1. Introduction

Bacteria form one of the largest groups of organisms occupying the planet, found in abundance across hosts and environments, thriving under some of the harshest conditions. The symbiotic relationship between bacteria and eukaryotes forms the basis of theories surrounding the acquisition of organelles and the effects of microbes on their host. Bacteria are highly effective colonisers, a trait which facilitates pathogenesis; however, bacteria are far more than pathogens with commensal bacteria being invaluable in maintaining human health. Bacteria often have short generation times and can rapidly expand their populations; they are very metabolically efficient making use of many molecules as food sources, and some populations, like the facultative anaerobes, can forgo the use of oxygen for metabolism to enhance their survival. Bacteria are also genetically dynamic making use of polymorphisms and efficient horizontal gene transfer to increase their repertoire of functions.

1.1 The Gut Microbiome

The term “microbiome” is used to define the collection of genomes belonging to the microorganisms in a particular niche, while the “microbiota” is used to refer to the organisms themselves (Prescott S.L., 2017). Environmental microbiomes such as the soil and water have been well understood, and these habitats have previously been used to classify bacteria. Humans are heavily colonised by a unique subset of bacteria, that exist in a commensal relationship with us, and can have profound effects on health and disease throughout the body. The human microbiome project represents a large-scale effort to consider the microorganisms resident in humans as a defined population and characterise them (Nelson K.E. *et al.,* 2010).

The gut microbiota is the community of bacteria that reside along the gastrointestinal tract (GIT). While they are referred to collectively, there are sub-niches or microenvironments with different microorganisms adapted to the range of environmental conditions along the GIT. There are defined communities in all the sections of the GIT including the stomach, duodenum, jejunum, proximal ileum, distal ileum, as well as further divisions along the colon. The abundance of bacteria increases in size along the GIT as conditions become less extreme, particularly in terms of pH, and become more favourable to bacterial life (Figure 1).

1.2 Residents of the Gut Microbiota

The residents of the gut microbiota are diverse and include members from several phyla including: Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria. While the community membership is evolutionarily diverse, there is a clear enrichment of bacteria from phylum Firmicutes as demonstrated in Figure 2. These species are believed to form the bulk of the gut commensal organisms that are required for a healthy gut. The Gram-positive Firmicutes had previously been under-sampled due to difficulties in culturing them. However, the methods described by Browne *et al.* (2016) have provided the tools necessary to probe this large phylum of bacteria, and will form the primary focus of this thesis.

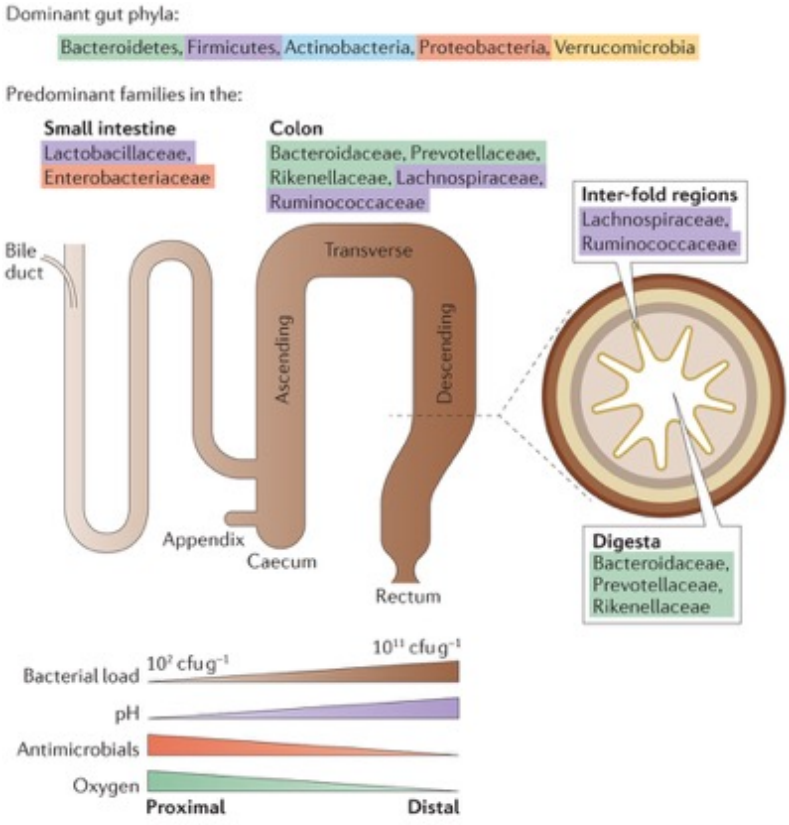

Figure 1.1. Distribution of bacterial phyla and environmental composition along the GIT. The bacterial load increases distally, as the environment becomes more favourable to anaerobes with increased pH and decreased oxygen. (Donaldson G.P. *et al.,* 2015)

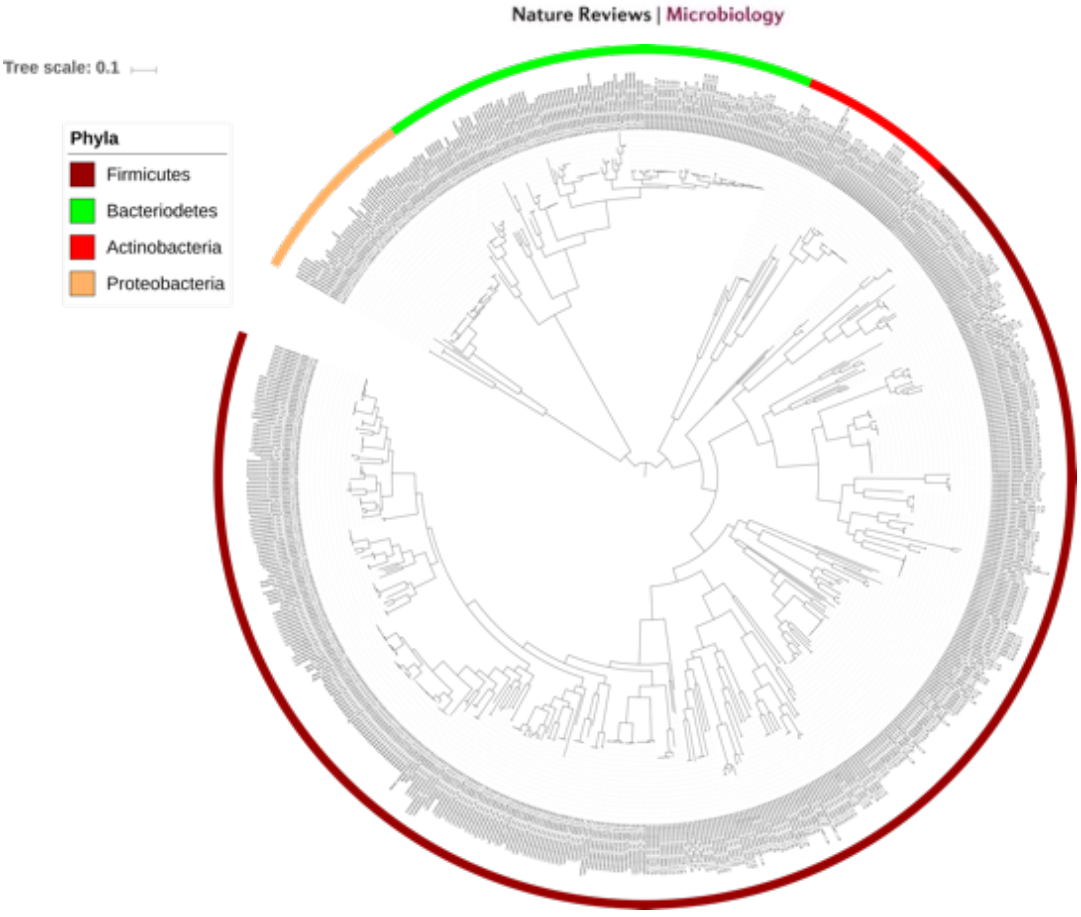

Figure 1.2. Phylogenetic tree of the Lawley lab human gut microbiota culture collection, including 653 genomes. A draft genome sequence is available for every isolate included in this tree. The outer ring indicates the phyla to which the various taxa belong. The phylum Firmicutes is well represented (Kumar N., *unpub*, [Accessed Feb 2017])

1.3 The Role of the Microbiota in Health and Disease

The human gut microbiota have been implicated in a variety of functions relating to health and disease. The symbiotic relationship is primarily beneficial; bacteria can mine fuel sources that the human body cannot readily digest for essential nutrients and supply nutrients like vitamins that we cannot synthesise. During the early development of germ-free (GF) techniques and animals, GF animals died while young due to malnutrition and vitamin deficiencies. The diet of a GF animal is supplemented with vitamins to ensure survival (Al-Asmakh M. and Zadjali F., 2015). There are however some instances where bacterial processing can cause dysbiosis. Bacterial metabolic productions have been found to be a part of our metabolome and may contribute to an individual's specific metabotype. This can have effects on how an individual processes food, altering susceptibility to lifestyle illnesses such as obesity; and how a patient processes medications which can have an impact on dosage and efficacy (Selwyn F.P. *et al.,* 2015). The microbiota have been implicated in the processing of common environmental toxins such as polycyclic aromatic hydrocarbons (PAHs) formed from incomplete combustion of carbon based fuels which encompasses sources from exhaust fumes through to burnt food. These compounds are processed and activated by the gut microbiota and the resulting in the formation of oestrogenic compounds that have a negative effect on the body (Van de Wiele T. *et al.,* 2005).

The gut microbiota also has a strong impact on systemic physiological and immunological processes. The physiology of the GIT is affected by the microbiota, GF animals have been shown to exhibit decreased rates of cell renewal, as well as disorganised villi development (Umesaki Y., 2014). In addition to changes in the GIT several studies that investigated the changes in development in germ-free mice, have shown stunted cardiac, lung and liver development (Al-Asmakh M. and Zadjali F., 2015). GF mice also have imbalances in fluid, which has a series of effects including: altered circulation as the volume of blood is affected, altered endocrine signalling, and changes to electrolyte balance (Al-Asmakh M. and Zadjali F., 2015).

The 'microbiota-gut-brain' axis is the proposed route through which signals are modulated between our guts and our behaviours and wellbeing, via the central nervous system. Signals are transduced through the vagus nerve as part of parasympathetic nervous system responses. The microbiota has been implicated in immune signalling and development, playing a role in the recruitment of intraepithelial lymphocytes (IELs), the induction of regulatory T cells (Treg) cells, and signalling processes such as fucosylation (Umesaki Y., 2014). Investigation of these pathways may explain the role of microbiota disruption in autoimmune diseases such as the inflammatory bowel diseases. The microbiota may also play a role in behaviour. Bacterial infection has been linked to anxiety (Foster J. A. and McVey Neufeld K. A., 2013), some probiotics are known to have effects on cognition (Davari S. *et al.,* 2013), and behavioural traits have been transferred through faecal transplant (Collins S. M. *et al.*, 2013). Changes to signalling through serotonin, and GABA related pathways have been observed alongside changes to the microbiota composition. (Foster J. A. and McVey Neufeld K. A., 2013). Germ-free mice, when compared to wild type mice, present changes in anxiety like behaviour,

memory and cognition, as well as social interactions (Luczynski P. *et al.,* 2016). These examples point to the microbiota playing a significant role in the healthy function of the body.

1.4 Plasmid Biology

Previous studies of the gut microbiome have primarily focussed on chromosomally encoded genes. However extrachromosomal elements such as plasmids form a valuable reservoir of auxiliary functions, and understanding the biology of the plasmids present in the gut will provide insight into the additional potential of the microbiota.

Plasmids were first characterised in 1952 by Lederberg and the definition redefined in 1965 to distinguish them from viruses (Tatum E. L. and Lederberg J., 1947). A plasmid is considered to be extrachromosomal DNA, that is able to self-initiate its replication and is often self-transmissible. Plasmids can be described and classified on the basis of their conserved elements, their structural conformations, and finally by their function.

Plasmids can have several conserved elements; all plasmids will be in possession of an origin of replication (ori). The ori is the region of DNA where replication begins. There the strands of DNA are cleaved and unwound by replication machinery. The ori has several key sequence features that are necessary to its function including: dnaA boxes, iterons, DNA methylation sites, AT rich regions and inverted repeats (del Solar G. *et al.*, 1998). These sequences are responsible for the specific binding of replication machinery, the regulation of replication, and marking the boundaries of mobile sequences such as transposons. The ori is also used to assign the plasmid to an incompatibility group. Plasmids of the same incompatibility (Inc) group will not coexist in a bacterium over many generations (Novick R. P.,1987). The Inc group modulates the recruitment of replication machinery, a bacterium simply recognises the presence of an inc group to initiate replication of plasmids in bacteria. Plasmids from the same inc group may not both recruit replication machinery. Other conserved elements of plasmids include functional genes, which are used to assign a functional class to plasmids. Finally, some megaplasmids (plasmids of size >100kb) will often encode their own replication machinery; however, this is not common and most plasmids will exploit the replication machinery of the host (Birge E.A., 2013).

Plasmids will take on one of three conformations depending on which stage in the replication cycle they are. As shown in Figure 3, the various plasmid conformations are supercoiled, circular, and linear; some plasmids will also inhabit intermediary conformations as they move throughout their replication cycle (Higgins N. P. and Vologodskii A. V., 2015). The native plasmid conformation is supercoiled and plasmids will ordinarily be found in this conformation in the host. The intermediary form between supercoiled and circular is supercoiled denatured. This conformation is often observed after experimental isolation of plasmids due to the disruption of bonds by processing chemicals. The next conformation is circular, plasmids are found in the relaxed circular conformation just prior to replication and in this conformation, the sequence is accessible to replication machinery. The

intermediary state between circular and linear conformations is the nicked open circular conformation. This conformation is found at the beginning of replication after the DNA has been cleaved at the ori by the replication machinery. The final conformation is linear, this conformation is found during replication, the DNA is laid out and read by the replication enzymes. However, there are a small subset of plasmids that exist naturally in the linear conformations using either hairpin loops or proteins to protect their ends (Dib J.R. *et al,* 2015)

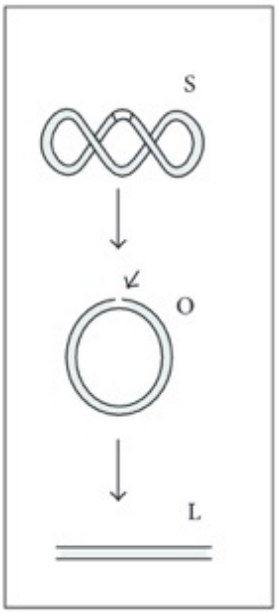

Figure 1.3. diagram of the conformations of plasmids, S- Supercoiled, O- Open, L- linear. (Somyoonsap P. *et al.,* 2013)

Plasmids can be divided into six categories based on the functional genes they encode: Fertility F plasmids, resistance R plasmids, Col Plasmids, Degradative Plasmids, Virulence Plasmids, and Integrative Conjugative Elements- ICEs. F plasmids contain the genes necessary for assembling the donor conjugation machinery, as well as insertion sequences (Griffiths A.J.F., *et al.,* 1999). Bacteria can only possess one F plasmid at a time. R plasmids encode resistance genes allowing bacteria to be resistant to a variety of threats, including heavy metals and different classes of antibiotics. R plasmids are probably the most investigated class of plasmid as antimicrobial resistance (AMR) and in particular mobilisable AMR, poses the greatest threat to humans. Col plasmids contain genes encoding for bacteriocins, proteins that can kill other bacteria, such as Colicin (Grohmann E., *et al.,* 2003). In the light of the growing AMR problem, colicin-like proteins are being investigated as potential novel anti-bacterial therapies. Degradative plasmids contain genes for enzymes that degrade atypical substances, such as the environmental toxins mentioned earlier. This allows bacteria to survive in a variety of conditions using these different substrates as alternative sources of energy, and the genes for these functions may be plasmid-encoded (Ogawa N. *et al.,* 2004). Virulence plasmids, contain pathogenicity loci that encode substances such as exotoxins, that allow a bacterium to be pathogenic (Moore R. *et al*., 2014). ICEs are genetic elements that possess machinery to be self-transmissible and integrate into or excise themselves from the chromosome at will (Johnson C. M. and Grossman A. D., 2015). Plasmids such as multidrug resistant plasmids or

megaplasmids may encode several cassettes or operons making them highly functional and advantageous to the bacterial population.

1.5 Plasmid Replication and Maintenance

The propagation of plasmids relies on their replication and transfer to other cells; either their daughter cells or other cells in the niche they inhabit. Plasmid replication is often accompanied by partition and stability genes. These genes ensure that plasmids are passed onto daughter cells. Circular plasmid replication follows two main strategies: opening of the ori primer-based replication, and single strand cleavage-based replication. The replication methods known as theta type and strand displacement fall into the former, and rolling circle replication employs the latter. Rolling circle replication is primarily used by plasmids found in Gram-positive bacteria, theta type by Gram-negative hosted plasmids, and strand displacement by plasmids in the IncP incompatibility group (del Solar G. *et al.*, 1998). Linear plasmids use two methods known as concatameric intermediates, and the protein primer mechanism (del Solar G. *et al.*, 1998).

Plasmids found in Gram-positive bacteria such as Firmicutes, which forms a major focus of this thesis, primarily undergo rolling circle replication Figure 4. Rolling circle replication begins at the double stranded origin of replication (dso). This sequence recruits the replication initiation protein (REP) to the binding site; this induces a conformation change in the sequence, exposing the nick site which is cleaved by REP. DNA helicase is recruited to the DNA strand and the DNA is unwound. Replication begins on the leading stand of DNA and proceeds until it is entirely displaced. On the lagging strand, replication begins at the single stranded origin of replication (sso) with the recruitment of a primer and DNA polymerases. To finish replication on both strands, the DNA is ligated and gyrase is used to coil the DNA (Khan S. A., 2005). Theta replication differs in that the DNA is not nicked, the two strands are separated, continuous replication occurs on the leading strand and discontinuous on the lagging strand (del Solar G. *et al.*, 1998).

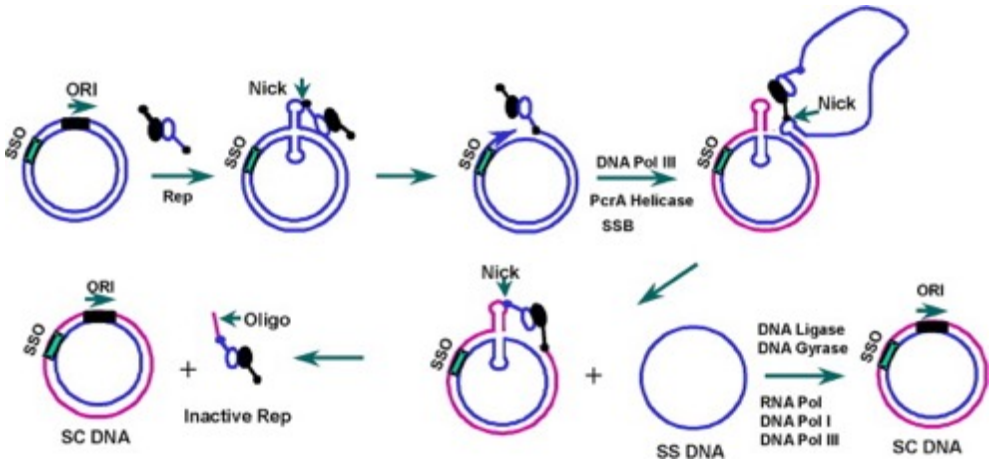

Figure 1.4. Summary of the process of rolling circle replication. (Khan S. A., 2005)

Plasmid partitioning is used to ensure low copy number plasmids survive in the population. Plasmids are moved to specific subcellular loci to increase the likelihood that the plasmids will be passed on to a daughter cell during division. The partition (par) system is formed of two proteins: a DNA binding protein, and an NTPase. The par systems are classified by the type of NTPase used for movement (Baxter J. C. and Funnell B. E., 2014). The par genes can also regulate plasmid incompatibility and plasmids with similar par genes can be paired (Pinto U. M. *et al.,* 2012). Plasmid stability is normally maintained through toxin and anti-toxin genes. Toxin genes are normally plasmid-encoded, and the anti-toxin may be encoded on the same or a different plasmid. Cells that contain both genes will be resistant to the toxin and cells without the anti-toxin gene will die. This ensures that only the daughter cells that have successfully acquired the plasmid(s) remain in a population. However, some bacterial cells have evolved addiction modules (chromosomally-encoded anti-toxin genes), to allow them to discard plasmids without succumbing to the toxin.

1.6 Horizontal Gene Transfer

The other method of plasmid dispersal is through horizontal gene transfer (HGT); this strategy allows bacteria to acquire new pieces of DNA that can increase fitness in a particular environment. The methods have been summarised in Figure 5.

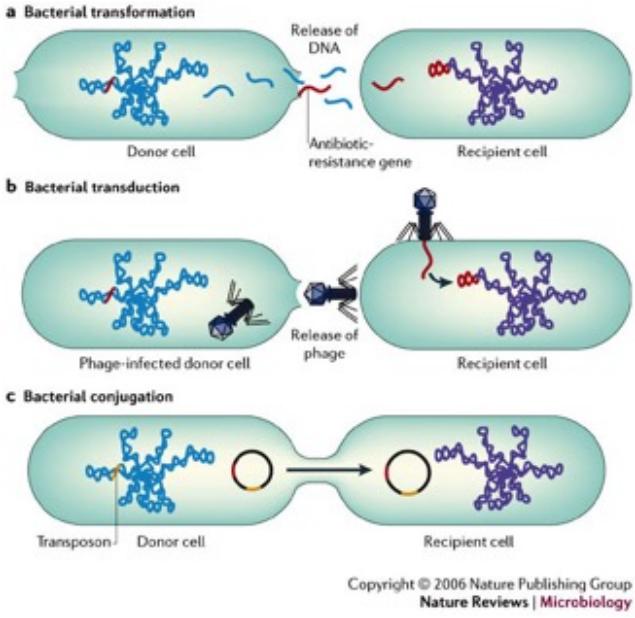

Figure 1.5. Figure summarising the methods of HGT: transformation (a), transduction (b), and conjugation (c). (Furuya E. Y. and Lowy F. D., 2006)

Transformation is the process by which genetic material is passed between bacteria in the surrounding environment. Bacteria hosting a plasmid are lysed and the DNA released into the surrounding environment; neighbouring bacteria can absorb these genetic elements. Transformation relies on the receptivity or competence of the recipient. A variety of stresses can induce bacteria to become competent, including a lack of nutrients and DNA damage. This induces bacteria to uptake environmental DNA as either an alternative nutrition source or as a template for DNA repair (Huddleston J. R., 2014).

Transduction is the acquisition of new DNA elements through viral infection. Bacteriophages are viruses that infect bacterial species. Bacteriophage infection is either lysogenic or lytic. Lysogenic infections involve the replication of viral genetic information alongside the host genome for several cycles before lysis. The viral genome may even become a stably expressed plasmid. However, some bacteria have developed a resistance to phage infections through the CRISPR-cas defence system.

Conjugation is the process by which DNA passes between bacteria through structures known as pili. A plasmid may be self-transmissible and contain all the genes needed for conjugation, or it may be mobilisable and require a helper plasmid or chromosomally-encoded conjugation machinery. Conjugation is an active process that allows the targeted and efficient spread of mobile genetic elements throughout a bacterial population. Plasmid transfer via a type four secretion system (T4SS) is a well-characterised method of conjugation: the system is comprised of DNA binding proteins, channel proteins, pili assembly proteins, and ATPases. Conjugation begins at a distinct site known as the origin of transfer (oriT), plasmid DNA is nicked at this site by the relaxosome which is formed from a relaxase and the DNA transfer and replication (Dtr) machinery. The resulting DNA-relaxosome complex, or nucleoprotein, is recruited to the membrane bound channel by a coupling protein. The nucleoprotein is bound to a trafficking ATPase, and it is proposed that the relaxase is also unfolded at this stage. The trafficking ATPases catalyse the formation of the transport complex which moves through the channel and is pumped through the pilus by ATP hydrolysis. The pilus adheres to the recipient cell, the nucleoprotein is released into the recipient and the relaxase re-circularises the DNA (Cabezón E. *et al.,* 2015).

1.7 Frequency of and Challenges to HGT

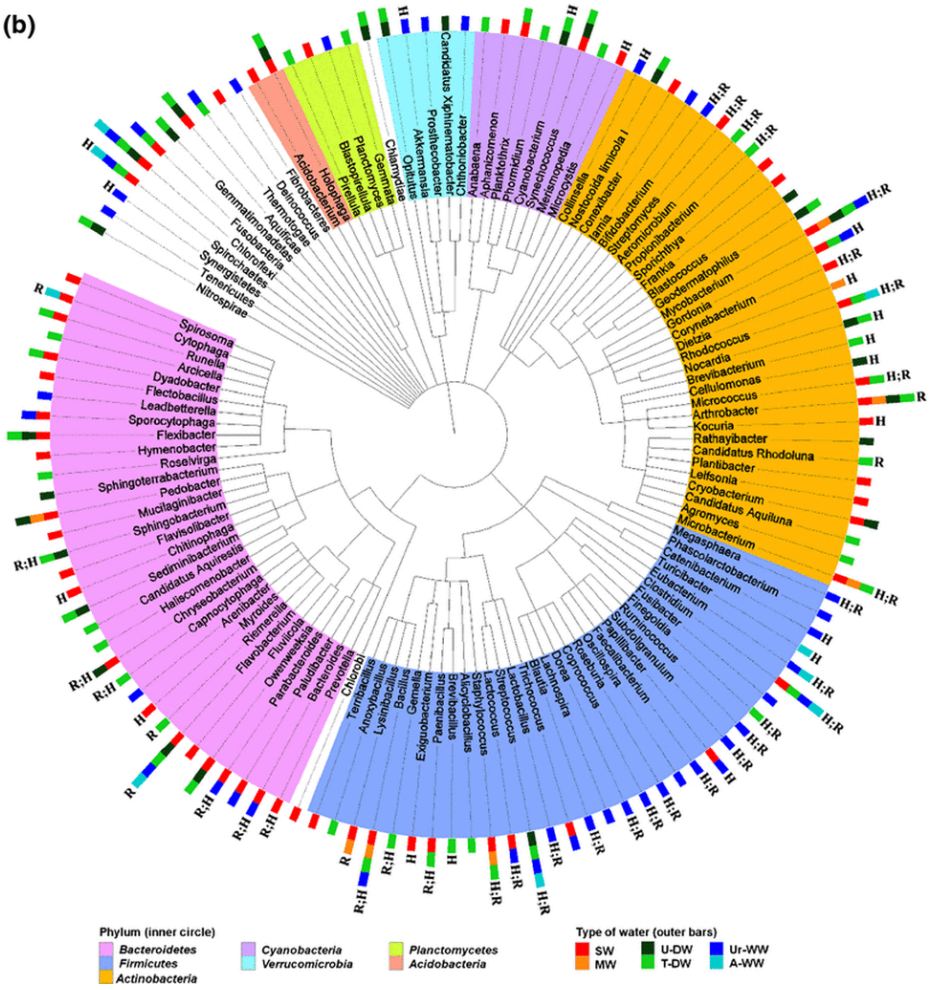

Figure 1.6. Phylogenetic tree displaying the distribution of bacteria in environmental water sources compared with the distribution of bacteria in the human microbiome (H), and the distribution of resistance genes (R). (Vaz-Moreira I. *et al.,* 2014)

Environmental microbiomes are more diverse than the human microbiota. As shown in Figure 6, the number of species that inhabit both humans and water sources is a fraction of the species solely found in water sources. The environment has been considered as a reservoir of plasmids and a donor for HGT, however this process does face some limitations. The first limiting factor is the stability of the environment; environmental ecosystems like the deep sea can remain undisturbed for thousands of years; comparatively the human microbiome is not a highly stable environment for long term evolution. It is frequently physically and chemically perturbed, and will ultimately only last several decades. The environment also affects the proximity and receptivity of bacteria; bacteria need to be in close contact for conjugation to occur, and environmental stressors determine transformation competence. HGT has been shown to occur when proximity is altered. For example, the microbiota of Japanese populations have acquired genes for the digestion of seaweed from ancestral marine bacteria, however these genes are not found in the microbiota of western populations where seaweed is not consumed to the same volume and frequency (Hehemann J-H. *et al.,* 2010).

Restrictions to HGT are not solely environmental, genetic restrictions play a large role. Plasmids are placed into incompatibility groups on the basis of their origin of replication. Incompatibility groups affect plasmid stability and maintenance over generations. Plasmids in the same group will not be co-

maintained over generations of replication because the bacterium will only recognise the ori sequence and perceive that it has multiple copies of one plasmid and not several different plasmids. This will lead to the plasmids being divided among the progeny. However, plasmids in different groups will both recruit replication machinery and be co-maintained over generations.

Finally, there are physical restrictions, all plasmids will have a cost of carriage that the bacterial cell has to pay in order to maintain the plasmid. The acquisition of a plasmid often leads to an increased lag phase, decreased maximal growth rate, and increased nutrient consumption (Baltrus D. A., 2013). Upon entry into the cell, integrative plasmids may cause genomic disruption by inserting into important regions of the chromosome such as essential genes or regulatory elements. Replication of the plasmid DNA requires host supplied building blocks – nucleotides and tRNAs – and machinery: polymerases and ribosomes. Moreover, in addition to requiring the use of ribosomes and polymerases, plasmids may not be codon optimised for the new host and would therefore create a large demand for tRNAs that are low in supply. The products of the plasmid sequence could cause cytotoxicity because they may be complex proteins requiring chaperones not possessed by the host, resulting in a high load of misfolded proteins (Baltrus D. A., 2013). The products may also disrupt cellular networks, altering cell signalling either through interactions with host proteins or through changes to the abundance of signalling molecules and metabolites (Baltrus D. A., 2013). It can be assumed that the plasmids in circulation exert a manageable cost of carriage because plasmids with such drastic effects would be quickly lethal. Plasmids also contain functional genes, and the plasmid-encoded functions could be the difference between dying and thriving in a particular environment. This explains why the plasmids in the human microbiome are enriched for AMR and metabolite processing. This theory proposes that there may be an upper limit on the size of plasmid that can readily circulate throughout the population. When applied to AMR this suggests that there is a point at which carrying a multidrug resistance plasmid in an environment where it is not being selected for would put bacteria at a disadvantage, thereby limiting the amount of resistance a bacterium can carry.

1.8 The Utility of Plasmids as Tools for Biotechnology

Plasmids provide bacteria with a wide variety of additional functionality under native conditions, but plasmids can also be exploited for a variety of applications in molecular biology, biotechnology and medicine.

Plasmids have been used as cloning vectors in molecular biology experiments. Bacterial plasmids can be engineered to contain a “multiple cloning site” (MCS), which is a short segment of DNA containing several restriction sites allowing for the addition of a variety of constructs, provided they contain compatible ends. This is useful to create constructs for the modification of both prokaryotic and eukaryotic cells. Competent cells of the bacterium intended for genetic manipulation can be derived in the laboratory and bacterial AMR genes provide a rapid and potent method of selecting for transformants.

Cloning vectors are primarily used to transform bacteria for vector propagation or protein production. During vector propagation the plasmid is engineered, transformed into bacteria, the bacterial population grown and harvested, and the plasmid purified for downstream uses. Bacteria may also be used as the producers of the protein coded on the cloning vector. Some origins of replication are highly replicated in bacteria such as the pUC ori (Lin-Chao S. *et al.,* 1992); these features can be used to increase production and create high yield production cells.

Plasmids can be used to transform cells other than bacterial cells, including mammalian, yeast, and fungal cells for a variety of functions. They can be used to turn a mammalian cell into a protein production cell. Protein production in mammalian or yeast cells might be preferred over production in bacteria for optimised yield and correct post translational modification. Mammalian protein sequences will not be codon optimised for bacteria and may place stress on the bacteria leading to inefficient production. The bacteria may also not possess the necessary folding machinery or localisation signals for complex proteins which could lead to cytotoxicity in the cells.

The primary application of plasmids in medicine is use in gene therapies. Gene therapy involves the delivery of genetic material to cells for a therapeutic effect either *in* or *ex vivo*. Plasmids have been used in a variety of ways in gene therapy. Plasmids are a widely used method in gene therapy; however there may be issues with delivery as naked DNA may be degraded before reaching the target (Figure 7). More commonly, plasmid DNA is used in conjunction with a vector. Viral vectors have been shown to be the most effective vectors and form the bulk of therapies currently in development (Figure 7).

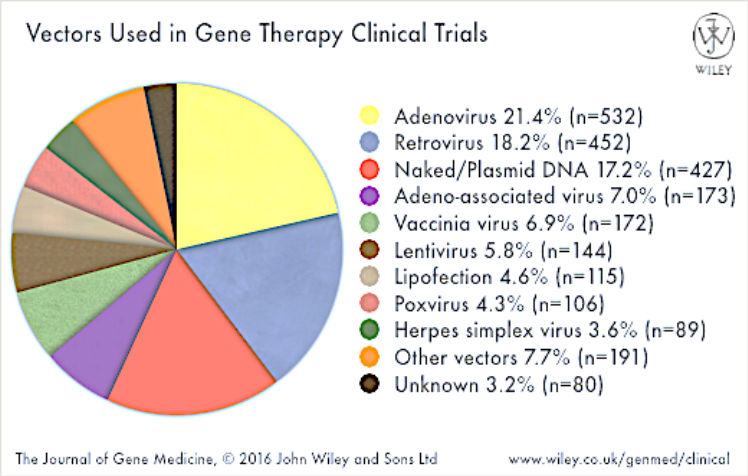

Figure 1.7. Figure displaying the frequency with which a variety of vectors are used in gene therapy trials. (Wiley Journal of Medicine, 2016)

1.9 Plasmids and the Gut Microbiome

The “gut mobilome” is the term used to describe the mobile genetic elements, including the plasmids, of the gut microbiota. These elements extend the functionality of bacteria in the gut; facilitating both their survival and symbiotic relationships with the host. Plasmids in the gut have not been widely sampled owing to difficulties with culturing and experimental isolation. There have been some

metagenomics studies including the study by Jones *et al.* (2010) where metagenomics data was probed for conserved sequences of the RelBE toxin anti-toxin system to observe the phylogenetic distribution of plasmids in the gut microbiome (Figure 8). This data shows the prominence of plasmids in bacteria from the Firmicute phylum, however a study examining the enrichment of plasmids across all members of the microbiota has not been conducted.

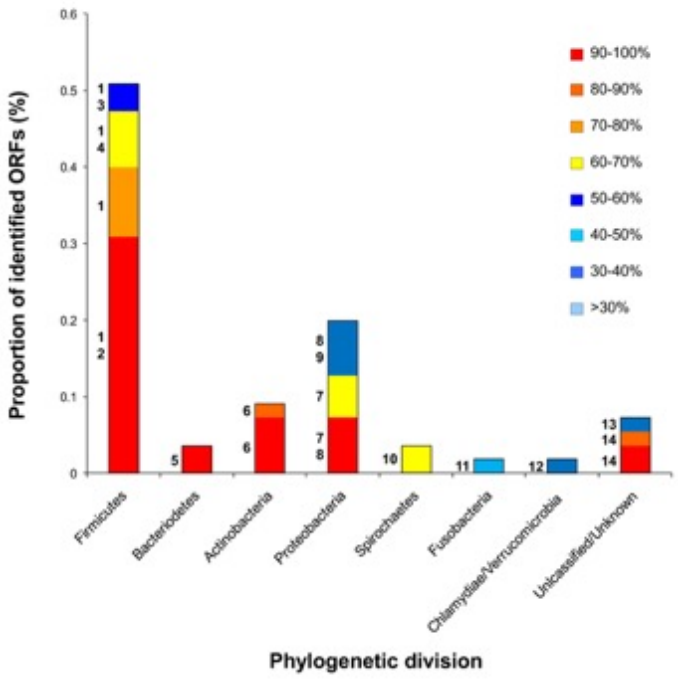

Figure 1.8. Graph displaying the proportion of RelE genes identified in gut metagenomic data in each phylum represented in the gut. (Jones B. V. *et al.,* 2010)

1.10 Plasmid Functions in the Gut Microbiome

The gut microbiome is often routinely exposed to doses of antibiotics through a variety of routes including through food and agricultural processes as well as prescribed courses. The agricultural industry routinely includes antibiotics in animal feeds and this exposes humans to antibiotics through the meat and increases the antibiotics in the environment through farm waste. This is the case of tetracycline, which is included in animal feeds; tetracycline resistance is prevalent in the gut and soil bacteria (Popowska M. *et al.,* 2012). There has also been a rise in the use of antibiotics for medicine, and studies have shown poor public knowledge of appropriate antimicrobial use. Subjects were unaware of appropriate conditions requiring antimicrobials, resulting in pressure on physicians to over prescribe (Gualano M. R. *et al.,* 2014). Subjects also exhibited a lack of knowledge in the proper administration of antibiotics leading to them not completing courses of antibiotics (Gualano M. R. *et al.,* 2014). This exposure allows commensals and pathogens alike to gain resistance to antibiotics. The gut is considered an AMR reservoir and HGT between pathogens and commensals may be responsible for some pathogens acquiring resistances; some species have already been implicated in the HGT of AMR cassettes.

The plasmids in the gut microbiota provide functionality other than AMR including metabolic processing and virulence factors. One key family of plasmids is the *Lactobacillus* megaplasmids that are found in several strains, including *L. salivarius* UCC118 that contains the plasmid pMP118. This plasmid contains several genes that allow *L. salivarius* to metabolise rhamnose, sorbitol, and ribose; produce bile salt hydrolases; and produce the bacteriocin Abp 118 (Li Y. *et al.,* 2007). Another important plasmid-encoded metabolic gene is C7D2 glucuronide transporter. C7D2 transports the products of β-glucuronidase enzyme activity and it has been implicated in Crohn's disease (CD). C7D2 and C7D2 expressing bacteria levels are elevated not only in CD patients, but in their unaffected immediate family members, making it a potential marker for CD predisposition (Gloux K. and Anba-Mondoloni J., 2016).

Plasmid-encoded genes have also been important for the virulence of bacteria in humans and animals. *C. perfringens* ordinarily exists in the gut microbiome at low levels, however it can be a potent pathogen and several pathogenicity loci are known to be contained on its plasmids including the CPE locus which causes food poisoning in humans, and the NetF locus which causes haemorrhagic gastroenteritis in dogs and necrotising enteritis in horses (Gohari I. M. *et al.,* 2016). Another example of plasmid-encoded virulence is the plasmid-encoded bacteriocin BAC21. *E. faecalis* contains the pPD1 plasmid that encodes *bac21* and its immunity gene. *Enterococcus faecalis* strains expressing this plasmid-encoded bacteriocin clear non-expressing strains and more readily colonise the niche. However, this plasmid is readily transferred through the population prompting the creation of resistant strains in addition to rapid colonisation of the introduced strain. The use of a conjugation-deficient strain allowed an introduced resistant strain to be effective killers without passing on the resistance, and this may be used as a strategy to clear harmful strains of bacteria (Kominneni S. *et al.,* 2015).

1.11 HGT in the Gut Microbiome

The full extent of HGT in the gut is not known; however there are some key examples of it occurring that will be discussed below. AMR is the major concern and has been observed in several cases including the transfer of carbapenem resistance between species of *Enterobacteriaceae*; the transfer of ampicillin resistance between species of *E. coli*; and the transfer of a multidrug resistance cassette to *Escherichia coli* from *Klebsiella pneumoniae* (Huddleston J. R., 2014). The C7D2 mentioned previously as being an indicator in CD was transferred from *Eubacterium eligens* into Firmicute species in the gut such as *Ruminococcus gnavus* (Gloux K. and Anba-Mondoloni J., 2016). The C*lostridium perfringens* toxicity plasmids all contain the 'transfer of Clostridia plasmids' (tcp) locus, meaning that they are conjugation competent (Gohari I. M. *et al.,* 2016). It has also been proposed that inflammation in the gut can contribute to increased HGT. Indeed, by monitoring HGT between *Salmonella* and *E. coli,* inflammation was found to cause the depletion of commensals and expansion of pathogenic populations, in turn increasing their proximity and facilitating HGT leading to increased resistance (Stecher B. *et al.,* 2012).

1.12 Plasmid Capture Strategies

Plasmids have been isolated from the gut microbiota for phenotypic and informatics analysis using a variety of endogenous and exogenous strategies. The development of new experimental and informatics techniques and tools will be crucial to further characterisation of the gut mobilome. The workflow of the methods discussed has been summarised in Figure 9.

Endogenous techniques are culture-dependent; bacterial isolates are cultured and purified from stool samples. The plasmids can be extracted from bacteria and bacterial DNA preps using a variety of techniques including Pulsed Field Gel Electrophoresis (PFGE), commercial plasmid kits, and alcohol precipitation. The plasmids can then be sequenced, and analysed using informatics tools to assemble and annotate the sequences. This method allows scientists to have the information about the bacterial host as well as the plasmid sequence data. This is often the preferred method and had been the standard protocol for isolating plasmids. The vast majority of bacteria in the gut were believed to be 'unculturable', however, the paper by Browne *et al* (2016) outlines a protocol to culture these bacteria and has led to the discovery of many new families, genera and species in the gut microbiota.

While it was thought that the vast majority of microbes in the gut were unculturable there was a need to develop alternative strategies or 'exogenous methods', to isolate plasmids from these species in a culture-independent manner and expand the range of species that could be investigated. Helper bacteria can be used to isolate bacteria and their mobile elements following the bi- or tri- parental methods outlined by Ó Cuív *et al* (2015). Alternatively, stool samples are collected, metagenomic DNA extracted, and metagenomic libraries sequences and analysed using software such as plasmidSPAdes. PlasmidSPAdes identifies plasmid sequences from whole genome sequence data by detecting sequences with a coverage that varies from the median coverage (Antipov D. *et al.,* 2016). Another strategy was developed by Jones B. V. *et al,* and is called the transposon aided capture (TRACA) protocol. Plasmids are isolated from metagenomic DNA preparations and the ori and selectable marker are modified using the Tn5 transposase system. The plasmids are then grown in an *E. coli* surrogate, purified, sequenced, and undergo bioinformatic analysis (Jones B. V. and Marchesi J. R., 2007). Finally, chromosome conformation capture and sequencing can be used isolate plasmid DNA in metagenomic samples. DNA is extracted from a metagenomic sample and all the DNA contained in each cell is fixed together; providing a complete picture of the genetic information in each bacterium. This allows extrachromosomal elements to be observed alongside the chromosome providing host information (Marbouty M. and Koszul R., 2015).

Figure 1.9. Flow diagram outlining the variety of processing routes of stool samples for plasmid analysis

1.13 Challenges to Plasmid Capture

There are several challenges that face both methods of isolation. Endogenous methods as previously stated are limited to species of bacteria that can be successfully cultured. Limiting discoveries to bacterial taxa that are readily cultured, which left the gut microbiota largely un-sampled until recently.

Exogenous sampling allows more species to be sampled but the data does not contain the host information, so further experiments are needed to gain host information post-sequencing; for example *in situ* hybridisation or informatic predictions based on sequence signatures. This challenge affects all exogenous methods, however each method faces specific challenges.

The TRACA method relies on modification of the plasmid and the compatibility of the surrogate. The surrogate may not be able to replicate the desired plasmid or may not be able to maintain it if there is incompatibility with any native plasmids. The genes on the plasmid may not be correctly expressed,

so phenotypic analysis may not be possible, and the mutagenesis step means that the plasmid is no longer in its native form, which may have effects on the expression of the cassette. The authors have also identified a potential size bias, as the experiment has only been successful in isolating smaller plasmids. It is not known if this is due to a true enrichment of small plasmids in the gut, or if it as a result of experimental bias (Jones B. V. and Marchesi J. R., 2007). However larger plasmids have previously been isolated from gut bacteria so it is expected that they would be sampled.

Identifying plasmid sequences from metagenomic databases using informatic tools such as plasmidSPAdes faces several bespoke challenges as well. The creators of the software have identified several key areas of bias. The software relies on identifying potential plasmid sequences based on variation from the median copy number so high copy number plasmids are more easily identified; however, this can lead to low copy number plasmids being excluded or false positive hits from chromosomal sequences. The authors also note that the script had difficulty identifying linear plasmids and would not provide an accurate picture of the proportion of linear plasmids present (Antipov D. *et al,* 2016).

Several sources of experimental bias affect both methods, such as that introduced during culture and DNA preparation. The growth phase at the time of harvest can affect plasmid copy number which can skew the results to portray a plasmid as being of the wrong copy number. DNA extraction and sequencing protocols have been known to significantly affect results particularly the observed diversity.

1.14 Hypothesis

Plasmids have been demonstrated in the more well-known members of the gut microbiome, and it stands to reason that similar and novel plasmids may be found in the more under-characterised members of the gut microbiome. The primary functions of these plasmids may be related to AMR, metabolic functions, and horizontal gene transfer.

1.15 Project Aims

Using a variety of endogenous and exogenous methods, this project aims to firstly, quantify the enrichment of plasmids in the Lawley Lab culture collection, with particular interest in phylum Firmicutes. Secondly, the mobility of genetic elements expressing AMR genes in the gut culture collection will be investigated.

1.16 Project Rationale and Identified Challenges

Plasmids provide additional functionality to bacteria; some of these functions have already been identified as having implication in health and disease. There are large populations of plasmids that are largely unsampled and uncharacterised, particularly the plasmids of species that are considered commensal and necessary for healthy function. Plasmids also provide tools with which to modify bacteria: plasmids found in the gut microbiota could be used to engineer commensal species to make

gene knock outs, or to produce proteins to better maintain gut health or treat disease. The phylum Firmicutes accounts for the largest group of bacteria that have remained uncultured (Browne H.P. *et al* 2016), therefore plasmids from these species will be considered of particular interest.

The majority of the challenges to the project have been outlined previously, including bias in the software, suitability of the plasmid surrogate hosts in the conjugation system, and experimental bias in DNA preparation for sequencing. However, this project makes use of both endogenous and exogenous strategies to try and identify plasmids in the gut microbiota. This approach provides a unique opportunity to observe the correlation between *in silico* prediction and *in vitro* validation. Plasmid sequences will be identified *in silico* using the plasmidSPAdes software; isolated experimentally using anaerobic culture techniques and plasmid preparation; used for genetic manipulation of commensal species; and sequenced and annotated.