318	Syntrophothermus lipocalidus DSM12680	GCA_000092405.1	known spore formers	0.66
319	Thermincola potens JR	GCA_000092945.1	known spore formers	0.70
320	Thermoanaerobacter ethanolicus CCSD1	GCA_000175815.1	known spore formers	0.80

321	Thermoanaerobacter pseudethanolicus ATCC33223	GCA_000192295.2	known spore formers	0.81
322	Thermosediminibacter oceani DSM16646	GCA_000144645.1	known spore formers	0.77
323	Thermus thermophilus	GCA_000091545.1	known non-spore formers	0.17
324	Turicibacter sanguinis	cultured in this study	Cultures from ethanol treated plates	0.56
325	Turicibacter sanguinis	cultured in this study	Cultures from untreated plates	0.52
326	Turicibacter sanguinis	cultured in this study	Cultures from ethanol treated plates	0.53
327	Turicibacter sanguinis	cultured in this study	Cultures from ethanol treated plates	0.53

Appendix 4 Publications from this thesis

Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. Nature. 2016.

Hilary P. Browne†, Samuel C. Forster†, Blessing O. Anonye, Nitin Kumar, B. Anne Neville, Mark D. Stares, David Goulding and Trevor D. Lawley

† These authors contributed equally

Transmission of the gut microbiota: spreading of health.

Nature Reviews Microbiology. 2017.

Hilary P. Browne, B. Anne Neville, Sam C. Forster and Trevor D. Lawley