Identification of Extrachromosomal Elements from Whole Genome Sequences of the Human Gut Microbiome to Investigate the Gut Mobilome and Resistome

Tapoka Thulisile Mkandawire
USN:303635303

University of Cambridge

Murray Edwards College

Table of Contents

List of Figures and Tables

Acknowledgements

Abstract

We are covered with beneficial microbial communities, termed microbiota, that play important roles in our health, sustenance and well-being. Pathological imbalances in our microbiota caused by things like antibiotic exposure and poor diet can directly cause, or predispose us to, a variety of diseases and metabolic syndromes. Our intestinal microbiota contains hundreds of uncultured bacterial species and the genetic complement of our microbiota, our microbiome, contains 150 times more genes than the human genome. The human microbiome is dynamic due to bacteria's ability to horizontally transfer large blocks of DNA between distantly related species. Horizontal gene transfer by mobile elements such as transposons and plasmids plays a central role in the evolution and functions of well-studied intestinal bacteria such as *Escherichia coli* and *Lactobacillus*. However, most of the human microbiota has never been cultured and characterized so the composition of mobile elements in the majority of the human microbiota remains poorly defined. For my MPhil Thesis, I analysed a large scale human commensal reference genome collection of 653 genomes, representing the phylogenetic diversity of the human intestinal microbiota, for the presence of extrachromosomal elements like plasmids and transposons. A combination of bioinformatics analyses and wet-lab validation methods have been employed to identify, isolate, and characterize elements from gut bacteria creating a catalogue of known and novel sequences. 240 genomes were predicted to contain extrachromosomal DNA, these elements are primarily small and high coverage, and predicted to contain resistance genes. A small number of megaplasmids were also detected, and a phylogeny was built to identify any plasmids with a broad host range. This database can be used as a reference for the computational isolation of extrachromosomal elements from whole genome sequencing data and metagenomic datasets. They will also form the foundation for developing tools for genetic manipulation of the novel and uncharacterized gut commensal microbes of the culture collection.

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.
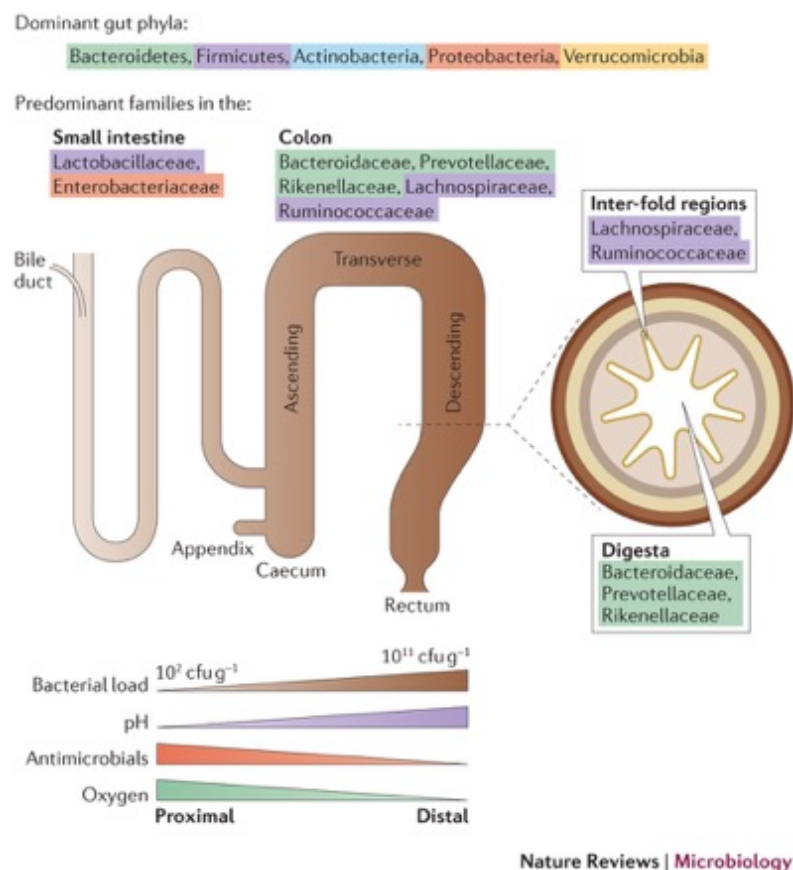
Figure 1.1. Distribution of bacterial phyla and environmental composition along the GIT. The bacterial load increases distally, as the environment becomes more favourable to anaerobes with increased pH and decreased oxygen. (Donaldson G.P. *et al.,* 2015)
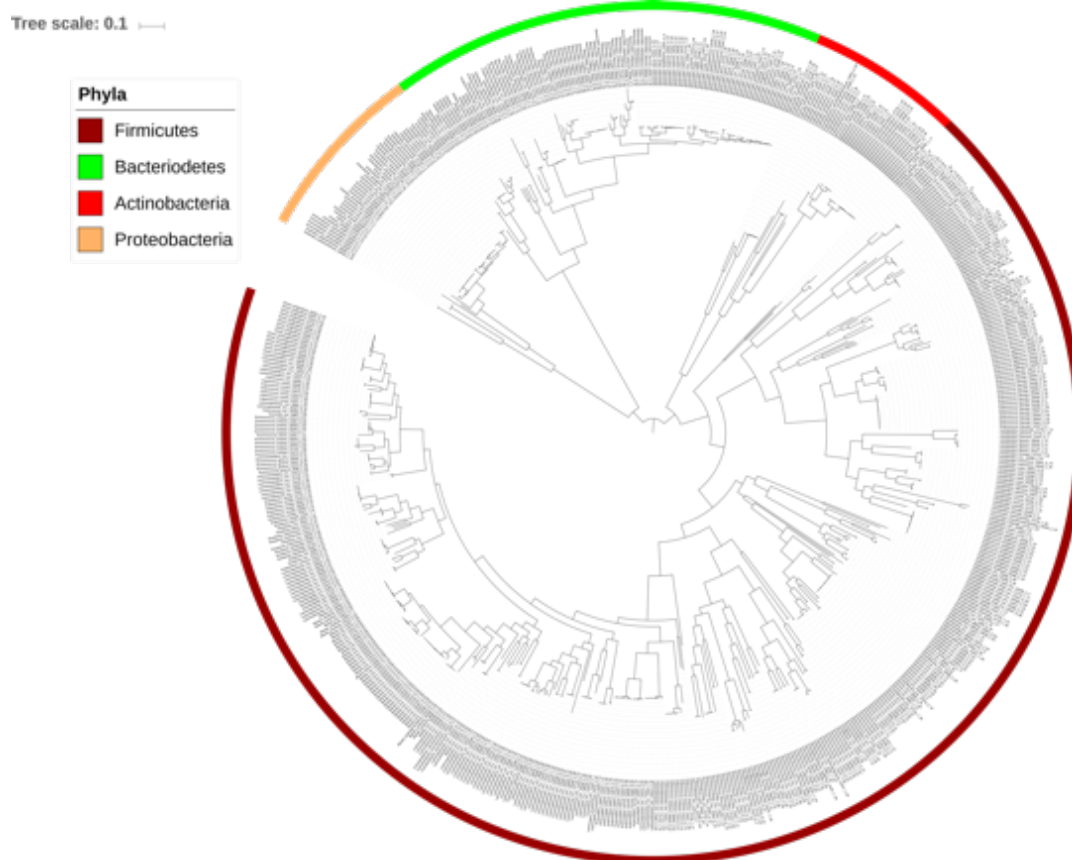
Figure 1.2. Phylogenetic tree of the Lawley lab human gut microbiota culture collection, including 653 genomes. A draft genome sequence is available for every isolate included in this tree. The outer ring indicates the phyla to which the various taxa belong. The phylum Firmicutes is well represented (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)
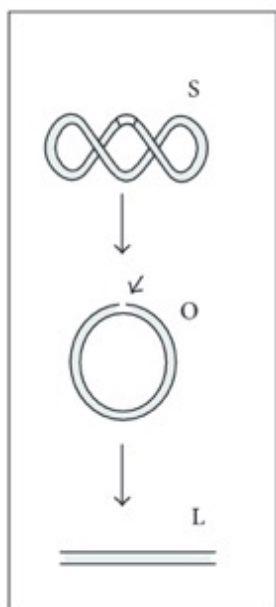
Figure 1.3. diagram of the conformations of plasmids, S- Supercoiled, O- Open, L- linear. (Somyoonsap P. *et al.,* 2013)

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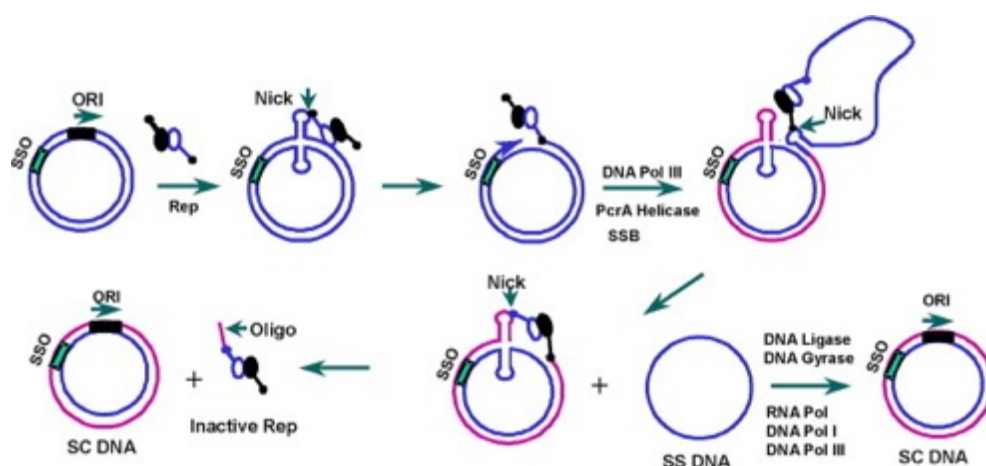
Figure 1.4. Summary of the process of rolling circle replication. (Khan S. A., 2005)

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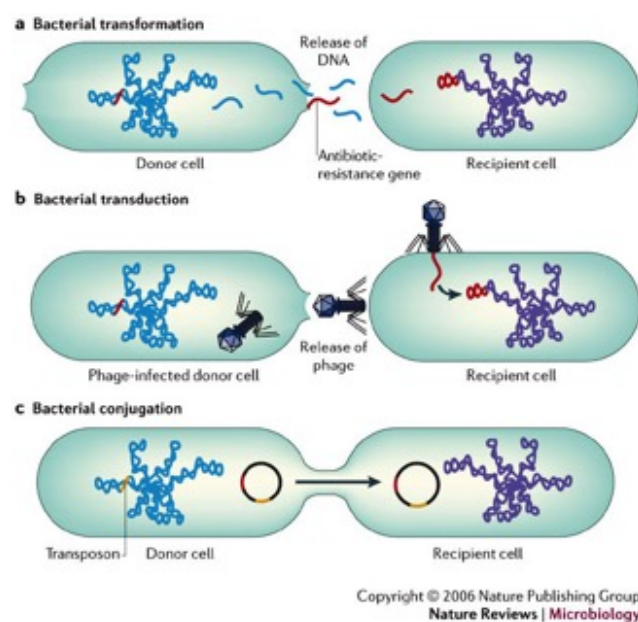
Figure 1.5. Figure summarising the methods of HGT: transformation (a), transduction (b), and conjugation (c). (Furuya E. Y. and Lowy F. D., 2006)

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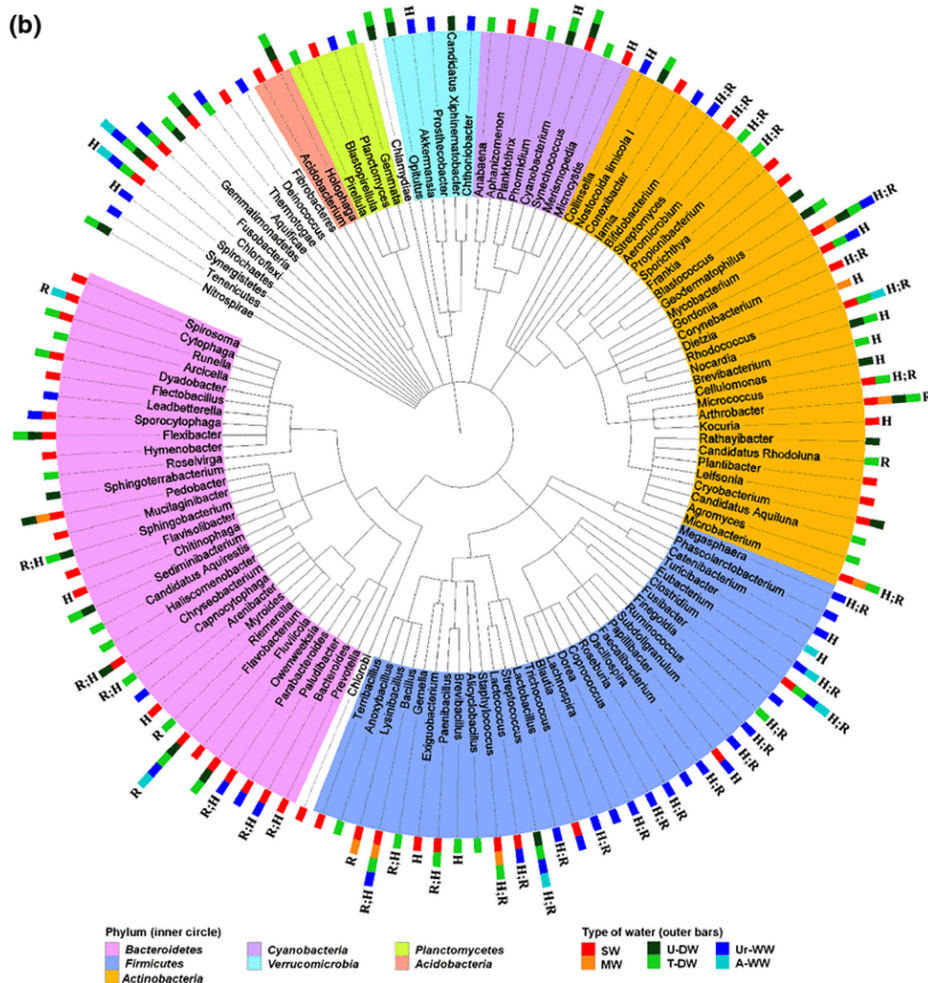
Figure 1.6. Phylogenetic tree displaying the distribution of bacteria in environmental water sources compared with the distribution of bacteria in the human microbiome (H), and the distribution of resistance genes (R). (Vaz-Moreira I. *et al.,* 2014)

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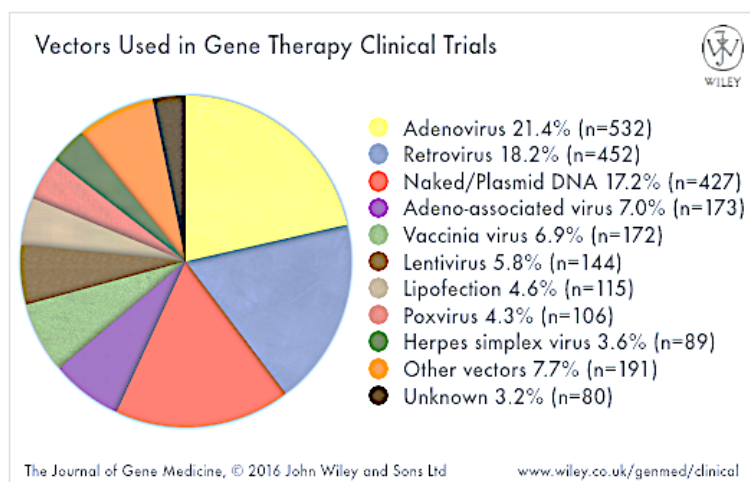
Figure 1.7. Figure displaying the frequency with which a variety of vectors are used in gene therapy trials. (Wiley Journal of Medicine, 2016)

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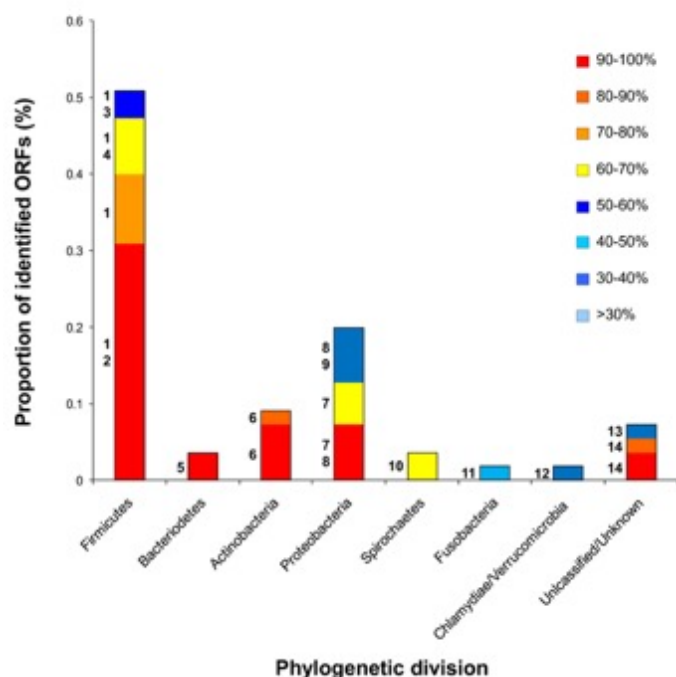
Figure 1.8. Graph displaying the proportion of RelE genes identified in gut metagenomic data in each phylum represented in the gut. (Jones B. V. *et al.,* 2010)

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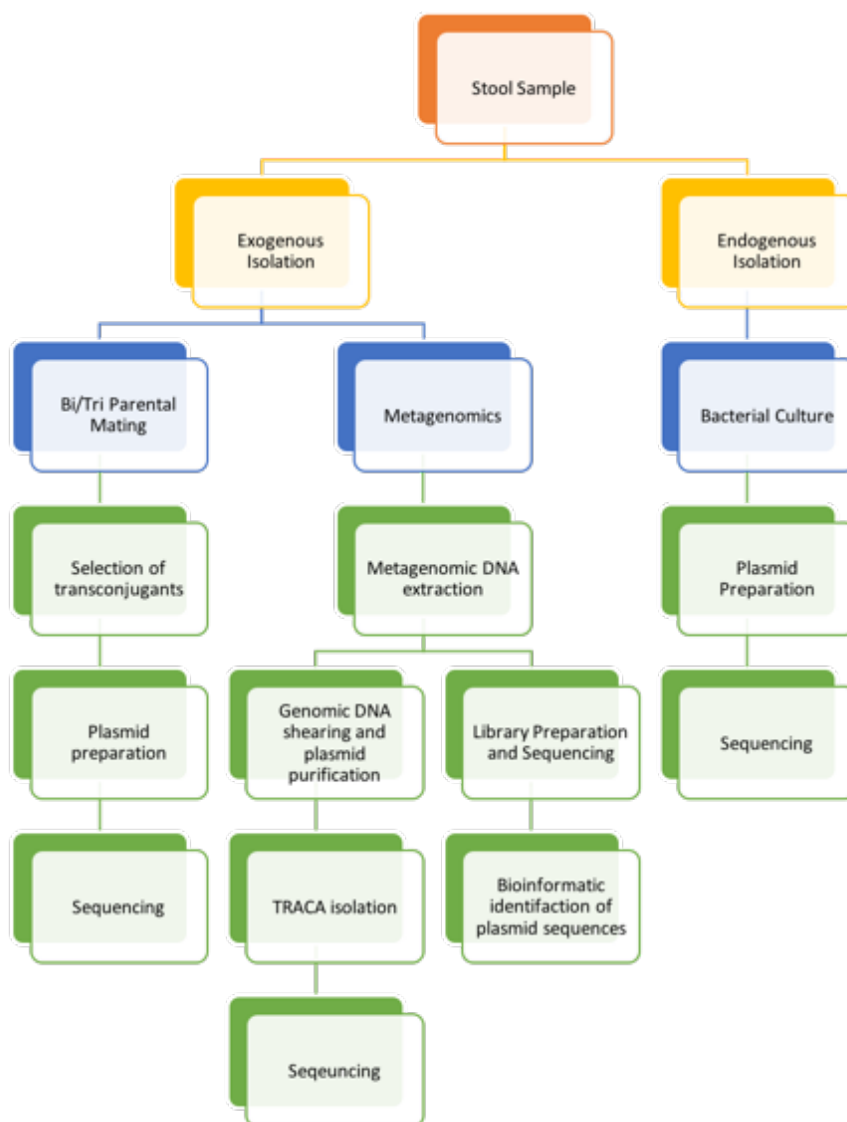
Figure 1.9. Flow diagram outlining the variety of processing routes of stool samples for plasmid analysis

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.

2. Methods

The Lawley Laboratory Culture collection was created from healthy donor faecal samples as described by Browne *et al.*, and this body of work makes use of the whole genome sequences (WGS) and the pure cultures of human gut microbiota species (Browne H. *et al* 2016).

The workflow is summarised in the flow diagram in Figure 1. WGS data were scanned with plasmidSPAdes for putative plasmid sequences. These sequences were BLAST (blastn) against the NCBI database to filter out well-known and well-characterised plasmids. The remaining plasmids were annotated and their annotations probed for functional genes such as mobility and AMR genes. The presence of the plasmids and their predicted phenotypes were then validated experimentally. In addition, growth curves were plotted, the isolates were 16s sequences validated, and plasmid DNA was extracted for long read sequencing.
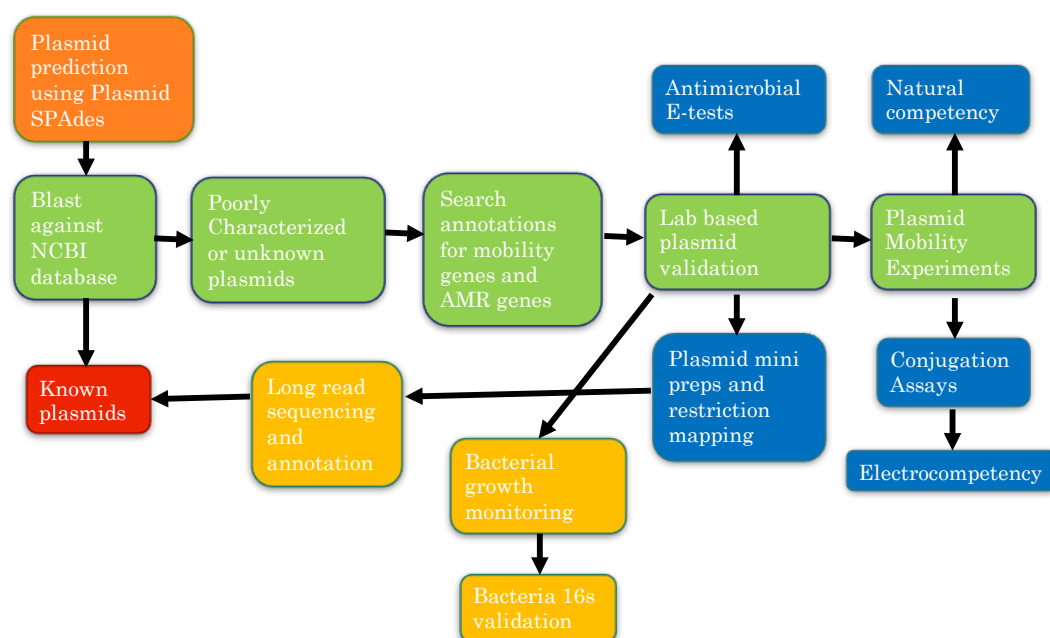
Figure 2.1. Workflow diagram summarising the strategy taken to probe the plasmids of the human gut microbiome.

2.1 Computational Plasmid Isolation, Phenotypic Predictions, and Phylogeny

PlasmidSPAdes, developed by Antipov *et al.* scans WGS data observing the coverage of the contigs (Antipov D. *et al,* 2016). Contigs are called as either chromosomal or plasmid contigs based on their deviation from the median coverage; plasmid contigs are then assembled to provide putative plasmid sequences. PlasmidSPAdes was run with the default parameters (*MaxDeviation* 0.3), as uniform coverage of genome was assumed.

The plasmid sequences were then BLAST (blastn) against the NCBI database using default parameters (*megablast*; *E threshold=10; word size= 28*) to isolate the well-characterised plasmids present in the culture collection.

The plasmid sequences were then annotated using the Sanger Pathogen Informatics Pipeline annotation tool. This tool searches the following databases to annotate sequences.

- RefSeq databases
- UniprotKB (bacteria/virus databases)
- Clusters
- Conserved domain database
- tigrfams
- pfam (A)
- rfam

Antibiotic Resistance Identification by Assembly (ARIBA) developed by Hunt *et al*, identifies resistances genes through alignment and targeted local assembly (Hunt M. *et al.*, 2017, bioRxiv). WGS were scanned for AMR genes using ARIBA with the Comprehensive Antibiotic Resistance Database (CARD- McArthur A.G. *et al,* 2013) and Short Read Sequencing Typing 2, Antimicrobial Resistance Gene Annotation (SRST2_ARGannot- Inouye M. *et al.,* 2014) databases. ARIBA was run using the default parameters as it is engineered to find the best match for a sequence. Putative plasmid sequences were scanned using BLAST against the CARD, SRST2_ARGannot, and ResFinder (Zankari E. *et al.*, 2012) databases at 60%, 70%, and 80% identity, the results were consistent across sensitivities and the results from the 70% identity screen are reported in Chapter 3.

Replicon typing was attempted using the PlasmidFinder Database (Carattoli *et al.,* 2014); plasmid sequences were BLAST (blastn) against the database at 70% identity. Plasmid sequences were also BLAST (blastx) against the database of sequences for *rep* domain PF01051, using an expected threshold of 10.

To build the phylogenies multiple sequence comparison by log-expectation (MUSCLE) was used to align the culture collection sequences to the RepA gene of an IncFII plasmid and to a RepE sequence of an unclassified plasmid, both from the culture collection. Muscle was run with default parameters (*maxiters 16*) to provide the most accurate alignment (Edgar R.C., 2004). FastTree was used to build the phylogenies using the generalised time-reversible model flag; this was selected as it is considered the most reliable general nucleotide evolutionary model (Price M.N. *et al,* 2010; Tavaré S. 1986).

2.2 Bacterial Culture and Growth Curve Plotting

Bacterial culture for aero-tolerant isolates was carried out under standard aerobic conditions (37°C, 5% CO_2). Aero-sensitive species were cultured under anaerobic conditions (37°C, Anaerobic gas

(ANO2) =10% H_2, 10% CO_2, 80% N_2). Bacteria were cultured in YCFA (Yeast extract-Casein hydrolysate-fatty acids) medium broth and plates (composition below).

To observe the growth curves bacteria were inoculated in 200µl of YCFA broth in a flat bottomed 96 well plate (Costar) and the optical density (OD) readings were recorded every 30 min over 24hr by the FLUOstar Omega Microplate Reader (BMG Labtech).

2.3 Growth Medium

YCFA medium (pH 7.45) consisted of (per 100ml):

Bacto Casitone	1.0g
Yeast extract	0.25g
NaHCO$_3$	0.4g
Glucose	0.2g
Maltose	0.2g
Cellobiose	0.2g
Mineral Solution I (3g K$_2$HPO$_4$, 1L d.H$_2$O)	15ml
Mineral Solution II (3g KH$_2$PO$_4$, 6g (NH$_4$)$_2$SO$_4$, 6g NaCl, 0.6g MgSO$_4$, 0.6g CaCl$_2$, 1L d.H$_2$O)	15ml
VFA mix (17ml Acetic acid, 6ml Propionic acid, 1ml n-Valeric acid, 1ml Isovaleric acid, 1ml Isobutyric acid, 26ml NaOH(10M))	0.62ml
Haemin Solution (0.28g KOH, 25ml Ethanol 95%, 100mg Haemin, ≤100ml)	1ml
Vitamin solution I (1mg Biotin, 1mg Cobalamin, 3mg PABA (P-amino benzoic acid), 5mg Folic acid, 15mg Pyridoxine, d.H2O ≤100ml)	100µl
Vitamin solution II (5mg Thiamin, 5mg Riboflavin, d.H$_2$O ≤100ml)	100µl (after autoclaving)
Resazurin (0.1%)	0.1ml
L-cysteine	0.1g
d.H$_2$O	≤100ml
Bacto Agar	1.6g

2.4 Species Validation by 16S rRNA gene PCR

The isolates were species validated using PCR of the 16S rRNA gene. PCR was carried out using the standard broad bacterial 16S rRNA gene primers (Sigma):

- Forward-Bact-7F (5'-AGA GTT TGA TYM TGG CTC AG-3')
- Reverse-Bact-1510R (5'-ACG GYT ACC TTG TTA CGA CTT-3')

and the GoTaq® Hot Start Polymerase kit (Promega). The reactions were completed in 25µl reaction volumes, and the master mix composition as follows: 1x Green GoTaq® Flexi Buffer, 25mM MgCl$_2$, 10mM dNTPs,10pM each forward and reverse primers, 1.25 U GoTaq® Polymerase, 2µl 1/10 dilution overnight culture, and sterile water up to 25µl. An *E. coli* sample and a sterile water sample were used as the positive and negative control respectively.

The PCR was run according to the protocol below on the DNA Engine Tetrad (MJ Research) with a constant lid temperature of 100ºC:

- 95ºC - 15mins
- 95ºC - 30secs \
- 58ºC - 30secs   → 30 cycles
- 72ºC - 2mins   /
- 72ºC - 8mins
- 10ºC - holding temperature.

The PCR product clean up and sequencing were provided by Eurofins Genomics. Sequence files were aligned to the 16S rRNA sequence database of the culture collection to identify the species matches.

2.5 Plasmid Extractions and Digests

Plasmid extraction was conducted using the QIAprep min prep kit (Qiagen) and the AquaPlasmid plasmid preparation buffer (MultiTarget), according to manufacturers' instructions. The Qiagen kit is a column based system and the upper size limit stated by the protocol is 12kb; the AquaPlasmid buffer is a column free application and suitable for plasmids of all sizes including megaplasmids. Extractions on a plasmid containing *E. coli* strain were done in parallel as a technical control.

Restriction sites were predicted using SnapGene (GSL Biotech), single cutters were preferred. Plasmid digests were set up in 50µl reaction volumes with 1U of enzyme (NEB), 1x buffer (NEB), and 500ng of DNA for 2 hours.

2.6 Plasmid Visualisation by Gel Electrophoresis

After isolation and digestion, plasmids were visualised by gel electrophoresis on 0.8% agarose in Tris-acetate EDTA (TAE). Gels were run at 75V for 1 hour; stained in TAE with 0.5 mg/L ethidium bromide for 30 mins; and visualised using the GelDoc-It^2® system (UVP).

2.7 Antimicrobial Sensitivity Testing

Antimicrobial sensitivity levels were tested using Etest strips (Biomérieux) on YCFA plates according to manufacturer's instructions.

2.8 Transformation

Natural competency can be induced by stimuli including DNA damage and starvation; starvation was used to attempt to induce natural competency in commensals (Huddleston J. R., 2014). Overnight bacterial cultures were filtered to collect nutrient deficient or spent media. This spent media was then mixed with fresh YCFA media in a series of ratios decreasing tenfold from 100% to 10% spent media. Bacteria were incubated in these mixes overnight with 0.5, 1.0, or 2.0µg of plasmid DNA. The cultures were plated on selection YCFA plates and growth observed.

Bacterial cells were harvested after 2hr, 6hr, and overnight incubations for the creation of competent cells. Electrocompetent cells were created by washing cells in ice cold 10% glycerol as described by NEB (NEB protocols, [*Accessed Apr 2017*]). Cells were resuspended in either 10% glycerol or SMP buffer (sucrose-magnesium-phosphate: 270 mM sucrose, 1 mM MgCl2, and 5 mM sodium phosphate, pH 6.5), then aliquoted and stored at -80ºC. Protoplasts were created as described by Rattanachaikunsopon and Phumkhachorn (2009). Cells were incubated in digestion buffer (50 mg/L mutanolysin, 0.5 M sorbitol, 0.01 M Tris-hydrochloride, pH 7.0) at 37 ºC for 1hr, washed and

resuspended in transformation buffer (0.5 M sorbitol, 0.02 M maleate, 0.02 M MgCl2, pH 6.5), then aliquoted and stored at -80ºC (Rattanachaikunsopon P. and Phumkhachorn P., 2009)

Electroporation was carried out in 2mm gap cuvettes (Flowgen Bioscience), using a MicroPulser Electroporation System (Bio-Rad). Aliquots of competent cells were thawed on ice, and incubated with 0.5, 1.0, or 2.0µg of plasmid DNA on ice for 20 min. Cells were transferred to a chilled cuvette and pulsed at 25µF and 1.2, 1.5, or 2.5 kV. Glass bead transformation was carried out using acid washed glass beads (Sigma-150-212µm) as described by Rattanachaikunsopon and Phumkhachorn. Protoplasts were thawed and added to the transformation mix (15% w/v Polyethylene glycol (PEG) 6000, 0.3% w/v acid washed glass beads); and the mix was agitated by vortex at the highest speed for 15 seconds (Rattanachaikunsopon P. and Phumkhachorn P., 2009). 1ml of 1% sucrose YCFA was added to the tube immediately after agitation, samples recovered for 3 hours at 37ºC anaerobically, and were plated on selection and control YCFA plates, as detailed below.

	YCFA	YCFA +Tetracyline
Electroporated +DNA	Positive	Positive
Electroporated	Positive	Negative
Thawed Competent Cells	Positive	Negative
Agitated +DNA	Positive	Positive
Agitated	Positive	Negative
Digested	Positive	Negative

2.9 Conjugation

Donor and recipient strains were cultured in YCFA broth anaerobically overnight. Culture OD readings were taken using the WPA Biowave (Biochrom) and the strains mixed in a 4:1 ratio, by volume, of donor to recipient. The culture mixes were immobilised onto Mixed Cellulose Ester Membranes (Millipore- 0.45µm pore size) using sterile syringes (BD Plastipak) and filter holders (Millipore). The membrane was placed onto YCFA plates and incubated anaerobically overnight. The membranes were then washed in 1ml of PBS and the bacterial suspension was plated onto selection and control YCFA plates, as detailed below. Conjugations using an E. coli mating pair were done in parallel as a technical control

	Nalidixic Acid	Tetracycline	Nalidixic Acid + Teracycline
Donor (D)	Negative	Positive	Negative
Recipient (R)	Positive	Negative	Negative
D+R Mix	Positive	Positive	Positive

	Erythromycin	Tetracycline	Erythromycin + Teracycline
Donor (D)	Negative	Positive	Negative
Recipient (R)	Positive	Negative	Negative
D+R Mix	Positive	Positive	Positive

	YCFA Anaerobic	YCFA Aerobic	YCFA+Tetracycline Anaerobic	YCFA+Tetracycline Aerobic
Donor (D)	Positive	Negative	Positive	Negative
Recipient (R)	Positive	Positive	Negative	Negative
D+R Mix	Positive	Positive	Positive	Positive

3. Bioinformatics Results

In this chapter I will present the results from the *in silico* screen for plasmids in the culture collection, the phylogeny and host range of the putative plasmids, and the predicted distribution of plasmid antimicrobial genes. Plasmids serve as a way for bacteria to easily facilitate the sharing of advantageous auxiliary functions between species and even between phyla. Plasmids with mobility genes are a good indication of instance where this sharing is being actively pursued. One example of genes that may be actively shared are the AMR genes. The increased use of antimicrobials from active courses to treat illness, to passive use in deodorants and hand sanitisers has increased the levels of exposure of pathogens to antimicrobials in particular members of the human flora. The gut in particular is proposed as a reservoir of AMR, capable of transmitting these genes to pathogens. Observing the predicted prevalence of AMR containing mobile elements, would allow for the investigation of the scope of this hypothesis, and potentially provide evidence to support it.

3.1 Frequency of Plasmid Detection Across Human Gut Microbiota Phyla

Bacterial whole genome sequences from the Lawley Lab Culture Collection were screened *in silico* using plasmidSPAdes (Antipov D. *et al*., 2016) to identify genomes containing predicted plasmid sequences. PlasmidSPAdes isolates and assembles plasmid contigs from whole genome sequence data based on contig coverage as plasmid coverage often differs from the median chromosomal coverage. This data is presented in Figure 1 as the number of genomes predicted to contain zero, one, two, or more plasmids in each phylum in the culture collection. 653 genomes were scanned; 60% are not predicted to contain any plasmids, and of the 240 genomes predicted to contain plasmid sequences, 70% are predicted to only contain one plasmid. Plasmids are predicted across all phyla in the culture collection phylogeny, with phylum Firmicutes containing the largest number of plasmids (151) and Bacteroidetes the most enriched phylum with a chi squared value of $\chi2=0.04$.
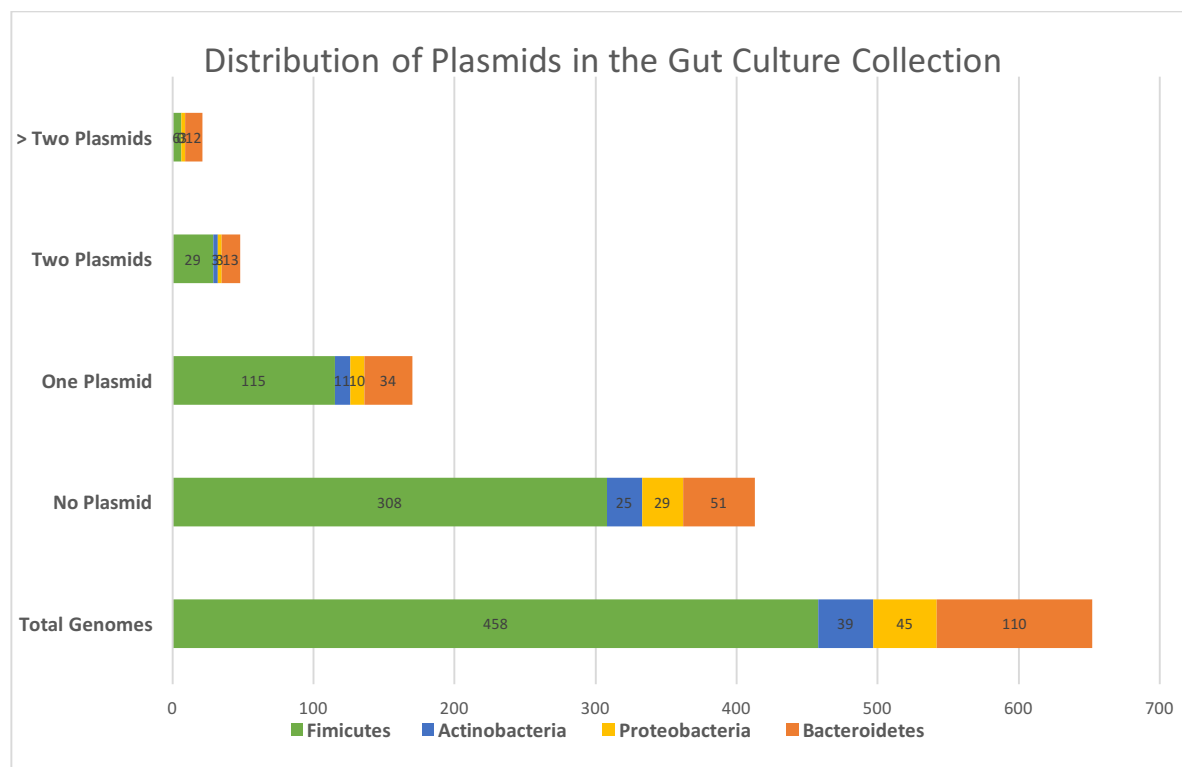
Figure 3.1. *The distribution of predicted plasmids in the 653 genomes of the Lawley Laboratory culture collection*. 40% of the genomes are predicted to harbour plasmid sequences; 70% of those genomes are only predicted to contain one plasmid.

The predictions from plasmidSPAdes were used to annotate the culture collection phylogeny, as displayed in Figure 2. Putative plasmids are widely distributed across the phyla in the culture collection, suggesting that there are important functions present on these elements facilitating survival in the gut environment.
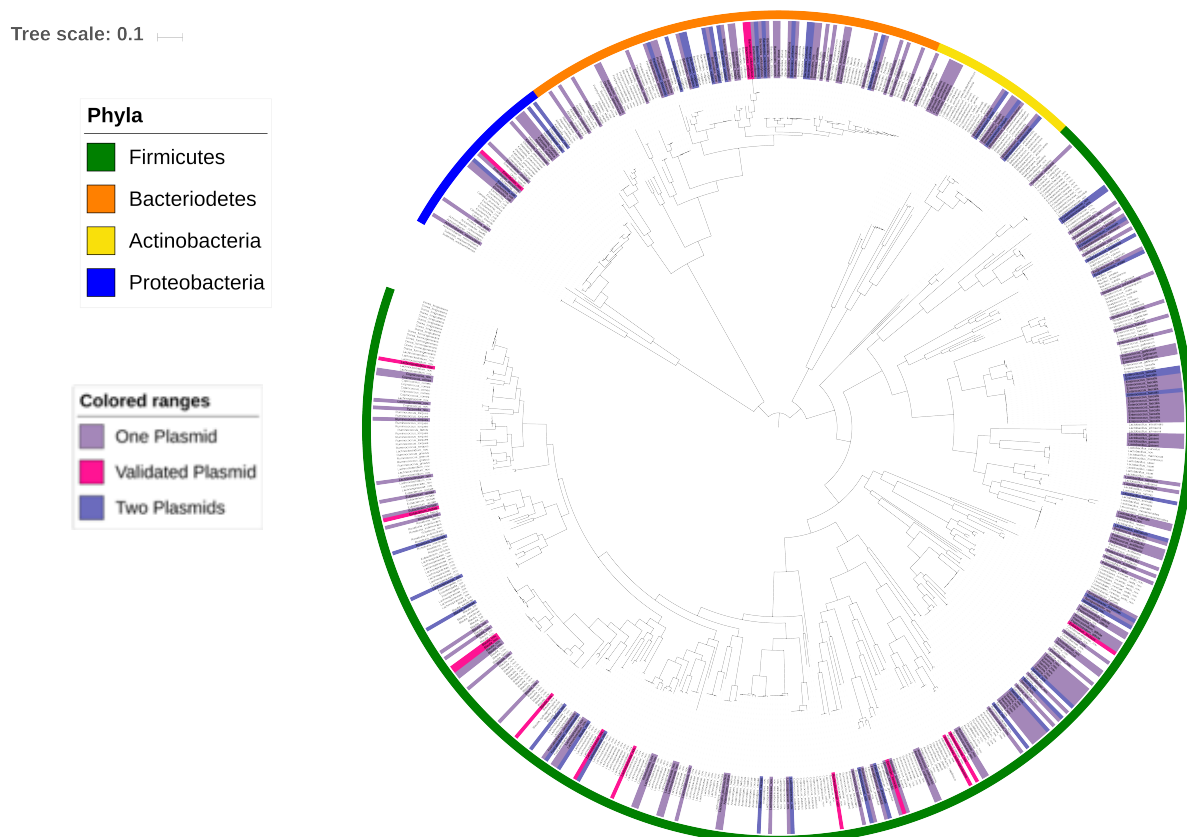
Figure 3.2. *Lawley Lab culture collection phylogeny with plasmid prediction annotations.* 653 bacterial whole genome sequences were scanned for the presence of plasmids. 240 genomes are predicted to contain plasmids: genomes predicted to contain one plasmid are annotated in pale purple, and genomes predicted to contain two plasmids are annotated in deep purple. Genomes where DNA has been experimentally isolated by mini prep and visualised on agarose are annotated in pink.

3.2 Size and Coverage and Distribution of Predicted Plasmids

In an attempt to understand the type of extrachromosomal elements in the culture collection, the plasmid size and coverage information from the plasmidSPAdes assembly graphs were used to plot the distribution of plasmids according to these traits; this data is presented in Figure 3. Plasmids size distribution is predicted to follow a bimodal curve, and plasmid coverage is often higher than that of the chromosome. The predicted plasmids of the culture collection are predominantly small, with 70% falling in the 1-10kb range. The majority of the predicted plasmids are high coverage elements with coverage between two and tenfold higher than the median. This could be an artefact of the prediction methods, or may give insight into the pressures in the gut environment shaping the distribution of plasmids by affecting factors such as carriage costs and the frequency of HGT.

29
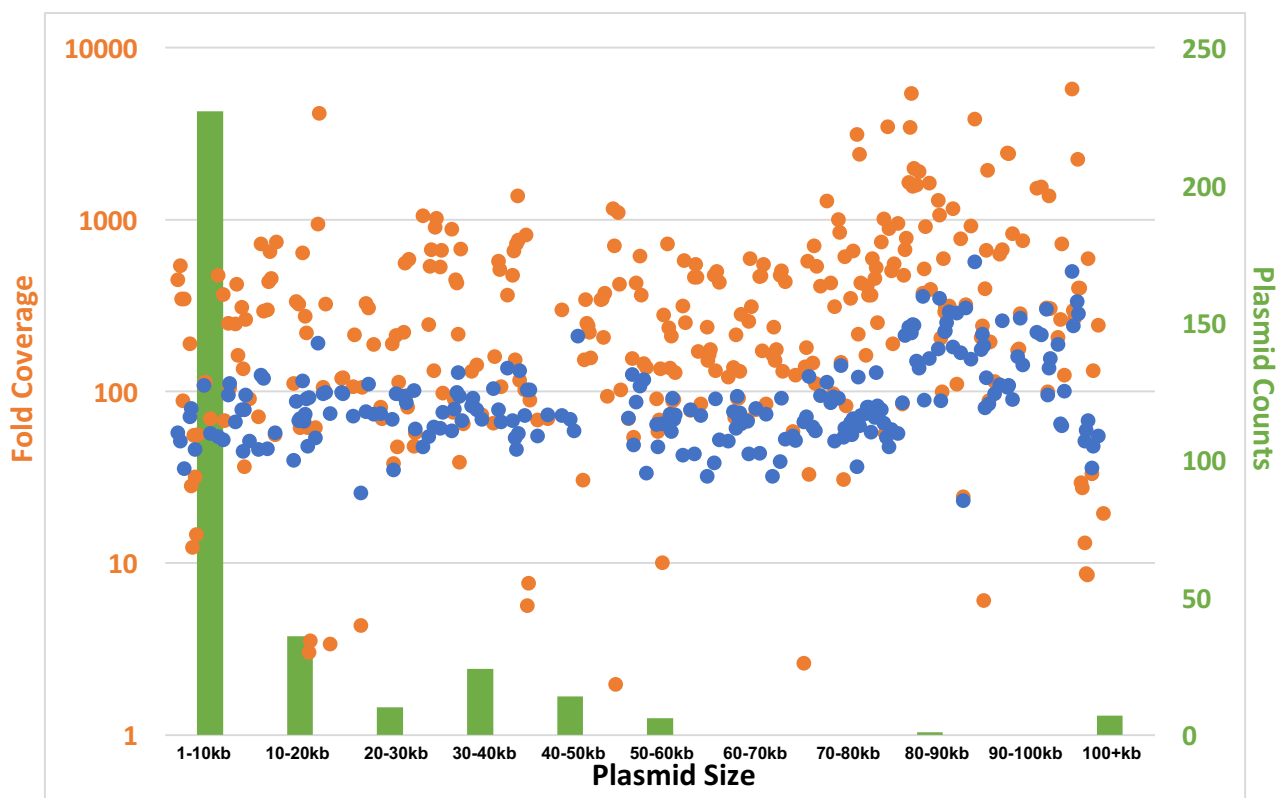
Figure 3.3. *Detected plasmid coverage compared to chromosomal median coverage and plasmid distribution by size*. Triaxial graph displaying predicted plasmid distribution by size and coverage. The green dataset plotted on the right-hand axis displays the plasmid counts for each size range. The blue and orange datasets plotted on the left-hand side show the median chromosomal coverage and plasmid coverage on a logarithmic scale to display the fold difference between them. The majority of plasmids are small, 1-10kb, and high coverage, 2-10 fold above median.

3.3 Plasmid Classification and Phylogeny

As discussed in Chapter 1 plasmids are classified either by function or by the inc grouping system, this can be done *in silico* through replicon typing. Replicon typing uses a computational alignment to associate plasmids to one of the known incompatibility groups based on the sequence of the plasmid's *rep* gene. Attempts to assign a replicon type to the predicted plasmid sequences using PlasmidFinder (Carattoli A. *et al.,* 2014) were unsuccessful, yielding only 19 hits against the 240 genomes predicted to contain plasmids. An alternative approach involved aligning the plasmid sequences with BLASTx against all the rep sequences on the pfam database for domain PF01051. This domain is common to both RepB and RepA and was the Rep protein domain that was the most frequently identified during plasmid annotation. All the sequences did yield hits to *rep* genes; however, the top hits were to uncharacterised plasmids. Therefore, the plasmids were not able to be provisionally typed by association.

Subsequently, annotated *rep* genes were used to infer a phylogeny of the culture collection genomes. Two trees were built, one using *repE* (Figure 4) as the most abundant rep gene in the culture collection, and the other using *repA* (Figure 5), which is more commonly considered a conserved or

standard rep gene. This would allow me to observe clusters of related plasmids in the absence of replicon typing. The *repE* gene was most frequently annotated in the culture collection. An alignment of the culture collection genomes to the *repE* sequence from one of the putative plasmids was used to build the phylogeny in Figure 4. This phylogeny shows clusters of related *rep* genes across the different phyla.
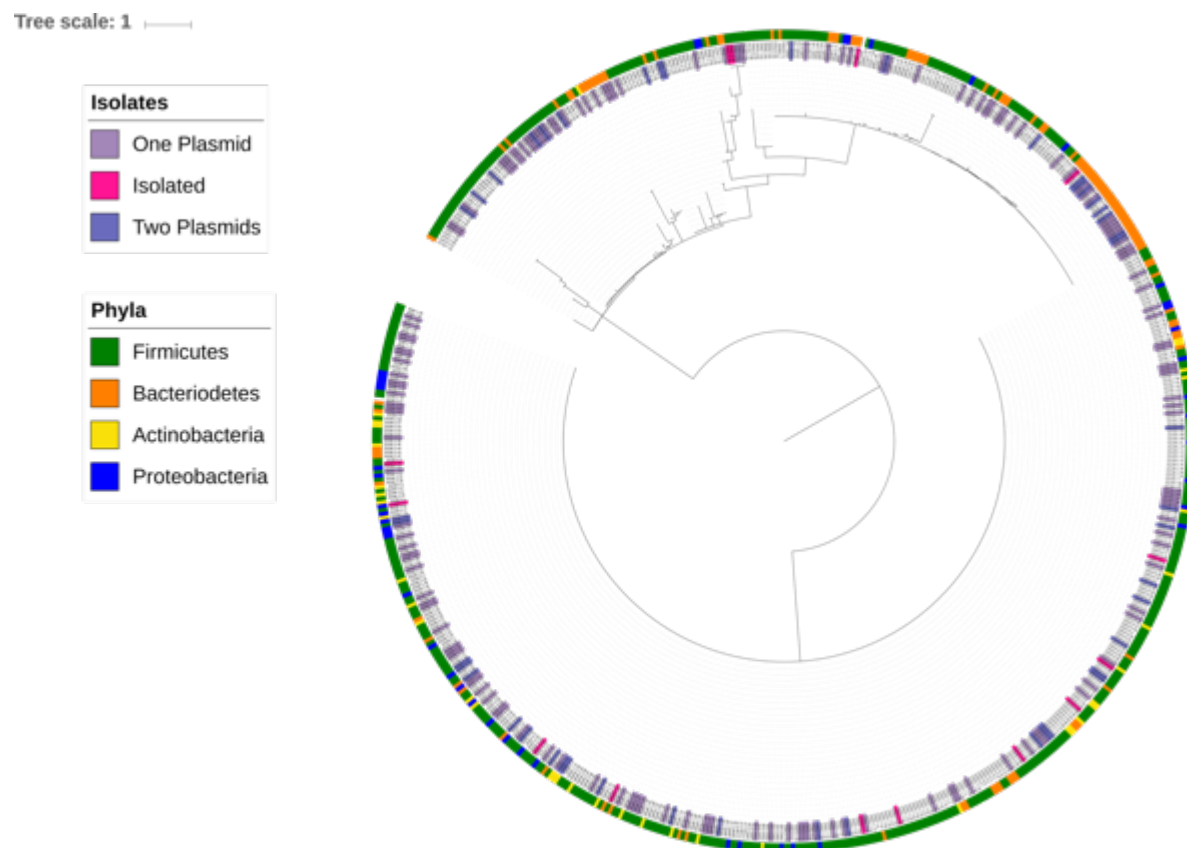
Figure 3.4. *Phylogenetic tree constructed using the repE gene*. A phylogeny of the 653 genomes in the culture collection was built from an alignment of the *repE* gene. Plasmids from all 4 phyla are clustered across the top third of the tree, displaying relatedness between the *rep* genes in those genomes. The shallow branch along the bottom third of the tree, may be due to an absence of the *repE* gene in these samples, of the presence of a very pervasive gene. Genomes predicted to contain one plasmid are annotated in pale purple, and genomes predicted to contain two plasmids are annotated in deep purple. Genomes where DNA has been experimentally isolated by mini prep and visualised on agarose are annotated in pink.

The *repA* gene is significantly less annotated and possibly less represented in the culture collection, however it is more commonly recognised as a conserved or core plasmid gene. The *repA* gene sequences were aligned across the culture collection and the resulting phylogeny presented in Figure 5; once again demonstrating clustering of the *rep* sequences across phyla.
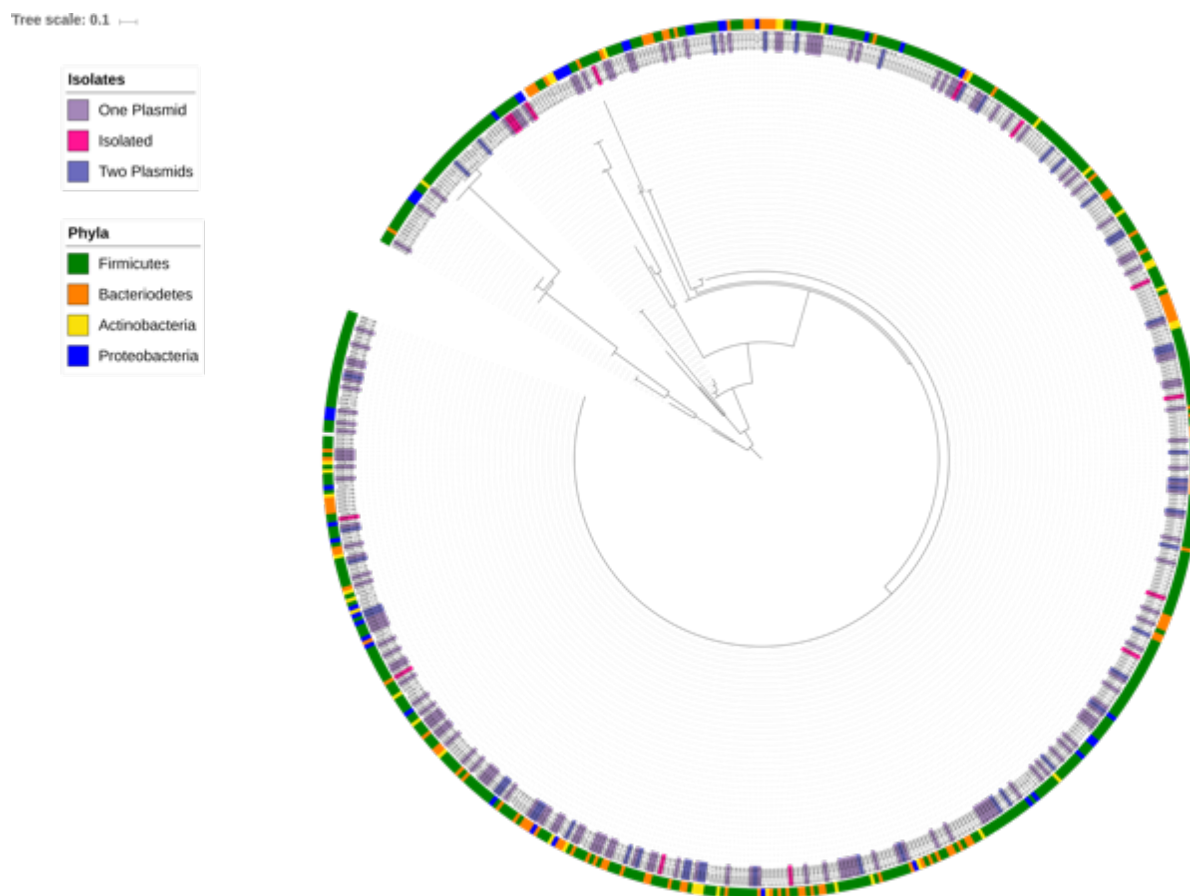
Figure 3.5. *Phylogenetic tree constructed using the repA gene.* A phylogeny of the 653 genomes in the culture collection was built from an alignment of the *repA* gene. Plasmids from all 4 phyla are clustered in a small segment of the tree, displaying relatedness between the *rep* genes in those genomes. The shallow branch is broader in this tree; and as the *repA* gene is not as abundantly annotated in the culture collection this may be due to an absence of the *repA* gene in these samples, or once again the presence of a very pervasive gene. Genomes predicted to contain one plasmid are annotated in pale purple, and genomes predicted to contain two plasmids are annotated in deep purple. Genomes where DNA has been experimentally isolated by mini prep and visualised on agarose are annotated in pink.

The clustering of the nodes along the top third of the tree in Figure 4 does not match the phyla-based clustering of the species tree; this incongruence suggests the sharing of plasmids across broad phylogenetic distances. These interactions were annotated on the culture collection tree to visualise the connections between genomes, producing the chord diagram in Figure 6. The chords (pink) show links between Firmicutes and Bacteroidetes, Firmicutes and Proteobacteria, and Bacteroidetes and Actinobacteria, indicating the possible presence of broad host range plasmids.
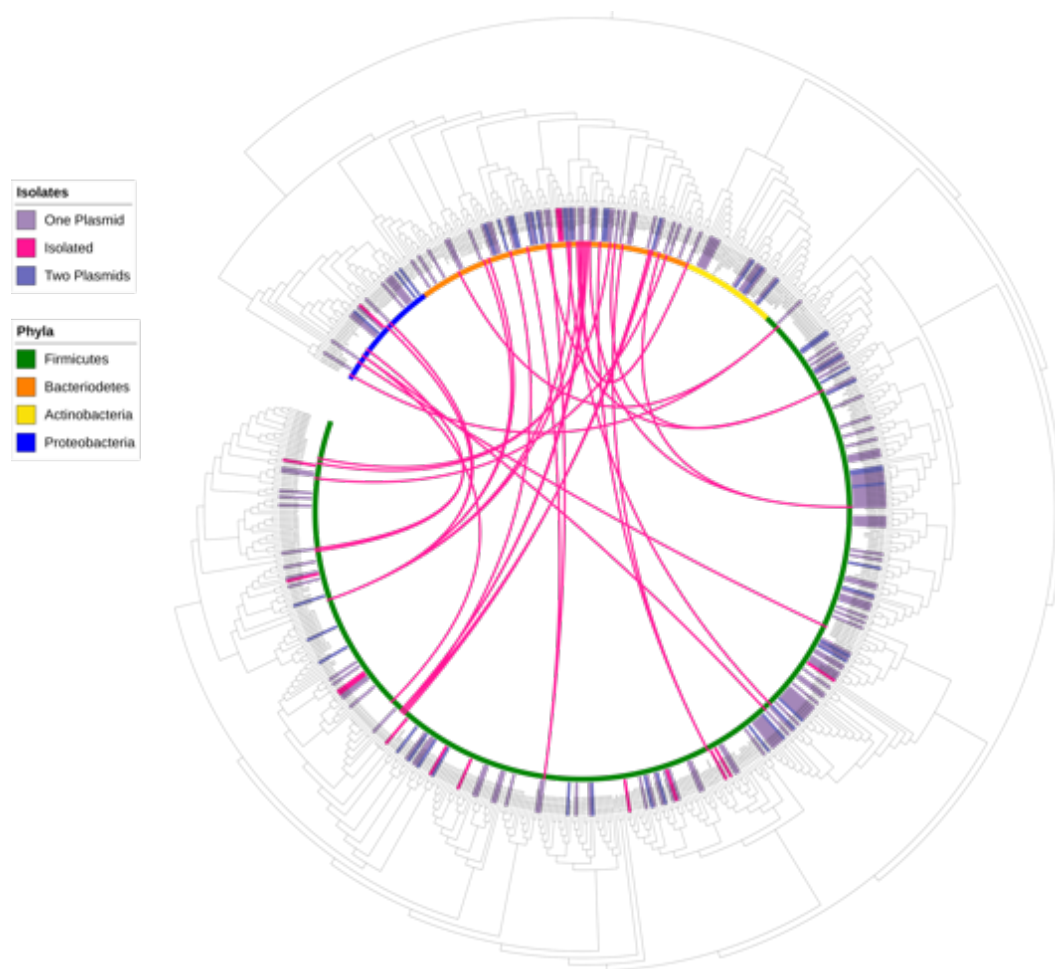
Figure 3.6. *Chord diagram displaying links between genomes containing closely related repE genes.* The chords (pink) display links between isolates with closely related *repE* genes, indicating potential instances of plasmid sharing between isolates of differing phyla and the presence of broad host range plasmids. Genomes predicted to contain one plasmid are annotated in pale purple, and genomes predicted to contain two plasmids are annotated in deep purple. Genomes where DNA has been experimentally isolated by mini prep and visualised on agarose are annotated in pink.

As with the *repE* phylogeny the clustering of the nodes in the *repA* tree does not match the phyla-based clustering of the species tree. These interactions were used to create a second chord diagram on the culture collection tree (Figure 7). In this instance, the chords show interactions between Firmicutes and Actinobacteria, Firmicutes and Proteobacteria, and Proteobacteria and Actinobacteria. A *Klebsiella* node is shown to have several connections to several *Blautia* species, including a species containing a predicted small resistance element, and one containing a mid-size conjugation plasmid. Both of these genetic elements were experimentally isolated, and these strains may be good candidates for demonstrating horizontal gene transfer *in vitro.*
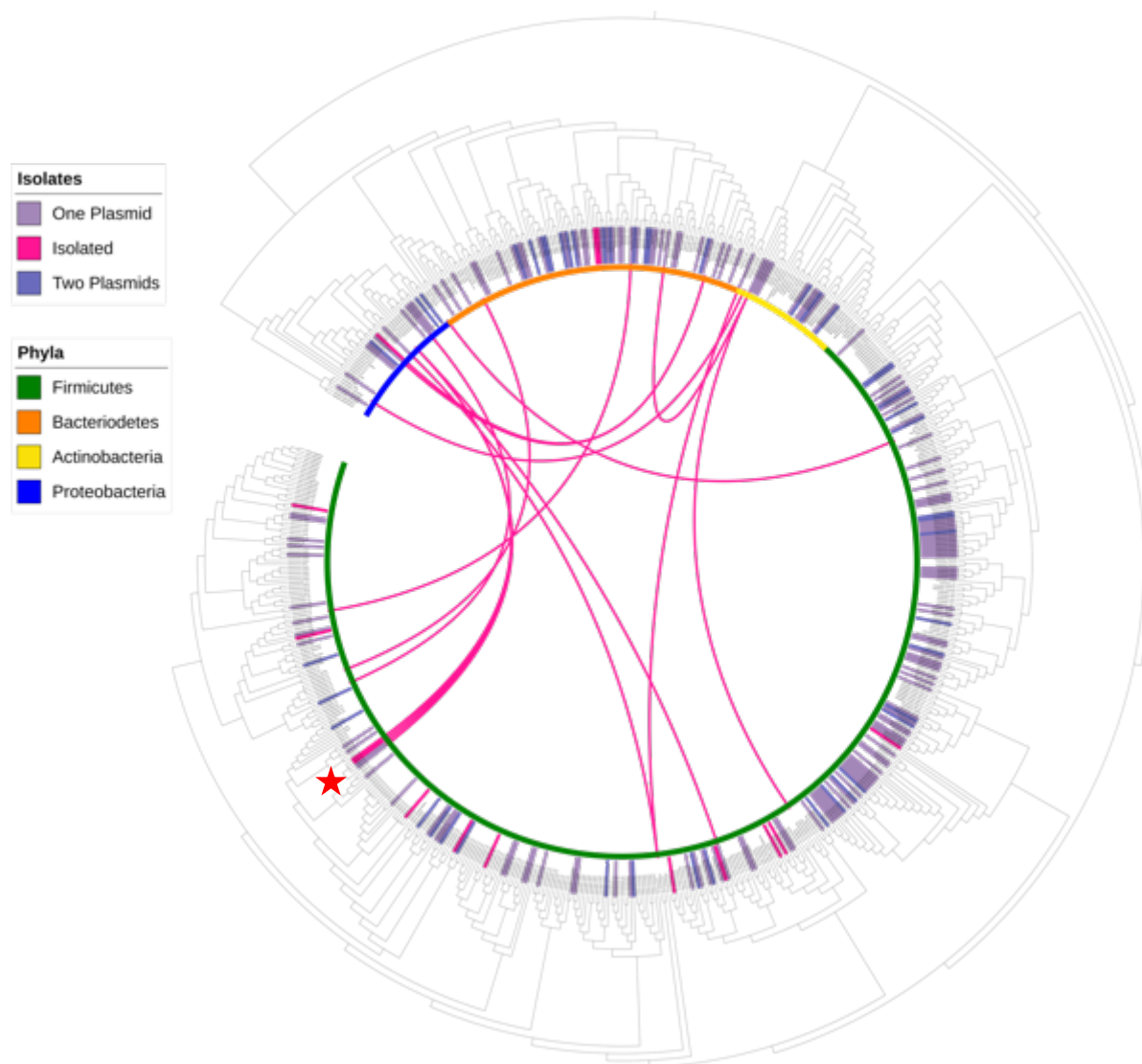
Figure 3.7. *Chord diagram displaying links between genomes containing closely related repA genes*. The chords (pink) indicate potential instances of plasmid sharing between isolates of differing phyla and the presence of broad host range plasmids. The *Blautia* species marked by a star are connected to a strain of *Klebsiella pneumonia*. The plasmids contained in the *Blautia* include a small resistance element, and a mid-range element with conjugation sequences, both of which have been visualised experimentally. Genomes predicted to contain one plasmid are annotated in pale purple, and genomes predicted to contain two plasmids are annotated in deep purple. Genomes where DNA has been experimentally isolated by mini prep and visualised on agarose are annotated in pink.

3.4 AMR Gene Distribution

The other method for classifying plasmids is based on function, and a function that readily leads itself to validation is AMR. In addition, once validated, the AMR phenotypes can be used as selectable markers when testing plasmid mobility. Examining the abundance of AMR on gut bacterial plasmids also begins to address the hypothesis of the gut microbiome as an AMR reservoir and location for HGT. ARIBA (Antibiotic Resistance Identification by Assembly, Hunt M. *et al*, 2017, bioRxiv) was used to predict the prevalence of AMR genes in the culture collection. AMR genes were predicted in the culture collection genomes from raw sequence reads by implementing ARIBA with the CARD and SRST2_ARGannot databases. AMR genes were also identified within the predicted plasmid assemblies using BLAST with the CARD, SRST2_ARGannot (SA), and ResFinder (RF) databases. (Zankari E. *et al.*, 2012; McArthur A.G. *et al,* 2013; Inouye M. *et al.,* 2014)

The incidence of AMR gene prediction was then compared between plasmid containing genomes and non-plasmid containing genomes, displayed as percentages in Figure 8. The result demonstrated almost equal populations of resistant and non-resistant bacteria: 53% resistant to 47% non-resistant with CARD, and 51% resistant to 49% non-resistant with SA. Scanning the genomes predicted to contain plasmids with the same databases demonstrated an increase in the proportion of reported AMR genes in the genomes predicated to contain plasmids 60% resistant to 40% non-resistant with CARD, and 55% resistant to 45% non-resistant with SA. The increase of AMR predictions in plasmid containing genomes compared to all genomes is significant when using the CARD predictions but not the SA predictions, with chi squared values of χ^2=0.035 and χ^2=0.36 respectively.

The incidence of AMR gene prediction in the culture collection was then compared to the prediction of AMR genes on the putative plasmid sequences. The plasmid assemblies were scanned using BLAST (blastn), against three databases at a similarity level of 70%. AMR genes were more frequently predicted to be on putative plasmid sequences than in the culture collection with 68% resistant to 32% non-resistant. The increase in predicted AMR genes on the putative plasmids sequences is significant when compared with the prediction on the whole genome sequences of the culture collection with both databases- chi squared values of χ^2=0.002(CARD) and χ^2=0.009(SA).

The correlation between frequency of mobility genes and AMR genes was also investigated. Plasmid annotations were searched for *mob* and *tra* genes and 90 of the 240 genomes predicted to contain plasmids also contain plasmid mobility elements. The proportion of AMR reported in this set of plasmids was compared to the amount of AMR reported in the culture collection. There is an increase in predicted AMR on the putative plasmid sequences with mobile elements with 75% resistant and 25% non-resistant. This result is significant when compared to the predictions made using the SA database χ^2=0.023, and falls just short of significance with the CARD database χ^2=0.062. These AMR predictions were validated on the species with experimentally isolated plasmids, and these results are discussed in the next chapter.
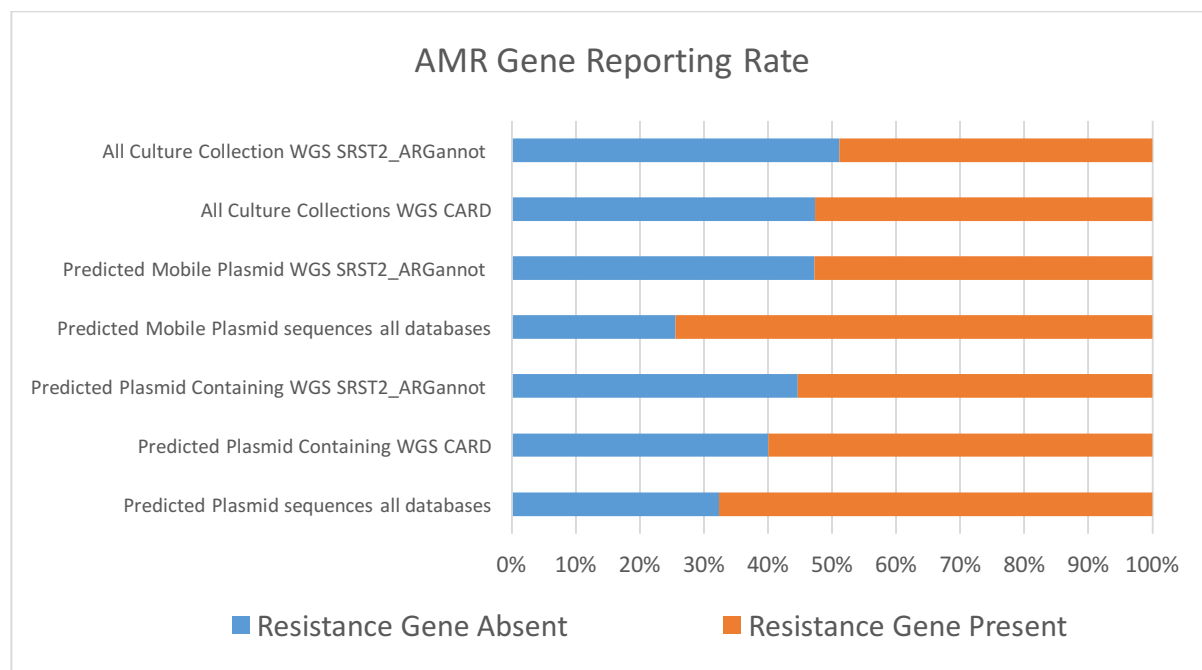
Figure 3.8. Percentage bar graph displaying the proportions of predicted resistance in the whole culture collection, predicted plasmid containing species, and on predicted plasmid sequences. Predictions were completed using the CARD and SRST2_ARGannot databases for the whole genome predictions, and using the CARD, SRST2_ARGannot, and ResFinder databases on the predicted plasmid sequences. There is a significant increase in the reporting rate of AMR genes in plasmid containing species compared to the rate in the culture collection when using the CARD database but not with the SRST2_ARGannot database. There is a significant increase in the reporting rate on the predicted plasmid sequences when compared to the reporting rate in the whole culture collection. When comparing the reporting rate from genomes with potential mobile elements, there is only a significant increase in the reporting rate from the predicted sequences themselves and not the whole genome sequence.

4. Experimental Results

In this chapter I will present the results from the experimental isolation of plasmids, the validation of the AMR predictions, as well as the transformation and conjugation attempts. The *in silico* predictions in the previous chapter revealed putative plasmids in one third of the culture collection, and among these predictions 38% are predicted contain mobility sequences, and 68% are predicted to contain AMR genes. Of particular interest are the elements predicted to exist in several phyla across the tree, hinting at the presence of broad host range mobilisable plasmids. Resistance and mobility phenotypes were tested experimentally and the results presented below.

4.1 Plasmid Isolation

Strains with predicted plasmids were cultured for plasmid isolation; strains of particular interest included Firmicutes strains with putative plasmids also predicted to contain plasmid-mobility sequences. Plasmid isolation was completed using two commercial kits; one for plasmids below 10kb supplied by Qiagen, one for plasmids greater than 10kb supplied by MultiTarget. Isolated plasmids were visualised on 0.8 % agarose TAE, to validate the presence of extracted extrachromosomal DNA (Figure 1).
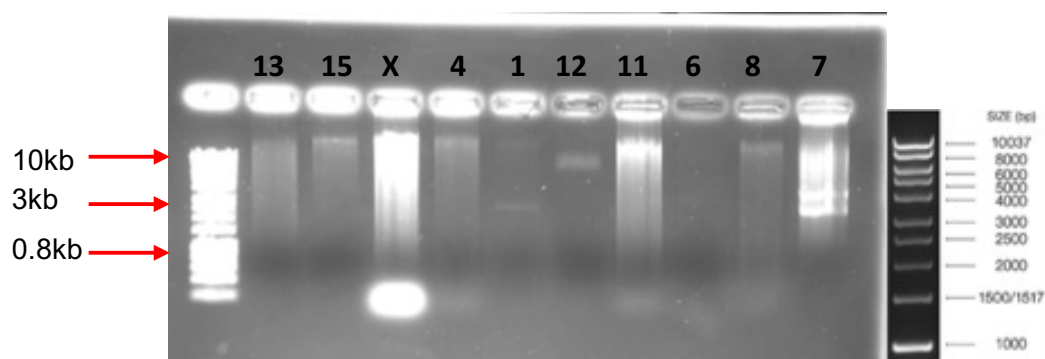
Figure 4.1. *Agarose gel displaying the isolated plasmids greater than 10kb.* DNA gel displaying undigested isolated DNA from plasmids greater than 10kb using MultiTarget's AquaPlasmid, plasmid preparation buffer. Ladder in lane 1 is the 10kb Hyperladder (Bioline) Sample identity is indicated by Table 1 row number, and predicted sizes are as follows:

- **13-** 102023
- **15-** 45070
- **X-** 16S validation failed
- **4-** 6131
- **1-** 12972 & 5302
- **12-** 133666
- **11-** 35136
- **6-** 60425
- **8-** 80318
- **7-** 12278

4.2 Species Validation

To be confident that the bacterial cultures which plasmids were being isolated from were consistent with the species identified in the predictions the identity of each isolate from which plasmid DNA had

been extracted was validated by 16S rRNA gene amplicon sequencing. The 16S rRNA gene sequences were aligned to the database of 16S rRNA gene sequences of the whole culture collection (Table 1). Only one isolate failed the screening and was identified as a different species to its expected identity, this was excluded from further screening.

	Results of the 16S Species Validation					
	Seq ID	**Species Name**	**Isolate Name**	**Matched Seq ID**	**Matched Species Name**	**Validation**
1	18048_2#80	*Erysipelotrichaceae nov.*	fCSP	18048_1#80.contigs_velvet.cleaned.fna|16S	*Erysipelotrichaceae nov. 18048_1#80*	Y
2	20287_6#28	*Blautia nov.*	E79_57	20287_6#62.contigs_velvet.cleaned.fna|16S	*Blautia nov. 20287_6#62*	Y
3	12718_7#90	*Bacteroides uniformis*	H1_6	21673_4#70.contigs_velvet.cleaned.fna|16S	*Bacteroides uniformis 21673_4#70*	Y
4	14207_7#59	*Intestinimonas butyriciproducens*	H5_60	14207_7#59.contigs_velvet.cleaned.fna|16S	*Intestinimonas butyriciproducens 14207_7#59*	Y
5	20287_6#22	*Ruminococcaceae nov.*	E72_40	20287_6#22.contigs_velvet.cleaned.fna|16S	*Ruminococcaceae nov. 20287_6#22*	Y
6	14207_7#7	*Eubacterium rectale*	H4_46	14207_7#7.contigs_velvet.cleaned.fna|16S	*Eubacterium rectale 14207_7#7*	Y
7	20298_2#57	*Bacteroides cellulosilyticus*	D85_115	GCF_000158035.cleaned.fna|16S	*Bacteroides cellulosilyticus GCF_000158035 (DSM 14838)*	Y
8	20427_4#28	*Blautia nov.*	F41_240	20298_3#81.contigs_velvet.cleaned.fna|16S	*Blautia nov. 20298_3#81*	Y
9	13470_2#56	*Lachnospiraceae nov.*	H2_21	13470_2#56.contigs_velvet.cleaned.fna|16S	*Lachnospiraceae nov. 13470_2#56*	Y
10	20287_6#48	*Blautia nov.*	F1_100	20287_6#62.contigs_velvet.cleaned.fna|16S	*Blautia nov. 20287_6#62*	Y
11	13470_2#93	*Lachnospiraceae nov.*	H4 _41	13470_2#93.contigs_velvet.cleaned.fna|16S	*Lachnospiraceae nov. 13470_2#93*	Y
12	20298_3#19	*Enterobacter nov.*	G8_180	20298_3#19.contigs_velvet.cleaned.fna|16S	*Enterobacter nov. 20298_3#19*	Y
13	13414_6#1	*Lachnoclostridium nov.*	H1_38	13414_6#1.contigs_velvet.cleaned.fna|16S	*Lachnoclostridium nov. 13414_6#1*	Y
14	20298_3#39	*Eubacterium limosum*	G14_207	20298_3#39.contigs_velvet.cleaned.fna|16S	*Eubacterium limosum 20298_3#39*	Y
15	13414_6#62	*Ruminococcaceae nov.*	H3_43	13414_6#62.contigs_velvet.cleaned.fna|16S	*Ruminococcaceae nov. 13414_6#62*	Y

Table 4.1. *Results of the 16S species validation.* Columns one and two contain the expected ID number and species name of each isolate. Columns four and five contain the ID number and species name the isolate has been identified as from 16S sequencing. The last column indicates if this is a correct (Y) match. Only one species failed and came back as a different species.

4.3 Plasmid Digests

To further validate that the extracted plasmids were consistent with the predictions the plasmids were linearised to measure their true size using separation by gel electrophoresis and this was compared to the size predicted for each plasmid. Plasmid restriction sites were predicted using SnapGene by GSL Biotech. The plasmids were linearized according to predicted restriction sites; enzymes predicted to have only one restriction site were used. The isolated plasmids under 10kb were all size verified through restriction digest (Figure 2), the larger plasmids experienced degradation upon digest (Figure 3).
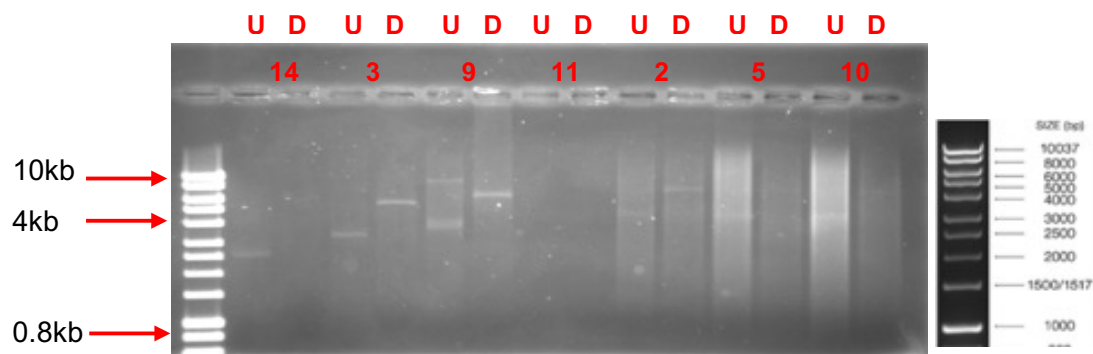
Figure 4.2. *Agarose gel displaying small plasmids in native (U) and digested (D) form.* Plasmids under 10kb were isolated using the QIAprep mini prep kit. The restriction sites were predicted using the SnapGene, the sequences were digested using restriction enzymes, and the linearised DNA run alongside the native DNA to observe the size of the sequences. Ladder in lane 1 is the 10kb Hyperladder (Bioline). Sample identity is indicated by Table 1 row number, and predicted sizes are as follows:

- **14-** 5498
- **3-** 5629
- **9-** 6018
- **11-** 6131
- **2-** 7296
- **5-** 7170
- **10-** 7170
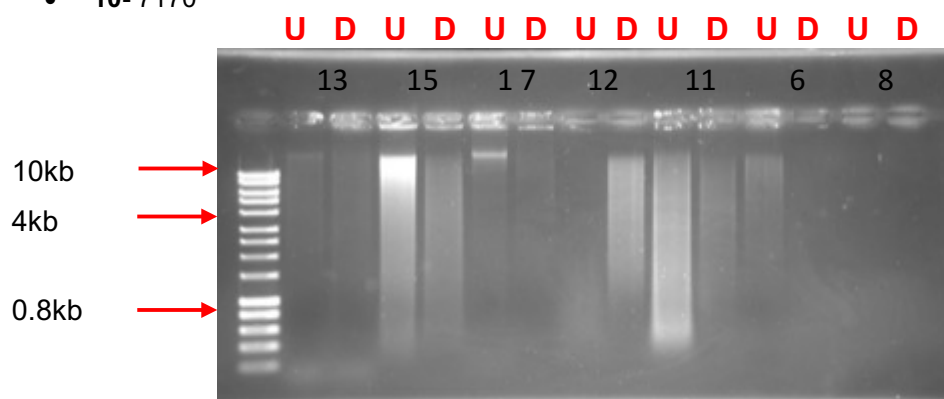
Figure 4.3. *Agarose gel displaying mid size and large plasmids in native (U) and digested (D) form.* Plasmids greater than 10kb were isolated using the AquaPlasmid buffer. The restriction sites were predicted using SnapGene, the sequences were incubated with the identified restriction enzymes, however the digested DNA appears degraded on the gel so the sizes have not been confirmed by restriction mapping. Ladder in lane 1 is the 10kb Hyperladder (Bioline), and sample identity is indicated by Table 1 row number

4.4 Antimicrobial Screening

The antimicrobial resistance predictions from screening with ARIBA (as described in chapter 3) were tested with Biomérieux Etest strips to validate phenotypic antibiotic resistance. The most commonly predicted resistance was tetracycline (tet) resistance, predicted in 45% of the putative plasmid sequences. Erythromycin (erm) sensitivity was also tested to identify selectable markers that could be used in the conjugation system. The European Committee on Antimicrobial Susceptibility Testing (EUCAST) epidemiological cut-off values (ECOFFS) were used as guidelines to classify strains as either resistant or sensitive. Suggested thresholds are 2-4mg/L for tet, and 4-8 mg/L for erm; the

levels of resistance are recorded in Table 2 in mg/L. The validations were used to select donors with tet resistance and erm sensitivity, the recipients were selected for the opposing profile. The majority of the predictions were observed to be accurate with 7 out of the 44 being incorrect.

Antibiotic Resistance Levels as Observed with Etest Strips							
			Tet (mg/L)		Erm (mg/L)		Validated
Seq ID	Isolate Name	Species Name	Predicted	Observed	Predicted	Observed	
12718_7#90	H1_6	*Bacteroides uniformis*	x	0.75	x	8	N
13414_6#1	H1_38	*Lachnoclostridium nov.*	✓	48	x	0.016	Y
13414_6#62	H3_43	*Ruminococcaceae nov.*	x	0.016	x	0.016	Y
18048_2#80	fCSP	*Gemmiger formicilis*	x	0.094	x	0.06	Y
20298_2#57	D85_115	*Erysipelotrichaceae nov.*	✓	48	x	0.5	Y
20298_3#19	G8_180	*Bacteroides cellulosilyticus*	x	2	x	8	N
20298_3#39	G14_207	*Enterobacter nov.*	✓	24	x	256	N
13470_2#56	H2_21	*Eubacterium limosum*	✓	16	x	0.016	Y
13470_2#93	H4_41	*Lachnospiraceae nov.*	✓	16	x	0.016	Y
14207_7#59	H5_60	*Intestinimonas butyriciproducens*	✓	48	x	0.016	Y
14207_7#7	H4_46	*Eubacterium rectale*	✓	12	x	0.016	Y
20287_6#22	E72_40	*Lachnospiraceae nov.*	✓	16	x	0.016	Y
20287_6#28	E79_57	*Ruminococcaceae nov.*	✓	48	✓	256	Y
20287_6#48	F1_100	*Blautia nov.*	✓	24	✓	0.016	N
20427_4#28	F41_240	*Blautia nov.*	✓	4	x	0.016	Y
13414_6#32	H2_22	*Blautia nov.*	x	0.016	x	0.032	Y
20287_6#76	F29_X14C	*Blautia nov.*	x	12	✓	0.016	N
20287_6#41	E92_84	*Blautia nov.*	x	0.016	✓	256	Y
20287_6#56	F9_115	*Blautia nov.*	x	0.016	✓	256	Y
20287_6#49	F2_101	*Blautia nov.*	x	0.016	✓	256	Y
20287_6#32	F83_63	*Blautia nov.*	x	0.016	✓	256	Y

Table 4.2. *Antibiotic resistance levels as observed with Etest strips.* EUCAST ECOFFs were used as a guide to classify strains as either resistant or sensitive. Species predicted to be resistant to the antibiotic are marked with a tick, and those predicted to be sensitive with a cross. The observed level of resistance is recorded in mg/L, and the last column indicates if both of the predictions from ARIBA were correct (Y) or incorrect (N). The validations provided a list of donor and recipient candidates.

4.5 Strain Growth Monitoring

The species used in this work are primarily novel commensals and their growth kinetics have not been previously determined. Bacterial population growth follows an S-shaped curve going through an exponential log growth phase, and forming a plateau in stationary phase. Generation times vary between species and will affect the efficiency of assays such as plasmid harvest and inducing

competency. The bacteria were incubated in YCFA and the growth curves were plotted to observe the growth patterns of commensals compared to well-characterised species, and to try and identify optimal timings for assays (Figure 4). Based on the results of these graphs 48hr cultures were preferred for plasmid harvests rather than overnight cultures. Additionally, knowing the generation times of the recipients allowed more suitable time points for creating competent cells to be identified.
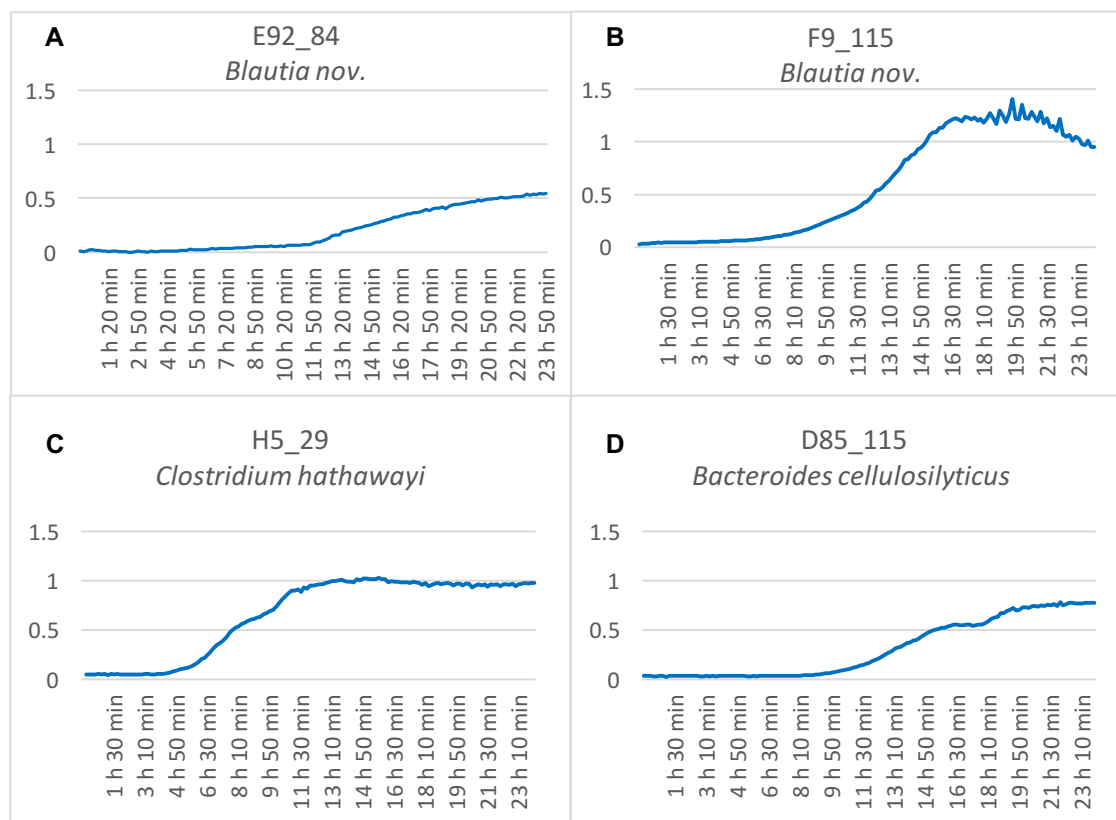
Figure 4.4 a-d. *OD growth curves for selected isolates.* A and B are candidate recipient strains, their growth curves display long lag phases which may be a factor in the creation of competent cells and the ability to participate in conjugation. C and D are plasmid containing strains, their growth curves demonstrate the point that different strains enter late stationary phase at varying times. These differences were used to optimise the timings of plasmid harvest and competent cell creation.

4.6 Plasmid Transfer

Plasmid transfer was attempted using a variety of protocols from the literature of bacteria phylogenetically similar to the selected isolates. Inducing natural competency was attempted first; natural competency is the innate potential of bacteria to uptake DNA from the environment when induced by a variety of stressors including DNA damage and starvation (Huddleston J. R., 2014). Plasmid DNA was isolated from all the strains in Table 1 for the transformation experiments. The recipient candidate was a novel *Blautia* species selected from a branch on the culture collection tree close to several of the donors. In addition, its genome annotation displayed predicted competence genes, and no antibiotic resistance genes; the resistance profile was validated by Etest. All the bacteria were incubated in different percentages of spent media. The spent media consisted of filtered

YCFA from overnight cultures of the same species mixed with normal YCFA broth in incremental ratios from 100% spent to 10% spent. This was to induce varying degrees of starvation, and potentially identify the correct level to trigger competence. The controls indicated that they survived the starvation but there was no growth on the selection plates.

Electroporation was attempted next; electroporation is the process by which bacterial cells are made porous by an electric pulse allowing DNA to move into the cell (Miller J.F., 1994). The recipient candidate was another novel Firmicute species selected experimentally due to its ability to survive the cell wall weakening procedure. Three electroporation buffers were tested, two caused the samples to arc, one allowed the bacteria to be shocked at an appropriate voltage. Once again, the controls indicated that bacteria survived the procedure, but there was no growth on the selection plates. The next attempt used protoplast transformation; protoplasts are formed by the enzymatic digestion of the cell wall (Rattanachaikunsopon P. and Phumkhachorn P., 2009). These protoplasts were then physically perturbed using glass beads and incubated with DNA to be taken up during cell wall repair. The candidate recipient was selected as mentioned above, with the same result– growth on the control plates only.

In addition to the transformation methods, conjugation was also attempted; conjugation is the procedure by which a bacterium directly transmits DNA to a neighbouring bacterium through an appendage called a pilus (Cabezón E., *et al* 2015). Species were selected to be donors from the list in Table 1, the final donors included two containing large plasmids with conjugation genes and two small plasmids containing mobility sequences. This would allow me to look at plasmids encoding their own conjugation machinery and plasmids relying on host conjugation machinery. The species chosen to be conjugation recipients were chosen based on their phylogenetic proximity to the plasmid-containing Firmicute donor species or their profile of selectable markers. The first iteration used antibiotic resistance negative Firmicute recipients that were selected for naladixic acid resistance.
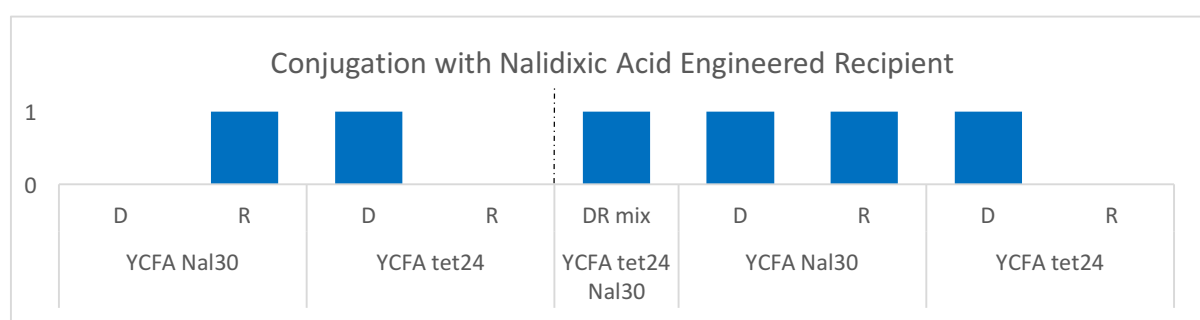
Figure 4.5. *Results of the conjugation between the donors (D) and naladixic acid engineered recipient (R).* 1 indicates growth and 0 indicates no growth. The left-hand side shows the bacterial profile before conjugation, and the right after conjugation. The donor and recipient both grew on the Naladixic Acid selection plate (YCFA Nal30), rendering the selection of trans-conjugants unsuccessful

Nalidixic acid resistance occurs as the result of a spontaneous point mutation in DNA gyrase genes (*gyrA, gyrB*) (Gellert M. *et al.,* 1977*)* When initially screened, the donor cells were nalidixic acid sensitive, however when plated during the test there was growth on the negative control, indicating that they probably acquired the mutation (Figure 5).

The second iteration used a selection of recipient candidates that were phylogenetically close to the donors and contained different resistance genes to the donor. There were several candidate Firmicutes identified that were predicted to contain only *erm* resistance; this however resulted in the exclusion of donors predicted to contain *erm* resistance in addition to *tet* resistance. The eligible donors were carried forward and mixed with the recipients. Growth was observed on the control plates but not the selection plates. In the case of the small plasmids this may indicate that the conjugation process did not occur due to lack of machinery in the host. In the case of both small and large plasmids the plasmids may not be compatible with or replication competent in the recipients (Figure 6).
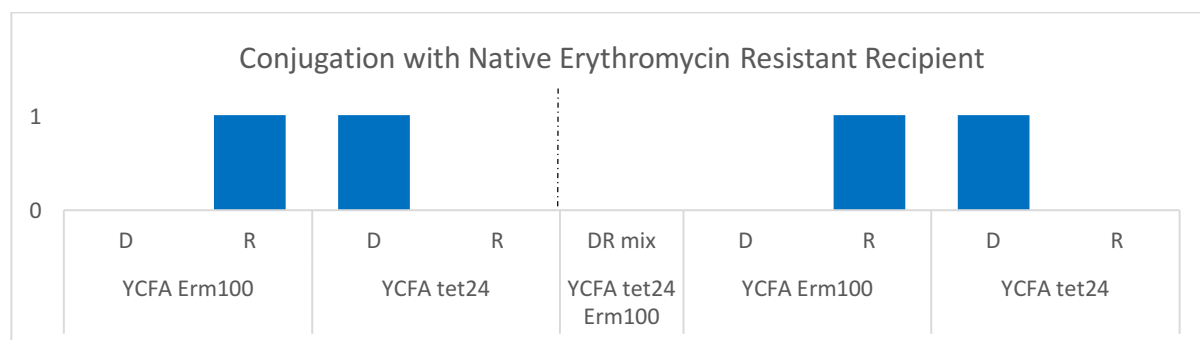
Figure 4.6. *Results of the conjugation between the donors (D) and erm-resistant bacteria (R).* 1 indicates growth and 0 indicates no growth. The left-hand side shows the bacterial profile before conjugation, and the right after conjugation. There was growth on the controls, but no growth on the selection plates (YCFA tet24 Erm100), meaning the conjugation was unsuccessful; this may be due to lack of conjugation machinery or incompatibility with the recipient.

The third iteration used *E. coli* and *Enterobacter cloacae* strains as recipient candidates to use aero-tolerance as a selectable marker. The recipient strains, while phylogenetically distant, were chosen as several of the plasmids showed potential broad host range (discussed in Chapter 3; Forster S. *et al, unpub, not shown*). Additionally, while the recipients could grow both aerobically and anaerobically, the donors were completely aero-sensitive allowing for easy selection between donors, recipients, and transconjugants. Donor and recipient strains were mixed and transconjugants were successfully observed for one of the donors. The levels of tetracycline resistance before- and after conjugation were verified by Etests (Figure 7), and plasmid DNA has been extracted for PCR validation and long read sequencing.
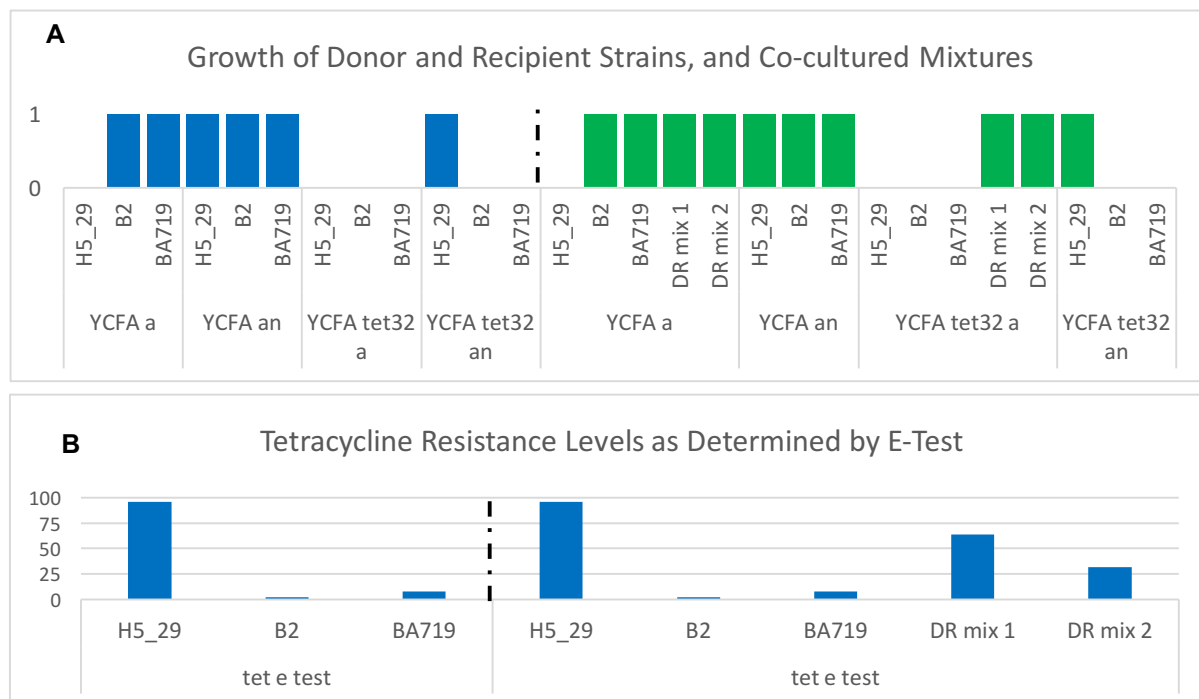
Figure 4.7 a-b. Results of the conjugation between donor H5_29 and aerobic bacterial recipients. The left-hand side shows the bacterial profile before conjugation, and the right after conjugation (a) 1 indicates growth and 0 indicates no growth. Growth of the donor and recipient mixes (DR mix 1 or 2) on the selection plate, YCFA tet32 in aerobic conditions (YCFA tet32 a), and the absence of donor or recipient growth under the same conditions indicates that conjugation was likely successful. (b) The levels of tet resistance in mg/L as determined by Etest before and after the conjugation show that the donor maintains its high level of resistance, the recipients remain sensitive, and the conjugation mixes display a boosted level of resistance.

5. Discussion

Plasmids are an important part of the bacterial genome and this thesis has shown their prevalence in the gut microbiome. The aims, as stated in the introduction, were to quantify the enrichment of plasmids in the culture collection and to investigate the distribution and functionality of AMR and mobility genes on the putative plasmids. This project successfully identified the distribution of putative plasmids and the prevalence of AMR genes in the plasmid-containing isolates, as well as the presence of mobility genes in a subset of plasmids. Investigating the spread of plasmids proved to be challenging; however, the results indicate that it may simply be a case of needing to find the right compatible pair.

5.1 Distribution of Plasmids in the Human Gut Microbiome Culture Collection

Plasmid distribution is predicted to follow a bimodal distribution with two peaks either side of the local minimum, usually around 20kb, as the standard population of plasmids will have large numbers of small mobilisable elements and bigger conjugative elements (Similie C. *et al.,* 2010). The distribution of putative plasmids in the gut microbiome doesn't follow the predicted bimodal model: the majority of the genomes present in the culture collections are predicted to contain only one plasmid, and these plasmids are predicted to be small, high-coverage elements. This may be a result of bias in the screening method, but alternatively this may reflect selection pressures in the niche.

The first explanation for observing primarily small, high-coverage plasmids is that this may be an artefact of the methods used to elucidate the plasmids. Antipov *et al. (*2016) identified that plasmidSPAdes struggles with the identification of plasmids with a coverage that is close to the median chromosomal coverage; since larger plasmids tend to be lower copy number they are more likely to have coverage similar to that of the chromosome. This suggests that plasmidSPAdes is not identifying the true number of plasmids across all size ranges in the gut microbiota. However, Jones and Marchesi (2007) observed similar results using the TRACA system to experimentally isolate plasmids from bacteria cultured from human stool samples. This could support the finding of primarily small elements as a true result, however, their experimental method may also be biased to isolate small elements over larger ones.

Arrendondo-Alonso *et al.* (2016) compared plasmidSPAdes to other completely automated platforms for computational plasmid isolation, giving them recall and precision scores. PlasmidSPAdes was compared to cBar (Zhou F. and Xu Y., 2010), Recycler (Rozov R. *et al.,* 2017), and PlasmidFinder (Carattoli *et al.,* 2014). PlasmidSPAdes had the highest recall score and was one of the software's able to identify novel plasmids. The study also identified the challenges with assembling large plasmids; plasmidSPAdes was flagged as being unable to assemble elements with their correct size,

however the median difference from correct size was 77bp. Ultimately, when compared to other completely automated methods of isolation plasmidSPAdes is the best for its purpose. Ideally plasmid isolation would make use of several automated platforms in conjunction with expert manual trimming, this however introduces high variability in predictions as trimming is dependent on operator expertise (Arrendondo-Alonso S. *et al.,* 2016, bioRxiv).

If the results presented reflect a genuine distribution pattern then this may serve to illustrate the cost of carriage mentioned in the introduction. It is proposed that there is a limit to the amount of extra DNA that can be carried, particularly in the absence of selective pressure. Considering that all the donors contributing to the culture collection are healthy and have not taken antibiotics for at least 6 months, it is likely there is only weak selection pressures in these donors' guts. For example, there is no primary need to maintain plasmids containing high carriage elements like multi-drug resistant cassettes. This point is further demonstrated by the fact that the most frequently identified resistance genes are tetracycline resistances; tetracycline is prevalent in the environment, primarily due to agriculture, and the donors would often be exposed to it (Popowska M. *et al.,* 2012).

5.2 Genomic Inference of Biological Functions: Plasmid Classification and AMR Distribution

Plasmid types exist for *Enterobacteriaceae, Pseudomonas*, *Staphylococcus*, and replicon typing strategies have been built around the replicons studied in these bacteria. The putative plasmid sequences were BLAST (blastn) against the PlasmidFinder database in an attempt to assign them to an incompatibility group. This strategy was not successful, yielding only 19 hits to the 240 plasmid-containing genomes in the culture collection. Similar strategies have been previously used to attempt to type Firmicute plasmids; for example, the work by Shintani *et al*. However, the plasmids classified were primarily of class *Bacilli,* with hits to only 4 out of the 90 plasmids from the *Clostridia* class in their dataset (Shintani M. *et al.,* 2015). The PF15101 *rep* gene domain was most frequently annotated in the putative plasmid sequences and the Pfam entry for this domain contains over 800 sequences. Using this extensive number of banked *rep* sequences I attempted to type the culture collection putative plasmids by aligning them to these reference sequences and looking at the incompatibility (Inc) groups of the plasmids they aligned to. However, the top BLASTx hits for the plasmids predicted in this study were to other uncharacterised plasmids. There is a large gap in the description of incompatibility groups and there is great scope for expanding our current knowledge of incompatibility groups to allow scientists to better understand all these novel extrachromosomal elements. This point is further illustrated by the phylogenies in Chapter 3: the trees are very low resolution, containing long shallow branches where presumably many genomes have not aligned to the classic *rep* genes. Despite current typing tools clearly only serving a small population of bacteria, aligning these genes does reveal distinct clusters. Classification of these groupings would allow for more educated experimental design around plasmid incompatibility and the ability to replicate in hosts.

The results of the AMR screen with ARIBA indicate a strong presence of AMR in the culture collection. Validating these predictions revealed them to be mostly accurate, with just 7 of the 44

tested predictions showing a different result; this may suggest the presence of non-functional genes, or genes that require induction. The gut microbiome is considered a reservoir for AMR and may contribute to the spread of AMR (Penders J. *et al.,* 2013). Mobility sequences are prevalent in the gut culture collection with 90 out of the 240 plasmid genomes containing plasmids with conjugation (*tra*) or *mob* sequences. In addition, 75% of these 90 are predicted to contain an AMR gene, further supporting the reservoir hypothesis. This may be an underestimation of the abundance of AMR in the gut as these predictions make use of known genetic variants to predict the presence of resistance; however, this would not indicate the presence of novel resistance genes or mechanisms of resistance. Novel resistance genes could be investigated by isolating the plasmids and transforming them into resistance free strains. These transformed bacteria can then be screened against a panel of antibiotics to observe if the addition of the plasmid confers any resistances that were not predicted during the *in silico* screen; this would be particularly useful for observing mechanisms such as novel efflux pumps.

The chord diagrams in Chapter 3 highlight potential broad range plasmids. The second chord diagram in particular identified plasmids in *Blautia* species that are shared with a *Klebsiella* strain. One of these elements is a small element containing tetracycline resistance, a toxin anti toxin system and plasmid recombination enzyme (Pre). The toxin anti toxin genes will help maintain the plasmid in the population; and the Pre protein is a mobilisable element, allowing the plasmid to be co-mobilised alongside conjugation competent plasmids (Priebe S. D. and Lacks S. A., 1989). In Chapter 4 plasmids have been demonstrated to be mobilisable, a plasmid from a health associated Firmicute was mobilised to Enterobacteriaceae species. The annotation of this plasmid revealed that it contained genes for citrate lyase, a component of the fatty acid synthesis chain; and genes for sorbitol transporters, in addition to conjugation genes and resistance genes. Citrate also known as E330 is a preservative found in many foods and household goods; and sorbitol is a sugar alternative found in sugar-free chewing gum and mints. Finding the genes for these processes on a mobile cassette that can move to a broad range of hosts, indicates potential adaptation to regularly consumed dietary nutrients. This is just one example of potentially many broad host range plasmids in the gut microbiota. The phylogenies in Chapter 3 display possible instances of plasmid sharing between the Firmicutes and all other phyla in the culture collection. These predictions can be used to set up more conjugation pairs and observe other instances of plasmid sharing *in vitro.*

A large portion of the Firmicutes are important health associated bacteria, this includes members of the genus *Blautia* which are important in the processing of polysaccharides (Eren A. M., *et al*., 2015) *Klebsiella* is a well-known pathogen and a major concern for the spread of AMR. It is on the WHO list of pathogens for which new antibiotics are critically needed; and recently a woman died from her infection with a strain of *Klebsiella* resistant to all available antibiotics in the United States (WHO, 2016; Chen L., *et al.,* 2016). The spread of plasmids is concerning not only because of the direct spread of resistance genes, but also the spread of genes related to adaptation to living in the gut. This means pathogens can more effectively colonise the gut and spend more time there, which can result

in chronic opportunistic infections as well as exposure to and acquisition of other extrachromosomal elements. One proposed theory regarding the formation of megaplasmids postulates that they are formed by the fusion of mini replicons (Zheng J., *et al.,* 2013). A plasmid with a broad host range *rep* is an ideal candidate for modular assembly of a megaplasmid through the addition of other functional cassettes; and the varying selection pressures of the gut may facilitate the assembly of plasmids with broad functionality.

To understand the breadth of functions present on the plasmids of the gut microbiota high quality sequences and detailed annotation are required. Long read sequencing of the isolated sequences would contribute to a better understanding of the functions encoded on these plasmids. Long read technologies can help resolve the problems with assembling repetitive regions and using the short read contigs as a reference or scaffold would help correct the errors that occur in long read sequencing, leading to more accurate annotations of functional genes and modification sites such as restriction sites. To investigate these annotations molecular cloning experiments could be used to identify the essential components of these novel plasmids and cassettes. This would allow for the creation of vectors for further genetic manipulation of gut commensals.

5.3 Transformation of Culture Collection Isolates

The difficulties with the experimental transformation may be due to a variety of reasons. The first strategy was to investigate natural competency. Natural competency is a highly coordinated process; in *Bacillus* species, it is controlled by the master regulator gene *comK* in response to a variety of environmental inputs that include nutrient availability and quorum sensing (Hamoen L.W. *et al.,* 2001). I attempted to use starvation to induce DNA uptake but was unsuccessful; this method may not have been the correct or sufficient stimulation to activate competence genes in the bacteria being tested.

The second strategy was to use electroporation. Plasmid transfer was previously tried with ruminuccocal species as described by Cocconcelli *et al.* (1992). The key difference between the protocol used by Cocconcelli and the protocol detailed in Chapter 2 is that the Cocconcelli experiment was carried out entirely under strict anaerobic conditions. During my procedure, some aerobic exposure was necessary, though the time spent in aerobic conditions was kept as short as possible. The bacterial cells did survive the protocol as there was growth on control plates, however the cells plated on selection did not grow. This suggests the stress of having to recover from aerobic exposure and the electrical pulse, in addition to the stress of maintaining a plasmid to survive antibiotic selection may have been too challenging to survival. The other issue highlighted by Cocconcelli *et al. (1992)* was plasmid and host compatibility: they inferred that one of their transformations may have failed as the plasmid could not replicate in the recipient. My candidate recipients were selected to be as phylogenetically close as possible but still may not have been suitable. A final possible source of the failure may have been ineffective competent cells. As previously mentioned, the growth phase of bacteria can affect competency: the growth curves of the potential recipients were plotted (Chapter 4) to observe growing times. Using this data cells were harvested at different time points and a variety of

competency-inducing methods were used, including ice washes and enzymatic cell wall degradation. This points to replication incompatibility being the key reason for the lack of transformants.

The final strategy was to use conjugation, which may also have been affected by plasmid and host compatibility. Another reason for difficulties may be due to the fact that conjugation relies on auxiliary machinery. This is normally encoded on large plasmids, and they are self mobilisable. In contrast, smaller plasmids, such as those identified in the culture collection, tend to contain mobilisable elements which only facilitate mobilisation using the hosts conjugation machinery or in conjunction with larger plasmids encoding their own machinery.

5.4 Applications of Identified Plasmid Sequences

The identification of multiple novel plasmids in commensal bacteria has many implications; one of the immediate applications of the screening is the creation of a database of plasmid sequences. This database can be used to identify extrachromosomal elements, particularly in metagenomic datasets, addressing one of the challenges to metagenomic analysis. As these sequences were isolated from whole genome sequences of pure cultures, this database would make it possible to make more accurate inferences of host-extrachromosomal DNA relationships in metagenomic datasets.

In addition to improving metagenomic data analysis, the database of plasmid sequences has broader applicability. Long read sequencing of these isolated elements, using the short-read sequencing data as a reference, will lead to a more accurate overall sequence assembly and better annotation of the genes. This high-resolution annotation then forms a foundation for future applications in genetic engineering. For example, the completed database will provide a collection of backbones for building vectors tailored to the microbiota. The vast potential applications of engineering the microbiota will be discussed in detail below

5.5 Plasmids and Plasmid Sequence Data as Genetic Tools

The microbiota are responsible for many functions in the gut and as presented in Chapter 1 these functions may be encoded on plasmids. Some functions include aiding in digestion, immunological and neurological signalling, and the regulation of bacterial growth. Digestion is aided through the fermentation of sugars and fibres; processing of short chain fatty acids; and butyrate production. Immunological signals to the gut epithelia include the stimulation of mucus production, the induction of Treg cells, and neurological signalling includes the production of molecules that stimulate the gut-brain axis. The growth of bacterial communities is regulated by the production of bacteriocins and other antimicrobial proteins to induce death, and products to regulate the formation of biofilms (Kali A., 2015).

With a wide range of information from the human body presented to the microbiota, the wide range of effects that they can produce, and high compatibility with the host, commensal bacteria are poised to be ideal sensors and effectors in the gut. The advances in techniques of synthetic biology have allowed the manipulation of well-characterised bacterial strains; e.g. developing engineered bacteria

and even isolating key genetic components to build cell-free systems for the production of proteins, metabolites, and therapeutics (Hodgman C.E. and Jewett M.C., 2012). The manipulation of commensal strains has been limited due to difficulties with culturing them for biological and genomic analysis. The development of endogenous and exogenous techniques as discussed in the introduction has led to increased isolation of commensal bacteria and their genetic information. With better understanding of both commensal bacteria and their mobile genetic elements, genetic engineering of these organisms is now ready to be realised. Commensal bacteria provide an advantage over traditional commercial strains by being a stable part of the host biological make up, able to exist in the niche over extended periods of time without causing a disease response, and likely contributing to host health.

One of the main outputs of bacterial engineering is the development of a DNA-based memory device; this would be a device with the ability to recognise input from its surroundings and execute pre-programmed responses. This would allow for a system with real time responses providing early diagnosis and treatment of conditions in the gut. Bacteria could be engineered to assist in the diagnosis of diseases by programming them for use as biomarkers and diagnostic sensors. For example, bacteria have previously been used as biomarkers in the screening of tumours as they can easily colonise tumour cells. These bacteria have been programmed with genes allowing them to produce fluorescence, bioluminescence, magnetic particles, or positrons for tumour imaging; in addition, they can infect in a tumour cell specific fashion enabling the monitoring of metastasis (Bernardes N. *et al.,* 2013). Engineering the native flora to produce these responses as a result of changes in the host that indicate cancer would provide a stable and long-term monitoring system and enable rapid diagnosis.

Beyond signals and diagnosis, bacteria can be programmed to elicit an active response that aides in combating an illness. Bacteria have been programmed to release substances in response to inflammation, neoplastic changes, infection, tissue damage – amongst other signals. Inflammation in the gut is a large concern since it features in many disease processes of this organ, including allergic responses, inflammatory bowel disease, and bowel cancers (Shen N. and Clemente J.C., 2015). Programming gut bacteria could lead to more comfortable lives allowing people to eat foods they previously couldn't, preventing and soothing painful flare ups, and minimising the amount of inflammatory damage faced by the gut epithelium – a major risk factor for the development of bowel cancer.

Furthermore, commensal bacteria could be used as a next generation strategy for vaccination against infection or on-going prophylaxis against infection, by engineering them to produce small molecules or peptides that stimulate antigen production. This would be particularly useful for the immunocompromised, giving an increased level of protection and allowing infection to be identified and treated faster (Braff D. *et al.,* 2016; Kali A., 2015). Other strategies being developed include using bacteria to produce small molecules that disrupt infection by altering signalling within the community.

Goh *et al* describe a method by which bacteria were programed to secrete cholera auto-inducers-CAI-1 and AI-2, small molecules that are used by *Vibrio cholera* in quorum sensing. By sensing its population density *V. cholera* can determine whether or not to secrete toxins. Goh *et al.* (2012) found that the synthetic bacteria were successful in inducing density-dependent quorum signalling, preventing the secretion of toxins by *V. cholera*. Strategies like this could be used to prevent opportunistic infections colonising niches on the human body.

Other small molecules that bacteria can be used to produce include anti-cancer agents. Cancer treatments have been developed using antibodies to deliver small molecules in a specific fashion. Scientists are now looking to use similar strategies with bacteria (Bernades N. *et al.,* 2013). This strategy would be especially potent to treat bowel cancers since it could be delivered in a highly localised manner using microorganisms already prevalent in the niche. In addition, bacteria may have an advantage over antibody therapy by overcoming several issues including: antibody degradation to their high gastrointestinal instability (Casanova E.B., 2013); the inability of antibodies to penetrate the centre of tumours; and their stimulation of the immune response that can result in cytokine release syndrome or hypersensitivity (Deng R. *et al.,* 2012; Elbakri A. *et al.,* 2010). Bacteria can easily 'infect' all the cells in a tumour and commensal species should not cause an immune response in the gut as the bacteria are familiar to the host.

5.6 Strategies to Engineer the Microbiota

The primary barrier to employing any of these therapies is difficulties in engineering commensal bacteria. Strategies and tools are defined for well-characterised bacteria like commercial stains of *E. coli,* but not for commensal species. Chapter 1 introduced us to the main challenges faced when studying commensals including the fact that many commensal organisms remained unculturable. This has been combatted with a variety of strategies including the work Browne *et al.* (2016), which allowed the development of the culture collection utilised in this study. The next hurdle is identifying genetic elements compatible with commensal species that can be used as vectors; the experimental and computational isolation strategies have been developed complement each other, and this thesis has aimed to illustrate that.

The next step is annotation and isolation of genes encoded by the isolated plasmids. Classification of replicons and functional annotation will provide insights into potentially novel functions encoded on plasmids, and isolation of the annotated genes will allow the building of synthetic gene circuits. Synthetic circuits enable the building of precise networks of programmed responses, including failsafe's, creating controlled genetically modified organisms (GMOs).

Three particularly pertinent types of gene circuit include dynamic, logic and communication circuits detailed in Figure 1. Dynamic circuits involve interlinked positive- and negative-feedback loops, which produce an oscillatory response. Figure 1 gives an example of positive arabinose signalling linked with negative *lac* repressor signalling. Each feedback loop is modulated by the abundance of small

molecule inducers and the reciprocal signalling produces oscillations in green fluorescent protein expression (GFP). This dynamic circuit system could be useful for developing bacterial prophylactics: bacteria would produce antimicrobials molecules in response to the positive disease signal, but this response could be switched off in the absence of a condition. Logic circuits make use of 'AND' and 'NOT' gates which can be used to build systems that follow a set of conditional rules. The system in Figure 1 requires the presence of both m1 'AND' m2 to inhibit the production of yellow fluorescent protein (YFP); systems like this could be used to build very specific biomarkers based on cell type as well as intracellular signals. Communication circuits can use quorum sensing as a method of population control: the example in Figure 1 uses the small molecule AHL to signal density; increased density and therefore increased AHL induces expression of a killer gene, reducing population density. Communication signals and circuits of this type are manipulated in the method of inhibiting *V. cholera* toxin secretion described above (Goh YL *et al.,* 2012).
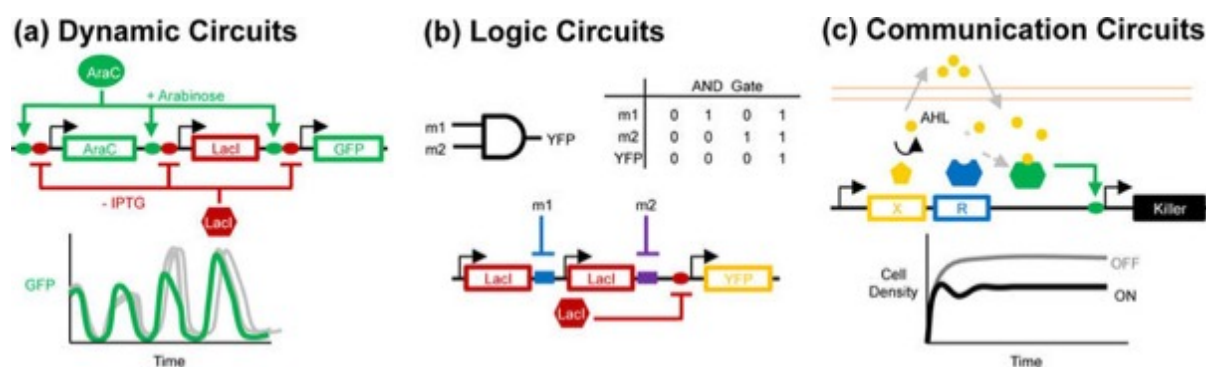
Figure 5.1. *Picture displaying the three key synthetic circuits: dynamic, logic, and communication circuits.* (a) Dynamic circuits employ interlinked positive and negative feedback loops to produce an oscillatory response. This example uses a positive arabinose signalling linked with negative *lac* repressor signalling. Each feedback loop is modulated by the abundance of small molecule inducers and the reciprocal signalling produces oscillations in green fluorescent protein expression (GFP). (b) Logic circuits use "AND" or "NOT" gates to build conditional rules for the system to follow. This example requires the presence of both m1 'AND' m2 to inhibit the production of yellow fluorescent protein (YFP). (c) Communication circuits use quorum sensing for population control. This example shows the small molecule AHL being used to signal density; increased density and therefore increased AHL induces expression of a killer gene, reducing population density. These circuits are assembled to build complex networks of responses, which are then often delivered on plasmid DNA (Chen Y.Y. *et al.*, 2012)

The final step is the actual engineering process and physical manipulation. As shown in Chapter 4, transformation of commensals is not always straightforward. As we continue to study and understand commensals we will be able to optimise methods of transformation, allowing us to develop *in vitro* and *ex vivo* bacteriotherapies.

For *in vivo* alterations, viral vectors have been shown to be highly efficient gene therapy vehicles, and are the most commonly used therapeutic vector (Wirth T. *et al.,* 2013). Investigation into the gut virome is a growing field and will likely lead to the discovery of new phages and the development of new tools and techniques for probing metagenomic datasets (Ogilvie L.A. and Jones B.V., 2015).

Plasmid delivery will be greatly enhanced by the creation of a database of gut phages, providing a catalogue of gut specific vectors that can be used to alter commensal bacteria. Phage based plasmid delivery would be advantageous as phages are often highly species specific making them great candidates for *in vivo* microbiome engineering. Phage-based strategies have been used as antimicrobial therapies, by delivering antimicrobial gene networks, or by inducing death directly. (Braff D. *et al.,* 2016). Second generation phage strategies include the use of phagemids: non-replicating viral capsules, capable of effectively delivering high copy number synthetic plasmids. Krom *et al.* (2015) developed this strategy for the delivery of non-lytic antimicrobials and to investigate alternatives to traditional antibiotics, finding that it successfully caused lysis free death of peritonitis causing *E. coli* in mice. The modular design of the system would allow for the easy production of phagemids targeting a variety of bacteria, which could induce transient product generation or stable integration of a new segment of DNA. Therefore, the database of plasmids from this study can contribute to a range of templates for designing phagemids to target the gut microbiota; and additional studies on novel plasmids and phages will extend the range of available templates.

5.7 Challenges to Bacteriotherapy

In vitro and *ex vivo* strategies will allow for the large-scale production of engineered microbes, but there are challenges to the delivery of these bacteriotherapies. One of the key challenges to bacteriotherapy is identifying a suitable route of administration. Probiotics are often used to bolster the gut microbiome and are generally administered orally, however the amounts of bacteria that survive the oral route of admission are very variable between manufacturers due to species robustness and delivery vehicle (Govender M. *et al.,* 2014). Probiotics primarily utilise lactic acid bacteria (LAB) as they are known for their tolerance of gastric acid; this includes the recently reported successful and proof-of-principle trial in India that definitively demonstrated the beneficial effects of probiotics (Panigrahi P. *et al.,* 2017)

The current strategy for delivering bacteriotherapies to the gut microbes is faecal microbiota transplantation (FMT); this therapy is primarily used for the treatment of recurring *Clostridium difficile,* however it is being investigated as a treatment for other conditions. This strategy typically involves the delivery of a native complex community of bacteria via enema, colonoscopy, or nasogastric tube, but could be adapted to introduce other bacterial populations including engineered commensals. One of the limitations to the use of this therapy is the public attitude to FMT. While the majority of the research surrounding attitudes has been conducted around recurring *C. difficile* infection (CDI) scenarios, the results are telling and have prompted investigation into the role language and the media can play in changing perceptions surrounding the procedure. Suggesting that moving away from terminology such as "the ick/yuck factor" when talking about the procedure will provide a more positive image and increase receptivity (Chuong K.H. *et al.,* 2015). Patients are reported to be receptive to the treatment if mandated by a physician, and if the contents are made more aesthetically pleasing, i.e. in a pill or clear solution. Many also indicated discomfort with route of delivery, preferring a colonoscopy to nasogastric tube delivery (Park L. *et al.,* 2017).

Once delivered the next major issue is the problem of stable engraftment: often therapeutic effects are lost as bacteria are pushed out of the community (Mimee M. *et al.,* 2016). Using engineered commensal species over commercial LAB should help as the bacteria would be part of the native community and not outcompeted. The ideal strategy, particularly for prophylactic bacteriotherapy, may be an *ex vivo* therapy where individuals provide a sample that is then engineered and supplied back to them. This would address another major patient concern, which is the origin and safety of donors. Another consideration is the biosafety of these engineered bacteria and any side effects they could have in the host and effects after being released in to the environment. Ideally, we should be able to induce the death of the engineered species in a specific manner and using a mechanism that is not likely to induce resistance. One solution may be to include a 'kill switch' in the engineering of the synthetic circuits, that will eliminate either the modified organism or the engineered DNA. (Sonnenberg J. L., 2015). Cailando and Voigt (2015) designed a CRISPR strategy that resulted in the degradation of the introduced synthetic plasmid once engineered organisms left their containment zone and were no longer exposed to specific amino acids; gut specific metabolites could be used to facilitate a similar strategy.

5.8 Non-Health Associated Uses of Engineered Microbiota

Engineering the microbiome can provide benefits other than direct improvement of an individuals' health, including environmental and economic benefits. According to the World Health Organisation (WHO) at least 2 billion people globally are using a water source contaminated with faeces and facing an increased risk of contracting a water borne disease (WHO, 2017). The BioVolt by Cambrian Innovation (MA, USA) is one example of a microbial fuel cell (MFC) being used to clean up water and create energy simultaneously. MFCs consist of bacteria that produce electrons as they respire, producing clean water, and enough energy to sustain themselves plus excess. Engineering strategies are already being utilised in this field with Kirchhofer *et al.,* (2017) providing a strategy for engineering to increase their electrical output. Stool bacteria could be engineered to conduct this procedure *in situ,* which would increase the energy output per litre of wastewater. A strategy like this would allow for easy environmental clean-up of water reservoirs, leading to increased population health and providing alternative energy sources.

5.9 Strategies for Investigating Engineered Microbiota

The examples above show that engineering the microbiome can provide solutions to a variety of health and non-health associated problems. However, the development of these solutions requires a variety of techniques in our "lab toolbox". We can now: culture gut bacterial species, scan the genomes for extrachromosomal elements, isolate these elements, and the techniques for manipulating the microbiota are being investigated. The next hurdle will be to develop techniques for *in situ* manipulations and culture. The current models being used include 2D and 3D cell culture, and gnotobiotic mouse models. Intestinal cells from patients, healthy donors, or induced pluripotent stem cells (iPSCs) are either grown in traditional 2D, or in 3D organoids. The organoid models allow for the

study of individual differences, as well as population-level differences, in an anatomically relevant model. Gnotobiotic animals are raised in a germ-free environment and inoculated with the desired community of bacteria. This model has been widely used for studying the microbiome, with a PubMed search for germ-free 'AND' microbiome revealing nearly 300 papers published in the last year with over 30 published in high impact journals.

The next generation of *in situ* testing tools may be the organ-on-chip models (Bhatia S.N., and Ingber D.E., 2014). These devices consist of a microchip with microfluidics channels engrafted with cells of the desired organ arranged appropriately within the 3D space. Once again, the channels can be engrafted with cells from patients, healthy donors, or iPSCs allowing different levels of variation to be studied. The channel is irrigated with the appropriate media; and these models can be used to study a wide variety of interactions. Mechanical forces can be modelled with the liquid flow, and interactions between various cells can be studied either by addition to the main channel or culturing in adjacent channels (Ingber D.E. 2016). This allows for a highly modular system where stimuli can be easily added and removed one at a time. This method has been used to culture bacteria with an intestinal epithelium and this system was maintained for several weeks – much longer than is possible with organoid cultures. In addition to bacteria, circulating immune cells can be added. This was used to demonstrate that the presence of an *E. coli* endotoxin only caused an irritable bowel disease (IBD) phenotype when circulating immune cells were present as well (Kim H.J. *et al.,* 2016). This is a highly tractable system that allows for highly flexible real-time investigation, and the use of cell cultures works towards the goals of the 3Rs to limit the use of animal models, which are currently imperative for these types of studies.

5.10 Summary

This body of work aimed to introduce and discuss the biology behind plasmids of the gut microbiome, investigate and validate the presence of plasmids in the microbiota, and provide scope for the future uses of these extrachromosomal elements. The plasmids observed in the microbiota provide insights into the prevalence extrachromosomal DNA and its function the gut, and present a starting point for the development of commensal engineering vectors.

Bibliography

Al-Asmakh, M. & Zadjali, F. (2015). Use of Germ-Free Animal Models in Microbiota-Related Research. *Journal of Microbiology and Biotechnology,* 25(10)**,** 1583-1588.

Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A. & Pevzner, P. A. (2016). plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics,* 32(22)**,** 3380-3387.

Arredondo-Alonso S., van Schaik W., Willems R., Schurch A., (2017), On the (im)possibility to reconstruct plasmids from whole genome short-read sequencing data, *bioRxiv preprint*

Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution,* 28(8)**,** 489-495.

Baxter J. C. Funnell B. E. (2014), Plasmid Partition Mechanisms, *Microbiology Spectrum*, 2(6), 1-20.

Bernardes, N., Chakrabarty, A. M. & Fialho, A. M. (2013). Engineering of bacterial strains and their products for cancer therapy. *Applied Microbiology and Biotechnology,* 97(12)**,** 5189-5199.

Bhatia, S. N. & Ingber, D. E. (2014). Microfluidic organs-on-chips. *Nature Biotechnology,* 32(8)**,** 760-772.

Birge, E.A., (2013) *Bacterial and Bacteriophage Genetics,* Springer-Verlag, New York

Braff, D., Shis, D. & Collins, J. J. (2016). Synthetic biology platform technologies for antimicrobial applications. *Advanced Drug Delivery Reviews,* 105**,** 35-43.

Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., Goulding, D. & Lawley, T. D. (2016). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature,* 533(7604)**,** 543-546.

Cabezon, E., Ripoll-Rozada, J., Pena, A., de la Cruz, F. & Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. *Fems Microbiology Reviews,* 39(1)**,** 81-95.

Caliando, B. J. & Voigt, C. A. (2015). Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nature Communications,* 6**,** 10.

Carattoli, A., Zankari, E., Garcia-Fernandez, A., Larsen, M. V., Lund, O., Villa, L., Aarestrup, F. M. & Hasman, H. (2014). In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial Agents and Chemotherapy,* 58(7)**,** 3895-3903.

Casanova Estruch, B. (2013). Safety profile and practical considerations of monoclonal antibody treatment. *Neurologia (Barcelona, Spain),* 28(3)**,** 169-78.

Chen, L., Todd, R., Kiehlbauch, J., Walters, M., Kallen, A., (2017) Notes from the Field: Pan-Resistant New Delhi Metallo-Beta-Lactamase-Producing Klebsiella pneumoniae — Washoe County, Nevada, 2016, *MMWR Morbidity and Mortality Weekly Report*, 66(33)

Chen, Y. Y., Galloway, K. E. & Smolke, C. D. (2012). Synthetic biology: advancing biological frontiers by building synthetic systems. *Genome Biology,* 13(2)**,** 10.

Chuong, K. H., O'Doherty, K. C. & Secko, D. M. (2015). Media Discourse on the Social Acceptability of Fecal Transplants. *Qualitative Health Research,* 25(10)**,** 1359-1371.

Cocconcelli, P. S., Ferrari, E., Rossi, F. & Bottazzi, V. (1992). Plasmid transformation of *Ruminococcus albus* by means of high-voltage electroporation. *Fems Microbiology Letters,* 94(3)**,**203-207.

Collins, S. M., Kassam, Z. & Bercik, P. (2013). The adoptive transfer of behavioral phenotype via the intestinal microbiota: experimental evidence and clinical implications. *Current Opinion in Microbiology,*16(3)**,** 240-245.

Cuiv, P. O., Smith, W. J., Pottenger, S., Burman, S., Shanahan, E. R. & Morrison, M. (2015). Isolation of Genetically Tractable Most-Wanted Bacteria by Metaparental Mating. *Scientific Reports,* 5**,** 11.

Davari, S., Talaei, S. A., Alaei, H. & Salami, M. (2013). Probiotics treatment improves diabetes-induced impairment of synaptic activity and cognitive function: behavioral and electrophysiological proofs for microbiome-gut-brain axis. *Neuroscience,* 240**,** 287-296.

del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M. & Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews,* 62(2)**,** 434-464.

Deng, R., Jin, F., Prabhu, S. & Iyer, S. (2012). Monoclonal antibodies: what are the pharmacokinetic and pharmacodynamic considerations for drug development? *Expert Opinion on Drug Metabolism & Toxicology,* 8(2)**,** 141-160.

Dib, J. R., Wagenknecht, M., Farias, M. E. & Meinhardt, F. (2015). Strategies and approaches in plasmidome studies uncovering plasmid diversity disregarding of linear elements? *Frontiers in Microbiology,*6**,** 12.

Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. (2016). Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology,* 14(1)**,** 20-32.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32(5)**,** 1792-1797.

ElBakri, A., Nelson, P. N. & Abu Odeh, R. O. (2010). The state of antibody therapy. *Human Immunology,* 71(12)**,** 1243-1250.

Eren, A. M., Sogin, M. L., Morrison, H. G., Vineis, J. H., Fisher, J. C., Newton, R. J., & McLellan, S. L., (2015), A single genus in the gut microbiome reflects host preference and specificity, *The ISME Journal*, *9*(1), 90–100.

Foster, J. A. & Neufeld, K. A. M. (2013). Gut-brain: how the microbiome influences anxiety and depression. *Trends in Neurosciences,* 36(5)**,** 305-312.

Funnell, B. E. (2005). Partition-mediated plasmid pairing. *Plasmid,* 53(2)**,** 119-125.

Furuya, E. Y. & Lowy, F. D. (2006). Antimicrobial-resistant bacteria in the community setting. *Nature Reviews Microbiology,* 4(1)**,** 36-45.

Gloux, K. & Anba-Mondoloni, J. (2016). Unique beta-Glucuronidase Locus in Gut Microbiomes of Crohn's Disease Patients and Unaffected First-Degree Relatives. *Plos One,* 11(1)**,** 17.

Goh, Y. L., He, H. F. & March, J. C. (2012). Engineering commensal bacteria for prophylaxis against infection. *Current Opinion in Biotechnology,* 23(6)**,** 924-930.

Gohari, I. M., Kropinski, A. M., Weese, S. J., Parreira, V. R., Whitehead, A. E., Boerlin, P. & Prescott, J. F. (2016). Plasmid Characterization and Chromosome Analysis of Two netF plus Clostridium perfringens Isolates Associated with Foal and Canine Necrotizing Enteritis. *Plos One,* 11(2)**,** 20.

Govender, M., Choonara, Y. E., Kumar, P., du Toit, L. C., van Vuuren, S. & Pillay, V. (2014). A Review of the Advancements in Probiotic Delivery: Conventional vs. Non-conventional Formulations for Intestinal Flora Supplementation. *Aaps Pharmscitech,* 15(1)**,** 29-43.

Griffiths, A.J.F., Gelbart, W.M., Miller, J.H., Lewontin, R., (1999), Bacterial Conjugation, *Modern Genetic Analysis*, W. H. Freeman, New York.

Gualano, M. R., Gili, R., Scaioli, G., Bert, F. & Siliquini, R. (2015). General population's knowledge and attitudes about antibiotics: a systematic review and meta-analysis. *Pharmacoepidemiology and Drug Safety,* 24(1)**,** 2-10.

Hamoen, L. W., Haijema, B., Bijlsma, J. J., Venema, G. & Lovett, C. M. (2001). The Bacillus subtilis competence transcription factor, ComK, overrides LexA-imposed transcriptional inhibition without physically displacing LexA. *Journal of Biological Chemistry,* 276(46)**,** 42901-42907.

Hehemann, J. H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M. & Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut-microbiota. *Nature,* 464(7290)**,** 908-U123.

Higgins, N. P., & Vologodskii, A. V. (2015). Topological Behavior of Plasmid DNA. Microbiology Spectrum, 3(2), 1-25

Hodgman, C. E. & Jewett, M. C. (2012). Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering,* 14(3)**,** 261-269.

Huddleston, J. R. (2014). Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infection and Drug Resistance,* 7**,** 167-176.

Hunt M., Mather A., Sánchez-Busó L., Page A., Parkhill J., Keane J., Harris S., (2017), ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads, *bioRxiv preprint*

Ingber, D. E. (2016). Reverse Engineering Human Pathophysiology with Organs-on-Chips. *Cell,* 164(6)**,** 1105-1109.

Inouye, M., Dashnow, H., Raven, L. A., Schultz, M. B., Pope, B. J., Tomita, T., Zobel, J. & Holt, K. E. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine,* 6**,** 16.

Johnson, C. M., & Grossman, A. D. (2015). Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annual Review of Genetics*, 49, 577–601

Jones, B. V. & Marchesi, J. R. (2007). Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nature Methods,* 4(1)**,** 55-61.

Jones, B. V., Sun, F. N. & Marchesi, J. R. (2010). Comparative metagenomic analysis of plasmid-encoded functions in the human gut microbiome. *Bmc Genomics,* 11**,** 17.

Kali A, Human (2015), Microbiome Engineering: The Future and Beyond. *Journal of clinical and diagnostic research*, 9(9), DE01-4

Khan, S. A. (2005). Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid,* 53(2)**,** 126-136.

Kim, H. J., Li, H., Collins, J. J. & Ingber, D. E. (2016). Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. *Proceedings of the National Academy of Sciences of the United States of America,* 113(1)**,** E7-E15.

Kirchhofer, N. D., Rengert, Z. D., Dahlquist, F. W., Nguyen, T. Q. & Bazan, G. C. (2017). A Ferrocene-Based Conjugated Oligoelectrolyte Catalyzes Bacterial Electrode Respiration. *Chem,* 2(2)**,** 240-257.

Kommineni, S., Bretl, D. J., Lam, V., Chakraborty, R., Hayward, M., Simpson, P., Cao, Y. M., Bousounis, P., Kristich, C. J. & Salzman, N. H. (2015). Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. *Nature,* 526(7575)**,** 719-722.

Krom, R. J., Bhargava, P., Lobritz, M. A. & Collins, J. J. (2015). Engineered Phagemids for Nonlytic, Targeted Antibacterial Therapies. *Nano Letters,* 15(7)**,** 4808-4813.

Kumar N. unpub, http://itol.embl.de/tree/193622021385733147363687, [Accessed Feb 2017]

Li, Y., Canchaya, C., Fang, F., Raftis, E., Ryan, K. A., van Pijkeren, J. P., van Sinderen, D. & O'Toole, P. W. (2007). Distribution of megaplasmids in Lactobacillus salivarius and other Lactobacilli. *Journal of Bacteriology,* 189(17)**,** 6128-6139.

Lin-Chao, S., Chen, W. T. & Wong, T. T. (1992). High copy number of the puc plasmid results from a rom rop-suppressible point mutation in rna-ii. *Molecular Microbiology,* 6(22)**,** 3385-3393.

Luczynski, P., Neufeld, K. A. M., Oriach, C. S., Clarke, G., Dinan, T. G. & Cryan, J. F. (2016). Growing up in a Bubble: Using Germ-Free Animals to Assess the Influence of the Gut Microbiota on Brain and Behavior. *International Journal of Neuropsychopharmacology,* 19(8)**,** 17.

Marbouty, M., & Koszul, R., (2015), Metagenome Analysis Exploiting High-Throughput Chromosome Conformation Capture (3C) Data, *Trends in Genetics,* 31(12), 673-682.

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., Bhullar, K., Canova, M. J., De Pascale, G., Ejim, L., Kalan, L., King, A. M., Koteva, K., Morar, M., Mulvey, M. R., O'Brien, J. S., Pawlowski, A. C., Piddock, L. J. V., Spanogiannopoulos, P., Sutherland, A. D., Tang, I., Taylor, P. L., Thaker, M., Wang, W. L., Yan, M., Yu, T. & Wright, G. D. (2013). The Comprehensive Antibiotic Resistance Database. *Antimicrobial Agents and Chemotherapy,* 57(7)**,** 3348-3357.

Miller, J. F. (1994). Bacterial transformation by electroporation. *Bacterial Pathogenesis, Pt a: Identification and Regulation of Virulence Factors,* 235**,** 375-385.

Mimee, M., Citorik, R. J. & Lu, T. K. (2016). Microbiome therapeutics - Advances and challenges. *Advanced Drug Delivery Reviews,* 105**,** 44-54.

Moore, R., Adams, V., Rood, J., McClane, B., Wisniewski, J., Li, J., Uzal, F., (2014), Virulence Plasmids of Spore-Forming Bacteria, *Microbiology Spectrum,* 2(6), 1-24

Muth, G., & Espinosa, M., (2003), Conjugative Plasmid Transfer in Gram-Positive Bacteria, *Microbiology and Molecular Biology Reviews*, *67*(2), 277–301.

NEB Protocols, www.neb.com/protocols/2012/06/21/making-your-own-electrocompetent-cells , [Accessed April 2017]

Nelson, K. E., Weinstock, G. M., Highlander, S. K., Worley, K. C., Creasy, H. H., Wortman, J. R., Rusch, D. B., Mitreva, M., Sodergren, E., Chinwalla, A. T., Feldgarden, M., Gevers, D., Haas, B. J., Madupu, R., Ward, D. V., Birren, B., Gibbs, R. A., Methe, B., Petrosino, J. F., Strausberg, R. L., Sutton, G. G., White, O. R., Wilson, R. K., Durkin, S., Gujja, S., Howarth, C., Kodira, C. D., Kyrpides, N., Mehta, T., Muzny, D. M., Pearson, M., Pepin, K., Pati, A., Qin, X., Yandava, C., Zeng, Q. D., Zhang, L., Berlin, A. M., Chen, L., Hepburn, T. A., Johnson, J., McCorrison, J., Miller, J., Minx, P., Nusbaum, C., Russ, C., Sykes, S. M., Tomlinson, C. M., Young, S., Warren, W. C., Badger, J., Crabtree, J., Markowitz, V. M., Orvis, J., Cree, A., Ferriera, S., Gillis, M., Hemphill, L. D., Joshi, V., Kovar, C., Wetterstrand, K. A., Abouellleil, A., Wollam, A. M., Buhay, C. J., Ding, Y., Dugan, S., Fulton, L. L., Fulton, R. S., Holder, M., Hostetler, J., Allen-Vercoe, E., Clifton, S. W., Earl, A. M., Farmer, C. N., Giglio, M. G., Liolios, K., Surette, M. G., Torralba, M., Xu, Q., Pohl, C., Wilczek-Boney, K., Zhu, D. H. & Human Microbiome, J. (2010). A Catalog of Reference Genomes from the Human Microbiome. *Science,* 328(5981)**,** 994-999.

Novick, R. P. (1987). Plasmid incompatibility. *Microbiological Reviews*, *51*(4), 381–395

Ogawa, N., Chakrabarty, A., Zaborina, O., (2004). Degradative Plasmids, *Plasmid Biology,* ASM Press, Washington, DC.

Ogilvie, L. A., Firouzmand, S. & Jones, B. V. (2012). Evolutionary, ecological and biotechnological perspectives on plasmids resident in the human gut mobile metagenome. *Bioengineered,* 3(1)**,** 13-31.

Ogilvie, L. A. & Jones, B. V. (2015). The human gut virome: a multifaceted majority. *Frontiers in Microbiology,* 6**,** 12.

Panigrahi P., Parida S., Nanda N., Satpathy R., Pradhan L., Chandel D., Baccaglini L., Mohapatra A., Mohapatra S., Misra P., Chaudhry R., Chen H., Johnson J., Morris J., Paneth N. ,Gewolb I., (2017) A randomized synbiotic trial to prevent sepsis among infants in rural India. *Nature,* 548 (7668), 407-412

Park, L., Mone, A., Price, J. C., Tzimas, D., Hirsh, J., Poles, M. A., Malter, L. & Chen, L. A. (2017). Perceptions of fecal microbiota transplantation for Clostridium difficile infection: factors that predict acceptance. *Annals of Gastroenterology,* 30(1)**,** 83-88.

Penders, J., Stobberingh, E. E., Savelkoul, P. H. M. & Wolffs, F. G. (2013). The human microbiome as a reservoir of antimicrobial resistance. *Frontiers in Microbiology,* 4**,** 7.

Pinto, U. M., Pappas, K. M. & Winans, S. C. (2012). The ABCs of plasmid replication and segregation. *Nature Reviews Microbiology,* 10(11)**,** 755-765.

Popowska, M., Rzeczycka, M., Miernik, A., Krawczyk-Balska, A., Walsh, F. & Duffy, B. (2012). Influence of Soil Use on Prevalence of Tetracycline, Streptomycin, and Erythromycin Resistance and Associated Resistance Genes. *Antimicrobial Agents and Chemotherapy,* 56(3)**,** 1434-1443.

Prescott S., (2017), History of medicine: Origin of the term microbiome and why it matters, *Human Microbiome Journal,* 4, 24-25

Price, M. N., Dehal, P. S. & Arkin, A. P. (2010). FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. *Plos One,* 5(3)**,** 10.

Priebe, S. D., & Lacks, S. A., (1989), Region of the streptococcal plasmid pMV158 required for conjugative mobilization, *Journal of Bacteriology*, *171*(9), 4778–4784.

Rattanachaikunsopon, P. & Phumkhachorn, P. (2009). Glass bead transformation method for gram-positive bacteria. *Brazilian Journal of Microbiology,* 40(4)**,** 923-926.

Selwyn, F. P., Cheng, S. L., Bammler, T. K., Prasad, B., Vrana, M., Klaassen, C. & Cui, J. Y. (2015). Developmental Regulation of Drug-Processing Genes in Livers of Germ-Free Mice. *Toxicological Sciences,*147(1)**,** 84-103.

Shen, N. & Clemente, J. C. (2015). Engineering the Microbiome: a Novel Approach to Immunotherapy for Allergic and Immune Diseases. *Current Allergy and Asthma Reports,* 15(7)**,** 10.

Shintani, M., Sanchez, Z. K. & Kimbara, K. (2015). Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology,*6**,** 16.

Somyoonsap P. Kitpreechavanich V. Pornbanlualap S. (2013), A sequence-specific nicking endonuclease from Streptomyces: purification, physical and catalytic properties. *ISRN biochemistry*, 2013, 287158

Sonnenberg, J. L. (2015). Microbiome Engineering. *Nature,* 518(7540)**,** S10-S10.

Stecher, B., Denzler, R., Maier, L., Bernet, F., Sanders, M. J., Pickard, D. J., Barthel, M., Westendorf, A. M., Krogfelt, K. A., Walker, A. W., Ackermann, M., Dobrindt, U., Thomson, N. R. & Hardt, W. D. (2012). Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proceedings of the National Academy of Sciences of the United States of America,*109(4)**,** 1269-1274.

Tatum E., Lederberg J., (1947), Gene Recombination in the Bacterium Escherichia coli. *Journal of bacteriology*, 53(6), 673-84

Tavaré S. (1986), Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *American Mathematical Society: Lectures on Mathematics in the Life Sciences*, 17, 57-86

Umesaki, Y. (2014). Use of gnotobiotic mice to identify and characterize key microbes responsible for the development of the intestinal immune system. *Proceedings of the Japan Academy Series B-Physical and Biological Sciences,* 90(9)**,** 313-332.

Van de Wiele, T., Vanhaecke, L., Boeckaert, C., Peru, K., Headley, J., Verstraete, W. & Siciliano, S. (2005). Human colon microbiota transform polycyclic aromatic hydrocarbons to estrogenic metabolites. *Environmental Health Perspectives,* 113(1)**,** 6-10.

Vaz-Moreira, I., Nunes, O. C. & Manaia, C. M. (2014). Bacterial diversity and antibiotic resistance in water habitats: searching the links with the human microbiome. *Fems Microbiology Reviews,* 38(4)**,** 761-778.

Wirth, T., Parker, N. & Yla-Herttuala, S. (2013). History of gene therapy. *Gene,* 525(2)**,** 162-169.

Wiley Journal of Medicine, http://www.abedia.com/wiley/vectors.php [Accessed Aug 2017]

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M. & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy,* 67(11)**,** 2640-2644.

Zheng, J., Peng, D., Ruan, L., & Sun, M., (2013), Evolution and dynamics of megaplasmids with genome sizes larger than 100 kb in the *Bacillus cereus* group, *BMC Evolutionary Biology*, *13*(262), 1-11.

Zhou, F. F. & Xu, Y. (2010). cBar: a computer program to distinguish plasmid-derived from chromosome-derived sequence fragments in metagenomics data. *Bioinformatics,* 26(16)**,** 2051-2052.