Mobile Mercury Resistance Transposons: Surveillance and Resistance Gene Cassette Variation in Wastewater

By

Alexander C. W. Pritchard *BSc* (Hons)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

December 2021

Acknowledgements

Choosing to continue my studies at the University of Nottingham for another four years has taken me on quite the journey. There have been many a peak and trough over the time spent trying to produce the work you are about to read. I have learned a lot, not just science but about myself. The most prominent of which is resilience.

First and foremost, my thanks go to my supervisor Dr Jon Hobman, whom without his trust in my ability, I do not think this thesis would resemble anything close to what it is now. From the beginning when I did not know which way was up, to the end of the writing process pestering him with many a question. I also thank you Jon for the opportunity to use your lab, so that I may have met such a wonderful group of people. This group was ever changing throughout my time here, but many a laugh, Star Wars joke or wild conspiracy theory was shared. Equally, this group has helped keep me somewhat sane and on the straight and narrow through dark times, when experiments failed repeatedly, and I came to my last tether.

I would also like to thank the BBSRC DTP for providing me with the opportunity to undertake this PhD programme, including a trip to Sydney as part of my placement. This allowed me to meet some wonderful people, Distinguished Professor Michael Gillings and his research group at Macquarie University were most accommodating.

I thank my parents, Rebecca Pritchard and Simon Goose, my sisters, Isabelle and Bryony, my grandparents, Marion and Derek Chapman, and Suzanne for your love, <u>PATIENCE</u> and unwavering support throughout. You have always had confidence in me, even when I did not.

Aside from these thanks, in some way I am thankful for the COVID-19 pandemic. It has highlighted the impact humans have had on this world and further brings to the forefront the growing challenges we as a planet face to tackle pathogens, be they viral, bacterial, or degenerative.

Abstract

In a wastewater environment, mercury resistance and other antimicrobial metal resistance genes have been observed despite their lack of use clinically. The hypotheses explored whether the change in populations to a rural wastewater treatment plant affects the abundance of Tn*21* and similarly identify potential co–occurrence antimicrobial resistance genes carried within or alongside Tn*21* and Tn*21*-like transposable elements. Finally, Tn*21* is known to be carried by a wide range of Gram-negative bacteria, however without being able to cross-link Tn*21* to the host it is not possible to identify in large scale samples which organisms may in fact carry the mobile element within an environmental sample. Results showed that large-scale population changes impacted the abundance of Tn*21* and the carriage of co-occurrent resistance genes. Wastewater treatment processing was also shown to reduce diversity of Tn*21* gene cassette arrays of the class I integron and therefore not remove the presence of antimicrobial resistance genes disseminating into the environment. The studies highlight the need for intervention within the wastewater treatment process.

Table of Contents

Tables

Figures

XIV

Abbreviations

A. salmonicida	*Aeromonas salmonicida*
Ag	Silver
Ag (I)	Silver (I) ion
AMR	Antimicrobial Resistance
Amp	Ampicillin
AST	Antibiotic Sensitivity Testing
ARG	Antibiotic resistance Gene
BLAST/ BLASTN	Basic Local Alignment Search Tool
b	Base(s)
bp	Base Pair(s)
Cam/Cat/Cm	Chloramphenicol
Cd (II)	Cadmium (II) ion
CFU	Colony-Forming Unit
C. freundii	*Citrobacter freundii*
Cu	Copper
Cu (I)	Copper (I) ion
Cu (II)	Copper (II) ion
dH_2O	Deionised Water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
EAggEC	Entero Aggeragrative *Escherichia coli*
E. coli	*Escherichia coli*
E. cloacae	*Enterobacter cloacae*
ESBL	Extended Spectrum Beta Lactamase
fmol	Femtomoles
g	Gram
H_2O	Water
Hg	Mercury
$HgCl_2$	Mercury Chloride
Hg (I)	Mercury (I) ion
Hg (II)	Mercury (II) ion
IR	Inverted Repeat
kb	Kilo Base(s)
KCl	Potassium Chloride
K. pneumoniae	*Klebsiella pneumoniae*
km	Kilometer
L	Litre(s)
LB	Luria-Bertani
mL	Millilitre(s)
mM	Millimolar
M	Molar

m	Meter
Mb	Megabase(s)
$MgCl_2$	Magnesium Chloride
MGE	Mobile Genetic Element
$MgSO_4$	Magnesium Sulfate
min	Minute(s)
MRD	Maximum Recovery Diluent
mV	Millivolts
NaCl	Sodium Chloride
ng	Nanogram
nt	Nucleotide(s)
µg	Microgram
µL	Microliter
µM	Micromolar
nM	Nanomolar
NADP+	Nicotinamide Adenine Dinucleotide Phosphate
NADPH	Reduced Nicotinamide Adenine Dinucleotide Phosphate
OD_{600}	Optical Density at 600 nm
ORF	Open Reading Frame
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
PFGE	Pulse Field Gel Electrophoresis
pmol	Picomoles
psi	Pounds Per Square Inch
QAC	Quaternary Ammonium Compound
^{r}	Resistant
RAST	Rapid Annotation Subsystem Technology
RND	Efflux-Resistance Nodulation
rpm	Revolutions per minute
RNA	Ribonucleic Acid
s	Second(s)
SDS	Sodium Dodecyl Sulfate
SGI	*Salmonella* Genomic Island
spp.	Species
Sul	Sulphonamide
TEs	Transposable Element(s)
Tet/TET/T	Tetracycline
Tris	Tris(hydroxymethyl)aminomethane
UTI	Urinary Tract Infection
UV	Ultraviolet
V	Volts
w/w	Concentration Weight per Weight
w/v	Concentration Weight per Volume

v/v	Concentration Volume per Volume
x g	Times Gravity
Zn	Zinc
Zn (II)	Zinc (II) ion
°C	Degrees Centigrade

Chapter 1: Introduction

1.1   An Introduction to Antimicrobials

According to the US Environmental Protection Agency, an antimicrobial is considered as an agent which is capable of killing or inactivating a microorganism, be it a bacteria, fungus, virus or parasite (US EPA, 2020). This definition covers a large range of compounds, including antibiotics, antifungals, biocides, antivirals, some metals and pesticides. Additionally, some antimicrobials have been shown to work synergistically in combination with other antimicrobials or antibiotics to enhance their potency (Kurenbach *et al.*, 2017). Antimicrobials, such as antimicrobial metals have existed for as long as microorganisms and maybe even before microorganisms. During the Great Oxidation Event, approximately 2.5 billion years ago, an increase in atmospheric oxygen caused many metal oxide compounds to become bioavailable in various oxidised forms, which became widespread throughout the biosphere (Rye and Holland, 1998). Bacteria were able to utilise some of these oxidised metals such as, iron, copper, manganese and zinc as cofactors for enzymatic redox processes such as aerobic respiration, which is more efficient than anaerobic respiration. These metals later became essential metals for almost all bacteria. Over time, homeostatic mechanisms including transport proteins and ion pumps, evolved to help maintain a stable concentration of these metal ions because when the concentrations became too high, they are toxic to the cells. Other bioavailable metals such as gold, mercury and silver have no known biological benefit to bacteria and are toxic even at very low concentrations (Hobman and Crossman, 2015).

1.2 Antibiotics

Over time, as cells evolved and propagated into the environment, natural selection occurred in varying environments within the biosphere. This caused competition for space and limited resources, such as nutrients and essential molecules needed for growth, respiration and proliferation. This competition led to the development of secondary organic metabolites, which accelerated into the production of some of the antibiotics seen today, modified by human synthetic production (Waglechner, McArthur and Wright, 2019). Some antibiotics, such as tetracycline, have been found in traces in Late Roman era tomb in Egypt (c. 350 AD) (Cook, Molto and Anderson, 1989). The first documented use of antimicrobial compounds was approximately 2000 years ago and there was also documented use of these compounds in 1550 BCE, using mouldy bread as a poultice on open wounds and medicinal soil were included in the list of remedies on the Eber's papyrus (Haas, 1999; Hutchings, Truman and Wilkinson, 2019). Penicillin, produced by *Penicillium chrysogenum*, was discovered in 1928 by Alexander Fleming and treptomycin was discovered in 1943 by Selman Waksman. Streptomycin is produced by the soil bacteria *Streptomyces spp.,* a genus of bacteria part of the *Actinomycetaceae* family, capable of specialising their cells to act like a tissue with the use of quorum sensing and colony motility when in the presence of other bacterial species (Jones and Elliot, 2017). Tetracyclines, also produced naturally by the same soil organisms, *Streptomyces spp.*, were first reported in 1948 (Nelson and Levy, 2011). Throughout the rest of the 20th century there was an increase in the reporting of novel classes of antibiotics until the later part of the century, where the discovery rate of novel antibiotics declined (Davies,

2006). Reports of resistance to these novel antibiotics and bacterial strains with resistance to multiple classes of antibiotics and antimicrobial metals followed (Figure 1.1).
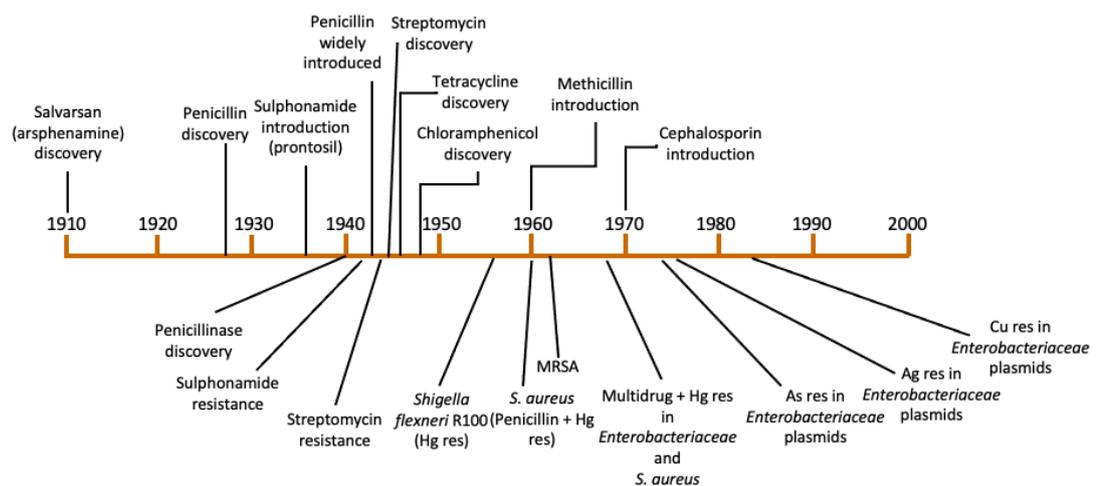
Figure 1.1 Timeline of antibiotic discovery and resistances reported.

Adapted from (Hobman, 2017). Key discoveries of antimicrobials are displayed above the timeline and key resistances to antimicrobials and antimicrobial metals are detailed below the timeline.

Due to the widespread production and use of antibiotics by humans, there has been an increase in the abundance and concentrations of antibiotics within the environment (Knapp *et al.*, 2010). Thanks to anthropogenic pollution, the presence of these antibiotics, biocides and other compounds such as oestrogen is well documented to have disseminated into all environments in low concentrations (Berglund, 2015). One study demonstrated that since the first reporting of antibiotic resistance in the 1940s there was an exponential relative increase in abundance of resistance genes to various antibiotic classes, including ß-lactamases, tetracyclines and erythromycins in the same locations over a 60-year period (Knapp *et al.*, 2010).

4

1.3   Biocides

Alongside the development and use of antibiotics in the 20th century, biocides were also developed. Like antibiotics, there are many classes of biocides which have different modes of action and are used for disinfection across industry, agriculture, healthcare and domestic settings. The first documented use of a biocide was a chlorinated handwash in the 19th century, since then biocides have become an integral part of infection control alongside the use of antibiotics (Jones and Joshi, 2021). There are many types of biocides including: quaternary ammonium compounds (QACs), biguanides, chlorine releasing agents and peroxygens (Table 1.1). The main advantage of using biocides over antibiotics is their broad range of uses. An example of this is the compound chlorhexidine. As with most biocides, the concentration of this biguanide dictates the efficacy and toxicity for its use in each situation it is deployed: in surface disinfection 0.5-4% (v/v), as an antiseptic 0.02-4% (v/v) and preservation lower concentrations between 0.0025-0.01% (v/v) are used (Maillard, 2005). Further uses of chlorhexidine include mouthwash and prescribed skin lotions. Similarly, biocides such as chlorine releasing agents and ozone are used daily to supply the human population with potable drinking water (Li, Zhu and Ni, 2011).

Biocide	Mode of Action
QACs	Destabilisation of cell membrane causing lysis
Biguanides	Disruption of cytoplasmic membranes membrane
Chlorine Releasing Agents	Oxidative damage, oxidises sulfhydryl groups preventing DNA and protein synthesis
Peroxygens	Generation of reactive oxygen species extra- and intracellularly
Ozone	Cell membrane oxidation of phospholipids and lipoproteins and lysis
Bisphenols	Cell membrane protein target inhibition preventing cell membrane synthesis
Acids	Denatures proteins in the cell membrane and in the cell, disrupts cell permeability
Bases	Bond breakage of proteins, such as disulfide bonds between amino acids

Table 1.1 Type of biocide and its corresponding mode of action.

Each type of biocide has a similar mode of action, causing oxidative stress or disruption of cellular membranes, ultimately resulting in cell death, however some biocides are application specific.

Currently, the increased usage of biocides and other antimicrobials across multiple disciplines caused debates as to whether biocide usage should change in the future. The indiscriminate use of low concentrations of compounds used as biocides, such as phenols, cationic compounds (eg. QACs) has raised concerns about increased resistance leading some professionals to conclude that their impact may be detrimental in the long term (Levy, 2001; Daschner and Schuster, 2004). This is due to the spread of these compounds at sub-lethal concentrations into the environment, creating selective pressures to drive the development of resistance to biocides. This stance is not new, the first report pointing out the limited future of biocides came in the 1960s (Ayliffe *et al.*, 1969). However, diverging opinions lead to the development of more biocides, each with their own efficacy/toxicity balance, temperature requirements. In the mid 1990s there were a limited number of biocides, and now over 700 biocides are commercially available (Levy, 2001).

1.3.1 Quaternary Ammonium Compounds (QACs)

One key class of biocides developed over the last century was QACs. QACs are used in food manufacture for cleaning due to their excellent non-tainting properties and industrially as phase transfer catalysts to help dissolve immiscible compounds and used domestically in fabric softeners and for other washing purposes (Fazlara and Ekhtelat, 2012). In fabric washing, QACs aid the softening of fabrics but also remove stains. QACs mode of action is what makes them so effective, QACs bind to phospholipids and proteins irreversibly, which prevents cell walls from absorbing molecules, often causing leakage of cell contents, and therefore killing the cells (Ioannou, Hanlon and Denyer, 2007; Gerba, 2015). QACs may also cause the compaction of DNA by binding the phosphate backbone, into a state where it can no longer be transcribed for protein synthesis (Zinchenko *et al.*, 2004). These effects can also be seen on viruses containing a lipid envelope. The virus can no longer enter a cell as it relies on the envelope for adsorption to the host cell. Interestingly, the efficacy of quaternary ammonium compounds varies greatly depending on the organism and QAC combination, suggesting some QACs used could be host specific rather than targeting all potential microbial contaminants (Eterpi, McDonnell and Thomas, 2009). Whilst this was identified as a problem for viral particles, it is also likely the case for other microorganisms. As with many of the mentioned uses of QACs, large volumes of water are often used alongside these compounds to dispel them or help with the cleaning process in the case of fabric softeners and are then drained into wastewater for treatment. This provides an environment where

biocides are at sub-lethal concentrations providing the right environment to select for resistance to them (Daschner and Schuster, 2004).

1.4   Antimicrobial Metals

Historically antimicrobial metals such as mercury, silver, copper and zinc have been used in ointments and salves to inhibit the growth of microorganisms from wound infections. Zinc compounds such as zinc oxide are still also used as a mild antiseptic to treat skin irritation and even as an anti-diarrhoeal animal feed (Hobman and Crossman, 2015). In the Roman era, 2000 years ago, Pliny the Elder documented the use of zinc compounds: 'calamina' were used in complex medicines to treat eye and dermatological infections, evidence of these calamina tablets were discovered in a Roman shipwreck (Giachi *et al.*, 2013). Around the same time (c. 202BCE) in Chinese medicine, gold was known to have been used in medicine to treat various conditions from, treating wounds, removing mercury from ears, treating smallpox, and even curing sore eyes (Huaizhi and Yuantao, 2001). Salvarsan (arsphenamine) first synthesised in 1908 a compound containing arsenic was used to treat a large range of ailments, from liver damage to syphilis and was injected straight to the affected area. Prior to this, mercury-based compounds were commonplace in medical treatment. Since then, many of these compounds, such as arsenic and mercury, have been shown to be detrimental to human health and are no longer used. However, some of these antimicrobial metals are still used today, such as silver nanoparticles and silver salts in wound dressings (Durán *et al.*, 2016). Gold, platinum and their compounds such as cis-platin, whilst not used as an antimicrobial in modern

medicine are still used in pharmacology as novel agents in radiotherapy for cancer treatment. Interest in silver has also been reinvigorated. This includes new uses for silver nanoparticles as therapeutics, cosmetics, refined antimicrobials, antifungals, renewable energy, nuclear reactors, production of electronics and even in textiles to increase freshness of clothing (Zhang *et al.*, 2016).

The development of increased occurrence of bacterial resistance to these metals has been observed. It is thought metal resistance gene cassettes were present in bacteria since the first need to survive in such an environment with extreme selection pressures of antimicrobial metals millions of years ago when the biosphere became mostly oxygenated, which in turn caused the metals to become oxidised and therefore bioavailable (Barkay *et al.*, 2010).

For essential metals such as zinc and copper, molecular systems have developed to help moderate the availability of these ions within the cells for essential processes. Although the use of metals as an antimicrobial has been noted throughout history, the exact specifics of the toxicity of these ions is less well understood. It is widely reported that the ions interact through their ability to form ligands with functional groups like thiols, causing inactivation of functional groups and displacing other metal cofactors. There have been attempts to categorise antimicrobial metals by toxicity and valency with various descriptions such as heavy metals or toxic metals used to try to describe toxicity (Duffus, 2002; Hodson, 2004). However, to define such antimicrobial metals, the understanding of Lewis acids is required.

1.4.1.1 Lewis Acids

The definition of a Lewis acid is a chemical species capable of accepting a pair of electrons and a Lewis base is a chemical species capable of donating a pair of electrons to complete their outer shell of electrons. From this definition of Lewis acids there are two sub-definitions: a hard acid, an acid species with a small atomic radius and therefore higher affinity for accepting electrons, and a soft acid, with larger atomic radii and a weaker affinity for electrons. As such, the biologically important divalent transition metals can be ordered by strength of affinity as a proxy for toxicity and used to predict the selectivity of metal ions to organic donor ligands (Ralph Pearson, 1963; Hobman and Crossman, 2015).

From these definitions, many metals can be described as soft Lewis acids, where some of the most toxic metals Hg (II), Cu (I), Ag (I), and Cd (II) are included. Toxic metals are found within the transition elements section of the Periodic Table. As they are soft acids, they are therefore much harder to displace, which explains why irreversible ligands may form and cell damage and cellular process interruption may occur. This is due to the lower affinity to electrons in the outer orbitals of the electronic arrangement around the atomic nuclei making them more readily polarised to form covalent bonds. A bond stronger than ionic bonds often formed by those that are hard Lewis acids.

1.4.1.2 Copper Toxicity

Copper, although a toxic metal at high concentrations, has gained biological importance due to its ability to interact with proteins and enzymes as a cofactor,

acting as an intermediate electron donor and acceptor in electron transport chains in key metabolic processes due to its ability to adopt multiple valence states. However, when the environmental concentrations exceed a certain level of Cu (II), it becomes toxic to the cells. For *E. coli*, concentrations above 1 mM of Cu (II) are considered toxic in simple growth media (Nies, 1999). In fact, the other valence state of copper, Cu (I) or the cuprous ion, is considerably more toxic than Cu (II) but is commonly higher in abundance due to the redox reactions taking place within cells (Solioz, 2016). Due to the development of the biological role for Cu (II), there exists a homeostatic mechanism for the ion to enter the cell and maintain a certain tolerable concentration range in which the cell may function optimally. When above the tolerable concentration, the Cu (II) ion is capable of interacting with many organic components in Fenton-like reactions, forming superoxides, hydroxyl free radicals and hydrogen peroxide, all capable of causing further damage to the cell (Grass, Rensing and Solioz, 2011). In higher local Cu (II) concentrations or even on solid surfaces, bacterial cell membranes can be ruptured, and reactive oxygen species may form causing DNA damage due to interaction with the phosphate backbone. The damage caused by reactive oxygen species formation promoted by Cu has been seen as a defence mechanism for mammalian immune cells, which release a copper promoted oxidative burst containing reactive oxygen species into phagosomes (Grass, Rensing and Solioz, 2011). However, whilst this has been demonstrated to be extremely effective, some invading microorganisms have developed resistance to this in the ongoing 'arms race' between competing taxa (German, Doyscher and Rensing, 2013; Gillings, 2014).

1.4.1.3 Silver Toxicity

In the environment, silver often exists in a monovalent cationic form of Ag (I). Humans are thought to have been using silver as an antimicrobial for approximately up to 7,000 years. Its first documented use was by Alexander the Great 2,000 years ago, as a means of making water potable (Burrell, 2003; Silver, Phung and Silver, 2006). There is no known essential requirement within cells for Ag and it is in fact the second most toxic metal to cells (Nies, 1999). Ag (I) is considered a soft acid and has a high affinity for thiol and phosphate ligands. Its ability to interact with thiol groups means disruption can be caused to membrane proteins (Holt and Bard, 2005). Once inside the cell, the Ag (I) ion is capable of interacting with functional groups of enzymes, resulting in interruption of respiratory processes by disrupting proton motive forces, causing membrane leakage and free radical formation, in turn slowing down cell metabolism and even causing cell death even at low concentrations (Lok *et al.*, 2006; Gordon *et al.*, 2010; Hobman and Crossman, 2015). This has even been documented to have effect irrespective of the presence of an oxygenated environment by creating reactive oxygen species within the bacterial cells, although the presence of aerobic conditions did enhance the bactericidal nature of Ag (I) (Park *et al.*, 2009).

1.4.1.4 Mercury Toxicity

Mercury is not just toxic to humans but to bacteria such as *E. coli*. It is the most toxic metal to life (Nies, 1999). Behaving in a similar manner to the Ag (I) ion, it has the

potential to cause oxidative stress due to its high affinity for sulfhydryl ligands in amino acids, causing loss of protein structure and function (Broussard *et al.*, 2002). The Hg-induced conformational changes which take place in proteins to interrupt cellular processes such as iron homeostasis, inhibit sensitive proteins within the organisms and generate reactive oxygen species, which may cause death. Mercury can also interfere with DNA by inactivating RecA in the same manner, preventing the DNA repair mechanism (Liebert, Hall and Summers, 1999; Nies, 1999; Lee and Singleton, 2004). Interestingly, mercury metal (Hg) is not toxic to cells but the mercuric ion (Hg (II)) is and so is the mercurous ion (Hg (I)) although this is less toxic. Due to its bioavailability, the Hg (II) ion builds up inside the cell and causes interruption of cellular action (Mirzaei, Kafilzadeh and Kargar, 2008). Not only is its mechanism of toxicity unusual, but so is the resistance mechanism to remove inorganic mercuric ions from the cell.

Mercury, the post-transition metal, is also toxic in organic forms, known as organomercurials. Such forms of organomercury include the methylmercury, dimethylmercury and other forms of alkylmercury. The most toxic of which is dimethylmercury. Such methylated forms of mercury can be produced naturally by anaerobic bacteria and can be bioaccumulated and biomagnified in food chains (Rimmer *et al.*, 2010). Organomercurials were previously known to be used as antiseptics, diuretics, seed dressings and spermicides (Craig, 1982).

1.5   An Introduction to Antimicrobial Resistance

The World Health Organisation defines antimicrobial resistance (AMR) as a change that happens to microorganisms when they are exposed to antimicrobials resulting in their survival (WHO, 2018). Resistance existed long before the discovery of chemotherapeutics, such as Salvarsan, by Ehrlich in 1908 and modern antibiotics, such as penicillin by Fleming in the 1928 (Ehrlich, 1908; Fleming, 2001; Kholodii *et al.*, 2003). Bacteria have naturally developed resistance as a mechanism to try and outcompete each other for resources or to occupy ecological niches that may be too toxic for other microorganisms, meaning resistance to these antimicrobials has evolved to compete with those cells producing antimicrobials. These resistance mechanisms vary, from simple efflux pumps, DNA protection mechanisms, and inactivating enzymes, to ribosomal protection mechanisms (Table 1.2) (Nikaido, 2009). In the 'pre-antibiotic era' resistance to antibiotics was less prevalent in human pathogens. This can be seen through examining culture collections from such times. One such collection that exists is the Murray collection, *Enterobacteriaceae* isolates which were collected from 1917 and 1954 from various geographical locations (Datta and Hughes, 1983). The collection helped identify the core machinery needed for plasmid transfer between strains with isolates ranging in plasmid number, between zero and seven, which have similarities to plasmid core genes seen today. What varies mostly between this collection and similar bacteria isolated now, however, is the presence of antibiotic resistance genes (ARGs) within the strains (Datta and Hughes, 1983; Baker *et al.*, 2015). Current clinically isolated strains may now be more likely to possess 3 or more ARGs conferring resistance to different classes of

antibiotics. These are known as multidrug resistant microorganisms (Kumar and Khan, 2015).

Class [a] (Example)	Introduced Clinically	Mode of Action	Mechanism of Resistance
Antibiotics produced from *Actinomycetes*			
Aminoglycosides (Kanamycin A)	1946	Protein synthesis: 30S ribosomal subunit	Mutation of binding site, increased efflux or modification
Tetracyclines (Tetracycline)	1948	Protein synthesis: 30S ribosomal subunit	Efflux
Amphenicols (Chloramphenicol)	1949	Protein synthesis: 50S ribosomal subunit	Efflux
Macrolides (Azithromycin)	1952	Protein synthesis: 50S ribosomal subunit	Methylation of binding site (*erm*)
Tuberactinomycin (Viomycin)	1953	Protein synthesis: 30S and 50S ribosomal subunits (binds to the inter subunit bridge B2a)	Antibiotic inactivation
Glycopeptides (Vancomycin)	1958	Cell wall synthesis: D-Ala-D-Ala termini of lipid II	Alteration of target
Lincosamides (Clindamycin)	1963	Protein synthesis: 50S ribosomal subunit	Efflux or mutation of 23S ribosomal subunit
Ansamycins (Rifampicin)	1963	Nucleic acid synthesis: RNA polymerase	Mutation of beta subunit of RNA polymerase and inactivation
Cycloserines (Seromycin)	1964	Cell wall synthesis: inhibition of alanine racemase and D-alanine-D-alanine ligase	Antibiotic target alteration
Streptogramins (Pristinamycin)	1965	Protein synthesis: 50S ribosomal subunit	Efflux
Phosphonate (Fosfomycin)	1971	Cell wall synthesis: MurA (UDP-GlcNAc-3-enolpyruvyltransferase)	Inactivation
Carbapenems (Meropenem)	1985	Cell wall synthesis: penicillin-binding proteins	Inactivation using carbapenemase
Lipopeptides (Daptomycin)	2003	Cell wall: cell membrane disruption	Efflux
Lipiarmycins (Fidaxomicin)	2011	Nucleic acid synthesis: RNA polymerase	Mutation of RNA polymerase
Antibiotics produced from other bacteria			
Polypeptides (Gramicidin A)	1941	Cell wall: increase the permeability of the bacterial cell membrane, by forming ion channels	-
Bacitracin (Bacitracin A)	1948	Cell wall synthesis: inhibition of	Alteration of antibiotic target

		dephosphorylation of C_{55}-isoprenyl pyrophosphate	
Polymyxins (Colistin)	1959	Cell wall: cell membrane disruption	Alteration of antibiotic target
Mupirocin Mupirocin)	1985	Protein synthesis: isoleucyl t-RNA synthetase	Mutation of isoleucyl t-RNA
Monobactams (Aztreonam)	1986	Cell wall synthesis: penicillin-binding proteins	Hydrolysis by beta-lactamase, alteration of penicillin-binding proteins and decreased permeability
Antibiotics produced from Fungi			
Penicillins (Amoxicillin)	1943	Cell wall synthesis: penicillin-binding proteins	Production of penicillin-binding proteins
Fusidic acid (Fuscidic acid)	1962	Protein synthesis: elongation factor G	Inactivation of antibiotic
Cephalosporins (Cefotaxime)	1964	Cell wall synthesis: penicillin-binding proteins	Production of penicillin-binding proteins
Pleuromutilins (Retapamulin)	2007	Protein synthesis: 50S ribosomal subunit	Mutation of 23S ribosomal subunit
Synthetic Antibiotics			
Sulphonamides (Sulfamethoxazole)	1936	Inhibition of folate pathway	Horizontal transfer of folP
Salicylatese (4-Aminosalicylic acid)	1943	Inhibition of folate pathway	Mutation of thyA
Sulfones (Dapsone)	1945	Inhibition of folate pathway	Mutation of folP
Pyridinamides (Isoniazid)	1952	Cell wall: inhibition of mycolic acid synthesis	Efflux or target alteration
Nitrofurans (Nitrofurantoin)	1953	DNA synthesis: DNA damage	Mutation of target site
Azoles (Metronidazole)	1960	DNA synthesis: DNA damage	Efflux
(Fluoro)quinolones (Ciprofloxacin)	1962	DNA synthesis: inhibition of DNA gyrase, and topoisomerase IV	Mutation of gyrA or increased efflux
Diaminopyrimidines (Trimethoprim)	1962	Inhibition of folate pathway	Target replacement
Ethambutol (Ethambutol)	1962	Cell wall: arabinosyl transferase inhibition	Mutation of target protein
Thioamides (Ethionamide)	1965	Cell wall: inhibition of mycolic acid synthesis	Mutation of target protein

Phenazines (Clofazimine)	1969	DNA synthesis: binds to guanine bases	Mutation of *gyrA*
Oxazolidinones (Linezolid)	2000	Protein synthesis: 50S ribosomal subunit	Mutation or methylation of oxazolidinone binding sites
Diarylquinolines (Bedaquiline)	2012	ATP synthesis: proton pump inhibition	-

Table 1.2 A broad overview of each class of antibiotic, mode of action and known resistance mechanisms.

[a] Classes are defined by origin, structure and/or mechanism of action, which distinguishes between bacitracin, colistin and daptomycin, for example. [b] Salicylic acids are naturally occurring, but this was not the source of this class of antibiotic. Adapted from (Hutchings, Truman and Wilkinson, 2019).

Since the start of large-scale commercial production of antimicrobials, humans have overused them, which has resulted in an acceleration in the occurrence of bacterial multidrug resistance in the environment and in a clinical setting, which ultimately threatens food security, human and animal health (Knapp *et al.*, 2010). Through the use of antibiotics in this manner, the process of natural selection has accelerated the emergence and dissemination of MDR bacteria, causing socioeconomic impact globally on many different industries (Ojo *et al.*, 2008).

These antimicrobial selection pressures as well as the use of refined metals and biocides in everyday life, such as amalgam dental fillings and quaternary ammonium compounds, may link metal resistance to these multidrug resistances through co-selection and co-occurrence (Hobman and Crossman, 2015). This has been seen in research where mobile genetic elements such as transposons and plasmids, capable of horizontal gene transfer, harbour metal resistance as well as multidrug resistance due to co-selection and co-occurrence, where both metal toxicity and certain

antibiotics were selective pressures and the organism has therefore developed resistance to both as a result (Madigan *et al.*, 2015; Pal *et al.*, 2017).

1.5.1   Mercury Resistance

In the case of copper, silver and zinc, a simple efflux system works to remove a high concentration of the toxic metal ions out of the bacterial cells (Figure 1.2). For mercury however, instead of using a general or metal-ion specific efflux system to remove the mercuric ion, a specific resistance system is used transport Hg (II) into the cytoplasm where the Hg (II) ion is enzymatically reduced to Hg by the enzyme mercuric reductase (MerA). This resistance mechanism is unique amongst bacterial metal ion resistances. In Gram-negative bacteria (Figure 1.3) the Hg (II) ion is thought to enter the periplasm via porins. MerP binds the Hg (II) ion, which is transported to the inner membrane protein MerT, which also helps protect the cell from Hg (II) toxicity (Morby, Hobman and Brown, 1995; Pal *et al.*, 2017). In addition to MerT, MerC is another inner membrane spanning protein which also acts as a Hg (II) and organomercurial importer into cells. MerC is carried by Tn*21* but no other mercury resistance operons such as Tn*501* (Sone *et al.*, 2013), and is indicative of a Tn*21*-family *mer* operon. After being imported into the cytoplasm, MerA (mercuric reductase) then catalyses the reduction of Hg (II) to Hg by oxidising NADPH to $NADP^+$ (Lian *et al.*, 2014). The reduced Hg (0) is volatile at room temperature and pressure and therefore leaves the cell by volatilisation.
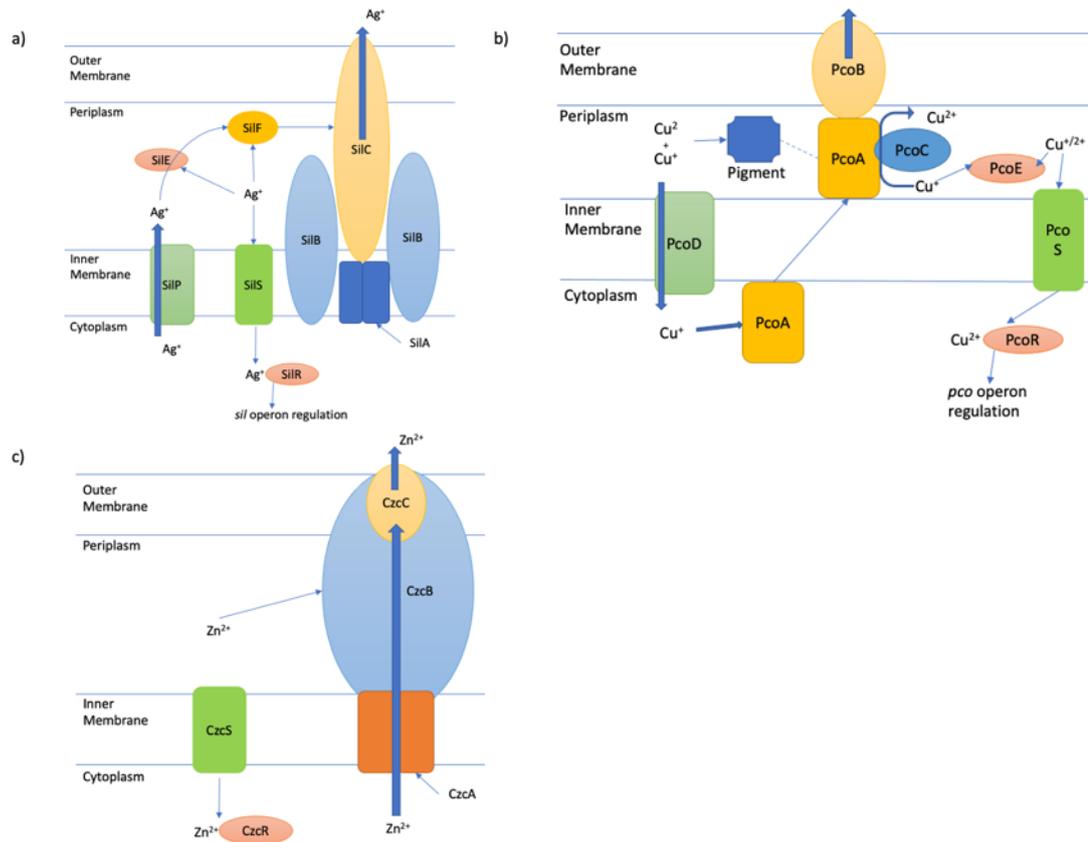
Figure 1.2 Resistance mechanisms to other antimicrobial metals a) silver b) copper and c) zinc.

All three metal ion resistance mechanisms use an ABC transporter and chaperones to export their respective toxic metal ions from within the cell cytoplasm or the periplasm. a) Silver resistance is achieved when Ag (I) ions enter the cell's periplasm, where they are detected by SilS, part of the two-component silver-responsive transcriptional regulation system, SilRS. SilRS regulates its own expression by SilR and SilE phosphorylation. SilRS also regulates expression of silCBA, silF and silP. SilCBA is a three-part resistance-nodulation-division (RND) effluxer. SilF is believed to act as an Ag (I) chaperone in the periplasm and SilP exports Ag (I) from the cytoplasm to the periplasm through the hydrolysis of ATP. SilE function is less well characterised but believed to act as a homologue to PcoE. b) The *pco* system varies to the other two systems as PcoABC does not span the inner and outer membranes, but rather PcoA acts as a chaperone to transport Cu (I) into the periplasm before it is exported from the cell completely. Cu (II) and Cu (I) enters the cell's periplasm through porins in the outer membrane and is detected by the two-component sensor kinase regulator system PcoRS, regulates the expression of pcoABCDRS operon and pcoE, from its own promoter, when PcoR is phosphorylated. PcoE acts as a copper binding protein in a similar manner to a siderophore, acting as a 'sponge' protein. PcoA, moves between the inner membrane and periplasm when loaded with Cu (I) from PcoD, an inner membrane spanning protein that imports Cu (I) from the periplasm to the cytoplasm. PcoA oxidises toxic Cu (I) to Cu (II) bound to PcoC the periplasm. PcoB is an outer membrane protein which helps export copper. c) Zinc resistance is achieved through the detection of increased zinc concentration within the periplasm is monitored by CzcS, the sensor protein of the CzcRS two-component regulatory

20

system. This results in CzcR activates its own transcription and the transcription of CzcCBA efflux pump proteins. Zn(II) is then exported from cellular cytoplasm and periplasm out of the cell using a proton motive force to power it. Panels a and b are adapted from figures in Hobman and Crossman (2015). Panel c is adapted from Ducret et al., (2020).
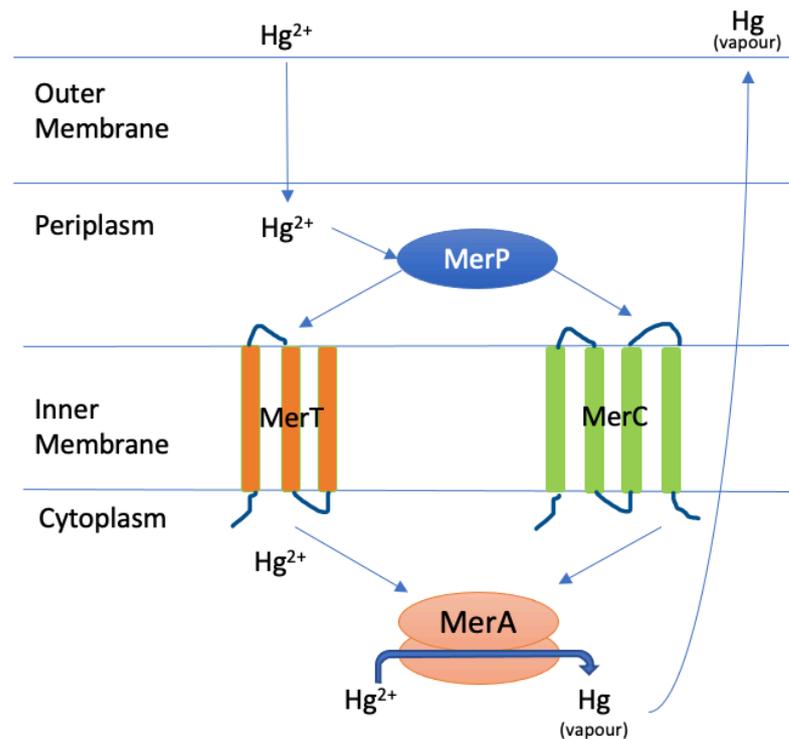
Figure 1.3 Model of the Gram-negative bacterial mercuric ion resistance mechanism from Tn*21*.

Adapted from Hobman & Crossman (2015). Hg (II) enters the periplasm via porins in the outer membrane, where they bind to cysteine residues in MerP. Hg (II) is then transferred to membrane located MerT and/or alternate importers MerC or MerF (not shown). Mercuric ions are transferred via paired cysteine residues in MerT and emerge in the cytoplasm. MerA then reduces Hg (II) to Hg0. This reduced form of Hg is volatile at room temperature and pressure and can leave the cell as mercury vapour. Expression of the mercury resistance structural genes is regulated by homodimeric MerR (not shown). MerD acts as a co-regulator of expression. Finally, MerE is reported to import organomercurial ions (not shown) (Sone *et al*., 2013).

1.5.1.1 Regulation of Mercury Resistance Genes

The *mer*, mercury resistance system is an operon which is regulated by MerR which is both a transcriptional repressor or activator, in the absence or presence of mercury, respectively. This system acts via steric hindrance where in the absence of mercury, MerR homodimers regulate access of RNA polymerase to the promoter region by binding to the *mer* operator motif between the -35 and -10 RNA polymerase recognition sites upstream of the transcriptional start site. This system is very tightly regulated as the -10 and -35 sites overlap *merR* gene so when MerR is bound no more MerR is transcribed, acting as an autoregulator. In the presence of mercury, Hg (II) binds MerR. The metalated MerR homodimers then undergo conformational change, bending the promoter region to allow underwinding of the operator DNA allowing RNA polymerase to bind the transcriptional start site (Figure 1.4) (Ansari, Chael and O'Halloran, 1992) .
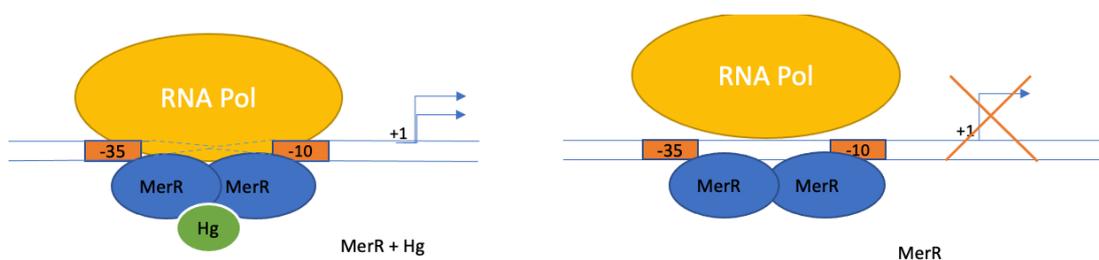
Figure 1.4 The interaction between MerR dimers and Hg (II) binding the -35 and -10 regions of the promoter.

Left: In the presence of Hg (II), MerR dimers regulating the winding of the -35 and -10 region of the promoter which is recognised by RNA polymerase. Upon binding the Hg (II), the promoter is underwound due to conformational change of the MerR homodimer to allow RNA polymerase to bind and transcribe the *mer* operon. Right: without Hg (II) the MerR homodimer prevents the binding of RNA polymerase to the promoter region so no transcription may occur.

Genes, such as these, encoding resistance are often capable of being transferred horizontally between organisms via numerous methods. In the case of mercury resistance, it is present in many forms, in the enteroaggregative *E. coli* strain 042 it is found chromosomally, or in the form of a mobile element like a transposon (Tn) such as the transposon Tn*21*, part of the Tn*3* family (Figure 1.5).
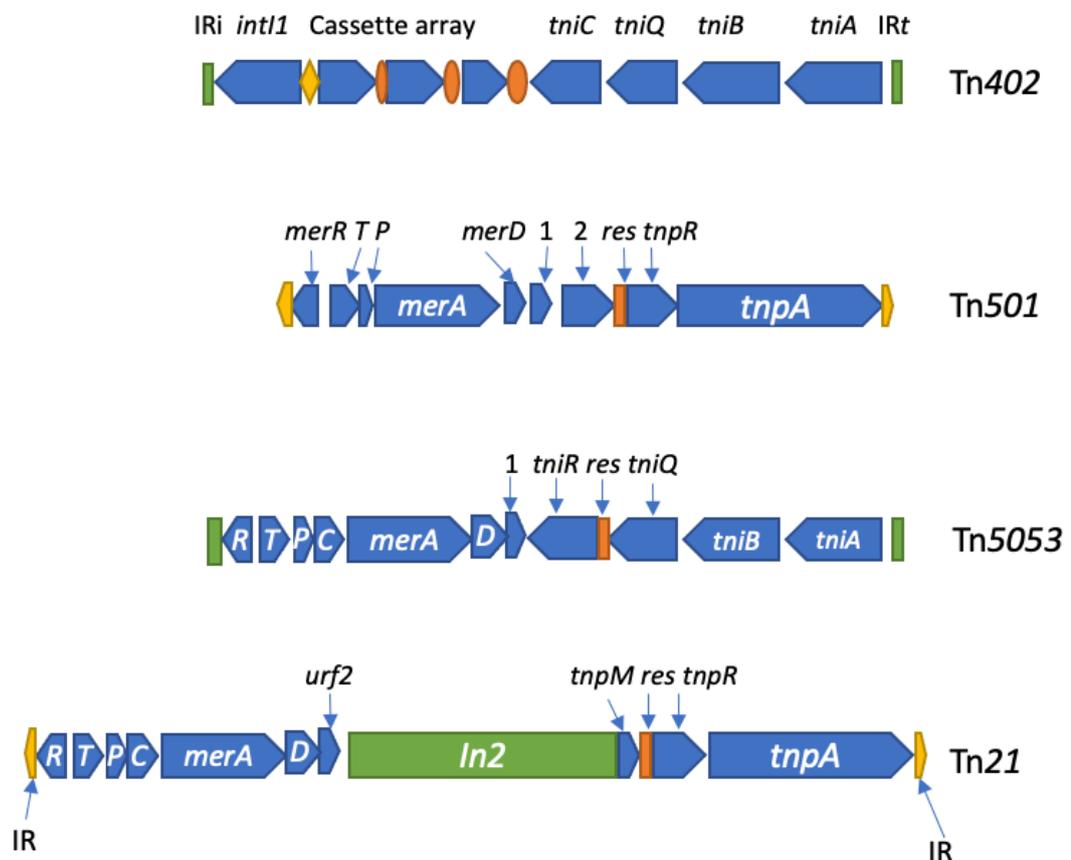
Figure 1.5 Tn*402*, the derivative of mobile class I integrons, containing a *tni* module.

Mercury resistance transposons Tn*501*, which does not contain *merC* in the *mer* operon, Tn*21*, containing the full *mer* operon and immobile class I integron and the reported Tn*21* ancestor Tn*5053* isolated from Russian permafrost.

1.6   Horizontal Gene Transfer

At its simplest, horizontal gene transfer (HGT) is the transmission of coding DNA from one organism to another without the use of reproduction. There exist three mechanisms for HGT: transformation, transduction and conjugation. The HGT process helps drive the natural development and evolutionary progression of microbial communities with the purpose of overcoming community selective pressures and surviving in their ecological niches. This natural phenomenon has aided not only the dissemination of alternate substrate use pathways to out compete other organisms within their environments, but also resistance genes and pathogenicity islands which are of concern from an anthropological and zoological perspective. HGT gives cells the opportunity to gain the ability to cause the most severe of effects to its host organism from the acquisition of a small number of genes by chance (De la Cruz and Davies, 2000). Coupling this with the uptake of resistance genes encoding resistances to a large spectrum of antimicrobials poses quite a large selective pressure on the hosts and their treatment of these potential infections.

It is clear that the spread of these accessory genes providing resistance to antimicrobials is widespread in the biosphere, with resistance genes found in the extremities of the globe, such as Antarctic soils (Van Goethem *et al.*, 2018). Many of these genes are carried within transposable elements (TEs), plasmids and integrons which act as vectors for the spread of these genes and gene cassettes to other members of the microbial community. Some genes may even be transferred by bacteriophage or taken up naturally (Arber, 2000; Burmeister, 2015)

When studying HGT and model organisms such as *E. coli* and *Salmonella* spp.*,* GC analysis and codon analysis can be used to identify potential origins of DNA. It is thought that approximately 18% of the bacterial genome was acquired by some form of HGT in order to distinguish them as the species they are today (Lawrence and Ochman, 1998). This highlights the key role that HGT plays within the evolutionary picture of microbial evolution.

1.6.1   Transformation

Transformation is the simplest form of HGT and is the uptake of naked DNA from the environment (Burmeister, 2015). Transformed DNA is often composed of small simple genes or gene cassettes. DNA from dead cells or circularised excised gene cassettes may be in the immediate vicinity of a live cell and transformed. In the case of a gene cassette, transformation may take place by a circular piece of DNA taken in, decircularised and is taken up into an integron using an integrase enzyme. This process of DNA integration results in a cell obtaining a fragment of DNA to help it outcompete its rivals in the environment and overcome other selective pressures (Arber, 2000; Dubnau, 2003). However, this simple piece of DNA could have no effect or negative effect on cell survival. Whilst uptake of DNA is a random event, so is the uptake of a gene giving selective advantage for the cell in that environment.

1.6.2   Transduction

Transduction is HGT where a bacteriophage packaged with host DNA infects a new host cell with DNA from the capsid (Burmeister, 2015). When bacteriophage enter

lytic phase, phage packaging proteins may load non-phage/host DNA fragments into viral capsids resulting in the transmission of host DNA to new host cells allowing uptake of resistance genes, virulence factors, pathogenicity islands and genomic islands (Chiang, Penadés and Chen, 2019). DNA may then be integrated into host genomes by phage integrases and then passed to the daughter cells, disseminating the new DNA into the microbial community.

1.6.3   Conjugation

Conjugation is arguably the most efficient way of exchanging larger fragments of DNA between cells horizontally, and even allows the exchange of DNA to daughter cells vertically. Conjugation uses conjugation machinery to allow for the transfer of DNA between a mating pair of bacteria (donor and recipient) via plasmid DNA, which may contain transposons, and sometimes integrative conjugative elements (ICE) systems; some of which are mobile without carriage on plasmids. In Gram-negative cells, conjugation is mediated by a retractable pilus from the donor cell which is assembled by the type IV secretory system (T4SS), known as the transferosome (Arber, 2000). The complex spans from the donor to the new recipient cell, in mating pair formation. *oriT* is nicked and ssDNA is formed. Plasmid ssDNA is then transferred and passed down the pilus into the recipient cell, then recircularised in the recipient cell and the complimentary strands reformed. The pilus then retracts, and mating is complete (Figure 1.6). Depending on the size of the plasmid, the time taken for this process varies from several minutes to hours (De La Cruz *et al.*, 2010; Raleigh and Low, 2013). This process is advantageous in terms of HGT, as host range of the T4SS is broad, and

cells can conjugate with multiple different host types and large fragments of genetic material may be shared at any one time.

However, for plasmids involved in conjugation, there exists the problem of plasmid incompatibility. IncF, IncH and IncP replicons are the most commonly observed conjugative plasmids in *E. coli.* They are often large plasmids containing resistance genes. Cells already containing an IncP plasmid cannot take up another IncP plasmid, the one present in the host cell will likely use a toxin/anti-toxin system such as Hok/Sok which means the plasmid will be destroyed if removed from the cell. These recombination mechanisms work because the half-life of the toxic protein is longer than the half-life of the antitoxin. Upon curing the plasmid from the host cell, toxin produced will not be removed by the antitoxin as there is no antitoxin present. As a result, the cell dies (Thisted and Gerdes, 1992).
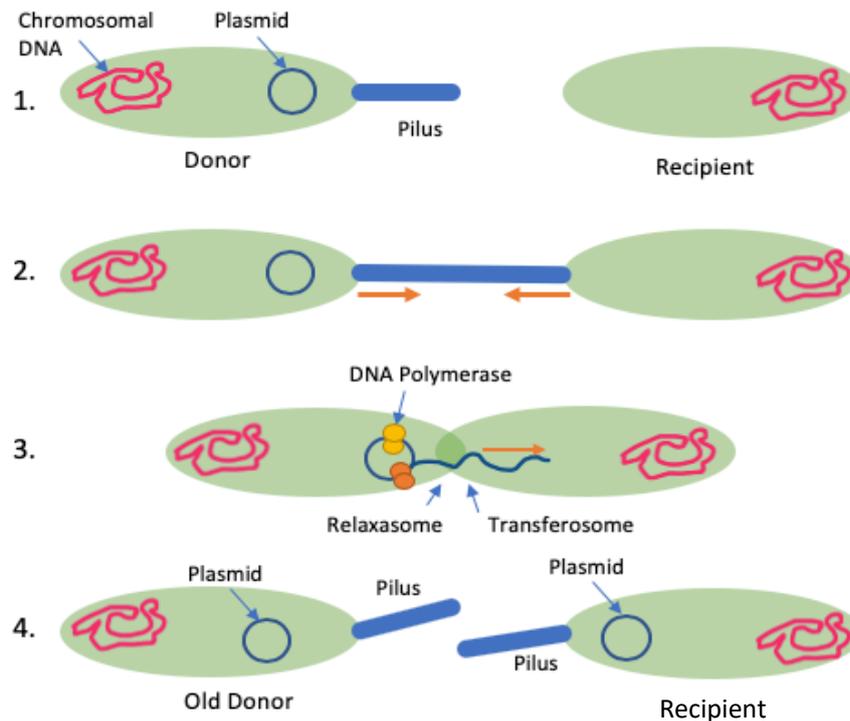
Figure 1.6 Horizontal gene transfer by conjugation.

A plasmid is transferred by the production of a Type IV pilus to allow the transfer of a plasmid through the pilus from the donor cell to the recipient cell. 1. A cell containing a transmissible plasmid capable of producing the T4SS extends the pilus when a nearby cell is detected. 2. The T4SS pilus connects with the recipient cell within the host range. 3. Provided the recipient cell does not already contain a plasmid of the same incompatibility group, the plasmid is replicated by rolling circle replication from the *oriT* site by DNA polymerase and the linearised replicated plasmid DNA is transferred across to the recipient cell. 4. Finally, the two cells are freed from each other by the retraction of the T4SS pilus.

Plasmids are also capable of losing their ability to conjugate which may have downstream effects. One such example is pMG101, a well-studied self-transmissible plasmid originally isolated from *Salmonella* Typhimurium in a burns unit in a hospital in Boston, USA in the 1970s (McHugh *et al.*, 1975). The large resistance plasmid carrying resistance to silver nitrate, as well as mercury (II) chloride, ampicillin, chloramphenicol, tetracycline, streptomycin and sulphonamides has since been

28

characterised to have two forms. One is a 168 kb IncHI plasmid, containing IncFIA/FIB/FII sequences (pMG101-B), and another form is a 383 kb plasmid, containing IncH1/H12A/FIA/FII/FIB sequences (pMG101-A). In a recent study we identified that pMG101-B had lost the ability to conjugate due to an ICE element interrupting a crucial part of the *tra* region (Hooton *et al.*, 2021). As a result, it was noted that the Tn*7* portion of the plasmid (33.7kB in length) containing silver resistance and copper resistance in the form of *tnsABCDE/silESRCFBA/pcoEABCDRS* had transposed into the chromosome as a potential means for self-preservation and greater chance of dissemination into the environment. This highlights the importance of the role plasmids play as a transport for the dissemination of transposable elements into new hosts. Many of these transposable elements using this system contain integrons which may allow for a greater selective advantage in pressurised selective microbial communities. Such examples include Tn*21*, Tn*7* and Tn*402*.

1.7 Co-selection and Co-occurrence

1.7.1 Co-Selection: Definition

Co-selection arises when there is a selective pressure which causes selection for a specific gene, but which causes multiple genes not related to the selective pressure to be preserved at the same time, thus increasing their presence within an environment at the same rate as the gene selected for. This is commonly seen in integrons where multiple open reading frames are found within a gene cassette or

multiple different cassettes may be seen within one integron (Bunny, Hall and Stokes, 1995). It can also be seen in other mobile elements like plasmids and transposons, such as the co-carriage of IS*26* and *bla*$_{NDM-1}$ (He *et al.*, 2015).

1.7.2   Co-Occurrence: Definition

Co-occurrence is the instance where multiple antimicrobial resistance genes (ARGs) are found within the same strain irrespective of whether they are carried by different mobile elements such as plasmids and transposons or chromosomally.

1.8   Integrons

At its simplest, an integron is a genetic element allowing the acquisition and dissemination of genes and gene cassettes into the host DNA for its own use. Whilst hundreds of families of integrons have been discovered, they have been grouped by similarity and percentage similarity (Gillings, 2014). From their wide presence and great taxonomic distribution across Gram-negative bacteria, they are thought to be ancient mechanisms allowing rapid adaptation to changing environments (Boucher *et al.*, 2007). Integrons exist both chromosomally, as immobile elements containing hundreds of gene cassettes mostly of unknown function, and in mobile elements, such as transposons, conjugative plasmids, and ICE elements. These often contain very few gene cassettes, but resistance genes are more readily retained for improved transmission across a microbial population (Chen *et al.*, 2003; Escudero *et al.*, 2015). Integrons are a primeval feature consisting of three key pieces allowing the capture of gene cassettes: the integron-integrase (*intI1*), a promoter region (*Pc*) and a

recombination site allowing those cassettes captured to join the DNA (*attI*) (Figure 1.7). Integrons are commonplace in bacterial genomes but are also seen to be present within transposable elements and large transmissible plasmids. With these three components, exogenous DNA can be integrated and expressed by a host cell. Gene cassettes are integrated into the DNA at the *attI* site, mediated by the expression of *intI1* and is then expressed using the Pc promoter (Cambray, Guerout, and Mazel, 2010; Boucher *et al.*, 2011; Gillings *et al.*, 2015).
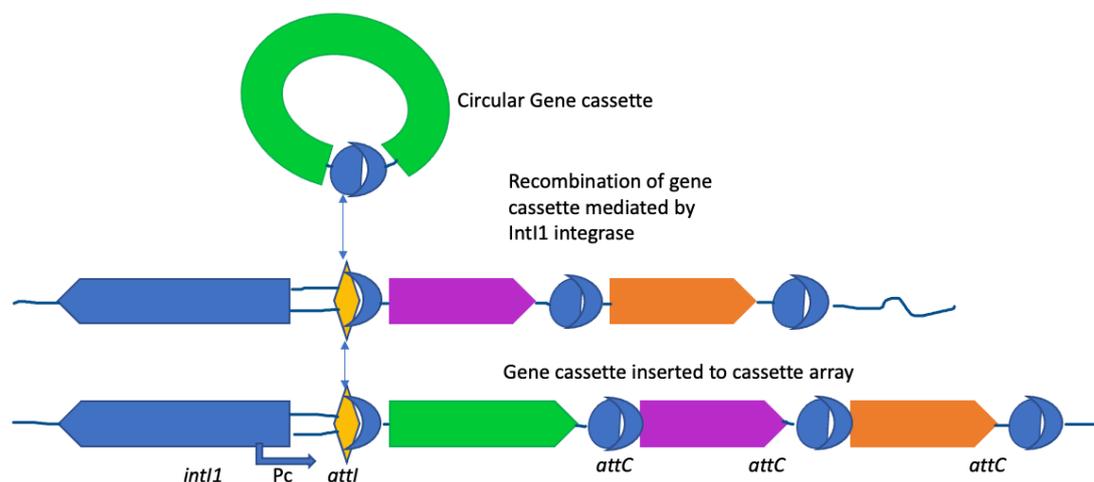
Figure 1.7 Integron structure and function.

Adapted from **(Gillings *et al.*, 2015)**. Gene cassettes are inserted downstream of the *intI1* gene and the Pc promoter at the *attI* site. Multiple gene cassettes may sit in a series forming a cassette array. The uptake of cassettes is a reversible process whereby gene cassettes may be excised in a similar manner.

There are 4 main classes of mobile integrons associated with antimicrobial resistance, all with slightly different characteristics and can be distinguished by their homology. Classes I, II and III are usually isolated from clinical isolates settings and

31

class IV found on the SXT element of *Vibrio cholerae* (Sorum, Roberts and Crosa, 1992; Hochhut *et al.*, 2001; Partridge, 2011).

Class I integrons are clinically relevant genetic elements which are on their own immobile (Labbate, Case and Stokes, 2009). They often carried by Tn*3* family transposons and have ties to the carriage of antimicrobial resistance genes (Recchia and Hall, 1995). Class II integrons are similar in that they are often associated to carriage within Tn*7* family transposons for their mobility via the *tnsABCDE* system (Xu *et al.*, 2009). Class III integrons contain similar overall structure to class I and class II integrons but are more closely related in function to class I integrons in their ability to capture and excise gene cassettes, however their target *attC* site differs and are less commonly observed within the environment (Arakawa *et al.*, 1995). These integrons are also associated with carriage of genes encoding ESBLs on IncQ plasmids (Arakawa *et al.*, 1995). Finally, there are class IV integrons. These integrons are typically located chromosomally and often possess larger cassette arrays than other classes of integrons, with many gene cassettes extending beyond antimicrobial resistance and pathogenicity genes; many of which contain ORFs which contain hypothetical and unknown proteins (Shibata *et al.*, 2003). Class IV integrons were firstly detected within *Vibrio spp*. predating the antibiotic era and have since been identified across many other genera (Rowe-Magnus *et al.*, 2001). From the cassettes arrays of the larger integrons, it has since been possible to determine the origin of gene cassettes by examining the folding mechanisms of the *attC* sites using co-variance models (Ghaly, Tetu and Gillings, 2021)

1.8.1 Class I Integrons

Class I integrons, the first of the five classes to be discovered, have been identified as a proximal marker for human environmental pollution (Gillings *et al.*, 2015). This type of integron, although ancient, is thought to have the most interaction with humans as they are found mainly within *Betaproteobacteria*, which make up a large proportion of commensal bacteria within the human gut. Class I integrons frequently possess gene cassettes that carry resistance or metabolic functions which allow the host to survive in the presence of a number of naturally occurring antimicrobials like streptomycin, or synthetic compounds such as quaternary ammonium compounds and have even been known to possess resistance gene cassettes to oxidative bursts. This resistance cassette is in fact one of the very rare known cassettes to possess its own individual promoter (Stokes and Hall, 1991). This is confirmed by the lack of prevalence of class I integrons in environments unaffected by humans in any way. One study showed 0.002% of bacterial cells isolated from unaffected soil possessed class I integrons compared to 5% of bacterial cells isolated from affected soil and biofilms (Hardwick *et al.*, 2008; Gaze *et al.*, 2011).

Dissemination of integrons is likely to happen in the following manner. When humans or animals have antibiotics administered orally, the gut microflora is exposed to them, causing a selective pressure to arise. This would select for the more resilient strains within the gut due to horizontal gene transfer of resistance genes (Yoon and Yoon, 2018). When an antibiotic is administered, some of the bacteria in the mammalian gut that possess a resistance gene to the antibiotic, and transient free gene cassettes that also contain resistance to the antibiotic, are shed in the faeces

33

and into the environment along with a low concentration of the antibiotic and in some cases its metabolites. One potential means of further transfer of integrons is that upon entering the sewage system that these resistant bacteria could potentially share these genes with others via horizontal gene transfer via the integron contained in a transposon, extra chromosomally: or via excision of circularised cassettes and transformation of them into non-pathogenic and pathogenic bacteria alike. From this spread of resistance genes, there is an increased chance of multidrug resistant pathogens entering the environment.

Integrons not only help this process by their horizontal transmission of resistance genes via plasmids, but also by their ability to transcribe the resistance genes using the SOS response as soon as there is an environmental stress (Guerin *et al.*, 2009). In one study, screening of clinical *E. coli* isolates from a hospital in Australia, found that 23% of isolates carried integrons, of which many were noted to carry multiple drug resistances (Christopher, 2013).

From the surveillance of integrons in certain environments, such as wastewater, soil in arable fields, food livestock, or vermin, as well as nosocomial isolates, it could be possible to discover which resistance genes are more prevalent in different areas and therefore allow an informed prediction of the pollution within the environment. Not all resistance genes are carried in the integron, but it provides a good marker for tracking the prevalence of anthropological pollution. This is due to the high gene cassette turnover that occurs in integrons. To put this into perspective, there are over 80,000 different compounds traded globally, all with potential to pollute the environment (Rockström *et al.*, 2009).

Wastewater has been shown to be a very good source for isolating resistance genes and integrons and to analyse resistance profiles in *Enterobacteriaceae*, and this has allowed the discovery of new resistance cassettes. Not only that, but wastewater has also allowed the characterisation of the integron in different environments and how different populations and their industries affect the wastewater (Pignato *et al.*, 2009; Pellegrini *et al.*, 2011; Moura *et al.*, 2012). Multiple studies have used wastewater as a platform to assess the abundance of contaminants, such as estrone and tetracyclines (Baronti *et al.*, 2000; De Cazes *et al.*, 2016) and resistance to many different antimicrobials which may not be detectable, or are in lower concentrations, such as mercury, silver, QACs and beta lactams (Gaze *et al.*, 2011; Cacace *et al.*, 2019; Yuan *et al.*, 2019). The prevalence of integrons within wastewater has also been targeted in previous studies to show the importance and need to increase the efficacy of wastewater processing to remove pollutants, (potentially pathogenic bacteria and resistance genes) before they are disseminated further into the environment (An *et al.*, 2018). In the 2018 study by An *et al.*, integron abundance was seen to only drop 10-fold from start to finish of wastewater treatment, whilst genetic diversity of the integrons was also seen to drop. In this environment it was noted that anywhere between 10^7 and 10^{10} copies mL^{-1} of integrons exist in the influent, so 'cleaning' and treating this wastewater, is only reducing the genetic load by a tiny fraction before being released to the environment.

Integron integrase genes possess a LexA binding region in the promoter; LexA is a transcriptional repressor governed by the SOS response (Aertsen and Michiels, 2006; Cambray, Guerout, and Mazel, 2010). Upon cell stress, such as antibiotic stress,

35

thermal stress, oxygen stress or other chemically induced stresses, LexA activates transcription of the integrase changing the recombination rate of gene cassettes. In turn, this allows for phenotypic changes in the cells. New cassettes can be taken up, expression of current cassettes and re-expression of cassettes positioned further down the cassette array from the Pc promoter by reshuffling (Collis *et al.*, 1993). There are three major outcomes from these recombination events. 1). The expressed cassette by the Pc promoter provides a positive impact to the cell stress allowing easier survival and is selected for. 2). The cassette expressed is toxic to the cell or not needed which reduces expression of the integron. This results in recircularization and excision of the cassette. 3). The final possible outcome is that there is no phenotypic effect. The cassette is neither advantageous nor disadvantageous to the environment. This means that the cell is facing a constant regulation of cassette uptake and the cassette turnover rate will match environmental conditions and the integron expression rate is maintained (Aertsen and Michiels, 2006). This tight regulation by the LexA stress response ensures that exogenous cassette acquisition and reordering of the cassette array is only used when it is absolutely necessary for survival (Cambray *et al.*, 2011).

1.8.2 Gene Cassette Structure

Gene cassettes are circularised promoterless DNA segments containing at least one open reading frame followed by a 59-base element, known as the *attC,* specifically recognised by the IntI integrase, for capture and integration of gene cassettes to the host DNA. It is thought that although rare, gene cassettes carrying multiple open

reading frames have arisen from the fusion of two gene cassettes in an excision event

where one 59-base element was deleted (Recchia and Hall, 1995). Multiple sets of

these open reading frames captured to the host DNA located adjacently are known

as a cassette array and are expressed all together upon activation of the Pc promoter.

The *attC* site is the most crucial part to the structure allowing integration by the host

integrase. Part of the 59-base element is found at the 5' end of the linearised element

(5'-TTRRRY-3') and a 'G' found right at the end of the 3' end forms the key recognition

for integration (Figure 1.8).

Figure 1.8 Structure of a linearised gene cassette showing separation of the 59-base *attC* site allowing recircularization upon cassette excision.

Adapted from (Recchia and Hall, 1995). The core sites (GTTRRRY) found at each end of the integrated cassette are shown. The solid line and arrow above the cassette indicate the extent of the 59-base element associated with the free, circular form of the cassette, and the dashed line and arrow denote those bases making up the 3' end of the functional composite 59-base element in the integrated form. The inverse core site (RYYYAAC) at the 5' end of the 59-base element is also shown. Short dashes represent the central region of the 59-base element.

Class I integrons have the ability to prioritise gene cassette expression. The integron

contains a Pc promoter, which allows the transcription of the gene cassettes

downstream that have been taken up, as they rarely possess their own promoter

region. It is not known why the majority of these cassettes do not possess individual

promoters. The Pc promoter can be induced when the cell is under stress as a last resort to try and survive the conditions by the SOS signal response cascade. By applying a selective pressure corresponding to a certain antibiotic resistance gene within the integron, it has been shown that the integron can re-shuffle the gene cassettes to favour the transcription of that cassette more over the other gene cassettes by recombination between *attC* and *attI* sites very much in the manner of Cre – *loxP* (MacDonald *et al.*, 2006). Not only does this mean that the recombination system allows uptake of genes or reshuffling, but also allows the removal of cassettes, that may not be used very much, or removal of useless DNA which could have possibly been taken up. Excision of cassettes is common, these cassettes are then likely to be taken up by other naturally transformed bacteria within the environment.

Gene cassettes provide a novel approach to microbial natural selection, allowing free transfer of small open reading frames between Gram-negative bacteria, that may be beneficial to survival within their community. This can accelerate the process of natural selection, which is already rapid in bacteria due to extremely short generation times. Upon replication, one daughter cell will contain a copy of the gene cassette and the other cell will not. Whichever survives and reproduces the most will therefore in effect increase the abundance of the cassette in question.

1.8.3   Mobility of Class 1 Integrons

It is widely accepted that class I integrons have played a key role in microbial evolution (Mazel, 2006). In more recent times, due to anthropological impacts, the

transmission of resistance genes to many antimicrobials; including antibiotics, detergents, metals and disinfectants has been seen in greater abundance (Knapp *et al.*, 2010). Not only can gene transmission be within strains but also between species. This allows the development of diversity and complexity (Ghaly *et al.*, 2017). The pools of antibiotic resistance genes found in integrons and the presence of integrons is likely to be influenced by human activity in their environment (Moura *et al.*, 2012; Gillings *et al.*, 2015). Due to the improvement of human infrastructure, medicine and industrial practices allow for constant development and variation of the resistance genes found in these integrons. Furthermore, integrons are seen to be a major contributor to the increased occurrence of antimicrobial resistance in environmental *Enterobacteriaceae* due to their abundance, which is well documented (Chainier *et al.*, 2017; Partridge *et al.*, 2018). However, within a local environment, it is not known whether the similarities between integrons from different isolates and their resistance profiles are the same.

It is thought that integrons developed from a recent common soil and freshwater non-pathogenic ancestor, a *Betaproteobacterium,* which possessed this element on the chromosome. This may have given rise to the spread of antibiotic resistance genes through the microbiome by a transposon from the Tn*402* family trapping a chromosomal class I integron carrying *qacE1* and *sul1* where the end terminus was deleted by the uptake of *sul1* (Kholodii *et al.*, 1995; Gillings, Boucher, *et al.*, 2008). This can be seen through such high conservation of the integrase enzyme sequence and the recognition sites (Gillings *et al.*, 2015). In the case of Tn*21*, the integron is thought to have inserted in the middle of the gene *urf2M* which may have contained

tnpM (Liebert, Hall and Summers, 1999). This was further confirmed by the discovery of Tn*5060*, a Tn*21* like element containing no integron (Kholodii *et al.*, 2003).

1.8.4 Dissemination Theory of Gene Cassettes

Recently, the increased prevalence of integrons has been driven by anthropogenic pressures from industry, manufacturing and medicine, and resistance gene cassettes have been seen to be disseminated widely among pathogens (Severino and Magalhães, 2002; Domingues, da Silva and Nielsen, 2012). These clinically important integrons have all been found within mobile elements such as transposons, conjugative plasmids and ICE elements (Escudero *et al.*, 2015). There exist thousands if not millions of different gene cassettes in the biosphere, most of which are uncharacterised (Partridge *et al.,* 2009). However, those that have been characterised have great potential for bacterial host environmental adaptation, including ABC transporter systems, periplasmic binding proteins and thiosulfate binding proteins (Rowe-Magnus *et al.,* 2001). Integrons and gene cassettes have been recovered from every environment sampled and identified in sequences from different major bacterial taxa, however their biogeography and diversity are less well studied. Recently, a study examined this by sampling environments in a spatial manner of increasing distance and comparison between different environments (Ghaly *et al.*, 2019).

The abundance and diversity of gene cassettes was measured using metagenomic analysis which allowed for identification of 27,000 novel ORFs, which is thought to be an underestimate due to the fluidity of integron turnover within a bacterial

environment due to selection pressures (Ghaly *et al.*, 2019). Interestingly, this study demonstrated that the majority of cassettes possess a limited abundance locally in each environment, suggesting gene cassette generation is a continuous, multi-faceted and dynamic process in any location.

Despite a limited abundance, multiple cassettes were identified across all of the separate locations, other geographic locations, and non-soil matrices, outside the samples in the study (Ghaly *et al.*, 2019). This suggests they were disseminated or potentially separately generated and required for survival within different locations. In fact, spatially between sample sets the abundance of individual gene cassettes decreased but then there were some noted to be present in every single sample. There are three factors which are key to the distribution of cassettes globally: time, dispersal and selection (Ghaly *et al.*, 2019).

Since cassettes originate in a single cell or single cells within a specific location in limited copies, to help overcome a selective pressure to become fitter for survival, it will be shared between cells in the immediate vicinity over time. Eventually successful cassettes will be dispersed potentially further afield via cassette mobilisation or by co-selection of whole cassette arrays on larger mobile elements via horizontal transfer like those previously mentioned. Although in relative terms this horizontal gene transfer is rare, the most documented occurrences of this are the dissemination of antimicrobial resistance genes (de La Cruz and Grinsted, 1982; Partridge *et al.*, 2009; Gillings, 2017).

In contrast, most integron classes are chromosomal and not mobile, relying on vertical transmission of cassettes in the short-term dispersal (Mazel, 2006). This is

supported by evidence for high spatial turnover of cassettes over small relative distances and suggests cassette occurrence and abundance forms part of a larger more conserved shared genome, which is drawn on between many genera of bacteria within the environment (Ghaly *et al.*, 2019). Thanks to the ability of bacterial spore formation in some genera, this allows for potential increased dispersal rates and can even be accelerated by anthropogenic input as previously mentioned (Lévesque *et al.*, 1994; Gaze *et al.*, 2011; Gillings *et al.*, 2015; Zhu *et al.*, 2017).

In summary, cassettes are thought to originate within specific environments and whilst they are abundant locally, cassette diversity varies greatly even at local levels, and some cassettes may disperse over time globally if they are selectively advantageous. In a wastewater environment, a hyperdynamic environment compared to soil due to the constant flow of water and varying inputs, there should be more cassette variation across even smaller distances but also greater similarities between influent and effluent despite the overall decrease in abundance due to the water treatment processes that go on.

1.9 Transposable Elements (TEs)

TEs are a family of mobile genetic elements capable of horizontal gene transfer, often containing resistance genes to metals, antibiotics and genes aiding metabolism. A TE can jump between genomes, plasmids and other mobile elements by inserting itself into larger pieces of DNA by recognising consensus sequences and inserting at that site. TEs have been identified in both Eukaryotes and Prokaryotes in large numbers. Originally, transposons in bacteria were first identified as genetic elements carrying

drug resistance genes, which were flanked by insertion sequence (IS) elements. However, now IS elements are also grouped under the term transposable element (Griffiths *et al.*, 2000). TEs also encode a gene for transposase, an enzyme responsible for insertion and extraction of the element from host DNA at the IS loci. There are four categories of bacterial TEs: IS, composite Tns, non-composite Tns (Tn*3* family), and transposable phage Mu (Dziewit *et al.*, 2012).

1.9.1 IS Elements

IS elements are the smallest and simplest form of TE and are transmitted using independent transposases within the elements (Figure 1.9). IS elements are abundant in bacteria, only carrying DNA required for transposition and regulation. They are typically made up of a single ORF encoding a transposase and flanked by two inverted repeat (IR) regions recognised by the transposase. IS elements may excise freely, causing target site duplication of 2-15bp direct repeats (Chandler and Mahillon, 2002). Doing so can readily cause mutation of genes from the wild type in which they may be inserted (Nevers and Saedler, 1977; Mugnier, Poirel and Nordmann, 2009). Due to their ORF containing a transposase, IS elements are also capable of mobilising non-autonomous TEs which do not possess their own transposases (such as miniature inverted repeat transposable elements (MITEs)).

The IRs of IS elements contain two distinct functional domains. A transposase recognition site and the second part is for strand specific reactions and cleavage for transposition (Ichikawa *et al.*, 1990).

DR	IR	ORF(s)	IR	DR

Figure 1.9 Structure of an IS element.

Direct repeats, short target sequence which is duplicated upon integration. IRs flank the ORF, which is normally short. The whole element is normally less than 2,500bp.

1.9.2 Composite Transposons

Composite Tns are a less abundant form of TE present in prokaryotes, are known to carry AMR genes, and are often flanked by ISs. They are formed by two identical IS elements that are in close proximity integrating within a genome. If the genes between the two TEs possess potential advantageous traits, they may be fused so they can transfer the genes to other bacteria simultaneously. The transposase of one IS element may cause transposition of the two IS elements simultaneously, including the genetic material between the two IS elements to a new host, via insertion to a plasmid or directly (Figure 1.10). The likelihood of this event occurring is low, but if it is selectively advantageous, more of the gene pool are likely to survive with this iteration of the IS element. The larger the distance between the two IS elements, the lower likelihood of successful transposition of the composite element (Wagner, 2006).

DR	IR	ORF(s)	IR	DR	Host DNA	DR	IR	ORF(s)	IR	DR

Figure 1.10 Composite transposon structure.

Two IS elements inserted to host DNA, flanking a portion of DNA, forming a composite transposon. This may confer resistance genes or even biosynthetic pathways.

The transposable phage Mu: part of the *Myoviridae* family; is a lytic and lysogenic phase phage, capable of transmitting resistances horizontally between bacteria. It was christened Mu after its mutator abilities when it was first reported (Foguel, Dunford and Schwartz, 1963). The phage itself encodes for a transposase and the phage DNA is flanked by an *attR* and *attL* site. Transmission of Mu phage also varies from other TEs as it does not necessarily cleave itself from the host DNA and insert into transmissible plasmids. Instead, Mu uses MuA and MuB proteins for recombination with the host genome and identification of integration site, respectively. It also requires two more proteins: IHF (integration host factor) and HU (histone binding protein) (Harshey, 2014). As mentioned, when talking about transduction (1.6.2), when entering the lytic phase in a phage life cycle, DNA is cleaved, and phage DNA packaged to capsids of phage particles. However, sections of host DNA are also packaged and potentially resistance genes may be transferred to new hosts, thus disseminating resistance through a bacterial community.

1.9.3   Non-Composite Transposons

The final type of TE and the most relevant to the dissemination of antimicrobial resistance is the non-composite Tn. They are the most abundant form of TE within bacteria (Szuplewska, Czarnecki and Bartosik, 2014). Unlike composite Tns, they do not rely on terminal IS elements for transposition as they possess their own transposase genes, transposase (*tnpA*), resolvase (*tnpR*), and their own inverted repeat sequences at either end of the transposable element. The TnpA is responsible

for the transposition of the DNA contained between long terminal inverted repeats of the non-composite Tn, the expression of TnpA is negatively regulated by TnpR (Casadaban, Chou and Cohen, 1982). Non-composite Tns are synonymous with Tn*3* family transposons. The ampicillin resistance Tn, Tn*3* was the first isolated Tn containing resistance to any antibiotic (Hedges and Jacob, 1974). Each characterised Tn from this family possesses that of the non-composite Tn with long 35-48 bp terminal IRs; similar mode of mobility; leaving a 5-bp duplication of adjacent sequence to the insertion site and by their transposition proteins (Ohtsubo, Ohmori and Ohtsubo, 1979; Kostriken, Morita and Heffron, 1981; Sherratt, 1989; Maekawa, Yanagihara and Ohtsubo, 1996). Further to this, non-composite Tns have been found to transpose to AT rich target regions, preferentially transposing from plasmid to plasmids over plasmids to chromosomes (Szuplewska, Czarnecki and Bartosik, 2014). Other examples of Tn*3* family Tns include, but are not limited to: Tn*3*, Tn*1000*, Tn*1721*, Tn*1*, Tn*501*, Tn*21*. The non-composite Tn*3* family transposons also have a *res* site, a 120 bp sequence, which contains promoter regions for the transposase genes *tnpA* and *tnpR* and a site for recombination to occur in the resolution of cointegrates. Its role is as a site recognised by the serine recombinase, TnpR, for the resolution of cointegrates. The *res* sites also contain multiple inverted recognition regions separated by varying lengths of spacer DNA for resolvase protein dimers to bind to (Nöllmann, Byron and Stark, 2005). Non-composite Tns are also known as Tn*3* family transposons and play a critical role in dissemination and development of resistance genes. Upon transposition, these elements are known to cause mutation within host DNA (Wagner, 2006). Across the Tn*3* family, structures of non-composite

Tns vary between two main layouts of their transposase genes and resolution (*res*)

sites, further dividing this group of mobile elements into two sub-families; the Tn*3*

sub-family and Tn*501* sub-family (Figure 1.11).
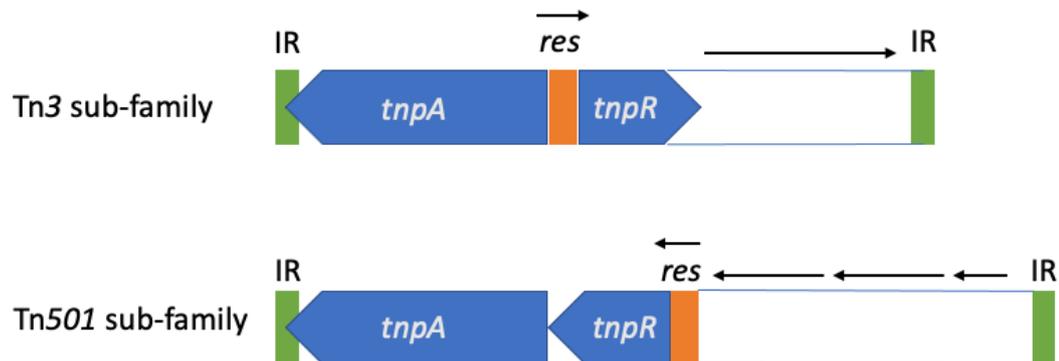
Figure 1.11 Structural differences between Tn*3* and Tn*501* sub-families of non-composite transposons.

Top: displays the layout of the Tn*3* sub-family of transposons indicated by the *res* site situated between the *tnpA* and *tnpR* genes. Further of note, the *tnpR* is oriented away from the *res* site, any other accessory genes carried within typically transcribe in the same direction (Tn*3*, Tn*1000*). Bottom: Transcription of *tnpA* and *tnpR* are in the same direction and the *res* site is located 5' of *tnpR* (Tn*501*, Tn*1721*, Tn*21*, Tn*551*). Adapted from Sherratt (1989).

Tn*3* family Tns have been identified throughout nature, commonly on plasmids

encoding antimicrobial resistance in Gram-negative bacteria. It is likely their

widespread dissemination is due to the indiscriminate use of antimicrobials by

humans in human and animal medicine. However, not all Tn*3* family elements are

found in Gram-negative bacteria, Tn*551* was first identified in *Staphylococcus aureus*,

suggesting there may be some cross-over between these TEs and their host range

(Heffron, 1983).

1.9.4   Mechanisms for Tn*3* Family Recombination

For the most part, Tn*3* family transposition is well regulated and replicative, with transposition frequency between 10^{-5} to 10^{-7}; this may be increased up to 100-fold for short periods in some instances under the right conditions (Sherratt, 1989). The most frequent and possibly the most biologically significant form of transposition identified within Tn*3* family transposons is from plasmid to plasmid resulting in both donor and recipient plasmids containing the transposable element. The intermolecular event in which transposition occurs is a two-step process requiring the transposase protein (from *tnpA*) then resolvase (from *tnpR*) (Arthur and Sherratt, 1979; Kostriken, Morita and Heffron, 1981; Kitts, Lamond and Sherratt, 1982).

Firstly, the donor transposon sequence ligates to the target sequence and ligates to produce a cointegrate, causing the fusion of donor and recipient replicons by two directly repeating copies of the transposon in a recombination event. The cointegrates can then be converted to produce products of transposition, by a resolvase mediated second site-specific recombination event using the *res* sites of the mobilising transposon. Of the two steps in the transposition process, less is known about the action of the first recombination event and how the transposase acts to cause the formation of the cointegrates (Arthur and Sherratt, 1979; Kitts, Lamond and Sherratt, 1982; Sherratt, 1989).

1.10 The Structure of Tn*21*

The Gram-negative non-composite mercury resistance transposon Tn*21,* belonging to the Tn*3* family, was first identified in the *Shigella flexneri* R100 plasmid isolated in Japan in the mid 1950s (Nakaya, Nakamura and Murata, 1960)*.* It is well studied as a model for the mercury resistance operon present in Tn*21* and the transposon itself (Pal *et al.*, 2017). In 2003 an ancestor to the Tn*21* was isolated in Russian permafrost, believed to be approximately ten thousand years old (Kholodii *et al.*, 2003). Tn*5060* at 8,667 bp is smaller than Tn*21* (19,672 bp) and is believed to be an ancestor of Tn*21* before the insertion of the integron. Since the discovery of Tn*21* it has helped understand the role of integrons (Figure 1.12).

As Tn*21* is a Tn*3* family transposon; it possesses 38 base pair terminal inverted repeats (IRs), flanking both ends of the transposon. It also contains the *tnpA* and *tnpR* genes, encoding transposase and a resolvase enzyme, respectively. A *res* site for recombination of cointegrates is also present. The Tn*21* subgroup differs slightly from other Tn*3* elements as both *tnpA* and *tnpR* are transcribed in the same direction and their products are 70% homologous to other transposases in the subgroup (Grinsted, De La Cruz and Schmitt, 1990).

Although, *tnpA* and the 38bp IRs are two defining characteristics of Tn*3* and Tn*21*, one of the 38 bp IRs contains the final 5bp of the *tnpA* in it, and the other containing the last 5 bp of *merR* present. This makes the two IRs directional, defining them as IR$_{mer}$ (IR$_L$)and IR$_{tnp}$ (IR$_R$) (Ward and Grinsted, 1987). In Tn*21* there is an open reading frame terminating at site I of the *res*. This proposed reading frame encodes a 116 amino acid protein called TnpM which is thought to enhance transposition and

49

suppress formation of cointegrates in Tn*21* and Tn*21*-like elements (Hyde and Tu, 1985). It was also shown to have the same effect on Tn*501* (Grinsted and Brown, 1984a)*.* However, despite the evidence produced in this study, a protein has still yet to be produced and characterised *in vivo* (Grinsted and Brown, 1984b; Hyde and Tu, 1985). Yet it is still unknown whether chemical stress may cause transposition of Tn*21*.

Commonly, variants of the Tn*21* transposable element (which possess the *mer* operon and a class I integron) have been found with varying integron contents. However, Tn*21* elements have been classically characterised as carrying conserved antimicrobial resistance genes: partial quaternary ammonium compound resistance (*qacEΔ1*), sulphonamide resistance (*sul1*) and aminoglycoside resistance (*aadA1*) (Liebert, Hall and Summers, 1999). It may be that mercury (II) ion presence in the environment can co-select for AMR due to Tn*21* carrying the integron with resistance genes, or because the *mer* operon is plasmid borne on a resistance plasmid carrying other antimicrobial resistance genes.

It was originally thought integrons were chromosome based mobile genetic elements, but it was later shown that they were capable of insertion into mobile elements such as plasmids and TEs. The integron in Tn*21* has lost its function of mobility due to the relatively recent insertion of IS*1326*, which deleted a portion of the 3' conserved sequence which disrupted the transposition module. As a result, the integron now relies on the larger mobile element to transmit itself between cells or by large conjugative plasmids (Brown, Stokes and Hall, 1996).

As a result, it has been proposed that Tn*21* evolved in the following manner (Figure 1.12). First, Tn*402*, a TE possessing a class I integron acquired *qacE* and the 3' end partially deleted (66 bp) due to integration of *sul1, orf5* and other gene cassettes, followed by the insertion of IS*1326* and IS*1353* causing the partial deletion of the *tni* module which allowed mobility of the integron. Next *aadA1* was acquired within the integron. This modified Tn*402* was then likely inserted to a potential Tn*21Δ* (a hypothetical ancestor to *Tn21*) backbone in the middle of the hypothetical protein of unknown function, *urf2M*. This is supported by the IRs identified at each end of the integron module, downstream of the *intI1* and upstream of *tniA* (Liebert, Hall and Summers, 1999).
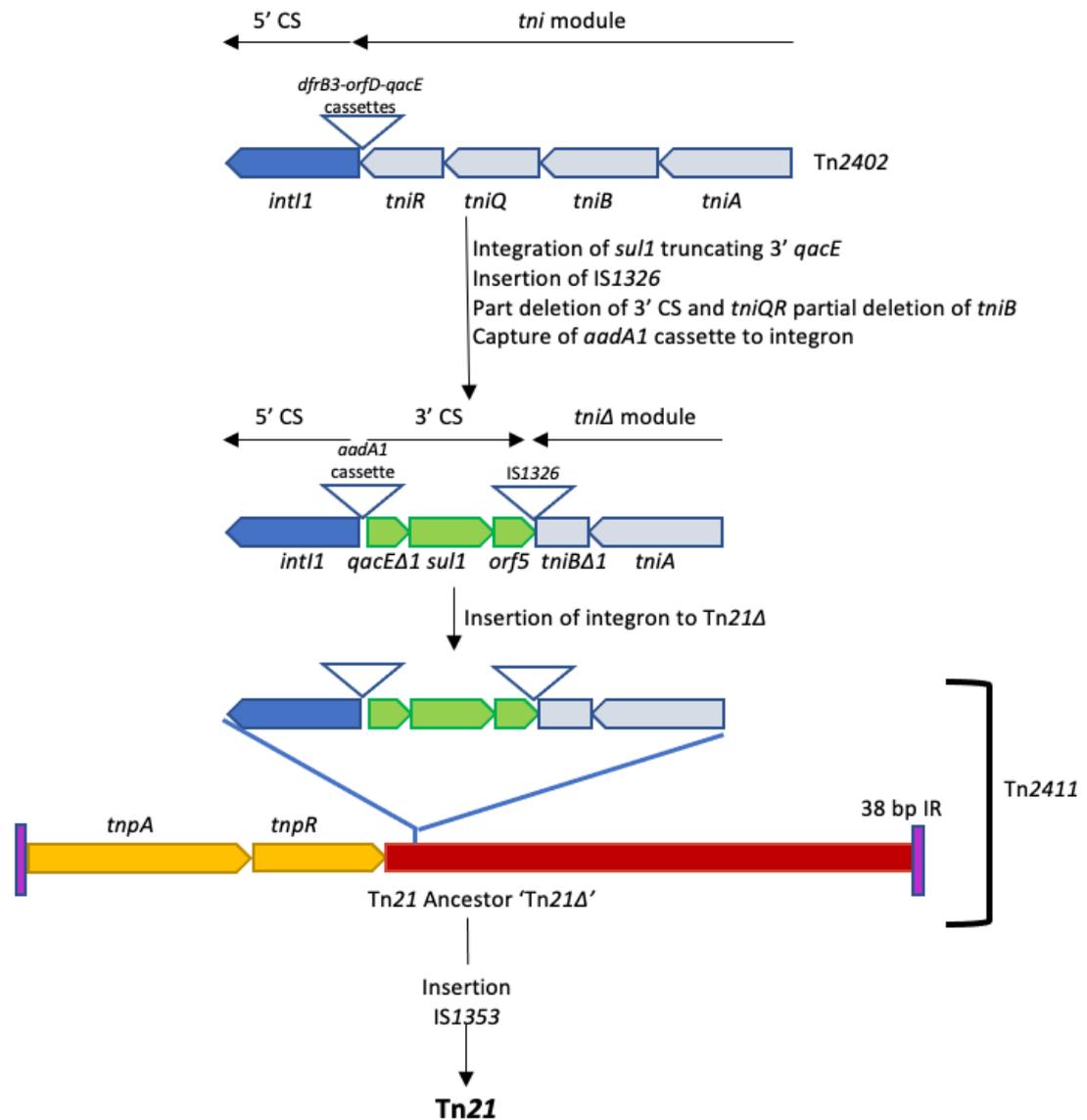
Figure 1.12 The proposed evolution of Tn*21* from Tn*402*.

Tn*402* underwent multiple expansion events resulting in the loss of *tniQR*, capture of *aadA1* and insertion of IS*1326.* The resultant transposon was able to insert to a Tn*21* ancestor known as Tn*21Δ* to form Tn*2411*. Finally, the insertion of Is*1353* resulted in the formation of the well-known transposable element, Tn*21*. Adapted from (Liebert, Hall and Summers, 1999).

Possessing such a resistance cassette like the *mer* operon (which is often found on a larger mobile genetic element) comes with a fitness cost: slower growth of cells and replication due to the size of the element (Melnyk, Wong and Kassen, 2015).

However, this is beneficial when in the correct environment, allowing not just survival but good growth. Like all resistances, there is an upper threshold to the resistance seen. Some strains harbouring this resistance are capable of surviving in up to 40 µM HgCl$_2$ (Freedman, Zhu and Barkay, 2012).

At present, mercury is rarely used as an antimicrobial and in fact there are directives in place to reduce the future use of mercury, including the banning of mercury mining and export within the EU (European Commission, 2017). However, mercury resistance is still rather widespread, probably due its past use, reported to have been used in medicine from at least the early First century (European Commission, 2017; Pal *et al.*, 2017) (1.4). This could be due to those mercury resistant bacteria possessing the class I integron harbouring multidrug resistance genes. The fitness cost for carrying such a large amount of excess DNA in order to take up AMR genes must be beneficial, and as suggested in meta-analysis, plasmid borne and mobile genetic element borne resistance does not carry a very high cost (Vogwill and MacLean, 2015).
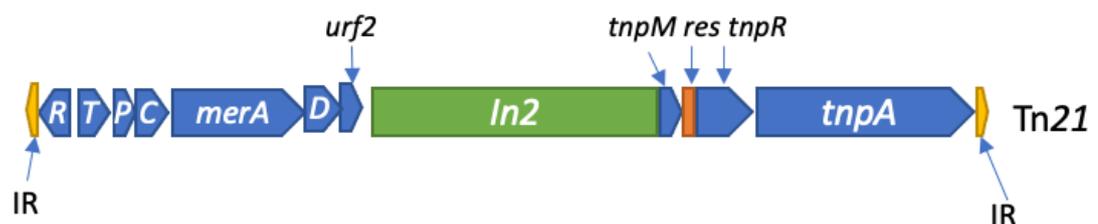
Figure 1.13 Classic structure of Tn*21*.

The Tn*21* is typically made up of three regions. The transposon genes, which facilitate the ability of Tn*21* to move between bacterial cells. The *mer* operon which possesses all the genes required to reduce the Hg (II) ion to the volatile and non-toxic Hg(0). And finally the integron, capable of taking up AMR gene cassettes, which separates the other two sections

either side with 25bp imperfect inverted repeats, marking where a Tn*402*-like transposon inserted to Tn*21Δ*.

1.10.1 Mobility of Tn*21*

In the 1980s a similar transposable element Tn*501*, a mercury transposon which is similar to Tn*21*, except missing the In2 region, *tnpM* and *merC* was shown to have induced mobility when exposed to HgCl$_2$ in low concentrations (Stanisich, Bennett and Richmond, 1977; Grinsted and Brown, 1984a). In Tn*501*, the *mer* operon transcription does not stop at the end of the *mer* operon, with transcription of the downstream *tnp* genes occurring (Figure 1.5). It is not currently known whether chemical stress, such as Hg (II) presence, causes transposition of Tn*21*. Whilst chromosomal copies of Tn*21* exist, such as in enteroaggregative *E. coli* strain 042, it is possible that Tn*21* may only be capable of transmission between plasmids, relying solely on self-transmissible plasmids to disseminate further into the environment. Interestingly, transposition of Tn*21*-like elements is temperature sensitive, with optimum frequency of transposition occurring at temperatures ranging between room temperature (25 °C) for Tn*21* and 30 °C for Tn*501* (Ubben and Schmitt, 1987; Turner, De la Cruz and Grinsted, 1990). Above these temperatures, the transposition rate of both transposable elements has been shown to decrease by two to three orders of magnitude. At 42 °C, transposition is virtually undetectable. Remarkably, it is thought that this is actually due to temperature sensitivity of formation of the transpososome or the transposase itself (Heffron, 1983).

However, if *Tn21* is like a classic Tn*3* family transposon, it is likely that transmission of the transposon targets AT rich regions and is most likely disseminated by

mobilising across plasmids. In some cases, non-composite transposons such as Tn*7* may transpose out of a plasmid especially if the plasmid it inhabits has lost self-transmissibility, resulting in transposition to the chromosome (Hooton *et al.*, 2021). Perhaps this is why Tn*21* was isolated in *E. coli* strain 042 within the chromosome. It is also just as likely that due to the large size of Tn*21* at just under 20 kb, has a transmission rate that is extremely low compared to other TEs. It is possible that antimicrobial selective pressures corresponding to resistance genes within the transposable element may cause transmission. For example, sulphonamide presence in the environment may cause increased transposition frequency of the Tn*21* elements carrying a *sul* gene. Another possibility is that the SOS response could result in its dissemination. It could be speculated that in the future the mercury resistance may be lost from this transposable element, whether it be down to lack of use, recombination or random mutation (Liebert, Hall and Summers, 1999; Hobman and Crossman, 2015). This is further supported by the reporting of Tn*21*-like transposons which do not possess *mer* genes (Grinsted, De La Cruz and Schmitt, 1990).

On the contrary, it was suggested that dental amalgams where mercury was in high concentration provided a specific reservoir for the co-selection for both antibiotic resistance genes and mercury resistance (Summers *et al.*, 1993). This significant finding provides a selective pressure where gut commensal and potential pathogens may be able to laterally transfer genes. However, it is extremely difficult to determine the exact reservoirs due to the multifactorial pressures in this gut interface, but it is likely to be a cause of co-selection (Stokes and Gillings, 2011).

Whatever the drivers for this selection, there is clear evidence that mercury resistant transposons have existed for a very long time in soil inhabiting bacteria, before the antibiotic era (Kholodii *et al.*, 2003; Barkay *et al.*, 2010). Many of which are closely related to the multidrug resistance transposons seen today in pathogenic isolates (Kholodii *et al.*, 2003). The use of mercury in dental amalgam has been linked to the increase of multidrug resistant isolates, which was at first found to be controversial for the dental community but then re-affirmed in a more recent study (Barkay, Miller and Summers, 2003; Roberts *et al.*, 2008). Co-selection is a key force in this linkage. It may be possible to see that this linkage could select both ways: antibiotic selective pressures may co-select for an increase in mercury resistant bacteria, and the increased presence of environmental mercury (human or naturally occurring) may increase the chances of antibiotic resistance gene retention (Barkay, Miller and Summers, 2003; Skurnik *et al.*, 2010). A good example of this is Tn*21*, which appears to be ubiquitous in the biosphere, having been isolated globally in clinical and environmental samples. However, relatives of this transposon were thought to have been prevalent before the dawn of commercialisation of clinical antibiotics (Kholodii *et al.*, 2003). Tn*5060* was isolated from Siberian permafrost and possessed a sequence closely related to the Tn*21*, however with the absence of the In2 region.

For co-selection to be viable as an explanation, co-selection works conversely, antibiotic stress causes selection for antibiotic resistance genes carried on the same mobile element, thus facilitating mercury resistance mobility. It may be that anthropogenic pollution is the driver for retention of the antibiotic resistance genes within these elements (Summers *et al.*, 1993; Skurnik *et al.*, 2010). This needs to be

studied further, population studies of biofilms and with varying selective pressures, to identify what may be driving co-selection where there are fewer factors obscuring solving this very important question.

In a wastewater environment, where there is expected to be a low environmental concentration of mercury, due to the reduction of usage of metals as antimicrobials by humans compared to the early 1900s and the industrially revolutionised global climate, underlying factors must be reason for the presence of this relatively rather large fragment of DNA.

1.11 Aims

Mercury resistance, Tn*21* and Tn*21*-like elements are found globally, in clinical and non-clinical settings, even though mercury usage in medicine has declined since the 1970s. Historical evidence has been used to understand Tn*21*-like element evolution, little is known about what drives their transposition (Liebert, Hall and Summers, 1999; Kholodii *et al.*, 2003). This deficit in knowledge highlights the gaps in our understanding of antimicrobial resistance gene mobility.

The resistance mechanisms and evolution of Tn*21* are well researched; little is known about its dissemination, or its relative abundance amongst bacterial populations. Integron research may provide some insights into this but, Tn*21* class I integrons likely only make up a small proportion of the integrons that are present in bacterial populations.

Similarly, antimicrobials, such as biocides and antibiotics, have largely replaced antimicrobial metals in medicine, thanks to the antibiotic era in the latter half of the 20th century, increasing use of small molecule biocides, however resistance to antimicrobial metals remain within bacterial populations, the reasons for which are less well understood. As soon as there is discovery and use of new antimicrobials, resistance arises, often within 5 years. The use of antimicrobials, whether synthetic or naturally occurring need to be more tightly regulated so that they may still be exploitable to treat infections further into the future as the global population increases and the demands of medicine are put under further strain. By understanding and careful surveillance of anthropogenic outputs such as wastewater it would be possible to learn what resistances within bacterial communities may be

selected for. Then it may be possible to develop ways to eliminate these communities before they are released into the environment. In order to help answer these problems and gain an insight into the effects of antimicrobials on bacterial populations, multiple areas must be investigated.

1. Tn*21* has clearly been disseminated globally and is found in various reservoirs, but the relative abundance within these communities is not known. They are well documented in Gram-negative bacteria, which make up a large proportion of the microbiome, what proportion of this carries Tn*21*? How are these mobile elements carried? How common are these mobile elements within a wastewater environment?

2. Co-selection has been demonstrated to work both ways, in the case of Tn*21*: is mercury resistance selected for indirectly due to presence of an antimicrobial in wastewater, or is mercury in the wastewater causing co-selection for the whole Tn*21* element and thus the class I integrons?

3. From reviewing the literature, it is possible to hypothesise that antimicrobial resistance genes present in the integron facilitate the retention of mercury resistance. As a follow up question, how does the co-occurrence of Tn*21*-like elements and other antimicrobial resistance genes vary across the wastewater treatment process and also compared to agricultural waste?

4. Little is known about the variation of gene cassettes of integrons found within Tn*21.* With only classic Tn*21* discussed within the literature. The variability of integron gene cassette arrays within Tn*21*-like mobile elements is not well studied, neither is it known how the gene cassettes within the integrons of

Tn*21*-like elements differ in their contents across different environments, or within the same microbial population.

5. Finally, Tn*21* transposition events have been demonstrated to be temperature sensitive, with the optimum being at 25 °C. Higher temperatures were shown to be detrimental to the transposition rate. With this information wastewater could be an ideal environment for these events to take place. However, it is not known whether chemical stresses effect the transposition rate in a similar manner. Tn*501*, a similar element demonstrated increased transposition when under the stress of Hg (II) it is possible that similar events may arise in Tn*21*. The SOS response may cause integron gene cassette reshuffling and gene cassette uptake. Is it possible that exposure to Hg (II) stress could induce transposition or does the same SOS response instigate the transposition of Tn*21*?

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Bacterial Strains

Strain	Characteristics	Use	Reference
E. coli MG1655	A K-12 derivative strain	Negative control for mercury resistance and recipient for larger plasmids	(Blattner *et al.*, 1997)
E. coli J53	Azi[r]	Negative control for mercury resistance and recipient for larger plasmids	(Yi *et al.*, 2012)
E. coli 042	EAggEC Cm[r] Tet[r] Sp[r] Tn*21*	Donor for Tn*21* conjugation	(Yamamoto *et al.*, 1992; Chaudhuri *et al.*, 2010)
E. coli J53 RFP kan[r]	Insertion of *mrfp1 kan*[r] downstream of *glmS*	Recipient strain for Tn*21* – pRP4-8 conjugation	This study

Table 2.1 Strains used in this study

2.1.2 Plasmids

Plasmid	Characteristics	Use	Reference
pDOC-C	*sacB SceI* recognition sites Amp[r]	Chromosomal insertion vector backbone	(Lee *et al.*, 2009)
pACBSR	*araC* λred Cm[r]	Induction of chromosomal insertion and curing of pDOC-R using λred system	(Lee *et al.*, 2009)
pDOC-R	*sacB mrfp1* Kan[r] Amp[r] *SceI* recognition sites	Vector for Chromosomal insertion of *mrfp1* and Kan[r]	This study
pULTRA	*mrfp1* Kan[r] Tet[r]	RFP containing plasmid for chromosomal insertion	(Mavridou *et al.*, 2016)
pRP4-8	Gent[r] Amp[r]	Conjugative vector Tn*21* transposition	(Quandt *et al.*, 2004)
pMG101A	IncF/IncHI Sul[r] Ars[r] Mer[r] Sil[r] Cu[r] Tel[r] Kan[r] Amp[r] Tet[r] Cm[r] SGI11	Positive control for Tn*21* presence	(Mchugh *et al.*, 1975)

Table 2.2 Plasmids used in this study

2.1.3 PCR Primers

All DNA oligonucleotides for PCR and sequencing were supplied by Sigma-Aldrich (Merck, Darmstadt, Germany), with 0.025 µmole synthesis, desalt purification, and provided at a concentration of 100 µM in nuclease-free H_2O. Working stocks of 10 µM were prepared using sterile HPLC grade H_2O, and stored at -20°C.

Primer Name	Primer Sequence 5' to 3'	Use	Reference
merA F	ACCATCGGCACCTGCGT	Mercury resistance *merA* identification	(Liebert *et al.*, 1997)
merA R	ACCATCGTCAGGTAGGGGAACAA		
merC F	CATCGGGCTGGGCTTCTTGAG	Mercury resistance *merC* identification	(Liebert *et al.*, 1997)
merC R	CATCGTTCCTTATTCGTGTGG		
merR F	ATCCGBTTCTATCAGCGCAAG	Mercury resistance regulator *merR* identification	(Pérez-Valdespino *et al.*, 2013)
merR R	ACGTCCTTNRGCTTGTGCTCG		
intI1 F	CCTCCCGCACGATGATC	Tn*21* Class 1 integron-integrase *intI1* identification	(Goldstein *et al.*, 2001)
intI1 R	TCCACGCATCGTCAGGC		
tnpA F	GCATGCACGCGGCGGCGTAGCTCGACGCTT	Tn*21* identification	This study
Tn*21* IR	CTCAGAAAACGGAAAATAAAGCACGCTAAG	Tn*21* identification and Forward primer for epicPCR	
sul1 F	CGCACCGGAAACATCGCTGCAC	Sulphonamide resistance *sul1* identification	(Pei *et al.*, 2006)
sul1 R	TGAAGTTCCGCCGCAAGGCTCG		
qacE1 F	TAGCGAGGGCTTTACTAAGC	Partial quaternary ammonium compound resistance *qacEΔ1* gene identification	(Wang *et al.*, 2007)
qacE1 R	ATTCAGAATGCCGAACACCG		
aadA1 F	AGCTAAGCGCGAACTGCAAT	Aminoglycoside *aadA1* resistance identification	(Zhu *et al.*, 2013)
aadA1 R	TGGCTCGAAGATACCTGCAA		
bla CTX F	CGCTTTGCGATGTGCAG	Beta lactamase CTX resistance identification	(Poirel *et al.*, 2001)
bla CTX R	ACCGCGATATCGTTGGT		
bla TEM F	GAGTATTCAACATTTCCGTGTC	Beta lactamase TEM resistance identification	(Zaniani *et al.*, 2012)
bla TEM R	TAATCAGTGAGGCACCTATCTC		
bla OXA-2 F	ACGATAGTTGTGGCAGACGAAC	Beta lactamase OXA-2 resistance identification	

Primer	Sequence	Description	Reference
bla OXA-2 R	ATYCTGTTTGGCGTATCRATATTC		(Hasman *et al.*, 2005)
bla SHV F	TTATCTCCCTGTTAGCCACC	Beta lactamase SHV resistance identification	(Weill *et al.*, 2004)
bla SHV R	GATTTGCTGATTTCGCTCGG		
pDOC-G_lin_F	TGGTACCCGGGATCCAAG	Linearisation of pDOC-G to insert *mrfp1*	This study
pDOC-G_lin_R	ACCGGTCAATTGGCTGGAG		
pDOC-rfp_HiFi_F	AGCTTGGATCCCGGGTACCATGGAATTCCTGCTGCGGAG	HiFi assembly of pDOC-R	This study
pDOC-rfp_HiFi_R	GCTCCAGCCAATTGACCGGTTTAAGCACCGGTGGAGTG		
pDOC-R-*glmS*-F	GCGCGGATCCAATCAAACATCCTGCCAACTCCATGTGACAAAC CGTCTGCATCGATAGAGTAT	Insertion of *glmS* homology region to pDOC-R for chromosomal insertion	This study
pDOC-R-*glmS*-R	CTCGAGGGCTACGAGAAGCAAAATAGGACAAACAGGTGACAA ATATCCTCCTTAGTTCCGCGC		
Integron F	GCCTGACGATGCGTGGAGACCGAAACCTTG	Tn*21* Class 1 integron amplification	This study
Integron R	CCGAACGTTCGGAGGCTCCTCGCTGTCCAT		
incP oriTF	CAGCCTCGCAGAGCAGGAT	Plasmid incompatibility group P identification	(Götz *et al.*, 1996)
incP oriTR	CAGCCGGGCAGGATAGGTGAAGT		
HS286	GGGATCCTCSGCTKGARCGAMTTGTTAGVC	All class I integron amplification	(Stokes *et al.*, 2001)
HS287	GGGATCCGCSGCTKANCTCVRRCGTTAGSC		
HS458RC V4	GTTGCTGCTCCATAACATCA	Class 1 integron reverse compliment and variation for Tn*21* epicPCR	Based on (Holmes *et al.*, 2003)
27F	AGAGTTTGATCMTGGCTC		(Marchesi *et al.*, 1998)
R1492	TACGGYTACCTTGTTACGACTT	*rrsA* gene amplification for sequencing and epicPCR	(Murrell, McDonald and Bourne, 1998)
R1542	AAGGAGGTGATCCAGCCGCA		(Hall *et al.*, 1999)
tnpM epic F	AATCACAAGCGTCCGGTTTGAC	Tn*21* identification fragment epicPCR andTn*21* cassette array PCR	This study
R1 – F2' 1	GAGCCAKGATCAAACTCTAATTGGTTGACGAAGCCGA	Bridging primer between *rrsA* and Tn*21* *tnpM* fragment	This study
R1 - F2' Tn*21*	GAGCCATGATCAAACTCTAAGCGTCGAGCTACGCCGCCGCGT GCATGC	Bridging primer between *rrsA* and Tn*21* *tnpA* fragment epicPCR	This study
27F Block F	TTTTTTTTTTAGAGTTTGATCMTGGCTC-3Csp	Blocking primers epicPCR prevent amplification of unfused product with 3 carbon spacers	This study
27F Block R	TTTTTTTTTTGAGCCAKGATCAAACTCT-3Csp		
Nanopore-1492R	TTTCTGTTGGTGCTGATATTGCTACGGYTACCTTGTTACGACTT	For four primer nanopore PCR Barcoding of epicPCR Amplicons	This study
Tn*21* F-nanopore	TTTCTGTTGGTGCTGATATTGCCTCAGAAAACGGAAAATAAA		

	GCACGCTAAG

Table 2.3 Table of primers used for end-point PCR reactions throughout the thesis.

2.1.3.1 Primers and Probes for qPCR

All qPCR fluorescently labelled probes were supplied by Eurofins Genomics (Ebersberg, Germany), with 0.01 µmole synthesis, HPLC purification,

and made up to 100 µM using qPCR probe dilution buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) supplied. For the 5' fluorophore FAM (ex

495 nm | em 520 nm), 3' quencher BHQ-1 (Black Hole Quencher one), working stocks of 10 µM were prepared in HPLC grade H_2O, aliquoted to

reduce freeze/thaw degradation, and stored at -20°C in a black out storage bag.

Primer Name	Primer Sequence 5' to 3'	Use	Reference
Tn*21* qPCR_forward	GGAAAATAAAGCACGCTAAGCC	Amplification of Tn*21* IR to *tnpA*	This study
Tn*21* qPCR-reverse	TGGGAGCACATCAACCTGAC		
Tn*21* qPCR_Probe	[FAM]-CGCTGCTGCGCCATAGGTAATCACGG-[BHQ1]	Probe binding Tn*21* amplicon	
*rrsA*_qPCR_forward	TTACTGGGCGTAAAGCGCAC	Amplification of *rrsA* qPCR control	This study
*rrsA*_qPCR_reverse	TCTACGCATTTCACCGCTACAC		
*rrsA*_qPCR_probe	[FAM]-CCGGGCTCAACCTGGGAACTGCAT-[BHQ1]	Probe binding *rrsA* amplicon	

Table 2.4 Table of primers and probe sets used for qPCR reactions for identification of Tn*21* (and Tn*21*-like elements) and the 16S RA gene, *rrsA.*

2.1.4 Media

All media components were sourced from Oxoid (Basingstoke, UK) and sterilised by autoclaving at 121 °C for 15 minutes at 15 psi unless otherwise stated.

2.1.4.1 Lysogeny Broth (LB) and Agar

10 g of tryptone, 5 g of yeast extract and 5 g of NaCl were dissolved in 1 L of reverse osmosis H_2O.

2.1.4.1.1 LB Agar

15 g/L of agar was added to LB broth before autoclaving. Where appropriate, antibiotics were added to the agar once it had cooled to 50 °C.

2.1.4.2 LB Broth (Miller)

10 g of tryptone, 5 g of yeast extract and 10 g of NaCl were dissolved in 1 L of reverse osmosis H_2O and autoclaved.

2.1.4.3 SOC (Super Optimal broth with catabolite repression)

20 g of tryptone, 5 g of yeast extract, 2 mL of 5 M NaCl, 2.5 mL of 1 M KCl, 10 mL of 1 M of $MgCl_2$ and 10 mL of 1 M of $MgSO_4$ were dissolved in 1 L of reverse osmosis H_2O and autoclaved. 20 mL of filter sterilised 1 M glucose was added per litre of SOC after autoclaving.

2.1.4.4 Super Optimal Broth (SOB)

20 g of tryptone, 5 g of yeast extract, 0.5 g of NaCl and 20 mL of KCl (250 mM) were dissolved in 950 mL of reverse osmosis H_2O, and the pH was adjusted to pH 7.0 with

NaOH (5 M). The volume was adjusted further to 1 L prior to being autoclaved with reverse osmosis H_2O. 5 mL of sterile $MgCl_2$ (2 M) was added after autoclaving, before use.

2.1.4.5 Tryptone Bile X-Glucuronide (TBX) Agar (Merck, Germany)

15 g of agar, 1.5 g of bile salts, 0.075 g of X-β-D-glucuronide and 20 g of peptone were dissolved in 1 L of reverse osmosis H_2O and autoclaved.

2.1.4.6 CHROMagar ESBL Agar (CHROMagar, France)

15 g of agar, 17 g of peptone and yeast extract and 1 g of Chromogenic mix were dissolved in 1 L of reverse osmosis H_2O and autoclaved. 0.57 g CHROMagar ESBL supplement was added after autoclaving once cooled to 50 °C.

2.1.4.7 Phosphate Buffered Saline (PBS)

8 g of NaCl, 0.2 g of KCl, 1.44 g of Na_2HPO_4, and 0.24 g of KH_2PO_4 was dissolved in 800 mL Milli-Q water, and pH adjusted to pH 7.4. The final volume was adjusted to 1 L with milli-Q water prior to autoclaving.

2.1.4.8 Maximum Recovery Diluent (MRD)

1 g of peptone, 8.5 g of NaCl was dissolved in 800 mL Milli-Q water, and pH adjusted to pH 7.0. The final volume was adjusted to 1 L with Milli-Q water prior to autoclaving.

2.1.5 Antibiotic Stock Solutions

Antibiotic stock solutions were prepared at 1000x working concentration.

2.1.5.1 Ampicillin

100 mg mL^{-1} of ampicillin sodium salt (Amp) (Sigma-Aldrich, UK, A0166) was dissolved in sterile milli–Q H$_2$O (Millipore), filter sterilized through a 0.22 µm filter (Sartorius, Epsom, UK) and stored in 1 mL aliquots at − 20 °C in sterile 1.5 mL microcentrifuge tube.

2.1.5.2 Cefotaxime

5 mg mL^{-1} of cefotaxime sodium salt (Sigma-Aldrich, UK, 219504) was dissolved in sterile milli–Q H$_2$O (Millipore), filter sterilized through a 0.22 µm filter, and stored in 1 mL aliquots at − 20 °C in sterile 1.5 mL microcentrifuge tubes.

2.1.5.3 Chloramphenicol

35 mg mL^{-1} of chloramphenicol (Chlor) (Sigma-Aldrich, UK, C0378) was dissolved in 100% ethanol and stored at -20 °C in sterile 1.5 mL microcentrifuge tubes.

2.1.5.4 Erythromycin

10 mg mL^{-1} of erythromycin (Ert) (Sigma-Aldrich, UK, E5389) was dissolved in 100% ethanol and stored at -20 °C in sterile 1.5 mL microcentrifuge tubes.

2.1.5.5 Gentamycin

7 mg mL^{-1} of gentamycin sulfate salt (Gent) (Sigma-Aldrich, UK, G1264) was dissolved in sterile milli–Q H$_2$O (Millipore), filter sterilized through a 0.22 µm filter, and stored in 1 mL aliquots at -20 °C in sterile 1.5 mL microcentrifuge tubes.

2.1.5.6 Kanamycin

50 mg mL^{-1} of kanamycin monosulfate (Kan) (Gibco, 11578676) was dissolved in sterile milli–Q H$_2$O (Millipore), filter sterilized through a 0.22 µm filter, and stored in 1 mL aliquots at -20 °C in sterile 1.5 mL microcentrifuge tubes.

2.1.5.7 Streptomycin

50 mg mL^{-1} of streptomycin sulfate salt (Strep) (Sigma-Aldrich, UK, S9137) was dissolved in sterile milli–Q H$_2$O (Millipore), filter sterilized through a 0.22 µm filter, and stored in 1 mL aliquots at – 20 °C in sterile 1.5 mL microcentrifuge tubes.

2.1.5.8 Tetracycline

12 mg mL^{-1} of tetracycline hydrochloride (Tet) (Sigma-Aldrich, UK, T7660) was dissolved in 100% ethanol and stored at -20 °C in sterile 1.5 mL microcentrifuge tubes.

2.1.5.9 Trimethoprim

10 mg mL^{-1} of trimethoprim lactate salt (Tri) (Sigma-Aldrich, UK, T0667) was dissolved in 10% ethanol and stored at -20 °C in sterile 1.5 mL microcentrifuge tubes.

2.1.6 Mercury (II) Chloride Solution

Mercury (II) chloride (HgCl$_2$) (Sigma-Aldrich, UK, 215465) at a concentration of 55 mM was dissolved in sterile milli-Q (Millipore) water at 1000X working concentration, filter-sterilised using a 0.22 µm PES filter (Sartorius) and stored in aliquots away from light at room temperature in sterile 1.5 mL microcentrifuge tubes.

2.1.7 Oligonucleotides

Oligonucleotides were supplied by Sigma as 0.05 µmole synthesis scale, desalt in water at a concentration of 100 µM and are shown in Table 2.3. Working stock solutions at 10 µM were prepared by dilution of the supplied stock of oligonucleotides using nuclease-free water (unless stated otherwise) were prepared and stored at -20 °C.

qPCR oligonucleotides and probes were supplied by Eurofins in a dehydrated form and are shown in Table 2.4. The oligonucleotides were resuspended in nuclease-free water to a concentration of 100 µM and working stock solutions of a concentration of 10 µM were prepared from this and stored at -20 °C. The probes used were dual labelled with a reporter, 6-carboxy-fluorescein (FAM), and quencher, Black Hole Quencher 1 (BHQ1). The probes were resuspended in 10 mM Tris-HCl; 1 mM EDTA; pH 8 to aliquots of 20 µM.

2.1.8 Restriction Enzymes

Restriction enzymes and buffers were supplied by New England BioLabs (NEB) and stored at -20 °C.

2.1.9 Buffers

All chemicals were by Sigma-Aldrich and were analytical grade unless otherwise stated.

2.1.9.1   50x TAE Buffer

242 g of Tris base, 57.1 mL of acetic acid (glacial) and 37.2 g of $Na_2EDTA \cdot 2H_2O$ was dissolved in reverse osmosis water and made up to a final volume of 1 L with reverse osmosis water (Ausubel et al., 2003).

2.1.9.2   1x TAE Buffer

100 mL 50x TAE buffer was combined with 4,900 mL of reverse osmosis water.

2.1.9.3   1 M Tris–HCl pH 7.5

121.14 g tris was dissolved in 800 mL of reverse osmosis water, adjusted to pH 7.5 with dilute 1 M HCl, and made up to a final volume of 1 L with reverse osmosis water, prior to autoclaving.

2.1.9.4   1 M Tris–HCl pH 8.0

121.14 g tris was dissolved in 800 mL of reverse osmosis water, adjusted to pH 8.0 with HCl, and made up to a final volume of 1 L with reverse osmosis water, prior to autoclaving.

2.1.9.5   10 mM Tris–HCl pH8.0 with 50 mM NaCl

1 mL of 1M Tris–HCl pH8.0 was added to 0.292 g of NaCl and made up to 100 mL with Milli-Q water prior to autoclaving.

2.1.9.6   0.5 M EDTA pH 8.0

186.1 g of EDTA was dissolved in 800 mL of reverse osmosis water, adjusted to pH 8.0 with NaOH pellets, and made up to a final volume of 1 L prior to autoclaving.

2.1.9.7   Tris-EDTA (TE) Buffer pH 8.0

10 mL of 1 M Tris-HCl pH 8.0, 2 mL of 0.5 M EDTA pH 8.0 and 982 mL of reverse osmosis water was combined, and autoclaved.

2.1.9.8   Lysis Buffer

2.5 mL of 50 mM Tris–HCl [pH 8], 5 mL of 50 mM EDTA, 0.5 g of *N*-lauroyl sarcosine, 100 µg mL^{-1} proteinase K and was made up to 50 mL with Milli-Q water. Used in pulse field gel electrophoresis.

2.1.9.9   Wash Buffer

1 mL of 20 mM Tris–HCl [pH 8] and 5 mL of 50 mM EDTA [pH 8.0] were added to 44 mL of Milli-Q water. Used in pulse field gel electrophoresis.

2.1.9.10 TK Buffer

2 mL of 1M Tris-HCl [pH 7.5] and 0.447 g of KCl was made up to 100 mL with Milli-Q water prior to autoclaving. Used in epicPCR.

2.1.9.11 STT Emulsion Oil

2.25 mL of Span 80, 200 µL of Tween 80 and 25 µL of Triton X-100 were made up to 50 mL mineral oil. Used in epicPCR.

2.1.9.12 ABIL Emulsion Oil

2 mL of ABIL EM90 oil and 25 µL of Triton X-100 were made up to 50 mL mineral oil. Used in epicPCR.

2.1.10 Solutions

2.1.10.1 30% D-Glucose ($C_6H_{12}O_6$) (% W/V)

150 g glucose (Sigma-Aldrich, G7528) was dissolved in reverse osmosis water, made up to a final volume of 500 mL with reverse osmosis water, and filter sterilized through a 0.22 μm filtration unit (Nalgene).

2.1.10.2 30% Sucrose ($C_{12}H_{22}O_{11}$) (% W/V)

150 g sucrose (Sigma-Aldrich, C0389) was dissolved in reverse osmosis water, made up to a final volume of 500 mL with reverse osmosis water, and filter sterilized through a 0.22 μm filtration unit (Nalgene).

2.1.10.3 10% Arabinose ($C_5H_{10}O_5$) (% W/V)

1 g L-Arabinose (Sigma-Aldrich, 10839) was dissolved in reverse osmosis water, made up to a final volume of 10 mL in reverse osmosis water, and filter sterilized through a 0.22 μm filter and stored at room temperature.

2.1.10.4 10% Glycerol (% V/V)

12.6 g (10 mL) glycerol (Sigma-Aldrich, G5516) was added to 90 mL reverse osmosis water, and autoclaved.

2.1.10.5 50% Glycerol (% V/V)

63 g (50 mL) of glycerol (Sigma-Aldrich, UK) was added to 50 mL reverse osmosis water, and autoclaved.

2.1.10.6 Water Saturated Diethyl Ether

200 mL of diethyl ether (Sigma-Aldrich, UK) was added to an excess of sterile Milli-Q water (greater than 200 mL). Prior to use the container was shaken to mix and separate the phases of diethyl ether and sterile Milli-Q water.

2.1.10.7 Water Saturated Ethyl Acetate

200 mL of ethyl acetate (Sigma-Aldrich, UK) was added to an excess of sterile Milli-Q water (greater than 200 mL). Prior to use the container was shaken to mix and separate the phases of ethyl acetate and sterile Milli-Q water.

2.1.10.8 70% (V/V) Ethanol

70 mL of ethanol was added to 30 mL of sterile Milli-Q water.

2.1.10.9 80% (V/V) Ethanol

80 mL of ethanol was added to 20 mL of sterile Milli-Q water. Used for MinION library preparation.

2.2   Methods

2.2.1   Growth of Bacterial Strains in LB Broth

Escherichia coli was routinely grown in sterile LB broth, with appropriate antibiotic(s), at 37 °C with shaking (200 rpm) for 16-20 hours. If a temperature sensitive plasmid was present, the growth temperature was adjusted accordingly.

E. coli was routinely grown on LB agar, with appropriate antibiotic(s), at 37 °C for 16–20 hours. If a temperature sensitive plasmid was present, the growth temperature as well as incubation time was adjusted accordingly. e.g. 30 ˚C for 40-48 hours.

Antibiotic stock solutions (Section 2.1.5) were added to sterile LB broth so that the final concentrations of antibiotics used were as follows; 50 µg mL^{-1} Kanamycin (kan$_{50}$), 100 µg mL^{-1} Ampicillin (amp$_{100}$), and 35 µg mL^{-1} Chloramphenicol (chlor$_{35}$).

2.2.2   Long Term Storage of Bacterial Strains

A single bacterial colony was inoculated into 5 mL of SOB broth and incubated at 30°C for 16-20 hr with shaking (250 rpm), overnight. 500 µL of an overnight culture was transferred into the Microbank^{TM} tube (Pro-Lab Diagnostics) and incubated at room temperature for 2-3 min, according to manufacturers' instructions. The liquid was aspirated off, and the Microbank^{TM} was stored at -80 ˚C. ESBL strains isolated using CHROMagar were stored at -80 °C in 25% glycerol solution by adding 500 µL of sterile 50% glycerol to 500 µL suspension of cells from a 10 mL of overnight culture in a 2 mL cryogenic storage tube (Nalgene).

2.2.3 DNA Manipulation and Purification

All DNA work was conducted using sterile filter pipette tips (Starlab, TipOne® Filter Tips). Gloves were always worn to reduce DNase and RNase contamination of sample.

2.2.4 Preparation of Electrocompetent *E. coli* Cells

10 mL LB broth (Lennox) was inoculated with a single colony and incubated at 37 °C, shaking at 200 rpm for 12-16 hours. 2 mL of the culture was then used to inoculate 200 mL of sterile LB (Lennox) in an Erlenmeyer flask and incubated at 37 °C, shaking at 200 rpm until an OD_{600} of between 0.5-0.6 was achieved. The culture was then incubated on ice for 5 min and centrifuged at 6,000 $x\,g$ for 10 min at 4 °C (Sigma 1–16K). The pellet was then resuspended in 100 mL of sterile 10% (v/v) glycerol at 4 °C and centrifuged at 6,000 $x\,g$ for 10 min at 4 °C. The bacterial pellet was then resuspended in fresh 100 mL of 10% glycerol solution at 4 °C twice. The washed cell pellet was then resuspended in 960 μL 10% glycerol solution and 40 μL aliquots of electrocompetent cells were stored in 1.5 mL microcentrifuge tubes at -80 °C.

2.2.5 Indole Testing

Pure cultures of bacterial isolates were grown on LB agar incubated at 37 °C for 16 hours. One to two drops of Spot Indole Reagent (Pro-Lab Diagnostics, UK) were dispensed onto the tip of a cotton swab. The saturated swab was then touched to the surface of a single colony from the pure culture on the LB agar medium. The

cotton swab was then observed for 3 min to identify the formation of a blue colour, indicating a positive result. A negative result showed no change in swab colour.

2.2.6 Oxidase Testing

Pure cultures of bacterial isolates were grown on LB agar incubated at 37 °C for 16 hours. Oxidase test strips (Merck, Germany) were then touched onto a single colony. The test strip was observed for colour change over the duration of 1 min. A positive reaction was indicated by colour change of the test strip from pink to dark blue. A negative reaction was indicated when no colour change observed over 1 min.

2.2.7 Wastewater Sampling

Samples were taken from a wastewater treatment plant in the East Midlands. Influent samples were taken from two different inlets to the treatment plant. One comes from a village and one comes from a University campus. Effluent samples were collected from an outlet pipe from the same site and water from the brook 10 m downstream of the outlet pipe from the wastewater treatment plant. From January 2020 permission was granted for sampling of effluent from the treatment plant itself. One further sample was taken from a drainage ditch, which appeared to contain grey water (Figure 2.1).
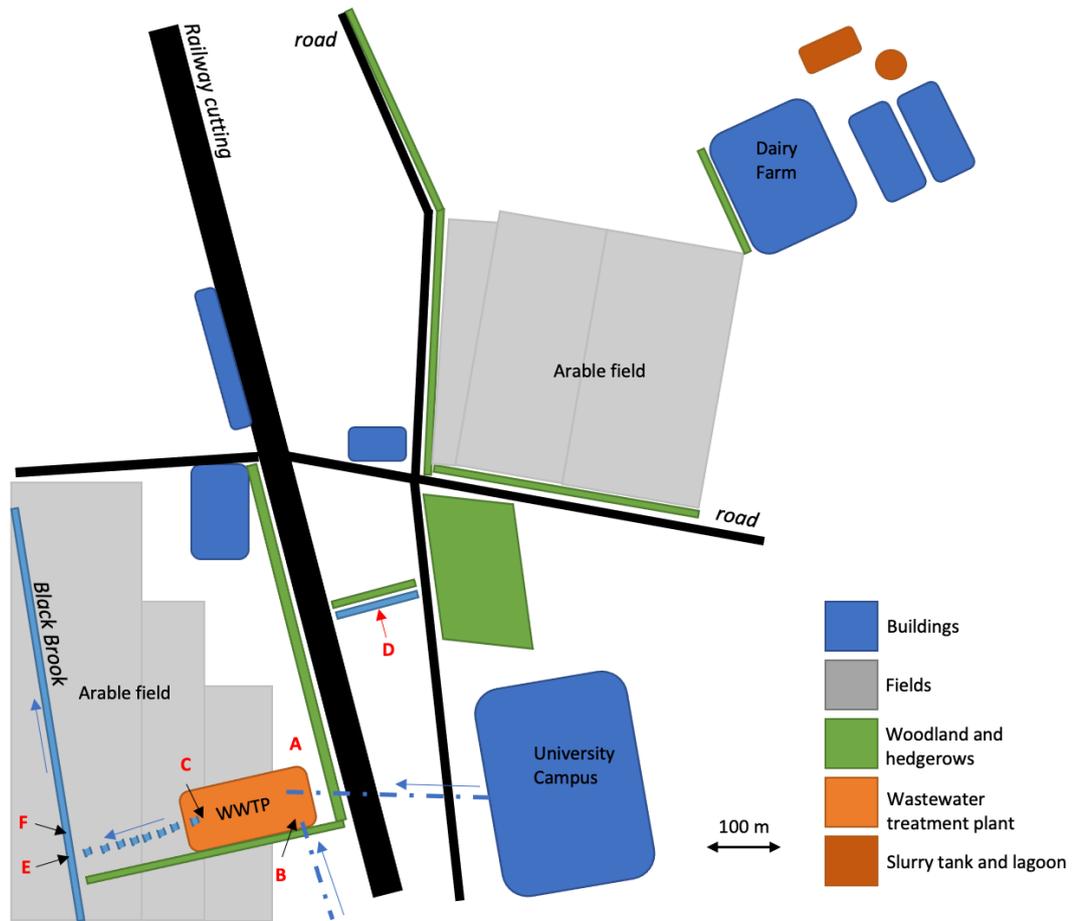
Figure 2.1 A map of the sampling sites, showing the location of each site in relation to each other.

A: Campus influent sample site B: Village influent sample site C: Effluent sample site D: Grey water contaminated ditch sample site E: Effluent pipe outlet to the Black brook sample site F: Black brook downstream of the effluent outlet sample site. (Modified from Swift *et al.,* 2019)

2.2.8 Bacterial Isolation

Wastewater samples were collected from the various sample locations (Figure 2.1).

Ten-fold serial dilutions of the wastewater were performed using with sterile

Maximum Recovery Diluent (MRD) (Sigma-Aldrich) 100 μL of each dilution was

spread plate in triplicate onto Tryptone Bile X-glucuronide (TBX) agar (Sigma-Aldrich).

The agar plates were then incubated at 37 °C for 24 hours. Putative *E. coli*, colonies

that grew containing blue/green pigment were then streaked for single colonies on TBX agar and incubated at 37 °C for 24 hours. Once a pure culture was obtained, the isolates were then inoculated onto Luria Bertani (LB) agar plates and incubated at 37 °C overnight.

For extended spectrum beta-lactamase (ESBL) *E. coli* isolation CHROMagar ESBL (CHROMagar, France) was used due to its chromogenic identification of *E. coli* (purple). 100 μL aliquots from serial dilutions of wastewater were spread onto these agar plates and incubated at 37 °C for 24 hours. Putative *E. coli* were streaked for single colonies on CHROMagar ESBL agar and incubated at 37 °C for 24 hours. Once a pure culture was obtained, the isolates were then inoculated onto Luria Bertani (LB) agar plates and incubated at 37 °C overnight.

Putative *E. coli* were confirmed with oxidase and indole tests. The indole test used was a RapID spot test (Remel, USA). The isolates were then screened for resistance to mercury (II) chloride. Bacterial isolates were streaked for single colonies on to LB agar plates supplemented with 25 μg mL^{-1} HgCl$_2$ and incubated for 24 hours at 37 °C. *E. coli* K-12 J53 pMG101 and enteroaggregative *E. coli* strain 042 were used as positive controls, for mercury resistance and *E. coli* K-12 MG1655 as a negative control strain. Confluent growth indicated a positive result. Mercury resistant isolates were then cryogenically preserved at -80 °C.

2.2.9   Replica Plating

From October 2019 until January 2020, a number of water samples were collected from the two influent sites and a further sample was collected in January 2020 of

effluent from the treatment plant. These samples were enumerated on TBX agar using serial dilutions as described above. Green colonies (glucuronidase positive organisms) were picked and plate purified for pure cultures. The remainder of the samples were centrifuged at 3,000 x g (Heraeus Megafuge 40 R, Thermo Scientific) for 15 min to pellet the cells. The supernatant was removed and 1 mL of MRD was used to resuspend the pellet before adding 50% (v/v) glycerol solution for cryogenic preservation at -80 °C.

The 336 colonies isolated were then screened using a replica plating technique. Each individual well in a series of 96-well plates was loaded with 200 μL of sterile LB broth and a single colony of each isolate was suspended in the media in the individual wells. A 96 solid pin multi-replicator (Merck, Germany) was then used to transfer cell suspensions onto different LB agar plates containing different selective agents: 25 μg mL^{-1} of mercury (II) chloride, 50 μg mL^{-1} of streptomycin, 10 μg mL^{-1} of trimethoprim, 5 μg mL^{-1} of cefotaxime and 12 μg mL^{-1} of tetracycline. Each condition was tested in triplicate. The agar plates were then incubated at 37 °C overnight and the presence and absence of growth was recorded.

2.2.9.1 Identifying Linkage Between Antimicrobial Resistances from Replica Plates

In order to determine the expected probability of a combination of resistances, first the occurrence frequency (F) of resistance to a single antimicrobial needs to be calculated.

$$F = \frac{number\ of\ isolates\ resistant\ to\ antimicrobial}{total\ number\ of\ isolates}$$

The probability of carriage of two antimicrobial resistance genes in a sample was then calculated as the multiplication of occurrence frequency of each individual antimicrobial selected. F_x and F_y represent two occurrence frequencies of different resistances to two different antimicrobial agents. F_n denotes that more than two occurrence frequencies can be multiplied to calculate predicted occurrence (P) of more than two antimicrobial resistances.

$$Predicted\ Occurrence\ =\ F_x\ *\ F_y\ *\ F_n$$

Linkage was then defined by occurrence frequency (F) divided by predicted occurrence (P).

$$Linkage\ value = \frac{F}{P}$$

If the value calculated was greater than 1, then the probability of occurrence of compound resistance to the multiple antimicrobials was unlikely to be due to chance.

2.2.10 Quantification of DNA

There were two methods used to quantify DNA concentration.

2.2.10.1 Method One: Quantification of DNA Concentration Using Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific)

The amount of DNA was quantified using a Nanodrop 1000 (Thermo Fisher Scientific) by measuring the optical density of 1 μL volume of a sample. 1 μL of nuclease-free water was used to initialise the spectrophotometer and then 1 μL of buffer or nuclease-free water with no DNA elution was used to blank the photometer.

2.2.10.2 Method Two: Quantification of DNA Concentration Using Qubit 3 Fluorometer (Invitrogen, USA)

The amount of DNA was quantified using the Qubit 3 (Invitrogen, USA). The machine was calibrated by making up 2 manufacturer provided standards: low dsDNA and high dsDNA. 10 μL of each standard was added to 190 μL of a 1 in 200 dilution of Qubit reagent in Qubit dsDNA HS Buffer (Q32851, Q32854) in 500 μL thin-walled PCR tubes. 1-10 μL sample was added to 199-190 μL respectively of a 1 in 200 dilution of Qubit reagent in Qubit dsDNA HS Buffer in 500 μL thin-walled PCR tube and quantified.

2.2.11 Crude DNA Preparation

DNA was prepared from bacterial isolates by boiling 1-2 colonies in 100 μL water for 10 min, snap freezing on ice for 5 min and then samples were centrifuged (Sigma 1-16K) at maximum speed (16,000 $x\,g$) for 7 min. The supernatant was recovered and transferred into a fresh microcentrifuge tube.

2.2.12 Genomic DNA Extraction for Sequencing

The Monarch Genomic DNA Purification kit (New England BioLabs, UK) was used for isolates needing sequencing, following the manufacturer's protocol for Gram-Negative bacteria. The DNA was eluted in 100 μL of sterile nuclease free water as required by Oxford Nanopore for best Sequencing results.

2.2.13 Agarose Gel Electrophoresis

A 1% agarose gel was made by adding 0.5 g of agarose (Sigma-Aldrich, A9539) to 50 mL of 1x TAE buffer, which was then dissolved by microwave heating. The agarose solution was left to cool to less than 50 °C, ethidium bromide was added to a final concentration of 0.2 µg mL^{-1} and the gel poured into a casting tray. After setting, 5 µL samples of the PCR product, or other DNA, was combined with 1 µL of 6x loading dye (NEB, B7024) and loaded into the agarose gel well. Either 5 µL of 100 bp ladder (NEB, N0551) or 1 kb ladder (NEB, N0552) was loaded as a reference. The agarose gel was electrophoresed at 80 V for 40-60 min. DNA was visualized on a UV trans illuminator (BIO-RAD Gel Doc XR+).

2.2.14 Environmental DNA Extraction for Metagenomics and qPCR

DNA was extracted from 250 µL 25% (v/v) glycerol suspension of cells using the Qiagen DNeasy Power Water kit (Qiagen, UK). The 5 mL filter lysis tube in the Power Water kit was swapped out for Qiagen Pathogen Lysis tubes (19092). The cell suspension was lysed with 1 mL of solution PW1 and vortexed using a FastPrep 5G bead beating grinder (MP Biomedicals, USA) and lysis system at 6 m s^{-1} for 45 s. The rest of the DNA extraction process was completed as described in the manufacturer's instructions in the Qiagen DNeasy Power Water kit (14900-100-NF). The extracted DNA was eluted with 100 µL of Elution Buffer.

2.2.15 Plasmid Purification

There were two methods of plasmid purification that were used depending on the amount of plasmid DNA required and the size of the plasmid itself.

2.2.15.1 Monarch Plasmid Miniprep Kit (New England BioLabs, USA)

A single colony from a freshly streaked and incubated LB agar plate was inoculated into 5 mL of sterile LB broth containing the appropriate antibiotic or antimicrobial selection and incubated 16 hours at 37 °C at 200 rpm. The cells were harvested by centrifugation of the culture at 4,200 $x\,g$ for 10 min (Hettish EBA 12R). The cell pellet was then resuspended in 200 µL Plasmid Resuspension Buffer (B1) by vortexing and the manufacturer's protocol for plasmid purification (#T1010) was followed. Plasmid DNA was eluted into 30 µL of pre-warmed (55 °C) DNA elution buffer (EB) (New England BioLabs, UK).

2.2.15.2 Qiagen Plasmid Midiprep Kit

A single colony from a freshly grown culture on an LB agar plate was inoculated into 100 mL of sterile LB broth containing the appropriate antibiotic or antimicrobial selection and incubated 16 hours at 37 °C at 200 rpm. The culture was then harvested by centrifuging the culture at 4,400 $x\,g$ (Beckman J2-21) for 10 min and the manufacturer's protocol for purification of a low copy number plasmid was followed and eluted into 300 µL of TE buffer at pH 8.

2.2.16 Polymerase Chain Reaction (PCR)

All PCR work was conducted using sterile filter pipette tips (Starlab, TipOne® Filter Tips). Gloves were always worn to reduce DNase and RNase contamination of sample. PCR reaction mixtures were set up on ice unless hot-start DNA polymerase was used. Unless otherwise stated all reaction mixtures contained 10 µL of stated DNA polymerase 2X master mix, 1 µL of forward primer (10 µM), 1 µL of reverse primer (10 µM), and x µL of template DNA (0.1-2 ng of plasmid DNA, or 50-100 ng gDNA), which were added to a PCR reaction mixture. The final reaction volume was made up to 20 µL with nuclease-free water in a 0.2 mL PCR tube. PCR was conducted in a thermocycler (BIO-RAD c1000).

2.2.16.1 Confirmation of *E. coli* (See Table 2.3)

To confirm that mercury resistant bacterial isolates were *E. coli*, a PCR was performed to amplify the 16S rRNA gene, (*rrsA*), using bacterial DNA (Section 2.2.11); Q5 2X Master Mix (New England BioLabs, USA), the primer pair '16S F' and '16S R'. The 16S PCR was made up to the final volume of 25 µL per reaction with sterile nuclease-free water (Thermo Fisher, UK). The PCR conditions were as follows: an initial denaturation of 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 67 °C for 15 s, 72 °C for 20 s with a final incubation at 72 °C for 2 min. PCR products were electrophoresed on a 1% 1X TAE agarose gel at 85 V for 1 hour. The resulting PCR product was purified and concentrated using a Monarch™ PCR and DNA Cleanup Kit (New England BioLabs, USA) and sent off for sequencing using Sanger sequencing

technology (Eurofins Genomics, Germany)**.** The sequenced PCR product was then identified using NCBI Blast.

2.2.16.2 *merA, C, R* and *intI1* PCR (See Table Table 2.3)

To confirm mercury resistance phenotype results for the isolates; a multiplex PCR was performed which screened for *merA, merC, merR* and *intI1* using the primers '*merA* F,' *merA* R,' '*merC* F,' '*merC* R,' '*merR* F ',' *merR* R,' '*intI1* F' and '*intI1* R' (Table 2.3). DreamTaq green 2X master mix (Thermo Fisher, UK) and these primers were made up to the final volume of 25 μL per reaction with sterile nuclease-free water (Thermo Fisher, UK). The following PCR conditions were used: an initial denaturation temperature of 95 °C for 5 min, followed by 30 cycles of 95 °C for 60 s, 63 °C for 60 s, 72 °C for 90 s with a final incubation at 72 °C for 10 min (Bio-Rad). PCR products were electrophoresed on a 2.2% 1X TAE agarose gel at 85 V for 1 hour.

2.2.16.3 Confirmation of Tn*21* (See Table of Table 2.3)

After confirming the isolates possessed the *mer* genes and *intI1*, Tn*21* needed to be identified. A pair of PCR primers were designed, targeting the highly conserved inverted repeat region of the Tn*21*,'Tn*21* IR' and part of the *tnpA* gene '*tnpA* F' (Table 2.3). Q5 2X master mix (New England BioLabs, UK) and these primers were made up to the final volume of 25 μL per reaction with sterile nuclease-free water (Thermo Fisher, UK). The following PCR cycling conditions were used: an initial denaturation of 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 71.5 °C for 15 s, 72 °C for 20 s

86

with a final incubation at 72 °C for 2 min. PCR products were electrophoresed on a 1% 1X TAE agarose gel at 85 V for 1 hour.

2.2.16.4 Identification of Classic Antimicrobial Resistance Genes Found Within Tn*21* Class I Integron (See Table 2.3)

To identify whether the Tn*21* positive isolates identified in 2.2.16.3 possessed any of the AMR genes associated with Tn*21*: the presence of *qacEΔ1, sul1* and *aadA1,* were tested for by PCR (Table 2.3). Q5 master mix 2X (New England BioLabs, USA) and these primers were made up to the final volume of 25 μL per reaction with sterile nuclease-free water (Thermo Fisher, UK). The following PCR cycling conditions for *sul1* were: an initial denaturation temperature of 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 66 °C for 15 s, 72 °C for 20 s with a final incubation at 72 °C for 2 min. For *aadA1*: an initial denaturation temperature of 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 67 °C for 15 s, 72 °C for 20 s with a final incubation at 72 °C for 2 min. For *qacEΔ1*: an initial denaturation temperature of 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 72 °C for 35 s with a final incubation at 72 °C for 2 min. PCR products were electrophoresed on a 1% 1X TAE agarose gel at 85 V for 1 hour.

2.2.16.5 Extended Spectrum Beta-Lactamase (ESBL) PCR (See Table 2.3)

Confirmed *E. coli* strains that were isolated from CHROMagar ESBL plates were also screened for the beta lactamase genes: *bla$_{CTX}$* ('*bla$_{CTX}$* F' and '*bla$_{CTX}$* R'), *bla$_{OXA-2}$* ('*bla$_{OXA-2}$* F' and '*bla$_{OXA-2}$* R'), *bla$_{SHV}$* ('*bla$_{SHV}$* F' and '*bla$_{SHV}$* R') and *bla$_{TEM}$* ('*bla$_{TEM}$* F' and '*bla$_{TEM}$* R'). Q5 master mix 2X (New England BioLabs, USA) and the primer pair for

bla_{CTX} or bla_{TEM} were made up to the final volume of 25 μL per reaction with sterile nuclease-free water (Thermo Fisher, UK). The following PCR conditions were used in a thermocycler (BIO-RAD c1000). For bla_{CTX}: an initial denaturation temperature of 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 65 °C for 15 s, 72 °C for 20 s with a final incubation at 72 °C for 2 min. For bla_{TEM}: an initial denaturation temperature of 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 62 °C for 15 s, 72 °C for 20 s with a final incubation at 72 °C for 2 min. For bla_{SHV} and bla_{OXA-2}, DreamTaq Green 2X Master Mix (Thermo Fisher, UK) and the primer pair for bla_{OXA-2} or bla_{SHV} were made up to the final volume of 25 μL per reaction with sterile nuclease-free water (Thermo Fisher, UK). The following PCR cycling conditions were used (BIO-RAD c1000): an initial denaturation temperature of 95 °C for 5 min followed by 30 cycles of 95 °C for 30 s, 55.5 °C for 30 s, 72 °C for 60 s with a final incubation at 72 °C for 5 min. PCR products were electrophoresed on a 1% 1X TAE agarose gel at 85 V for 1 hour.

2.2.17 Antimicrobial Susceptibility Testing (AST)

All isolates that possessed ESBL or mercury resistance genes were screened further for antibiotic susceptibility via disc diffusion assay, using the following antibiotics under the Clinical and Laboratory Standards Institute protocol (Clinical and Laboratory Standards Institute, 2018): Ampicillin 10 μg (AMP), Amoxicillin-clavulanic acid 20 μg and 10 μg (AMC), Cefoxitin 30 μg (FOX), Ceftazidime 30 μg (CAZ), Cefpodoxime 10 μg (CPD), Aztreonam 30 μg (ATM), Imipenem 10 μg (IPM), Streptomycin 10 μg (S10), Oxytetracycline 30 μg (OT), Ciprofloxacin 5 μg (CIP), Nalidixic Acid 30 μg (NA), Trimethoprim-sulfamethoxazole 1.25 μg and 23.75 μg

(SXT), Chloramphenicol 30 µg (C), Nitrofurantoin 300 µg (F) and Azithromycin 15 µg (AZM). All of which were sourced from Pro-Lab Diagnostics.

2.2.18 Mercury Resistant isolate *merA* Phylogenetic Tree Production

The *merA* PCR amplicons of all the bacteria containing mercury resistance genes, were purified using DNA Clean Up Kit (Monarch). The purified PCR products were then sent away with the corresponding merA F and merA R primers for Sanger sequencing (Eurofins Genomics). Forward and reverse reads generated from Sanger sequencing were aligned to produce 2X coverage. The sequences produced for each isolate were then aligned using the MAFFT multiple sequence alignment tool and a bootstrapped phylogenetic tree (100 bootstraps) was produced (Katoh and Standley, 2013).

2.2.19 Construction of *E. coli* Strains Containing a Chromosomal Copy of the Red Fluorescent Protein, *mrfp1*

The two *E. coli* strains: K-12 J53 RFP kan[r] and J53 RFP Azi[r] kan[r] were constructed from the strains J53 and J53 Azi[r] using the gene doctoring method (Lee *et al.*, 2009). This was in order to use the new strains as chromogenic and antibiotic selectable recipient strains for conjugation and mercury transposition experiments.

2.2.19.1 Construction of Red Fluorescent Protein Genome Engineering Plasmid pDOC-R

Plasmid pDOC-G (Lee *et al*, 2009), containing the *gfp* gene, was linearised using Q5 Master Mix 2X (New England BioLabs, UK) and 'pDOC-G lin_F' and 'pDOC-G lin_R' primers made up to the final volume of 25 μL per reaction with sterile nuclease-free water (Thermo Fisher, UK). This PCR reaction removed the *gfp* gene under the following conditions: an initial denaturation temperature of 98 °C for 30 s followed by 34 cycles of 98 °C for 10 s, 64 °C for 15 s, 72 °C for 180 s with a final incubation at 72 °C for 2 min. The plasmid pUltra-RFP-KM (Mavridou *et al.*, 2016), containing the gene *mrfp1,* kanamycin resistance and a biofab promoter*,* was amplified using Q5 Hot Start Master Mix 2X (New England BioLabs, UK), 'pDOC-rfp_HiFi_F' and 'pDOC-rfp_HiFi_R' primers made up to the final volume of 25 μL per reaction with sterile nuclease-free water (Thermo Fisher, UK) under the following conditions: an initial denaturation temperature of 98 °C for 30 s followed by 34 cycles of 98 °C for 10 s, 66 °C for 15 s, 72 °C for 60 s with a final incubation at 72 °C for 2 min (Figure 2.2).
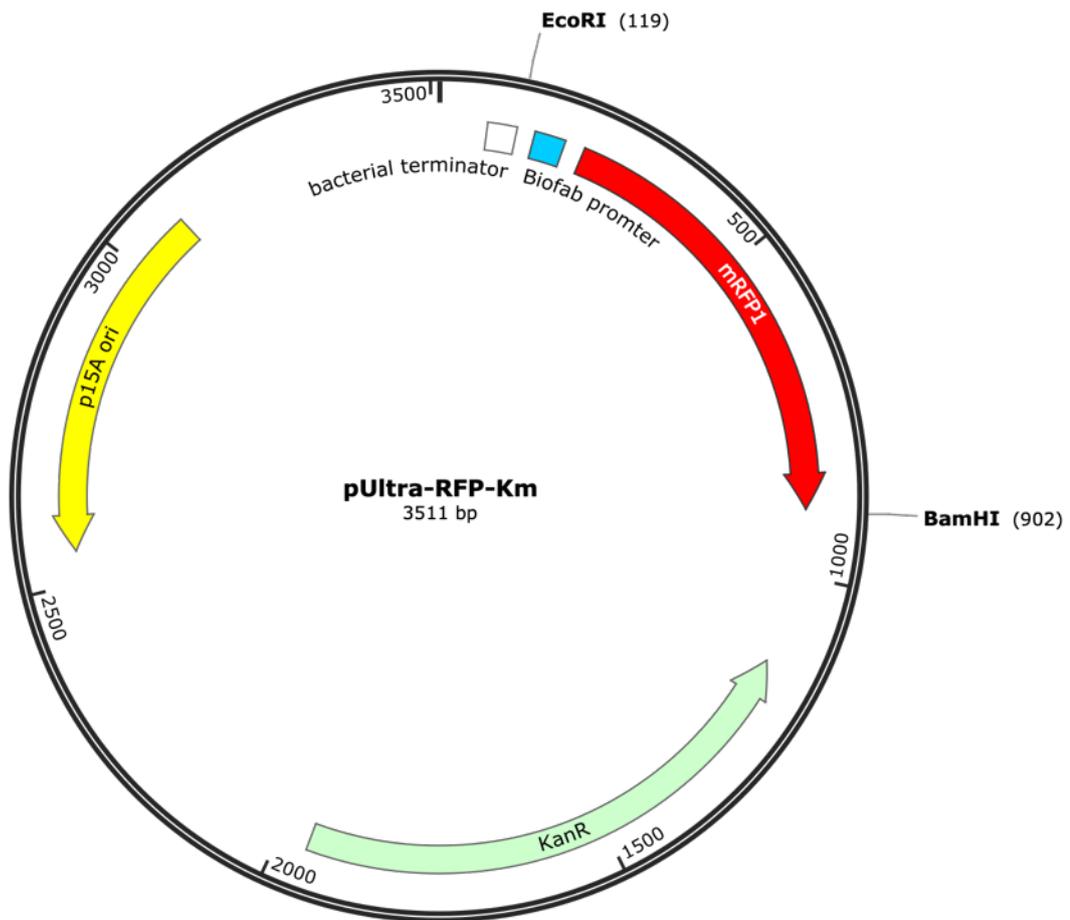
Figure 2.2 Plasmid map of pUltra-RFP-Km

Both PCR products were then purified using a Monarch™ PCR and DNA Cleanup Kit (New England BioLabs, UK) and eluted in 20 μL of sterile nuclease-free water (Thermo Fisher, UK). The concentration of DNA was measured using a NanoDrop 1000 (2.2.10.1).

A NEBuilder® HiFi Assembly Master Mix 2X (New England BioLabs, UK, E2621), 0.016 pmol ends of linearized plasmid vector, and 0.032 pmol ends of the insert were combined and made up to 20 μL with sterile HPLC grade H_2O. The sample was mixed

in a 0.2 mL PCR tube and placed in a thermocycler (BIO-RAD c1000) and heated to 50°C for 20 min.

40 μL Electrocompetent *E. coli* MG1655 cells (Section 2.2.4) and 2 μL of HiFi master mix containing pDOC-R were then transferred to a pre-chilled 2 mm electrode gap electroporation cuvette (Flowgen Bioscience, FBR-102). The transformation mixture was incubated on ice for 5 min. Cells were electroporated in an electroporator (Eppendorf 2510) at 2.5 V with a time constant of 4 ms. Immediatetly after electroporation, 960 μL SOC media was added to the electrocompetent cells and incubated at 37 °C for 60 min. Various dilutions of the transformation mix were spread plated onto LB agar supplemented with 50 μg mL^{-1} kanamycin, and the spread plates were incubated for 16 hours at 37 °C.

Single colonies of transformants were picked and streaked for single colonies onto selective LB agar and incubated for 16 hours at 37 °C. Single colonies were then inoculated into 5 mL of sterile LB broth (Miller) and incubated overnight at 37 °C and 180 rpm in a shaking incubator. Plasmid DNA was then extracted using a Monarch^{TM} Plasmid Miniprep Kit (New England BioLabs, T1010) and eluted into 20 μL nuclease-free water (Thermo Fisher, UK).

2.2.19.2 Construction of pDOC-R *glmS*

The *mrfp1* and *kan^r* genes from pDOC-R were amplified using PCR primers 'pDOC-R-glmS-F' and 'pDOC-R-glmS-R' and Q5 Hot Start Master mix 2X (New Englnd BioLabs, USA) were used to insert *glmS* homology region and restriction sites for *XhoI* and *BamHI* to the DNA fragment (Figure 2.3). This amplification was performed under the

following conditions: an initial denaturation of 98 °C for 30 s followed by 34 cycles of

98 °C for 10 s, 55 °C for 15 s, 72 °C for 60 s with a final incubation at 72 °C for 2 min.

Some of the PCR product was electrophoresed on a 1% w/w 1X TAE agarose gel to

confirm presence of the amplification product (2,088 bp).
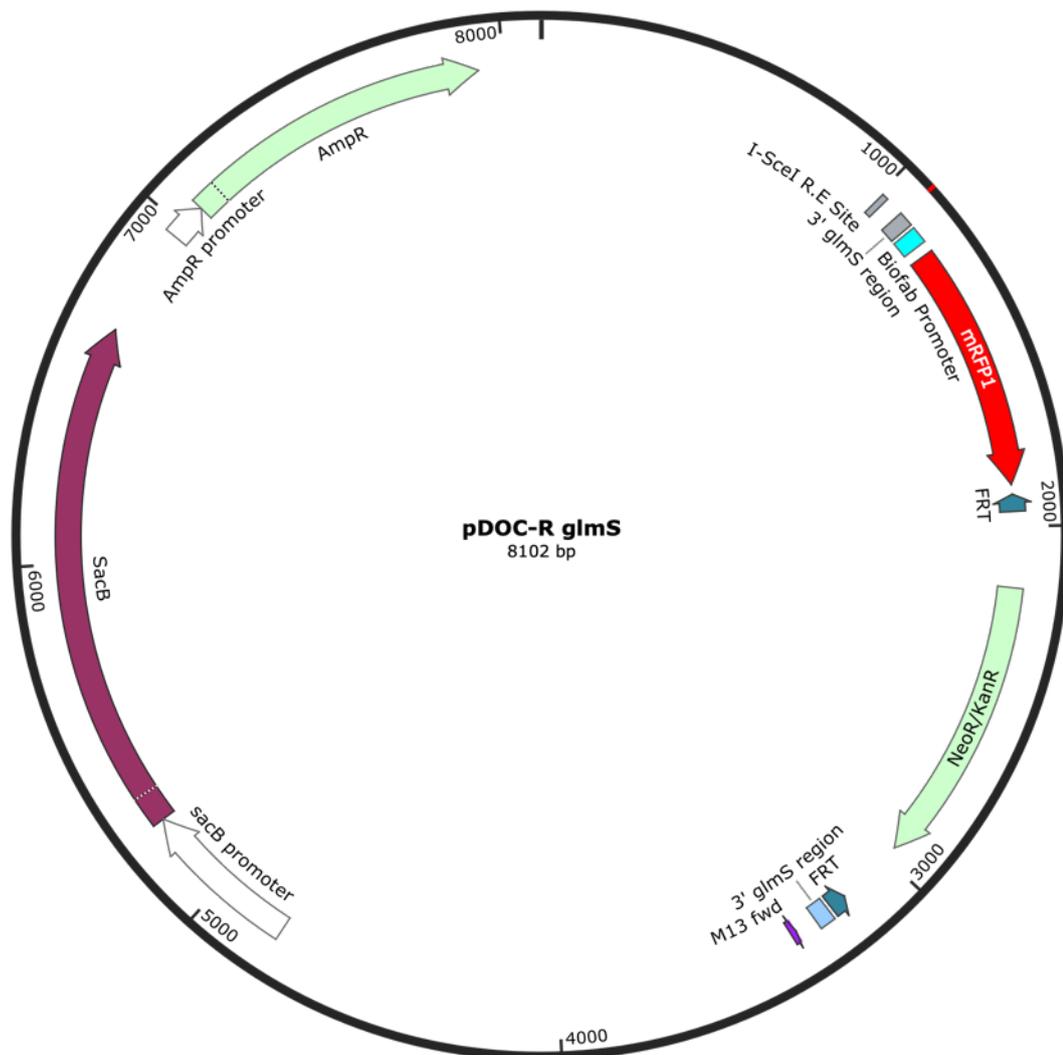
Figure 2.3 Plasmid Map of pDOC-R *glmS*

Plasmid map of pDOC-R *glmS* showing the relevant genes.

2.2.19.3 Isolation of pDOC-C

100 mL of sterile LB broth (Miller) was inoculated with *E. coli* MG1655 pDOC-C and incubated for 16 hours at 37 °C, shaking at 180 rpm. A Qiagen Plasmid Midi kit (Qiagen, 12143) was used to isolate the plasmid using the manufacturer's instructions for a low copy number plasmid.

2.2.19.4 Restriction Enzyme Digestion and Ligation of pDOC-C and *mrfp*1::*glmS*

The following enzyme digestions were used on pDOC–C and *mrfp1::glmS*: pDOC–C only with XhoI (negative control), pDOC-C only with *BamHI* (negative control), pDOC– C with XhoI and BamHI and finally *mrfp1::glmS* with *XhoI* and *BamHI*. The digestions were incubated at 37 °C for 2-3 hours and then 6X Loading Dye (New England BioLabs) was added and incubated at 65 °C for 20 min to inactivate the enzymes.

A ligation mixture using T4 Ligase (New England BioLabs) was made with a vector: insert ratio of 1:7 as well as two controls: pDOC-C BamHI only and pDOC-C XhoI only. The ligations were incubated at room temperature for 30 min and then heat inactivated at 65 °C for 10 min.

2.2.19.5 Transformation of pDOC-C *mrfp1::glmS* to *E.coli* MG1655

5 μL of each ligation mixture was added to 40 μL of Electrocompetent *E. coli* MG1655 cells were then transferred to a pre-chilled 2 mm electrode gap electroporation cuvette (Flowgen Bioscience, FBR-102). Each transformation mixture was

94

electroporated in an electroporator (Eppendorf 2510) at 2.5 V with a time constant of 4 ms. 960 μL of SOC was added immediately to the electroporation mixture and incubated for 1 hour at 37 °C with shaking (200 rpm). Various dilutions of the cells were spread onto LB agar supplemented with 50 μg mL^{-1} kanamycin (100 μg mL^{-1} ampicillin for controls) and incubated overnight at 37 °C. The red colonies that grew on the kanamycin selective plates were picked, streaked for single colonies on LB agar supplemented with 50 μg mL^{-1} kanamycin, and incubated for 16 hours at 37 °C. The same colony was inoculated to 5 mL LB broth (Miller) and incubated for 16 hours at 37 °C with shaking (180 rpm). A Monarch^{TM} Plasmid Miniprep kit was then used to isolate the plasmid from the broth culture and eluted to 20 μL nuclease-free water (Thermo Fisher).

2.2.19.6 Transformation of pACBSR to *E. coli* J53 and J53 Azi^r

50 ng of pACBSR was added to 40 μL of Electrocompetent *E. coli* K-12 J53 and 40 μL of Electrocompetent *E. coli* K-12 J53 Azi^r cells. Each reaction mixture was transferred into a pre-chilled 2 mm electrode gap electroporation cuvette (Flowgen Bioscience, FBR-102). Each transformation mixture was electroporated in an electroporator (Eppendorf 2510) at 2,500 V with a time constant of 4 ms. 960 μL SOC was added immediately to the cells and incubated for 1 hour at 37 °C. Various dilutions of the transformed cells were spread onto LB agar supplemented with 35 μg mL^{-1} chloramphenicol and incubated overnight at 37 °C. Transformed colonies were then streaked for single colonies onto non-selective LB agar and incubated for 16 hours at 37 °C.

2.2.19.7 Preparation of Electrocompetent *E. coli* K-12 J53 pACBSR and J53 Azi[r] pACBSR

Single colonies of *E. coli* K-12 J53 pACBSR and J53 Azi[r] pACBSR were inoculated into 10 mL of LB broth (Lennox) and incubated at 37 °C and 180 rpm for 16 hours. The overnight culture was then diluted 100-fold into 3 mL LB broth (Lennox) and incubated for 2-4 hours at 37 °C and 180 rpm until an OD_{600} 0.2-0.6 was reached. The culture was then placed on ice and 1mL each culture was pelleted at 20,000 *x g* for 1 min at 4 °C (Sigma 1–16K). The supernatant was aspirated, and the cell pellet was resuspended in 1 mL of chilled sterile 10% (v/v) glycerol. This was then centrifuged again at 20,000 *x g* for 1 min at 4 °C and the cell pellet was resuspended in 1 mL of chilled sterile 10% (v/v) glycerol solution. This was then centrifuged again at 20,000 *x g* for 1 min (Sigma 1–16K) at 4 °C and the cell pellet resuspended in 40 µL of chilled sterile 10% (v/v) of glycerol solution. Some of the cell suspensions were stored at -80 °C for long term storage.

2.2.19.8 Transformation of pDOC-C *mrfp1::glmS* into *E. coli* K-12 J53 pACBSR and J53 Azi[r] pACBSR

50 ng of pDOC-C *mrfp1::glmS* was added to each 40 µL aliquot of Electrocompetent *E. coli* J53 pACBSR and J53 Azi[r] pACBSR cells. Each reaction mixture was transferred into a pre-chilled 2 mm electrode gap electroporation cuvette (Flowgen Bioscience, FBR-102). Each transformation mixture was electroporated in an electroporator (Eppendorf 2510) at 2,500 V with a time constant of 4 ms. 960 µL SOC was added immediately to the transformation mixtures and were then incubated at 37 °C and

180 rpm for 1 hour. Various dilutions were spread to LB agar supplemented with 50 μg mL^{-1} kanamycin and 35 μg mL^{-1} chloramphenicol and incubated for 16 hours at 37 °C. Transformed colonies were picked and streaked for single colonies on the dual selective media, again incubated for 16 hours at 37 °C.

2.2.19.9 Chromosomal Insertion of *mrfp1::glmS* into *E.coli* K-12 J53 and J53 Azi^r

Single colonies of each transformed strain were inoculate into 5 mL of sterile LB broth supplemented with 50 μg mL^{-1} of kanamycin, 35 μg mL^{-1} of chloramphenicol and 5% sucrose and LB agar supplemented with 50 μg mL^{-1} of kanamycin and 35 μg mL^{-1} of chloramphenicol and incubated overnight at 37 °C. A fresh sucrose sensitive colony was picked and inoculated to 1 mL LB broth supplemented with 50 μg mL^{-1} of kanamycin, 35 μg mL^{-1} of chloramphenicol and 0.5% glucose and incubated at 37 °C for 1 hour. The cells were then centrifuged at 13,000 x g for 1 min (Sigma 1–16K) and the cell pellet was resuspended in 1 mL of LB broth containing 0.5% (w/v) arabinose and incubated at 180 rpm and 37 °C for 4-5 hours. Various dilutions were spread to LB agar supplemented with 50 μg mL^{-1} of kanamycin and 5% (w/v) of sucrose. These plates were incubated at 30 °C for 16 hours.

Transformants were tested for by picking single colonies and inoculating them onto LB agar supplemented with 100 μg mL^{-1} of ampicillin to test for loss of pDOC and LB supplemented with 35 μg mL^{-1} of chloramphenicol to test for loss of pACBSR.

To confirm chromosomal insertion of the *rfp* gene had taken place, a PCR was performed to amplify the overlap between the insertion site. 'glmS F' in the *glmS* gene and 'rfp R' which is midway along the *mrfp1* gene. When amplified this gives a

444 bp product. Transformants were PCR screened using DreamTaq green master mix 2X (Thermo Fisher) and the 'glmS F' and 'rfp R' primers were used under the following PCR conditions in a thermocycler (BIO-RAD c1000): an initial denaturation of 95 °C for 2 min followed by 35 cycles of 95 °C for 30 s, 61 °C for 30 s, 72 °C for 180 s with a final incubation at 72 °C for 5 min. The PCR products were electrophoresed on a 1% w/w 1X TAE agarose gel at 85 V for 1 hour. Upon imaging, these all showed to have chromosomally inserted *mrfp1* and the kanamycin resistance gene into the two J53 strains. The J53 (non-azide resistant) RFP strain was sequenced by short read sequencing (MicrobesNG, Birmingham, UK) as described in section 2.2.27.

2.2.20 Induction of Transposition of Tn*21*

Single colonies of donor strain (D) and *E. coli* MG1655 RP4-8 were inoculated into 5 mL LB broth (Miller) and incubated at 37 °C and 180 rpm for 16 hours. 1 mL of each culture was then centrifuged at 16,000 $x\ g$ for 1 min (Sigma 1–16K), the supernatant was removed, and the cells resuspended in 1 mL of sterile MRD. This was then centrifuged again at 16,000 $x\ g$ for 1 min (Sigma 1–16K) and resuspended in 1 mL of sterile MRD. The suspension was then centrifuged at 16,000 $x\ g$ for 1 min (Sigma 1–16K) and the cells resuspended in 500 µL of sterile MRD. The OD_{600} of the cell suspensions was measured. An OD_{600} between 0.3 and 0.5 for each suspension was used. 50 µL of each cell suspension was added to each other so there was approximately 1:1 cell ratio of each strain. This was mixed by gently pipetting. This 100 µL of the mating mixture was then pipetted onto nonselective LB agar and not spread. This was incubated at 37 °C for 16 hours.

The lawn of mating cells was then scraped off the agar plate into a 1.5 mL microcentrifuge tube using a 5 µL inoculum and resuspended in 1 mL MRD and vortexed thoroughly to end conjugation. The cells were then centrifuged at 16,000 $x g$ for 1 min (Sigma 1–16K) and the supernatant removed by aspiration. The conjugation mix was then resuspended in 1 mL of sterile MRD. Serial dilutions of the conjugation mix were made and 100 µL of each dilution was spread plated onto LB agar supplemented with 10 µg mL^{-1} of gentamycin (RP4 selective marker) and a selective marker corresponding to the recipient strain (50 µg mL^{-1} of streptomycin for isolate EVAL397 and 10 µg mL^{-1} of erythromycin for isolate BPW2-4) and incubated for 16 hours at 37 °C. Single colonies were then picked and re-streaked for single colonies on the same dual selective LB agar. These plates were then incubated overnight at 37 °C.

The RP4-8 trans-conjugants were inoculated into 1 mL of sterile LB broth (Lennox) aliquots with various concentrations of HgCl$_2$ (

Table 2.5), 5 mL of sterile LB broth (Lennox) was inoculated with the *E. coli* J53 RFP kan^r recipient strain and both were incubated for 20 hours at 30 °C and 180 rpm shaking. 100-fold dilutions of each HgCl$_2$ supplemented overnight culture was inoculated into 5 mL of sterile LB (Lennox) broth and incubated at 25 °C for 6 hours with weak agitation. 200 µL of this culture was then added to 800 µL of new recipient strain and incubated for a further 6 hours with no agitation at 25 °C. Each culture was then serially diluted and 100 µL of each dilution was plated onto LB agar

supplemented with 10 µg mL^{-1} of gentamycin and 50 µg mL^{-1} of kanamycin to select for J53 RFP kan^r RP4-8 cells. These plates were incubated overnight at 37 °C. The cells were then scraped into a sterile microcentrifuge tube containing 100 µL of sterile MRD and vortexed thoroughly to end mating before being placed on ice. These were then serially diluted and plated onto LB agar supplemented with 10 µg mL^{-1} of gentamicin and 50 µg mL^{-1} of kanamycin and incubated overnight at 37 °C. A plate count was then performed from the dilutions and then colonies were patch plated onto non-selective LB agar and the previously mentioned dual selective LB agar, looking for red colonies growing. Positive isolates were then screened for the presence of Tn*21* using PCR (2.2.16.3). Successful transconjugants were then cryogenically stored in 25% (v/v) of glycerol solution at - 80 °C (2.2.2). Selected transconjugants were then sequenced using long-read sequencing to confirm insertion (2.2.25).

Compound	Stock	1	2	3	4	5	6	7	8
Mercury (II) Chloride (µg mL^{-1})	50000	0	1	2	4	8	16	32	64

Table 2.5 Concentrations of HgCl$_2$ used for transposition induction.

2.2.21 Pulse Field Gel Electrophoresis for Plasmid Separation

Pulse field gel electrophoresis (PFGE) was used to resolve the sizes of plasmids found in short read assembly of sequenced isolates (2.2.27), where according to the method of Hooton *et* al., (2011). A single colony of each isolate was inoculated to

10 mL of sterile LB Broth, according to Miller. These were incubated at 25 °C for 48 hours and shaking at 180 rpm. The cultures were then centrifuged at 4,400 x g for 5 min (Hettish EBA 12R). The supernatant was removed, the pellets were re-suspended in 1 mL of sterile LB Broth and 55 µL of cell suspension was transferred to a 1.5 mL microcentrifuge tube. 5 µL of 20 mg mL^{-1} of Proteinase K (Sigma Aldrich) was added followed by mixing 50 µL of molten 1.2% (w/w) PFGE grade agarose (Bio-Rad, USA) in TE buffer. The suspension was mixed thoroughly and then set in PFGE plug moulds (Bio-Rad, USA). The plugs were then incubated overnight at 55 °C shaking at 300 rpm in 1 mL of lysis buffer. The plugs were then washed three times for 1 hour in 1 mL of wash buffer at 55 °C with shaking at 300 rpm. A 3 mm slice of each plug was then loaded to a 1% PFGE grade agarose gel (100 mL 1X TAE buffer) along with PFG lambda ladder (New England BioLabs, USA). The lanes were sealed with 1% PFGE grade agarose. A CHEF-DR II system was used for the run with a 10 - 30 s switch time over 18 hours at 6 V cm^{-1} and 1 X TAE running buffer was circulated at 14 °C. The gel was then stained using ethidium bromide (1 µg mL^{-1}) and DNA was visualized on a UV trans illuminator (BIO-RAD Gel Doc XR+).

2.2.22 qPCR Quantification of Tn*21* in the Wastewater Environment

2.2.22.1 *rrsA*, a Control Reference Gene

rrsA was chosen as a gene of reference for comparison of gene abundance in qPCR due to its presence in every bacterial cell and genetic stability. This reference gene has also been widely used as a reference strain in many qPCR studies (Zhou *et al.*,

2011; Peng *et al.*, 2014). Bacterial populations have been calculated to possess on average 5 copies of the *rrsA* gene per cell according to the Ribosomal RNA Operon Copy Number Database (rrndb) (Klappenbach *et al.*, 2001).

2.2.22.2 Preparation of Samples

Environmental DNA was extracted from bacterial cells collected from a wastewater treatment plant that had been cryogenically preserved in 25% (v/v) sterile glycerol solution at -80 °C (2.2.2) using a Qiagen Power Water kit with Pathogen Lysis Tubes (2.2.14) and eluted using 100 µL of Qiagen elution buffer. The harvested DNA was quantified using a Qubit 3 (2.2.10.2). DNA from each extraction was diluted using nuclease free water to achieve equal concentrations of 15 ng µL^{-1}.

2.2.22.3 qPCR Using the Luna® Universal qPCR Master Mix

qPCR was conducted with each sample, with three biological repeats. *rrsA* and Tn*21* were analysed in all 6 samples. Luna® Universal qPCR Master Mix (NEB, M3003S) components were thawed at room temperature, re-suspended by vortexing (Vortex-Genie 2, Scientific Industries) and placed on ice. 10 µL of Luna® universal probe one-step reaction mix 2X, 0.8 µL of forward primer (10 µM), 0.8 µL of reverse primer (10 µM), 0.4 µL probe (10 µM), and template DNA (15 ng) was combined, and made up to a final volume of 20 µL using nuclease-free water in a 0.2 mL PCR tube (Xtra-clear, I1402-8200). qPCR was conducted in a real-time PCR cycler (Qiagen Rotor-Gene Q) using the following thermocycler conditions with green fluorescence (FAM). Initial

denaturation 90 °C 60 s and 40 cycles of denaturation at 95 °C 15 s and extension at

60 °C 30 s.

2.2.22.4 qPCR Calibration Curve

A standard curve was produced to analyse the efficiency of the qPCR reaction. DNA

extracted from each sample was diluted from 10^0 to 10^{-3} and run under the same

conditions as (2.2.22.3). 100% efficiency was expected, but 90-110% efficiency is

widely accepted due to pipetting errors in dilutions of DNA, qPCR reaction set up and

the presence of inhibitors.

2.2.22.5 Determining Relative Abundance of Tn*21*

Relative abundance is commonly used in order to determine how many copies of a

target gene are present in a bacterial population compared to a 'housekeeping', or

well conserved gene. In this case the relative abundance of Tn*21* in the bacterial

wastewater environment was determined using this method.

C_q is the cycle number value at which fluorescent intensity of a target probe exceeds

an arbitrarily appointed threshold of background fluorescence. Calculation of relative

abundance makes two assumptions. Firstly, efficiency of gene amplification is close

to 100% and secondly that the reference gene and target genes have similar

efficiencies.

C_T values	Condition A (Treated)	Condition B (Untreated)

Reference gene (*rrsA*)	A_{ref}	B_{ref}
Target (Tn*21*)	A_{gene}	B_{gene}

Table 2.6 Definition of terms for determining relative abundance in qPCR.

$$Relative\ Abundance = 2^{-\Delta\Delta Cq} = 2^{-[(A_{gene}-A_{ref})-(B_{gene}-B_{ref})]}$$

2.2.23 Emulsion Paired Isolation Concatenation PCR (epicPCR)

The method used here was based on the works of (Spencer *et al.*, 2016; Hultman *et al.*, 2018). epicPCR is a screening process whereby environmental cells can be screened for target genes which are simultaneously fused to a housekeeping gene, which may define the taxonomic group the cell belongs to. Epic-PCR links resistance genes to the organism/bacterial taxa that carries the gene, therefore overcoming a major disadvantage of PCR of metagenomics samples, which does not. The method described below focussed on optimising the technique for long-read sequencing technology (Figure 5.9).

For each PCR reaction, a parallel pseudo-epicPCR of metagenomic DNA from each sample was used (2.2.14).

2.2.23.1 Pre-treatment of Cells

An aliquot of each cryogenically preserved sample was thawed and centrifuged for 1 minute at $12,000\,x\ g$. The supernatant was then removed, and the pellet resuspended in 30 μL sterile nuclease free water.

2.2.23.2 Acrylamide Bead Polymerisation

30 µL of cell suspension, 100 µL of sterile nuclease-free water, 100 µL of 30% Acrylamide/Bis-acrylamide 29:1 solution (BIO-RAD) and 25 µL of 10% ammonium persulfate (APS) were added to a 2 mL round-bottom safe-lock microcentrifuge tube (Eppendorf, Germany) and vortexed gently. 600 µL of STT emulsion oil was added and vortexed at max speed for 30 s. 25 µL of tetramethylethylenediamine (TEMED) was then added and vortexed for a further 30 s at max speed and then left to incubate at room temperature for 90 min.

2.2.23.3 Emulsion Oil Removal

800 µL water saturated diethyl ether was added to the polymerised solution and mixed by inverting and agitating the tube contents until a visible precipitate and an aggregate of acrylamide formed. The top phase was then removed. 1 mL of sterile water was then added to the tube and mixed by flicking and inverting. The sample was then centrifuged at 12,000 x g for 30 s (Sigma 1–16K). The top phase was then removed. 1 mL of sterile water was then added, and the process repeated 6-7 times until the upper phase was completely transparent and no STT emulsion oil was present. The remaining water was then removed without disturbing the acrylamide bead layer and the beads resuspended in 1 mL of sterile TK buffer. The bead suspension was then run through a 35 µm cell strainer (Falcon®, VWR, 352235) and then transferred to a sterile 1.5 mL microfuge tube.

2.2.23.4 UV Fluorescence Microscopy

100 μL of a 10-fold dilution of each bead sample was made in 1.5 mL microcentrifuge tubes and then 0.5 μL of 10,000 X Sybr Green stain was added to each sample and then visualised using an inverted microscope (Axiovert 135 TV, Zeiss) to identify samples containing 1-2 cells present per 100 beads (bead size 10–20 μm).

2.2.23.5 Fusion PCR

At first, the Tn*21* target region used for detection by PCR amplified the *attI* of the class I integron to t*npM*. In this reaction 53.5 μL of PCR master mix (Q5 Hot Start Master Mix 2X, 10 μM of 'R1542' primer, 10 μM of 'HS458RCv4' primer and 1 μM of 'R1 – F2' 1' primer) was added with 46.5 μL of the sample into a round-bottomed 2 mL safe-lock microfuge tube (Eppendorf, Germany). Four sterile 2 mm glass beads were added to each tube with 900 μL of ABIL emulsion oil, and tube contents were emulsified by vortexing for 1 min at max speed. The emulsion was then split into approximately equal volumes into 16 0.2 mL PCR tubes and heated in a thermocycler. Initial denaturation temperature was 98 °C 30 s followed by 35 cycles of: 98 °C for 5 s, 59 °C for 30 s and 72 °C for 120 s. A final incubation was performed at 72 °C for 5 min. The contents of the tubes were then pooled into a sterile 2 mL round-bottom safe-lock microfuge tube and 2 μL 50 mM EDTA pH 8 was added.

However, due to problems with amplification, different primers were used instead of 'HS458RCv4'. The following modified PCR reaction mixture and thermocycling conditions were used. 60 μL of PCR master mix (2x Q5 Hot Start Master Mix, 10μM of 'R1542' primer, 10 μM of 'Tn21 F' primer, 1 μM of 'R1 – F2' Tn*21*' primer) was

added with 40 μL of sample, totalling 100 μL to a 2 mL safe-lock microfuge tube. Four sterile 2 mm glass beads were added to each tube with 900 μL of ABIL emulsion oil and emulsified by vortexing for 1 min at max speed. The emulsion was then split into approximately equal volumes to 16 0.2 mL PCR tubes and heated in a thermocycler. The emulsion was then split into approximate equal volumes to 16 0.2 mL tubes and heated in a thermocycler. Preheating step 80 °C for 30 s, initial denaturation temperature 98 °C for 30 s and 32 cycles of: 98 °C for 15 s, 59 °C for 30 s and 72 °C for 60 s. A final incubation was performed at 72 °C for 5 min.

2.2.23.6 Breaking the ABIL Emulsion

The contents of the 16 0.2 mL microfuge tubes were pooled in sterile 2 mL round-bottomed microfuge tubes. 1 mL of water saturated diethyl ether was added to each sample and vortexed gently and then centrifuged at 13,000 $x\,g$ for 1 min (Sigma 1–16K) to separate the phases. The upper phase was then removed, 50 μL sterile nuclease-free water added and then the process repeated. 1 mL water saturated ethyl acetate was added and vortexed before centrifuging for 1 min at 13,000 $x\,g$ (Sigma 1–16K). The upper phase containing ethyl acetate was removed. 1 mL water saturated diethyl ether was then added and vortexed before centrifuging for 1 min at 13,000 $x\,g$ (Sigma 1–16K). The supernatant was removed and repeated once more. The samples were then left to dry for 5 to 10 min to evaporate the remaining diethyl ether residues. 100-150 μL bottom phase was then recovered to a fresh 1.5 mL microfuge tube and a Monarch™ DNA and PCR Clean Up Kit (New England BioLabs, UK) and eluted with 20 μL of nuclease-free water.

2.2.23.7 Initial Blocking PCR

92 µL of PCR master mix (5X GC Buffer, 10 mM of dNTPs, 32µM of 'BlockF 27F' primer, 32 µM of 'BlockR 27F' primer, Phusion DNA Polymerase (New England BioLabs) was added to 8 µL of purified fusion PCR product as a template, split into four 0.2 mL PCR tubes and incubated in a thermocycler: initial denaturation temperature of 98 °C for 30 s and 30 cycles of: 98 °C for 10 s, 60 °C for 30 s and 72 °C for 90 s and then a final incubation was performed at 72 °C for 5 min).

For the optimised version of the epicPCR process, Q5 Polymerase Master Mix 2X was used: 32 µM of 'BlockF 27F' primer, 32 µM of 'BlockR 27F' primer and 8 µL of DNA was used, nuclease free water was used to make up each sample reaction to 100 µL before splitting to 25 µL quadruplicates and incubating in the thermocycler. Initial denaturation temperature of 98 °C for 30 s and 39 cycles of: 98 °C for 10 s, 59 °C for 30 s and 72 °C for 60 s and then a final incubation was performed at 72 °C for 5 min. The products for each quadruplicate were then pooled and purified using a Monarch™ DNA and PCR Clean Up Kit (New England BioLabs) and eluted into 20 µL of nuclease-free water.

2.2.23.8 Nested and Final Blocking PCR

92 µL PCR master mix (Q5 Hot Start master mix 2X, 3 µM of 'HS458RCv4' primer('Tn21 F' for revised version), 3 µM of 'R1492' primer, 3.2 µM of 'BlockF 27F' primer, 3.2 µM of 'BlockR 27F' primer) was added to 8 µL of purified blocking PCR product as a template and split to four 0.2 mL tubes and incubated in a thermocycler:

initial denaturation temperature of 98 °C 30 s and 35 cycles of: 98 °C for 10 s, 59 °C

for 30 s and 72 °C for 60 s and then a final incubation was performed at 72 °C for

5 min.

The product was then visualised on a 1% w/w 1X TAE agarose gel at 100 V for 1 hour.

The rest of the product was then purified using a Monarch™ DNA and PCR Clean Up

Kit (New England BioLabs) and eluted into 20 µL of nuclease-free water.

2.2.24 epicPCR Product Library Preparation for MinION Sequencing

Two methods were used to prepare the DNA libraries of the epicPCR samples and

their parallel pseudo-epicPCR of metagenomic DNA PCR samples for MinION

sequencing.

2.2.24.1 Four-Primer PCR Method for Oxford Nanopore Library Preparation

The eluted DNA was transferred to a 1.5 mL DNA LoBind tube (Eppendorf, Germany)

and the DNA concentration was adjusted to achieve 30 ng µL^{-1} with nuclease-free

water and mixed by flicking. The suspension was then pulse centrifuged to collect the

suspension at the bottom of the tube.

Barcoding was attempted using the Four Primer PCR protocol (Oxford Nanopore

Technologies, 2018). 30 ng of DNA, 10 µM of 'Tn21F-nanopore' primer, 10 µM of

'nanopore-1492R' primer, 10 µM of BP 01-12 from the SQK-PBK004 kit (Oxford

Nanopore Technologies), LongAmp Hot Start Taq Master Mix 2X (New England

BioLabs) and nuclease-free water to make up to 50 µL were combined in a 0.2 mL

thin-walled PCR tube. The contents were then mixed by flicking and centrifuged briefly.

The reaction mix for each sample was then incubated in a thermocycler: initial denaturation temperature of 94 °C for 60 s and 5 cycles of: 94 °C for 30 s, 55 °C for 30 s, 65 °C for 150 s then 30 cycles of: 94 °C for 30 s, 62 °C for 30 s and 65 °C for 150 s and then a final incubation was performed at 65 °C for 5 min.

The reaction mix was then transferred to a 1.5 mL DNA LoBind centrifuge tube (Eppendorf) and 40 µL of resuspended AMPure XP beads (Agencourt) was added to bind the DNA. The tubes and their contents were then incubated on a rotator mixer (Grant-bio, PTR-60) for 5 min at room temperature. The reaction mixture was then centrifuged briefly before being placed onto a magnet. The supernatant was aspirated once a pellet had formed. 200 µL of 80% (v/v) ethanol was used to wash the pellet whilst remaining on the magnet and then ethanol was removed by aspiration. This wash process was repeated once more.

The tubes were then centrifuged briefly, placed back on the magnets, residual ethanol was removed by aspiration, and the pellet allowed to dry. The tubes containing the pellets were then removed from the magnets and the pellets were resuspended in 10 µL of 10 mM Tris-HCl pH 8.0 with 50 mM of NaCl and incubated for 2 min at room temperature. The tubes were then centrifuged briefly and placed back on the magnets until the eluate was clear and colourless, and free of magnetic beads. The 10 µL of eluate was transferred into a sterile 1.5 mL DNA LoBind tube (Eppendorf).

This procedure for library preparation with PCR introduced too much error within each sample, reducing the quality of reads and reduced the ability to assign reads to a barcode so was abandoned.

All the barcoded samples were then pooled together to the desired ratio and diluted to 50-100 fmol of PCR product to 10 µL in 10 mM of Tris-HCl pH 8.0 with 50 mM NaCl. 1 µL RAP (Oxford Nanopore Technologies, UK) was to the 10 µL amplified DNA library, mixed by flicking and then centrifuged briefly to collect at the bottom of the tube. The mixture was then incubated for 5 min at room temperature and then stored on ice until ready to be loaded to a flow cell for sequencing.

2.2.24.2 Native Barcoding of epicPCR and Parallel Pseudo-epicPCR Metagenomic DNA Samples

To avoid the introduction of error from four-primer PCR, native barcoding was used (EXP-NBD104), as a less invasive technique which involves enzymatic treatment of purified DNA to first repair DNA ends and ligate barcodes to amplicons (Oxford Nanopore Technologies, 2019).

2.2.24.3 End Preparation of epicPCR and Parallel Pseudo-epicPCR Metagenomic DNA Samples

100-200 fmol of purified amplicon DNA in a total volume of 48 µL of nuclease-free water was used at the start of each sequencing preparation run. Each sample was added to a 0.2 mL thin-walled PCR tube with 3.5 µL of NEBNext FFPE DNA Repair Buffer, 2 µL of NEBNext FFPE DNA Repair Mix, 3.5 µL of Ultra II End-prep reaction

buffer and 3 µL of Ultra II End-prep enzyme mix (New England BioLabs). The reaction was mixed gently by flicking and then centrifuged briefly before incubation in a thermocycler at 20 °C for 5 min then 65 °C for 5 min.

60 µL of resuspended AMPure XP magnetic beads (Agencourt) was added to the tube, mixed gently and then incubated for 5 min at room temperature on a rotator mixer (Grant-bio, PTR-60). The incubated mixture was then centrifuged briefly (Star Lab Mini Fuge PLUS) and placed on a magnet until a pellet formed. 200 µL of 80% (v/v) ethanol was used to wash the pellet whilst remaining on the magnets and then ethanol was removed. This wash process was repeated once more.

The tubes were then centrifuged briefly (Star Lab Mini Fuge PLUS), placed back on the magnet, residual ethanol was removed, and the pellet allowed to dry. The tubes containing the pellets were then removed from the magnets and the pellets were resuspended in 25 µL of nuclease-free water and incubated for 2 min at room temperature. End prepped DNA was then quantified using a Qubit 3 (2.2.10.2).

2.2.24.4 Native Barcoding of epicPCR and Parallel Pseudo-epicPCR Metagenomic DNA Samples

For each sample, a unique barcode was assigned to it. In a clean, sterile 0.2 mL thin-walled PCR tube: 22.5 µL of end prepped DNA, 25 µL Blunt/TA Ligase Master Mix (New England BioLabs, USA) and 2.5 µL Native Barcode (EXP-NBD104) were added. The tube contents were mixed by flicking and briefly centrifuged (Star Lab Mini Fuge PLUS). The contents were incubated at room temperature for 10 min.

50 µL of resuspended AMPure XP magnetic beads (Agencourt) was added to the tube, mixed gently and then incubated for 5 min at room temperature on a rotator mixer (Grant-bio, PTR-60). The incubated mixture was then centrifuged briefly and placed on a magnet until a pellet formed. 200 µL of 80% (v/v) ethanol was used to wash the pellet whilst remaining on the magnet and then ethanol was removed. This wash process was repeated once more.

The tubes were then centrifuged briefly, placed back on the magnets, residual ethanol was removed, and the pellet allowed to dry. The tubes containing the pellets were then removed from the magnets and the pellets were resuspended in 26 µL of nuclease-free water and incubated for 2 min at room temperature. 1 µL of barcoded DNA was then quantified using Qubit 3 (2.2.10.2).

Equimolar amounts of each barcoded sample were pooled in a final volume of 65 µL nuclease free water in a clean, sterile 0.2 mL thin-walled PCR tube along with: 5 µL of Adapter Mix II, 20 µL of NEBNext Quick Ligation Reaction Buffer (5X) and 10 µL of Quick T4 DNA Ligase. The mixture was mixed briefly by flicking and centrifuged briefly (Star Lab Mini Fuge PLUS) before incubation at room temperature for 10 min.

50 µL of resuspended AMPure XP beads (Agencourt) was added to the tube, mixed gently and then incubated for 5 min at room temperature on a rotator mixer (Grant-bio, PTR-60). The incubated mixture was then centrifuged briefly and pelleted on a magnet. 250 µL of short fragment buffer was used to wash the pellet and resuspended by flicking before being placed back on the magnet. A pellet formed and the wash buffer was aspirated. This wash process was repeated once more. The pellet was then resuspended in 15 µL elution buffer and incubated for 10 min at

room temperature. The bead suspension was then pelleted back on the magnet and the 15 µL elution buffer containing the pooled DNA was extracted and decanted to a sterile, clean 1.5 mL Eppendorf DNA LoBind tube.

2.2.24.4.1 Loading of DNA Library to MinION Flow Cell

1 µL of epicPCR sequencing library DNA was then quantified using the Qubit 3 (2.2.10.2). 12 µL of DNA library was added to 37.5 µL of sequencing buffer and 25.5 µL of loading beads before adding dropwise to a primed flow cell. Sequencing was run using MinKNOW (v 21.02.1) until 1 million reads were collected. Base calling was performed using the high accuracy parameter and detection of EXP-NBD104 barcodes was used.

2.2.24.5 Analysis of epicPCR Reads

The demultiplexed reads from Oxford Nanopore sequencing were trimmed using cutadapt v3.4 to remove the bridge primer fusion and Tn*21* portion of each read (Martin, 2011). The trimmed reads were then screened for operational taxonomic units (OTUs) using EPI2ME software Fastq 16S platform with the following paramaters. Minimum quality score 10, minimum read length 800 bases, for BLAST: minimum identity 80%, maximum error value 10^{-10} (conservative).

2.2.25 MinION Genomic DNA Library Preparation and Sequencing

For individual isolates, the Oxford Nanopore protocol for genomic DNA by ligation was used (Oxford Nanopore Technologies, 2017a). For multiple samples that were

114

sequenced at once, the Native Barcoding Genomic DNA protocol was followed. DNA end preparation and barcode ligation was described in section 2.2.24.2. Equal concentrations of each sample were pooled together (Oxford Nanopore Technologies, 2017b). For both preparation methods, DNA was quantified using the Qubit 3 at the end of each step. The MinION R9.5.1 flow cell was primed as per the Oxford Nanopore Protocol and a maximum of 100 fmol DNA was loaded to the MinION flow cell. The sequencing was run using MinKNOW (v 21.02.1) until approximately 50X coverage for each isolate was achieved assuming the average genome size to be 5 Mb. The flow cell was washed as per the manufacturer's instructions for storage at 4 °C.

2.2.26 Long Read Genomic Sequence Assembly

All sequencing runs were operated through the proprietary MinKNOW graphical user interface (GUI) software provided by Oxford Nanopore Technologies for the express purpose of MinION sequencing. Guppy, a proprietary base calling software for MinION, was used in real time to generate bases identities from DNA depolarising the membrane of the sequencing pores. Guppy simultaneously de-multiplexed sequenced reads using the native barcodes and then used to quality assess reads into passed and failed reads. Passed reads were then pooled together to one file and quality checked in FastQC (Andrews, 2010). Good quality reads were then assembled through Unicycler (Wick *et al.*, 2017). Long-read only assemblies were assembled in the Unicycler pipeline of miniasm assembly and polished using Racon polishing (Li, 2016; Vaser *et al.*, 2017). Those that also had an Illumina sequence were also

assembled using Unicycler, as a hybrid assembly using the long reads as a backbone and short reads to confirm sequences and build scaffolds. The pipeline for this was also miniasm assembler followed by racon polish. Short reads were then used for a second round of sequence polishing called Pilon used to close gaps between contigs. Polishing uses multiple DNA fragments which overlap to remove base errors in contigs.

2.2.27 Short-read Sequencing of Tn*21* Positive Wastewater and Slurry Isolates

Some of the bacterial strains isolated from the wastewater or from the EVAL-Farms collection (a study that collected over 1000 isolates from a dairy unit approximately 800 m NE to the wastewater treatment plant: Figure 2.1) that possessed mercury resistance or Tn*21*-like determinants and different antibiotic resistance profiles, these were prepared by the protocol provided by MicrobesNG (University of Birmingham) and sent to Microbes NG for short read Illumina MiSeq sequencing, assembly and annotation.

Assembly and annotation were repeated when the results were made available. Paired end reads were generated and checked for quality using FastQC, trimming of poorer quality reads was performed with trimmomatic (Andrews, 2010; Bolger, Lohse and Usadel, 2014). The paired end reads were then assembled using SPAdes v3.14.1 [parameters: -k 21,33,55,77,99,127 –careful] (Bankevich *et al.*, 2012). The assemblies were then screened for quality using QUAST (Gurevich *et al.*, 2013). Annotations were then performed with RASTtk annotation (Aziz *et al.*, 2008).

2.2.28 Post Assembly Processing of Short, Long and Hybrid Sequence Assemblies

Assemblies of each isolate were then screened for sequence type using Multi Locus Sequence Typing (MLST) and core genome sequence type, cgMLSTFinder1.1 (Larsen *et al.*, 2012; Clausen, Aarestrup and Lund, 2018; Gadou *et al.*, 2018; Zhou *et al.*, 2020) to identify similar or clonal sequences. Once sequence types were obtained, the Contig Scaffolding tool using Algebraic Rearrangements (CSAR) tool was used to generate scaffolded assemblies by aligning contigs against known sequence types if possible (Chen and Lu, 2018).

The PlasmidFinder tool was used in order to identify potential plasmid incompatibility groups in assemblies of each isolate, these were then manually checked to ensure continuity (Carattoli *et al.*, 2014). A clonal long-read sequence containing similar plasmid incompatibility determinants of similar sizes as some short-read isolates was used to map plasmids against each other using Circoletto [parameters: default] to identify identical plasmids (Darzentas, 2010).

Isolate sequence assemblies were also screened for acquired resistance genes and virulence gees using the ResFinder 4.1 tool and VirulenceFinder 2.0 (Camacho *et al.*, 2009; Joensen *et al.*, 2014; Zankari *et al.*, 2017; Bortolaia *et al.*, 2020; Tetzschner *et al.*, 2020).

2.2.28.1 Database Tn*21* Acquisition and Phylogeny

The Tn*21* and Tn*21*-like sequences included in this study were found on NCBI nucleotide using the keyword 'Tn21' as the search term, and the filter of sequence range 10,000-10,000,000 bases. This search generated 4155 results; these were

117

further filtered using the following criteria. Results were excluded if they were a partial sequence, or if there were no *mer* genes, no *tnp* genes and no class 1 integron-integrase observed within the sequences. This yielded 140 compatible Tn*21* and Tn*21*-like transposons with this study. The remaining sequences were then trimmed to only the Tn*21* sequence region.

The final database sequences and assembled Tn*21* and Tn*21* from the environmental isolates of this study were then aligned using MAFFT multiple sequence alignment tool [parameters: iterative refinement method - G-INS-i, scoring matrix 200PAM / k=2] (Katoh and Standley, 2013). The multiple sequence alignment was then used to produce bootstrapped phylogenetic tree (100 bootstraps) of all the mobile elements to identify relationships between.

2.2.29 Environmental *Tn21* Class I Integron PCR

Tn*21* class 1 integrons were amplified using primers '*tnpM* F' and '*qacEΔ1* R' to amplify Tn*21* specific integrons and their contents from wastewater bacterial DNA, extracted using Qiagen Power Water kit (2.2.14).

2.2.29.1 PCR for Sequencing by Illumina Nextera

45 µL PCR master mix (2X Q5 DNA polymerase master mix, 10 µM of '*tnpM* F', 10 µM of '*qacEΔ1* R', made up to 45 µL using sterile nuclease-free water) (Table 2.3) was added to 5 µL of template DNA. The reaction mixtures were then incubated in a thermocycler: initial denaturation temperature of 98 °C for 30 s and 30 cycles of: 98 °C for 10 s, 64 °C for 15 s and 72 °C for 150 s and then a final incubation was

performed at 72 °C for 5 min. 5 µL of PCR product was mixed with 1 µL of 6X loading

dye (no SDS) (New England BioLabs) and electrophoresed on a 1% w/w 1X TAE

agarose gel at 100 V for 1 hour. The rest of the product was then purified using a

Monarch^TM DNA and PCR Clean Up Kit (New England BioLabs) and eluted to 20 µL

nuclease-free water. The purified DNA was then library prepped and sequenced

externally using Illumina Nextera short read (DeepSeq).

2.2.29.2 Assembly and Annotation of Environmental Tn*21* Class I integrons

The short reads were demultiplexed by DeepSeq. Reads were then processed by

quality checking with FastQC and trimmed accordingly with trimmomatic

[parameters: -phred33 ILLUMINACLIP:adapters.fa:2:30:10 LEADING:3 TRAILING:3

SLIDINGWINDOW:4:15 MINLEN:30] where 'adapters.fa' is a fasta-formatted file

containing all commonly used Illumina adapter sequences (Andrews, 2010; Bolger,

Lohse and Usadel, 2014). If only one end of paired reads had acceptable quality, it

was used as a single read during assembly. Surviving paired-end reads and single

reads were assembled together using SPAdes v3.14.1 [parameters -

k 21,33,55,77,99,127 --only-assembler --careful] as described in (Ghaly *et al.*, 2021)

(Bankevich *et al.*, 2012). Annotation was performed using RASTtk, visualisation and

figure generation was performed using SnapGene (Aziz *et al.*, 2008).

2.2.29.3 Identification of Gene Cassettes and Taxa

Gene cassettes were identified and validated by searching for *attC* sites using

HattCI v1.0b, which uses a hidden Markov model to detect nucleotide sequence of

each core motif against *attCs* (i.e. R" - spacer" - L" - loop" - L' - spacer' - R') (Stokes *et al.*, 1997; Pereira *et al.*, 2016). attC-taxa.sh was then used to determine the taxonomic origin of the *attC* sites, this uses covariance models based on structural alignments and the cmsearch tool of Infernal v1.1.2 [parameters: --cpu 8 --notrunc --nohmm] to predict the folding structure of the *attC* (Ghaly, Tetu and Gillings, 2021).

Chapter 3: Quantification of Tn*21*, Genotypic and Phenotypic Identification of Mercury Resistant Organisms in Wastewater.

3.1   Introduction

Tn*21* is regularly found in environments that are directly impacted by humans, either because the environment is polluted with waste from medical, industrial and lifestyle products, which may contain antimicrobial compounds, toxic metal compounds and pharmaceuticals, or human pollution, such as faeces. The widespread distribution of Tn*21* is most likely due to its mobility and carriage on self-transmissible plasmids, as well as its ability to acquire antibiotic resistance cassettes (Liebert, *et al*., 1999). Tn*21* confers mercury resistance from the *mer* operon (*merRTPCADE*), which is a well conserved operon providing a pathway for the import of toxic mercury (II) ions to the cellular cytoplasm for reduction to mercury vapour by mercuric (II) reductase (*merA*). A class I integron is also located in the mobile element (Figure 1.13). Expression of the class I integron integrase is regulated by LexA in the SOS response. Under cell stress, novel gene cassettes may be captured and integrated to a cassette array (Guerin *et al.*, 2009). Antimicrobial resistance genes have been observed amongst the uptake of novel gene cassettes. The gene cassettes captured and kept by the integron may be advantageous to the survival of the host if exposed to a new environmental stressor, such as an antibiotic. Multiple studies have since identified the presence of intact antibiotics, such as sulfamethoxazole and trimethoprim, used to treat urinary tract infections, cleaning compounds such as quaternary ammonium compounds, painkillers and estrogenic compounds in wastewater (Nies, 1999; Gardner *et al.*, 2012; Cruz-Morató *et al.*, 2013; Venkatesan and Halden, 2015; Karkman *et al.*, 2018). Tn*21*-like transposable elements which are similar to the originally isolated Tn*21* found in plasmid R100, have been identified to carry

sulphonamide resistance genes such as *sul1* or *sul2*, aminoglycoside resistance, alternate folate reductases and beta-lactamases.

Due to the difficulties measuring mercury, less is known as to the concentration of mercury (II) found in waste waters compared to the antibiotic compounds mentioned previously. Quantifying other antimicrobials is also difficult, first they require extraction and are often measured with High Performance Liquid Chromatography. According to the Water Framework Directive (WFD), Hg is classed as a priority hazardous substance and therefore may not exceed 0.05 μg L^{-1} annually in rivers. 50% of Hg in wastewater is thought to have originated in dental amalgam fillings, waste disposal from industry or polluted soils (Bender, 2008). This may aid the selection for such an archaic resistance operon and therefore occurrence of genetic element such as Tn*21*.

3.1.1   Aims and Objectives

Due to the presence of antimicrobials such as sulphonamides, folate inhibitors and quaternary ammonium compounds in this environment, wastewater would be likely to be a prime location for Tn*21* to proliferate before disseminating into the environment. Therefore, the relative abundance of Tn*21* in this environment needs to be quantified in order to determine the potential impact Tn*21* may have post water treatment by disseminating resistance genes to the wider environment. To determine this, abundance of Tn*21* needs to be determined pre and post water processing, by comparing the influent and the effluent at different time points across a university term and holiday period. Whilst there is no legislation in the UK to

remove microorganisms or antimicrobial resistance genes from wastewater, the treatment process is targeted on measures such as reducing biological oxygen demand and ammonia (Water Resources England and Wales, 2017).

1. As a result, an indirect effect of wastewater treatment processes on bacterial diversity and abundance was predicted. When the university term starts and the local human population increases, an increase in diversity between the diversity and abundance of Tn*21*-like transposons should be observed.

2. From examining the antibiotic sensitivity of mercury resistant bacteria by using AST, it may also be possible to look at trends in co-occurrence of antimicrobial resistances within the wastewater and Tn*21*-like mobile elements.

3. This chapter therefore describes the determination of abundance of Tn*21*, isolation and antibiotic resistance characterisation of mercury resistant *Enterobacteriaceae* isolated from a wastewater treatment site across multiple time points, and sample locations throughout the treatment process.

4. Further to this, phylogeny of the *merA* gene in mercury resistant isolates will be compared to identify further variation between organisms. The chapter also aims to identify linkage between occurrence of resistance to different classes of antimicrobials within isolates.

3.2 Methods and Results

3.2.1 Determining the Concentration of Mercury (II) Within Wastewater Locations

Due to the COVID-19 pandemic, lab work ceased for four months, with limited work able to be performed for another month, and sampling of wastewater was halted due to studies detecting the virus within wastewater and the water treatment company would no longer allow sampling. The quantification of these mercury concentrations would have allowed for an insight as to whether mercury pollution was driving the presence of Tn*21*-like elements or whether it may be something else driving the selection of the transposons.

3.2.2 qPCR Validation: Calibration Curve

In order to determine the efficiency of the qPCR reactions, a calibration curve was used where a known concentration of template DNA is serially diluted ten-fold to 10^{-3} dilution. qPCR reactions are performed for each dilution and the C_q values are determined for each reaction. Theoretically, total PCR product should double every cycle meaning a C_q difference of 3.322 is expected between each ten-fold dilution. The C_q values were plotted against the nucleic acid dilution and linear regression analysis was performed to determine the R^2 value. Efficiency of between 90 and 110% is expected and widely accepted to account for human error, such as pipetting, and mild inhibition within the reaction.

A calibration curve was produced for both *rrsA* and the Tn*21* primer and probe set (Tn*21* qPCR_forward, Tn*21* qPCR_reverse and Tn*21* qPCR_Probe_0,

rrsA_qPCR_forward, rrsA_qPCR_reverse and rrsA_qPCR_probe). A known concentration of control sample DNA from J53 pMG101-A was used (1 ng μL^{-1}). Ten-fold dilutions were made up to 10^{-3} dilution and reactions performed separately with both primer and probe sets. The C$_q$ values were recorded and plotted (Figure 3.1).
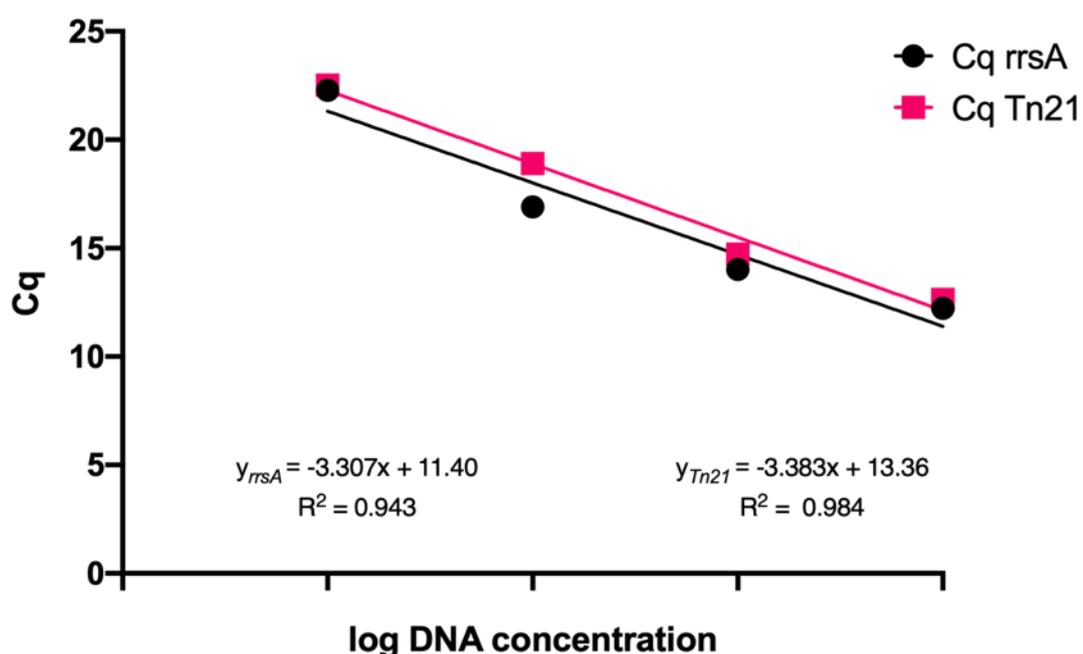
Figure 3.1 qPCR Calibration plot of primer/probe set *rrsA* and Tn*21*.

Green fluorescence (510 nm detection) was recorded after each cycle, and C$_q$ (quantification cycle) value appointed. For Tn*21* and rrsA, R^2 0.984 and 0.943, efficiency values of 96% and 101% were determined respectively through Q-Rex software (Qiagen, Hilden Germany).

E. coli J53 pMG101 was grown at 37 °C, with shaking at 200 rpm in LB broth. DNA was extracted from the bacterial cells, using the Monarch Genomic DNA prep kit (NEB) when an OD$_{600}$ 0.8 of the culture was achieved. DNA was then diluted to 1 ng μL^{-1} and quantified with Qubit 3 (2.2.10.2). qPCR was then conducted using the primers and probe for Tn*21* (Tn21 qPCR_forward, Tn21 qPCR_reverse and Tn21

qPCR_Probe_0) and *rrsA* (rrsA_qPCR_forward, rrsA_qPCR_reverse and rrsA_qPCR_probe) (Table 2.4). Ten-fold dilutions of 10^0 to 10^{-3} DNA template were used. Green fluorescence was recorded after each cycle and C_q values were assigned. PCR efficiency of the Tn*21* and *rrsA* targets were calculated. R^2 values and PCR efficiency of 0.984 and 96% and 0.943 and 101% were calculated respectively.

3.2.2.1   qPCR: Estimating Relative Abundance

To determine if there was a variation between abundance of Tn*21* between different wastewater sample points and dates, qPCR was used. Total DNA was extracted from each sample, quantified and diluted to 1 ng μL^{-1}. qPCR was then performed on all samples using *rrsA* and Tn*21* TaqMan primer/probe sets in technical triplicate.

3.2.2.2   *rrsA* as Validated as a Reference Gene

Figure 3.2 shows fluorescence of *rrsA* against cycle number. The data plotted included J53 pMG101 (control), and 6 other samples: campus influent, village influent and effluent. Each of these had two samples for two different time points, pre and during term time. Each reaction was performed with three technical repeats. The graph shows there is little variation between the C_q values of *rrsA*, except in J53 pMG101, this has a lower C_q value as it comes from a pure cell culture not wastewater where the copy number is slightly more varied between cell species. Each triplicate of the technical repeats were within 0.4 cycles of the others. The range in C_q values between the variables was 2, this variation is partly expected due to fluidity of the microbiome flowing through the wastewater treatment plant with slightly varying

127

rrsA copy numbers. Normalising the abundance data and correcting using The Ribosomal RNA Operon Copy Number Database (rrndb) will combat this (Stoddard *et al.*, 2015).
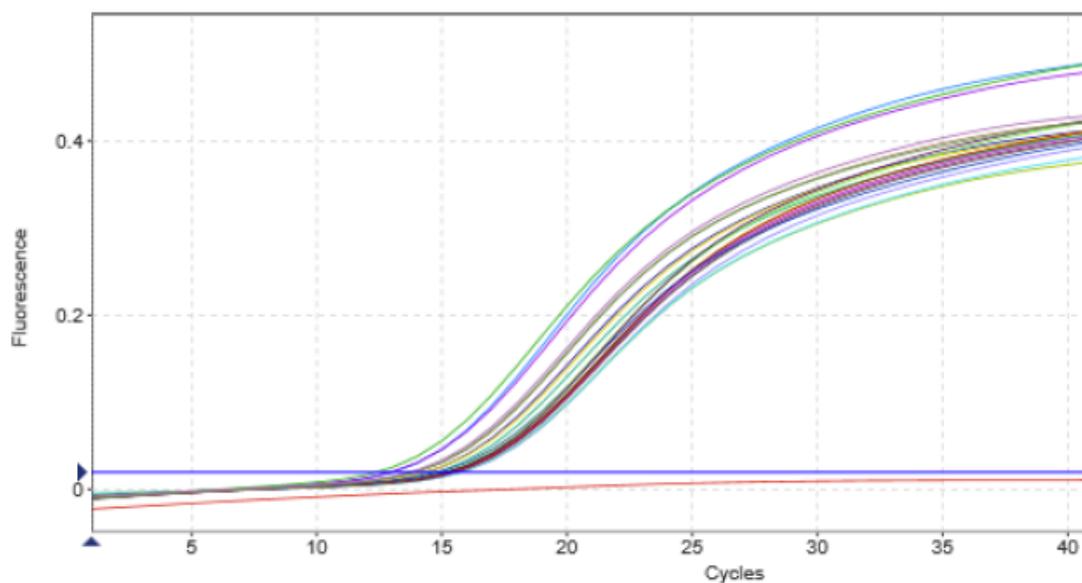
Figure 3.2 qPCR plot of rrsA DNA.

Normalised fluorescence (RFU) is plotted against the cycle number to identify the C_q values of each sample. This data was used to identify the number of copies of *rrsA* within each reaction. 1 ng μL^{-1} DNA extracted from each sample and control strain *E. coli* J53 pMG101 and amplified with primers (rrsA_qPCR_forward, rrsA_qPCR_reverse and rrsA_qPCR_probe). 3 technical repeats for each sample and the control.

 After validating the qPCR reaction, it was possible to resolve how frequently Tn*21* occurred in the wastewater samples. The relationship between concentrations of *rrsA* DNA and Tn*21* DNA was evaluated. All calculations of relative abundance are in relation to abundance of *rrsA*. An assumption was made for the mean copy number per cell of *rrsA* of in wastewater was estimated at 4.1 using the rrndb (Stoddard *et al.*, 2015). It is also of note that as with all reference genes used in qPCR, that there are limitations due to copy number disparity between bacterial species.

Firstly, the abundance data allowed quantification of Tn*21* in each wastewater environment. The influent possessed between 83 and 93-fold less copies of Tn*21* when compared to *rrsA*. The effluent however between 93 and 96-fold less copies compared to *rrsA* (Table 3.1). Initially, it is quite clear that wastewater treatment processing has an effect on the relative abundance of Tn*21*. At both time points, before term time and during term time, the wastewater treatment process was shown to remove between 48 and 55% of Tn*21* from the bacterial population. [It is worth noting due to the global COVID-19 pandemic, a third time point sample later in the term (approximately week commencing 23/03/2020) would have been collected to identify whether the Tn*21* abundance would have returned to the levels seen before term time]. The lab was closed from 18/03/2020-late July 2020.

	Pre-Term time			During Term time		
	Campus Influent	Village Influent	Effluent	Campus Influent	Village Influent	Effluent
Mean	0.0750	0.0702	0.0398	0.1630	0.1712	0.0798
Standard deviation	0.0063	0.0107	0.0109	0.0102	0.0103	0.0021
Mean (rrndb corrected)	0.0183	0.0171	0.0097	0.0398	0.0417	0.0195

Table 3.1 Mean relative quantities of Tn*21* in wastewater.

The samples were isolated at three locations: campus influent, village influent and effluent pipes to a wastewater treatment plant. Samples collected pre and during university term times. Adjustment for rrndb was calculated by dividing mean relative abundance by 4.1.

Relative abundance of Tn*21* in the influent varied between time points according to whether the university nearby was in term time or not (Figure 3.3). An approximate two-fold increase in Tn*21* was observed from pre-term time to term time samples

across influent and effluent as the local population increased due to an influx of students from around the UK and overseas. Notably, across both time points there was no statistical significance at P<0.05 seen between campus and village influent samples within time points.
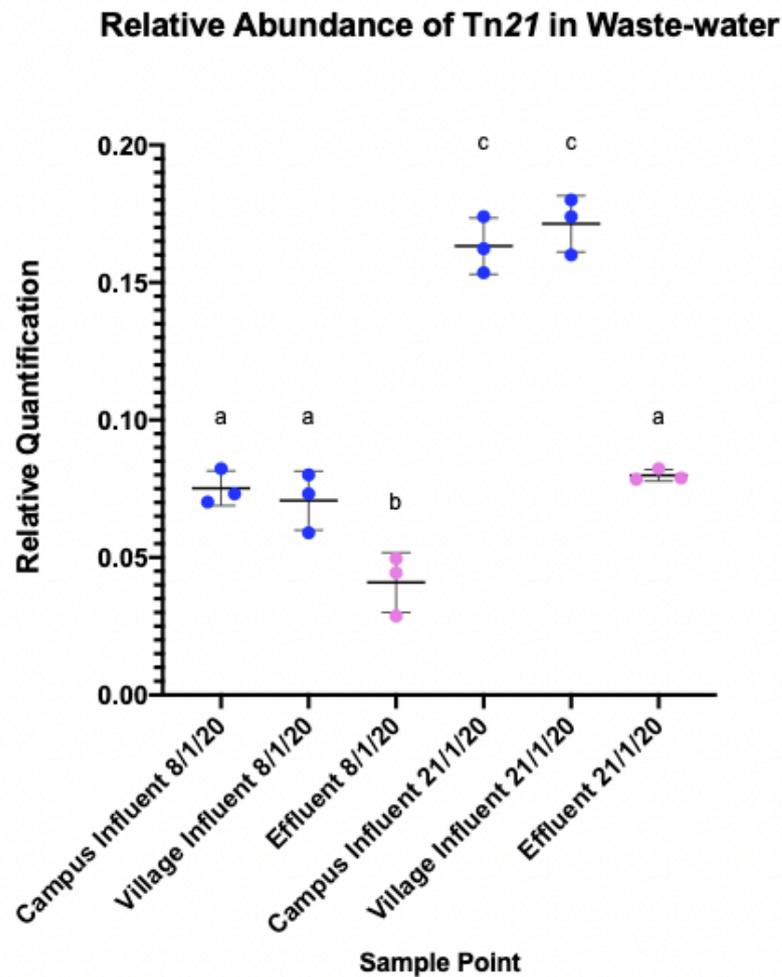
Figure 3.3 Relative abundance of Tn*21* across influent and effluent over term time and pre-term time.

The means and standard deviations are represented by the bars. Blue dots indicate influent samples, pink indicates effluent samples. Multiple one-way ANOVA tests were conducted on all datasets and statistical significance was noted where P<0.05. a, b and c respectively refer significant difference between each sample set.

The wastewater treatment process was noted to influence the relative population of Tn*21* in the wastewater samples. Across both time points, an approximate two-fold reduction was observed. Figure 3.4 clearly demonstrates this change alongside the individual values displayed in Figure 3.3. This reduction in Tn*21* abundance corroborates my hypothesis that bacterial genetic diversity would be reduced over the course of treatment of wastewater.
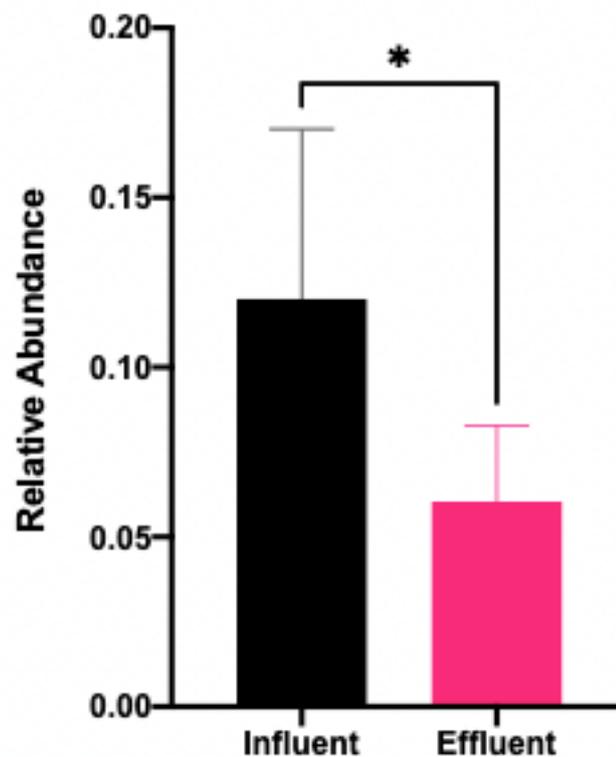
Figure 3.4 Relative abundance of Tn*21* across combined influent and effluent samples

Influent refers to the combined relative abundance of campus and village influent samples across the two before and during term time points. Effluent refers to the combined relative abundance of effluent samples across the two before and during term time points. Relative abundance of Tn*21* calculated from the qPCR Cq values using the formula RQ = $2^{-\Delta\Delta Cq}$. Mean relative abundance and standard deviation were plotted. * P≤0.05 (two tailed, unpaired, t-test) comparison to effluent.

131

Finally, the qPCR allowed for identification of differences in Tn*21* abundance between term time and pre-term time as a whole. When looking at the grouped data, it aligned with the individual data in Figure 3.3. Figure 3.5 confirms that there was a statistical difference between the abundance of Tn*21* in wastewater before and after term starts. A higher proportion of Tn*21* is released to the environment after wastewater treatment in term time. The relative abundance the Tn*21* pre-termtime influent is similar to that of during term time effluent. When the two samples were compared showed there to be no statistical difference between the influents in a two-way unpaired t-test where P ≤0.05 (Figure 3.3).
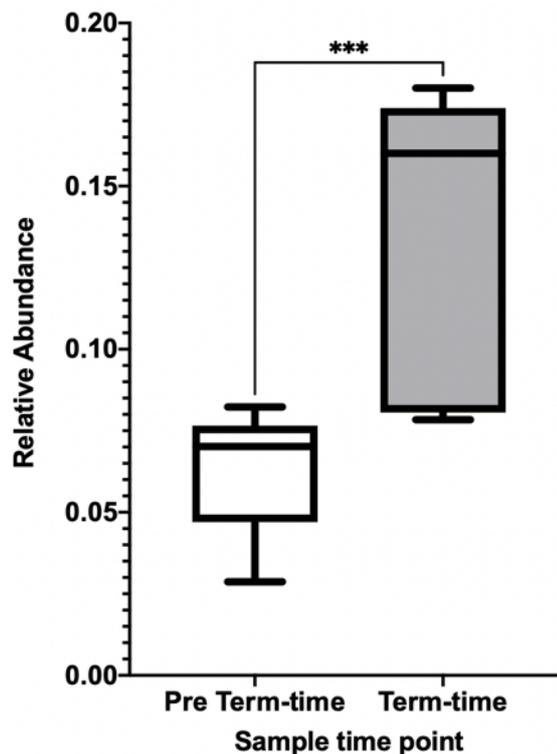
Figure 3.5 Box and violin plot of relative abundance of Tn*21* in wastewater samples before and during term time.

Datasets of all the influent and effluent samples from each time point were pooled to generate pre-term time and term time datasets. Relative abundance was plotted with mean

and whiskers as maxima and minima. *** p ≤0.001 (two tailed, unpaired, t-test) comparison to pre-term time.

Overall, the data supports the theory that wastewater treatment processing will influence Tn*21*. Secondly, an influx of a diverse student population to a local area caused an increase in the presence of Tn*21* and Tn*21*-like transposable elements in wastewater which are being released into the environment.

3.2.3 Isolation of Mercury Resistant *Enterobacteriaceae*

In order to isolate *E. coli*, tryptone bile X-glucuronide (TBX) agar was used as a selective medium. 3-4% of *E. coli* strains are glucuronidase negative, the most important of which is enterohaemorrhagic *E. coli* O157:H7. X-glucuronide is cleaved by glucuronidase to produce 5,5'-dibromo-4,4'-dichloro-indigo, a green/blue pigment, and glucuronide. The chromogenic agar therefore allows for easy identification of non O157 *E. coli* strains for further testing. However, it quickly became apparent that not all glucuronidase positive colonies with the same colony morphologies were *E. coli*, but rather *Enterobacteriaceae* and *Aeromonad* families were isolated. The intended use of TBX media is for enumeration and detection of *E. coli* in food and animal feed.

TBX media was also used to enumerate presumptive *E. coli* and glucuronide negative bacteria within the wastewater samples collected. Total colony counts on TBX were recorded as well as presumptive *E. coli* colonies, displaying green colony colour. Colony counts were also recorded between time point samples, and sample locations (Table 3.2).

Sample location and time point	Colony Count Phenotype on TBX agar/ CFU mL^{-1}		
	Green	White	Total
Campus Influent 10/10/19	3.67×10^4	6.90×10^5	7.27×10^5
Village Influent 10/10/19	1.50×10^5	4.40×10^5	5.90×10^5
Campus Influent 08/01/20	7.43×10^5	9.33×10^4	8.37×10^5
Village Influent 08/01/2020	5.80×10^4	3.57×10^4	9.37×10^4
Campus Influent 21/01/20	7.57×10^4	1.59×10^5	2.35×10^5
Village Influent 21/01/20	9.33×10^4	1.58×10^5	2.51×10^5
Effluent 21/01/20	6.23×10^3	5.37×10^3	1.16×10^4

Table 3.2 Cell counts of presumptive *Enterobacteriaceae* from wastewater treatment plant over different time points.

Green denotes glucuronidase positive organisms and white denotes glucuronidase negative organsims.

CHROMagar was used to selectively isolate *Enterobacteriaceae*. Isolate identities were initially confirmed by using Kovacs reagent for the metabolism of indole and testing for oxidase test strips were used to identify the presence of oxidase. Bacterial isolate identities were further confirmed by 16S rRNA sequencing (Eurofins Genomics). The 16S sequencing confirmed that although TBX agar is specified for the isolation of *E. coli*, other *Enterobacteriaceae* are capable of growing and matching the morphology of glucuronidase positive *E. coli*, such as *Citrobacter* and *Enterobacter.* This discovery turned out to be a positive for this study as Tn*21*-like transposons tend to be found in gammaproteobacteria, not just *E. coli*.

Tukey's multiple comparisons test	Significant Difference	P Value
Campus Influent 10/10/19 vs. Campus Influent 21/01/20	*	0.0303
Village Influent 10/10/19 vs. Campus Influent 21/01/20	*	0.0135
Campus Influent 08/01/20 vs. Campus Influent 21/01/20	*	0.0446
Village Influent 08/01/2020 vs. Campus Influent 21/01/20	**	0.0032
Campus Influent 21/01/20 vs. Village Influent 21/01/20	**	0.0056
Campus Influent 21/01/20 vs. Effluent 21/01/20	**	0.0024

Table 3.3 Statistical outcomes from comparison of total CFU mL^{-1}.

Sample time points and sampling locations * $p \leq 0.05$, ** $p \leq 0.01$ (two tailed, unpaired, t-test). There is a statistically significant difference between sample collection time points between term time and non-term time. There is statistically significant difference between the campus and village during term time.

Colonies that had a blue/green appearance on TBX agar, glucuronidase positive isolates, were then screened for phenotypic resistance to Hg (II). Of those screened, 17.5% of them demonstrated phenotypic resistance to 25 µg mL^{-1} Hg (II) on LB agar. Over two years isolating bacteria from wastewater samples from this wastewater environment, a total of 572 individual *Enterobacteriaceae* were isolated: 16.4% of those isolates screened had phenotypic mercury resistance. presence of these genes was tested using PCR in order to confirm presence of Tn*21* (*merA, merC, merR* and *intI1*) (Table 2.3). This process was later streamlined to identify Tn*21-like* transposable elements by amplifying the IR-L region of the Tn*21* insertion sequence and *tnpA* gene as these are also highly conserved sequences. A multiplex PCR was used in order to amplify *merA, merC, merR* and *intI1* genes. The product sizes were 1230 bp, 250 bp, 218 bp and 280 bp respectively (Figure 3.6, Figure 3.7).
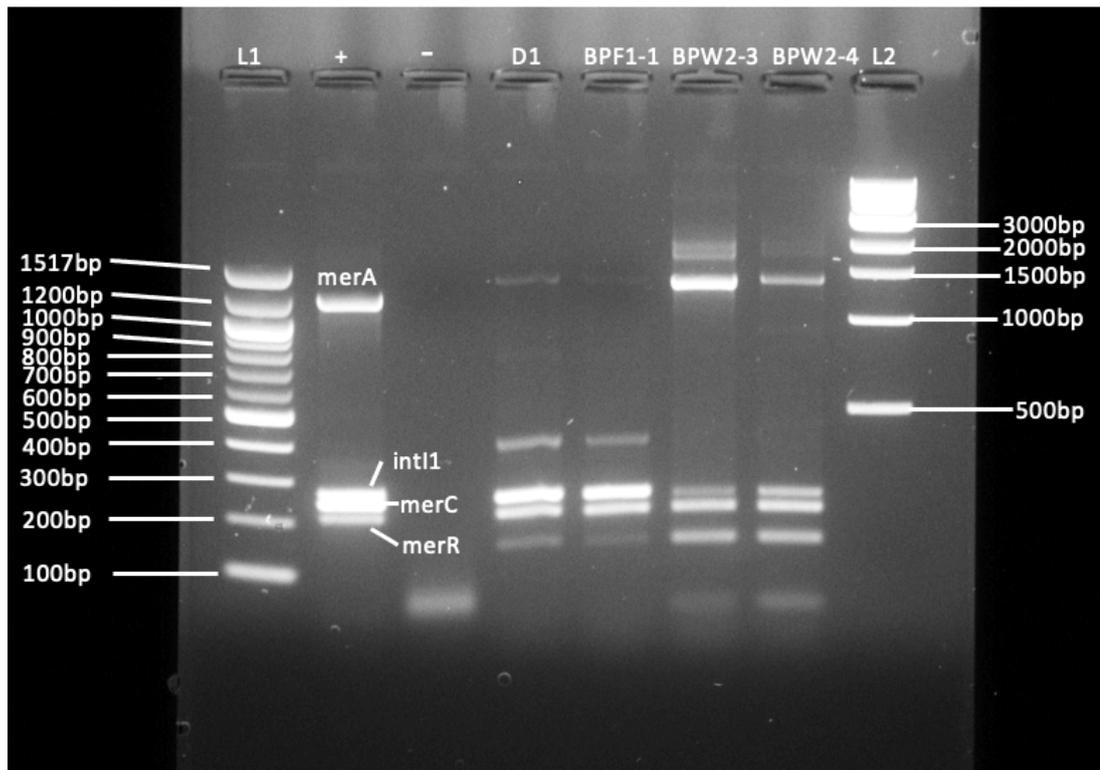
Figure 3.6 Ethidium Bromide 1.0% agarose gel image of *merACR* and *intI1* multiplex PCR.

L1 NEB quick-load 100 bp ladder, + plasmid prep of pMG101 which possesses a Tn*21*, - Negative control of nuclease free water, D1 wastewater isolate from a drainage ditch contaminated with grey water. BPF1-1, BPW2-3 and BPW2-4 wastewater isolates collected in a brook downstream of the wastewater treatment plant effluent pipe.
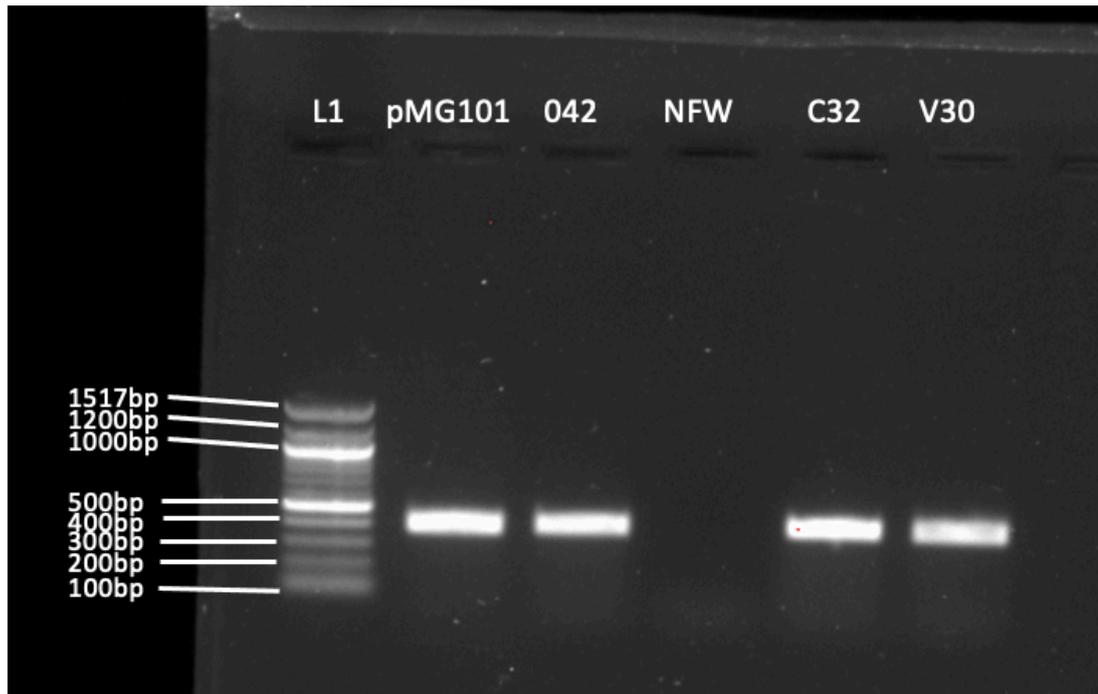
Figure 3.7 Ethidium Bromide 1.0% agarose gel image of Tn*21* IR-L to *tnpA* by PCR amplification.

Product size expected is 392 bp. pMG101 and 042, were used as positive controls. NFW, nuclease-free water negative control. C32 and V30 are wastewater isolates from campus and village influents respectively.

3.2.3.1   Isolation of *mer* Operon Carrying *Pseudomonads*

As mentioned in 3.2.3, TBX agar was not as selective as described by the manufacturers. In fact, *Pseudomonads* were also identified on the media and further screened for phenotypic resistance to mercury (II) chloride. These were then screened further for Tn*21*, *merA*, *merC*, *merR* and *intI1* by PCR.

3.2.3.2   *merA* Phylogeny of Wastewater Isolates

PCR products of mercury resistant isolates containing Tn*21-like* transposons were sent to Eurofins for Sanger sequencing of the *merA* gene as this key gene in

mercury (II) reduction (and hence resistance) is highly conserved. The size of the *merA* amplicon sent for Sanger sequencing was 1230 bp, (the full gene length is 1707 bp.) A forward (merA F primer) and reverse sequencing read (merA R primer) was performed on each sample and the sequences aligned to give 2X coverage (Table 2.3). A phylogenetic tree was produced comparing *merA* genes from *E. coli* strain 042 (containing chromosomally inserted Tn*21*), *mer* genes isolated from wastewater, a local dairy herd and sheep footbath. Plasmids R100 (originally characterised *merA* gene), pMG101-A and pMG101-B (Hooton *et al*, 2021) were also included in the phylogenetic tree. Bootstrapping to the value of 100 was also performed (Figure 3.8). The phylogenetic tree produced four distinct clades of *merA* genes with the majority of classes contained within two clades. The remaining two clades were made up from five isolates. The relationships of isolates within classes were varied; including the time points at which they were isolated and the locations they were sampled. Most notably, 042, R100 (ORI), pMG101A and B existed within the same clade, alongside isolates from downstream of a wastewater treatment plant, a sheep foot bath and sheep foot from a neighbouring farm. Within the same clade were also isolates from a dairy unit on the same neighbouring farm, isolated between 2012 and 2014. The bootstrapping values gave great confidence when determining the four classes seen in Figure 3.8.
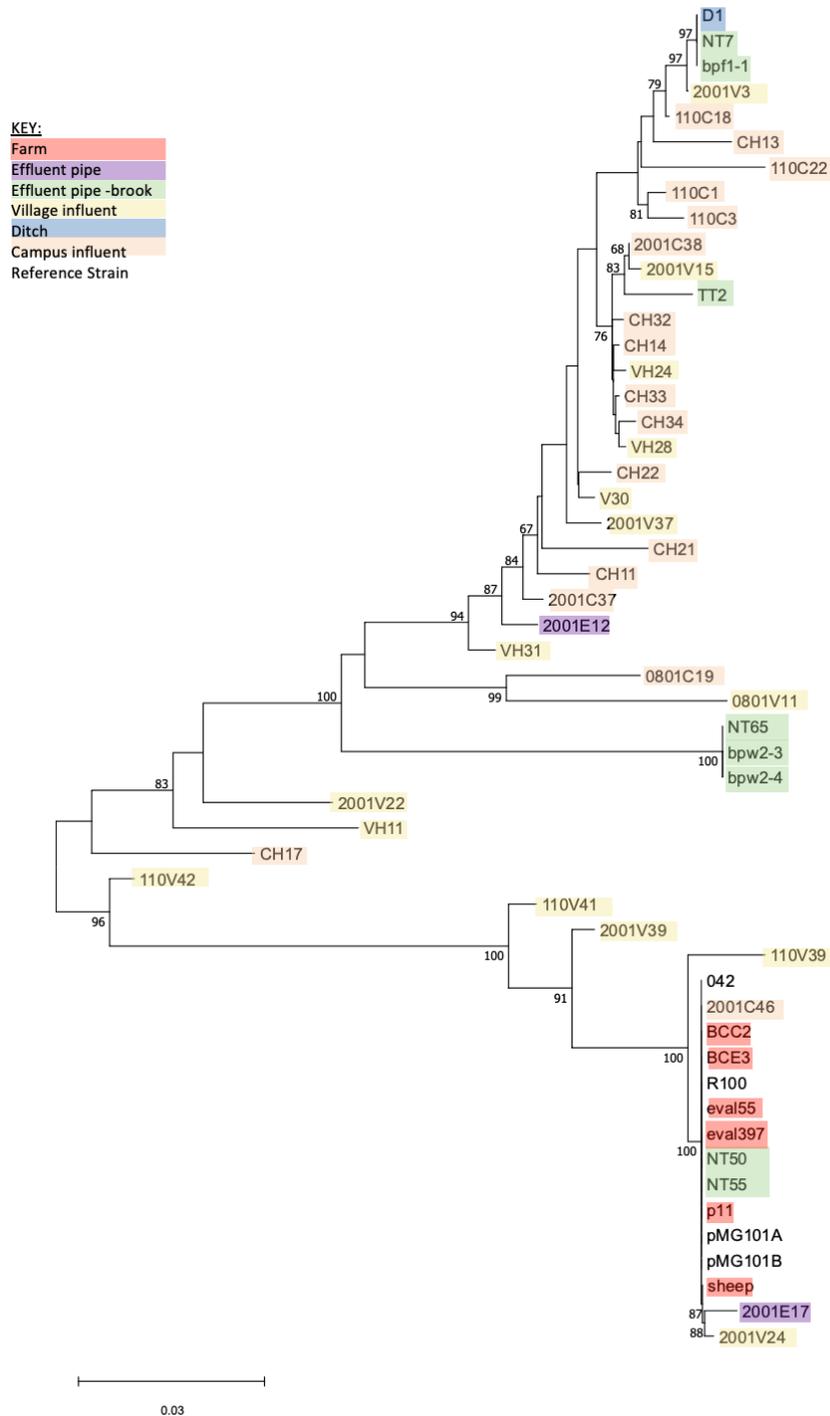
Figure 3.8 Phylogenetic tree of *merA* genes of isolated bacteria from wastewater and surrounding areas.

A MAFFT alignment was performed and the aligned sequences were processed using bootstrapping (100 times) and Akaike Information Criterion was performed to generate the tree (Lefort, Longueville and Gascuel, 2017). Bootstrap values less than 65 are not shown. (Appendix shows isolate origins and antimicrobial sensitivity data).

139

Due to the limitations of Sanger sequencing, only approximately 1000 bases were consistently sequenced between all isolates. A consensus sequence was determined from the area covered (Figure 3.9). Despite being a very well conserved operon, the *merA* gene shows more variability than expected. Most of the variations between sequences are constrained to the same regions. The variable regions were between 5 and 10 bases in length. When examining the consensus sequence however, only one variable region was noted at four bases in length, between bases 468 and 472 (Figure 3.9). Other differences between sequences were single nucleotide polymorphisms (SNPs) also often noted in similar regions. Up to 100 bases may be lost in the Sanger sequencing process.

```
ATNGCNNCNACNNCGCCGACNATCNNGCGCNNGNNGCTGCTGGCCCAGCA        50
GCAGGCCCGNGTCGANGAACTNCGNCANGCCAAGTACGAAGGCATCNTGG       100
ANGGCAATCNNGCNATCACNGTNCTGCACGGNNNNGCNCGNTTNAAGGAC       150
NANCNNANCCTNATCGTNNNNNTNAACGANGGNGGCGAGCGCGTNGTGNN       200
NTTCGACCGCTGCCTGNTCGCCACNGGNGCNAGCCCGGCNGTNCCGCCGA       250
TTCCNGGCNTGAAAGANNCNCCNTACTGGACTTCCACNGANGCNCTGGNN       300
AGCGANACNATTCCNNANCGCCTNGCCGTNATNGGCTCNTCNGTGGTGGC       350
GCTGGANCTGGCGCANGCNTTNGCCCGNCTNGGNNNNAAGGTNACGNNCC       400
TGGCNCGCANNACNNTGTTCTTCCGNGAAGACCCNGCNATNGGCGANGCN       450
GTNACNGCCGCNTTCCGNNNNGANGGNATCNAGGTNNNGGANNACACNCA       500
NGCCAGCCAGGTCGCNNATNTCAATGGTGAAGGACGGNGAATTCGTGCTN       550
ACCACNNNGCACGGNGAANTCGCGCCGACNAGCTGCTGGTCGCCACNGG       600
CNGNNCNCCNAACACNCGCANNCTGGCANTGGANGCGNCGGGNGTNNCNN       650
TCANNNCGCANGGNGCNATCGTCATCGACNNNGGCATGCGNACNAGNNNN       700
NNNNACATCTACGCNGCNGGCGACTGCACCGACCAGCCGCAGTTCGTCTA       750
TGTGGCGGCAGCGGCCGGCACNCGNGCNGCGATCAACATGACNGGCGGNG       800
ANGCGGCCCTGNACCTGACCGCNATGCCGGCCGTGGTGTTCACCGACCCG       850
CANGTNGCNACCGTNGGCTACAGCGAGGCGGAAGCNCANCANGACGGNAT       900
CNANACNGANAGTCGCNNGCTNACNCTGGANAACGTGCCGCGNGCGCTNG       950
CCAACTTCGACACNCGCGGCTTCATCAANCTGGTNNTNGANGAAGGNAGC      1000
GGACGNCTNATCGGCGTGCANGCNGTGGCCCCGGAAGCGGGNGAACTGAT      1050
NNNNNNNN                                                1058
```

Figure 3.9 Consensus sequence of *merA* covered from Sanger sequencing of environmental isolates.

53 *merA* sequences were aligned to produce a consensus of an internal section from 493 bases from the translational start site of *merA* to 1544. The threshold for an absolute value was determined by base presence being greater than 90%.

3.2.3.3   Antibiotic Resistance Profiling of Mercury Resistance

All bacteria isolated that were phenotypically resistant to mercury (II) were screened

for resistance to 16 different antibiotics representing 14 classes of antibiotics.

(2.2.17). A multidrug resistant organism is defined as an organism displaying

resistance to 3 or more classes of antibiotics (Magiorakos *et al.*, 2012). By this

definition 20.2% of all phenotypically mercury resistant isolates displayed multi-drug

resistance (Figure 3.10). All isolates displaying 7 different resistances were isolated

from different locations at different dates but possessed resistances to the same

antibiotics. Similar patterns in the resistome were noted for other isolates.

141
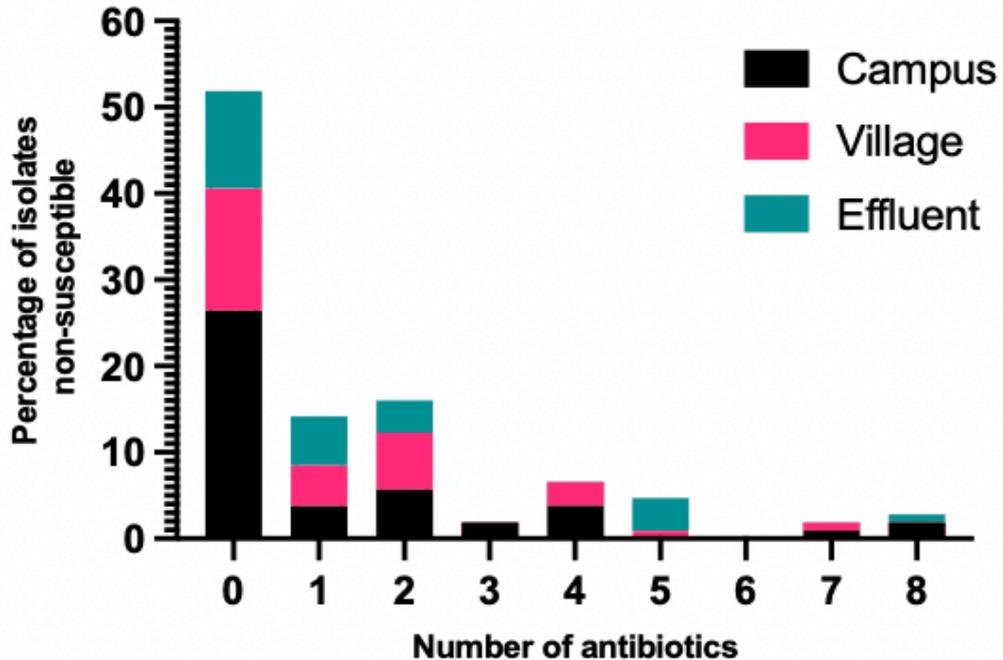
Figure 3.10 Percentage breakdown of mercury (II) chloride resistant isolates and the number of antibiotics they are non-susceptible to.

Campus influent n = 47 village influent n = 32 effluent n = 27.

Greater than 50% of all mercury resistant isolates did not possess resistance to any of the 14 classes of antibiotics screened, meaning some isolates may not possess Tn*21*. The remaining 43% of isolates with multiple resistances followed trends, showing similar resistance patterns. Of all *Enterobacteriaceae* isolated, 48.6% were non-susceptible to at least one form of antibiotic alongside being resistant to mercury (II) chloride.

A two-tailed unpaired t-test with 95% confidence interval was performed to identify any difference in the number of susceptibilities to antibiotics between the three populations, campus influent, village influent and effluent. There was no statistically significant difference when comparing any of the populations (Table 3.4).

	Campus vs Influent	Campus vs Effluent	Village vs Effluent
P value	0.9202	0.6189	0.68
Statistical significance	ns	ns	ns
t value	0.1006	0.4996	0.4146
df	77	72	57

Table 3.4 Statistical test results from two-tailed unpaired t-test comparing the number of non-susceptibilities to antibiotics between the three isolation populations.

The mercury resistant isolates were also categorized by how many isolates possessed resistance to each antibiotic screened for. The screened isolates were most commonly non-susceptible to ampicillin, aminoglycosides, tetracycline, aztreonam (monobactam) and the combination of amoxicillin and clavulanic acid (Figure 3.11). The term non-susceptible encompasses both resistant and intermediate sensitivity to antimicrobials. The most common of which was non-susceptibility to sulphonamides (25.2%). Notably, no isolates possessed resistance to carbapenems such as imipenem. The number of isolates non-susceptible to chloramphenicol, aztreonam, nitrofurantoin and the third generation cephalosporins cefotaxime and ceftazidime were relatively low in comparison to the other classes of antibiotics screened. Occurrence was less than 5% for each of these resistances. These data may suggest there is likely no link between mercury resistance and extended spectrum beta-lactamase (ESBL) and extended spectrum cephalosporinase (ESC) producing bacteria. This should be tested further given the previous isolation of beta lactamase genes such as oxacillinases in Tn*21* operons.
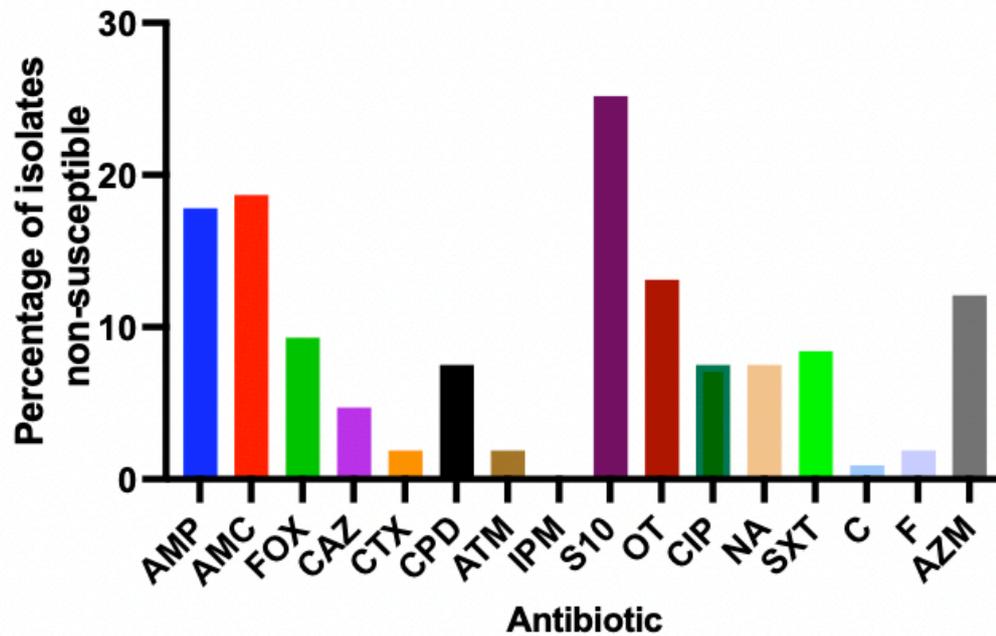
Figure 3.11 A breakdown of percentage of mercury resistant isolates displaying non-susceptibility to each antibiotic.

107 *Enterobacteriaceae* isolates were screened for resistance to each antibiotic using CLSI standard methods. Concentrations of each antibiotic can be found in section 2.2.17. Breakdowns of each isolate AST data can be found in Figure 8.1, Figure 8.2 and Figure 8.3.

The susceptibility data also showed that there was a reduction in the proportion of isolates non-susceptible to first generation beta lactams and those non-susceptible to third generation cephalosporins meaning co-carriage of ESBL and ESC genes was likely uncommon in mercury resistant bacteria. Susceptibility to antibiotics were defined by the breakpoints of CLSI M100 document. Intermediate resistance was considered as a zone of clearing measured on Muller-Hinton agar greater than that of resistance and less than the diameter of sensitivity.

144

3.2.4 ESBL Selection of Wastewater Bacteria to Isolate Tn*21* in Wastewater

Due to the growing emergence of clinically and environmentally isolated ESBL *Enterobacteriaceae* there may be a link between the two populations, perhaps a co-selection or co-occurrence may play a role in the dissemination of these resistances. This was tested by a round of selection for resistance to cephalosporins using CHROMagar ESBL first and then screening for resistance to mercury (II) chloride on the isolates that grew on CHROMagar ESBL, as screening together proved too toxic on isolates containing both resistances in the preliminary trial.

Bacteria isolated from both campus and village influents on CHROMagar ESBL were screened for resistance to the same 16 antibiotics. None of the isolates from CHROMagar displayed resistance to mercury (II) chloride.

Despite the lack of resistance to mercury (II) chloride identified, all of the isolates displayed multi-drug resistance. Only one isolate possessed resistance to three or less classes of antibiotics. All isolates displayed non-susceptibility to five or more antibiotics (Figure 3.12). Notably, the majority of isolates were not sensitive to between 6 and 9 antibiotics.
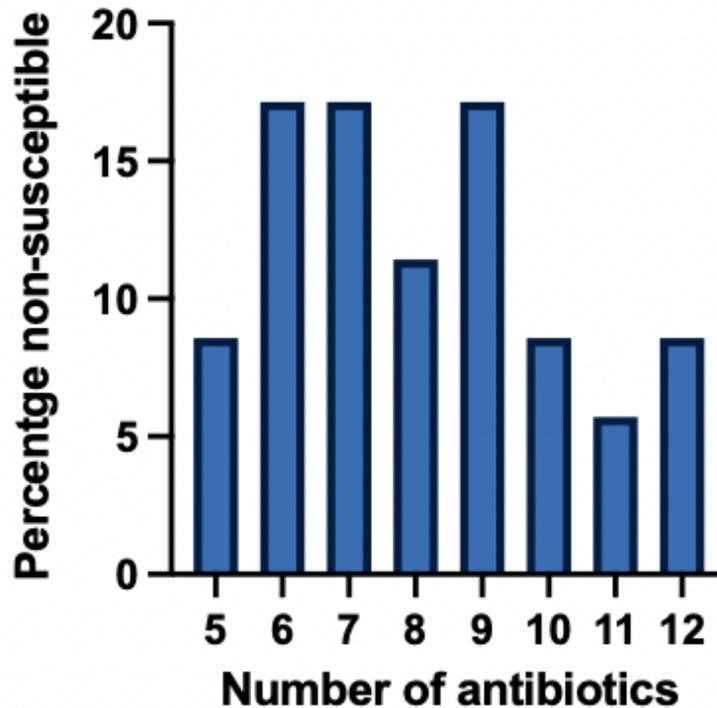
Figure 3.12 Percentage breakdown of ESBL isolates non-susceptibilities.

n = 35. 16 classes of antibiotics were screened for resistance as per section 2.2.17.

Due to the selective agent for CHROMagar ESBL, ESBL supplement, identification of resistance to at least two classes of antibiotics, penicillins and third generation cephalosporins was expected. Figure 3.13 shows 100% of isolates displayed resistance to ampicillin, cefotaxime and cefpodoxime. Despite screening for resistance to 3rd and 4th generation cephalosporins, none of the isolates displayed resistance to the combination of clavulanic acid, a beta lactamase inhibitor, and amoxicillin.
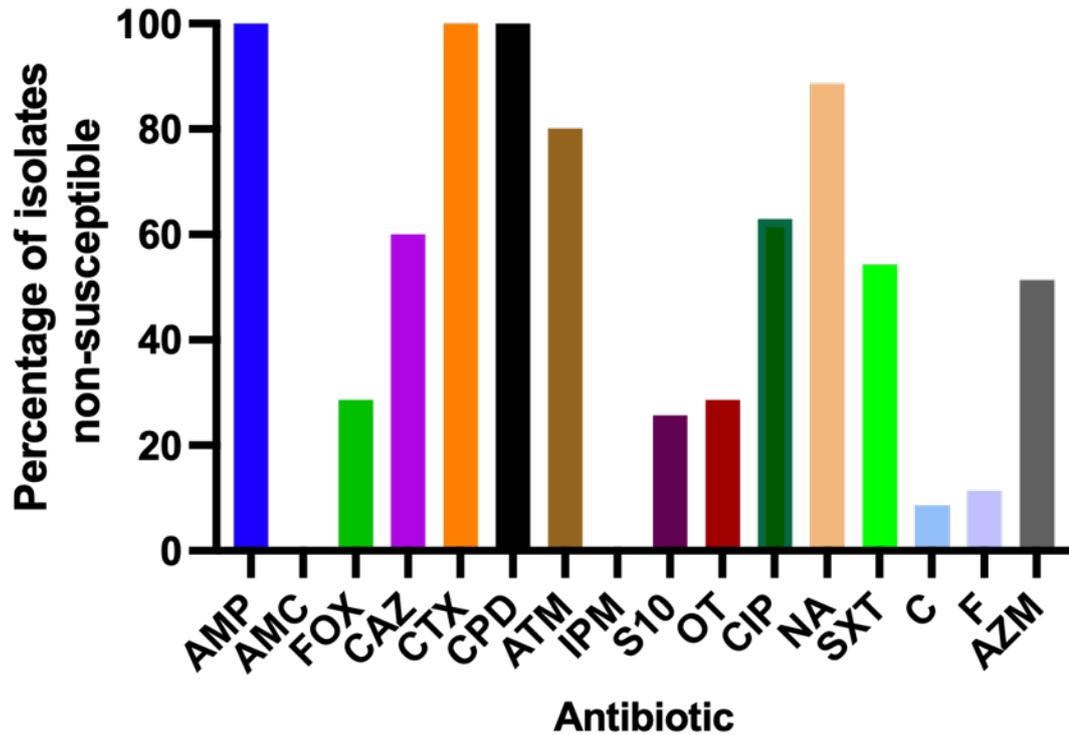
Figure 3.13 Breakdown of resistance patterns of CHROMagar ESBL agar isolates displaying non-susceptibility to each of the 16 antibiotics screened for.

n=35. Concentrations of each antibiotic can be found in section 2.2.17. Breakdowns of each isolate AST data can be found in Figure 8.4

Resistance to monobactam, quinolones, fluoroquinolones and folates were widespread. Similar to mercury the resistant isolates screened in 3.2.3.3 from the same environment, chloramphenicol, nitrofurantoin and imipenem sensitivity was observed in the bacteria isolated from CHROMagar ESBL. Folate resistance and macrolide resistance was also quite commonly observed with more than 50% resistant to each antibiotic (51.4% and 54.3% respectively). In both mercury resistant isolates and ESBL isolates from wastewater, 25% of isolates were non-susceptible to aminoglycosides.

This data was then added to another data set of isolates screened for susceptibility to antibiotics after being isolated on CHROMagar from the same source (Figure 3.15). These isolates were also screened for resistance to mercury (II) chloride, none of them displayed any resistance. Similar to Figure 3.13, non-susceptibility to macrolides (50%), aminoglycosides (40%), folates (48%), tetracycline (33%), fluoroquinolones (54%) and quinolones (74%) were widespread throughout all isolates Figure 3.15. Susceptibility to chloramphenicol and nitrofurantoin was also very common, 12% and 18% were non-susceptible respectively.
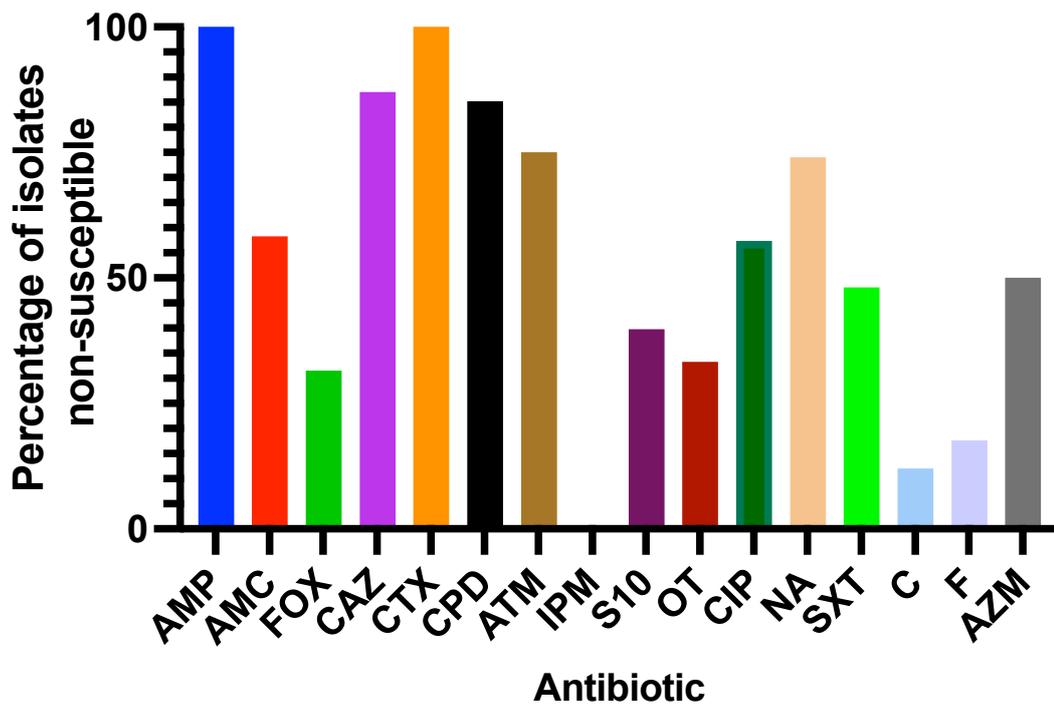
Figure 3.14 Breakdown of population ESBL isolates from multiple datasets displaying non-susceptibility to each antibiotic screened.

Each isolate was identified using CHROMagar from the same wastewater treatment plant. n = 108. Concentrations of each antibiotic can be found in section 2.2.17.

The CHROMagar selection of non-susceptible ESBL isolates was less efficient at isolating bacteria resistant to cefoxitin. 30% of isolates were non-susceptible, much lower than all other beta-lactams, cephalosporins and monobactams tested.

The breakdown of non-susceptibilities of strains between campus and village influents shows village influent to possess more bacteria that are non-susceptible to a greater number of antibiotics. However, the campus influent group possessed the greatest number of non-susceptibilities in a single isolate (13). Isolates originating from the village influent possessed more non-susceptibilities to antibiotics (8.80 3 s.f.) whereas campus influent possessed less (8.61 3 s.f.). Non-susceptibility of 6 to 9 antibiotics accounted for 57% of all the isolates screened for antibiotic sensitivity.
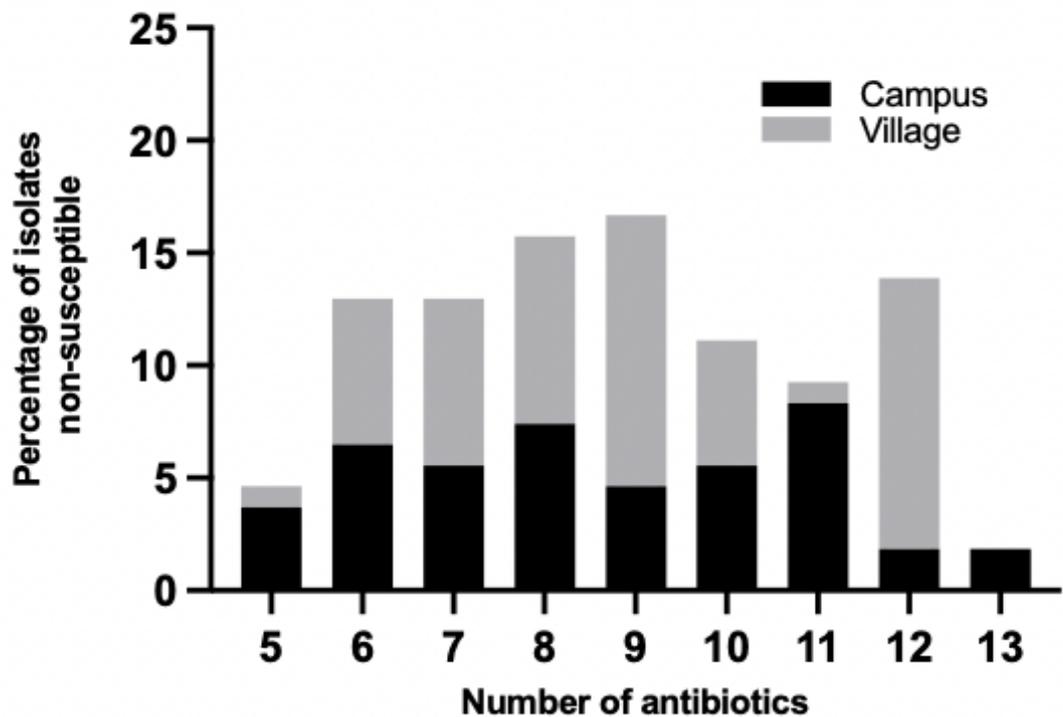
Figure 3.15 Percentage breakdown of ESBL isolates from each influent source and the number of antibiotics they are non-susceptible to.

Campus influent n = 49, village influent n = 59.

An unpaired t-test with 95% confidence interval was performed to compare the number of non-susceptibilities between isolates originating from campus influent and village influents (same dataset as Figure 3.15). After confirming the two datasets were normally distributed using a D'Agostino & Pearson normality test; an unpaired, two tailed t-test was performed to compare the resistance patterns of the campus influent and village influent isolates: $t = 0.5467$ (2, N = 108), $p = 0.5857$, meaning there was no significant difference between the two resistance patterns between the two populations of ESBL producing isolates.

3.2.4.1   Screening ESBL Isolates for ESBL Genes

ESBL producing isolates were screened for four commonly observed resistance genes which confer resistance to extended spectrum beta-lactams: bla_{CTX}, bla_{SHV}, bla_{TEM} and bla_{OXA-1} genes. Notably bla_{SHV} presence is rarely identified in the UK so presence of this gene was not expected in any of the isolates that were screened. Co-occurrence of multiple ESBL genes was also identified and plotted together on Figure 3.16. bla_{SHV} was not identified in any isolates as expected.

bla_{CTX} was co-occurrent with both bla_{TEM} and bla_{OXA} but this was approximately five-fold less than individual occurrence of bla_{CTX} and bla_{TEM}, 3.2-fold for bla_{OXA}. Co-occurrence of bla_{TEM} and bla_{OXA} was more frequently (5.7%) observed than co-occurrence of either bla_{CTX} bla_{TEM} and bla_{OXA} bla_{CTX}. Three of the isolates tested for presence of these four genes did not carry any of them. However, this is entirely

possible as the ESBL genes screened for are four out of hundreds of ESBL genes previously characterised.
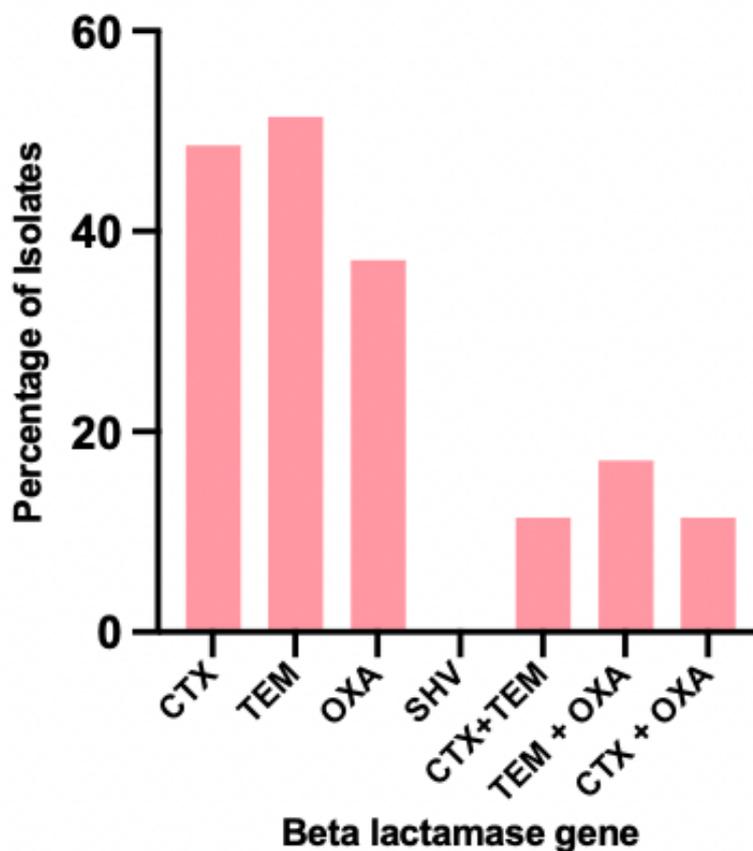
Figure 3.16 Percentage of ESBL producing isolates from both campus and village influent screened for *bla* genes. n=35.

3.2.5 Co-occurrence of Mercury (II) Chloride Resistance and Other Resistance Genes

Bacteria from wastewater that were isolated on TBX agar were screened via replica plating for resistance to, mercury (II) chloride, trimethoprim, streptomycin, tetracycline and cefotaxime. Bacteria were isolated from different time points, influent locations, and effluent. Streptomycin, an aminoglycoside, was selected as resistance genes to this class are commonly observed in class I integrons in Tn*21-like* transposable elements. Similarly, trimethoprim resistance in the form of folate

reductase gene cassettes have been identified within class I integrons. Tetracycline

was shown to be abundant in the wastewater at these wastewater sites (unpublished

data, Garduno Jimenez) and a Tn*3*-like transposable element conferring resistance to

tetracycline has been associated in large self-transmissible plasmids to be adjacent

to Tn*21-like* transposons. Finally, cefotaxime was used to identify resistance and

linkage to co-occurrence of extended spectrum beta-lactam and mercury resistance.

Sample Location	Percentage of Isolates resistant / %				
	$HgCl_2$	TET	CTX	STREP	TMP
Campus	25.00	20.14	16.67	15.97	5.56
Village	18.06	12.50	20.14	20.83	6.94
Influent (TOTAL)	21.53	16.32	18.40	18.40	6.25
Effluent	25.00	25.00	14.58	22.92	10.42
Total	22.29	17.77	18.07	19.28	6.93

Table 3.5 Percentage occurrence resistance to each antimicrobial in each sample location.

Influent (TOTAL) describes the combined frequency of the influent samples, from both village
and campus influents.
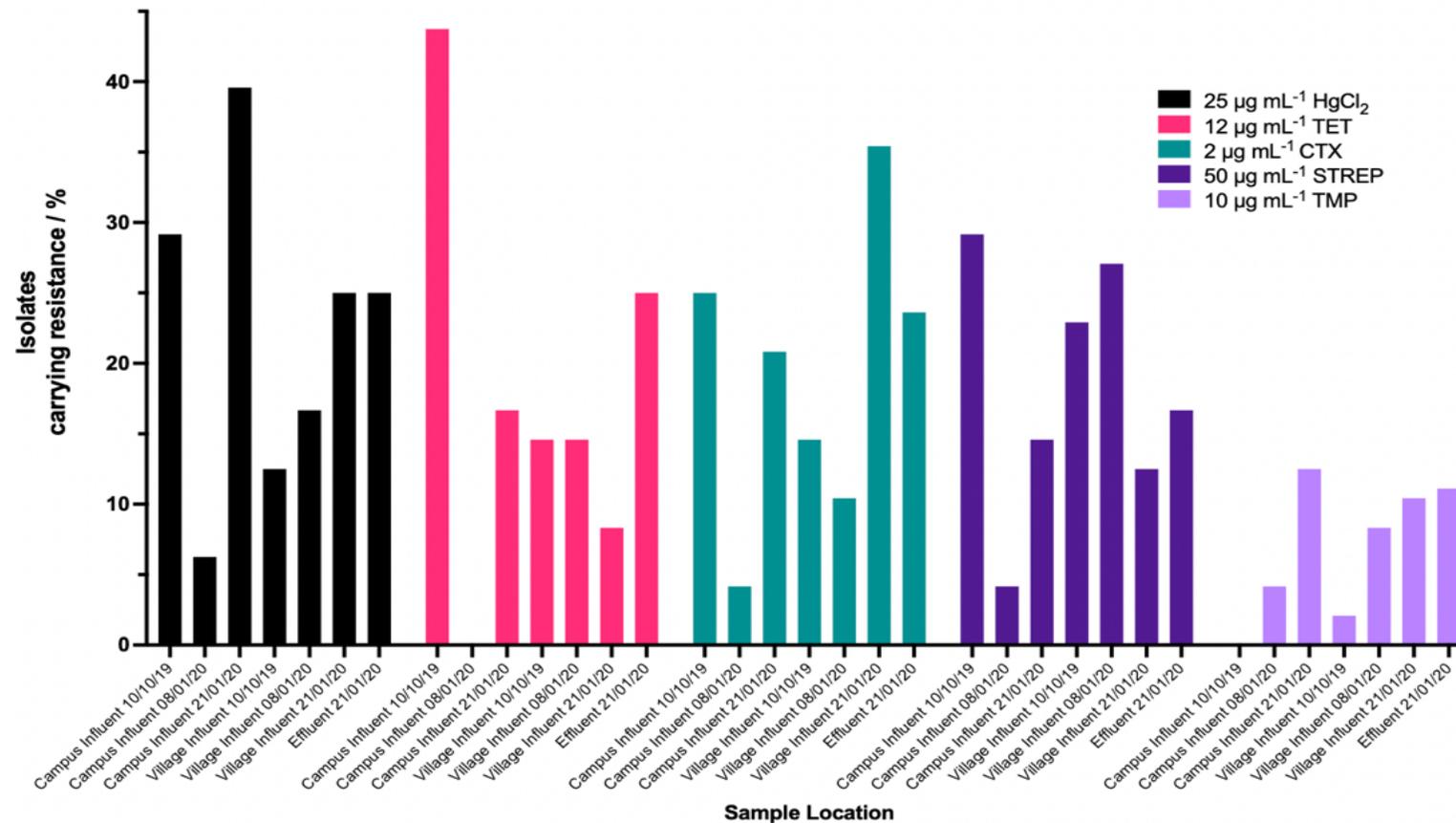
Figure 3.17 Replica plating results for each wastewater sample.

n=48 biological repeats per sample. Frequency of resistance to each antimicrobial was broken down per sample and displayed in different colours for visual comparison. For each antimicrobial, between village sample dates there was little variation in the percentage of isolates resistant. Between the campus influents, there was increased occurrence of resistance to each antimicrobial when the university term was in session.

In Figure 3.17, except for campus influent 08/01/2020, presence of resistance to each antimicrobial agent identified were consistent across samples. Influent samples were pooled and split to sample location, campus and village influent. A drop in overall resistance frequency in the population for this sample was expected as a large proportion of the human population at the campus at this time had left for Christmas break. This was confirmed first using a Chi-squared test (χ^2 = 103.7 df = 45, p < 0.0001) showing there was a difference between the datasets. Chi-squared testing was used in this instance and in the following comparisons as the data collected was binomial. For each isolate screened, values for resistance (1) and sensitivity (0) were assigned for each antibiotic. Resistance was defined as growth on correspondingly supplemented LB agar media.

Firstly, comparison of the populations and occurrence of each resistance was made. This dataset involved multiple Chi-squared tests, the p values generated were followed up using Benjamini and Hochberg correction to prevent false discoveries (Table 3.6). The significant difference in the data was shown to be generated by the sample '08/01/2020 Campus Influent', supporting the hypothesis that change in population would influence the resistance to antimicrobials observed in wastewater.

Datasets compared	Statistical Significance	P Value
08/01/2020 Campus Influent vs. 10/10/2019 Campus Influent	***	0.0008
08/01/2020 Campus Influent vs. 21/01/2020 Campus Influent	***	0.0008
08/01/2020 Campus Influent vs. 08/01/2020 Village Influent	*	0.0288
08/01/2020 Campus Influent vs. 10/10/2019 Village Influent	*	0.0452
08/01/2020 Campus Influent vs. 21/01/2020 Village Influent	**	0.0035
10/10/2019 Campus Influent vs. 21/01/2020 Campus Influent	*	0.0452
10/10/2019 Campus Influent vs. 08/01/2020 Village Influent	*	0.0452
10/10/2019 Campus Influent vs. 10/10/2019 Village Influent	*	0.0797
10/10/2019 Campus Influent vs. 21/01/2020 Village Influent	**	0.0068
21/01/2020 Campus Influent vs. 08/01/2020 Village Influent	ns	0.3174
21/01/2020 Campus Influent vs. 10/10/2019 Village Influent	ns	0.1445
21/01/2020 Campus Influent vs. 21/01/2020 Village Influent	ns	0.7300
08/01/2020 Village Influent vs. 10/10/2019 Village Influent	ns	0.9704
08/01/2020 Village Influent vs. 21/01/2020 Village Influent	ns	0.1730
10/10/2019 Village Influent vs. 21/01/2020 Village Influent	ns	0.1730

Table 3.6 Chi-square test with Benjamini and Hochberg correction of each sample location and date.

The post hoc test was performed to identify which sample populations carrying resistance profiles to antimicrobials may be different to the others. ns denotes no statistical significance, * p ≤ 0.05, *** p ≤ 0.001, **** ≤ 0001.

Further to this, it was possible to screen for overall resistance frequency across sample locations and dates (effluent was excluded from this as there was only one sample date). A Chi-square test was used for this, returning no statistical significance over the two sample locations (χ^2 = 7.080 df = 9, p = 0.6288) (Table 3.7).

Antimicrobial Agent	Statistical Significance	P value
HgCl$_2$	ns	0.3793
Tetracycline	ns	0.3793
Cefotaxime	ns	0.5589
Streptomycin	ns	0.4785
Trimethoprim	ns	0.6264

Table 3.7 Chi-square test with Benjamini and Hochberg correction comparing resistance frequency of each antimicrobial between campus and village influents across all sample dates combined.

No statistical significance was identified in frequency for each antimicrobial across influent samples.

Similarly, a Chi-square test was used to determine statistical significance in overall resistance frequency across the dates for each of the campus influent samples (χ^2 = 71.08, df = 18, p < 0.0001). Benjamini and Hochberg false discovery rate test was used to account for multiple Chi-squared tests (Table 3.8).

Campus Influent Sample Dates	Statistical Significance	P Value
10/10/2019 vs. 08/01/2020	***	0.0002
10/10/2019 vs. 21/01/2020	*	0.0241
08/01/2020 vs. 21/01/2020	***	0.0002

Table 3.8 Chi-square test with Benjamini and Hochberg correction of overall resistance frequency of Campus wastewater influents collected at different sample dates.

The post hoc test was performed to identify which sample date caused the significant difference in frequency of the resistances to antimicrobials screened. ns denotes no statistical significance, * p ≤ 0.05, *** p ≤ 0.001.

The corresponding multiple comparisons for each antimicrobial across sample dates of each was also performed using Chi-square tests with Benjamini and Hochberg false discovery corrections (Table 3.9).

Campus Influent Sample Date and Antimicrobial Compared	Statistical Significance	P Value
08/01/2020 25 µg mL^{-1} HgCl$_2$ vs. 10/10/2019 25 µg mL^{-1} HgCl$_2$	**	0.0084
08/01/2020 25 µg mL^{-1} HgCl$_2$ vs. 21/01/2020 25 µg mL^{-1} HgCl$_2$	ns	0.3028
10/10/2019 25 µg mL^{-1} HgCl$_2$ vs. 21/01/2020 25 µg mL^{-1} HgCl$_2$	***	0.0008
08/01/2020 12 µg mL^{-1} TET vs. 10/10/2019 12 µg mL^{-1} TET	***	0.0008
08/01/2020 12 µg mL^{-1} TET vs. 21/01/2020 12 µg mL^{-1} TET	**	0.0084
10/10/2019 12 µg mL^{-1} TET vs. 21/01/2020 12 µg mL^{-1} TET	**	0.0084
08/01/2020 5 µg mL^{-1} CTX vs. 10/10/2019 5 µg mL^{-1} CTX	**	0.0084
08/01/2020 5 µg mL^{-1} CTX vs. 21/01/2020 5 µg mL^{-1} CTX	*	0.0227
10/10/2019 5 µg mL^{-1} CTX vs. 21/01/2020 5 µg mL^{-1} CTX	ns	0.6272
08/01/2020 50 µg mL^{-1} STREP vs. 10/10/2019 50 µg mL^{-1} STREP	**	0.0050
08/01/2020 50 µg mL^{-1} STREP vs. 21/01/2020 50 µg mL^{-1} STREP	ns	0.1145
10/10/2019 50 µg mL^{-1} STREP vs. 21/01/2020 50 µg mL^{-1} STREP	ns	0.1145
10/10/2019 10 µg mL^{-1} TMP vs. 21/01/2020 10 µg mL^{-1} TMP	ns	0.1745
08/01/2020 10 µg mL^{-1} TMP vs. 10/10/2019 10 µg mL^{-1} TMP	ns	0.1745
08/01/2020 10 µg mL^{-1} TMP vs. 21/01/2020 10 µg mL^{-1} TMP	*	0.0214

Table 3.9 Chi-square test with Benjamini and Hochberg correction comparing resistance frequency of each antimicrobial across three sample time points for campus influent.

The post hoc test was performed to identify which sample comparisons showed statistical significance. ns denotes no statistical significance, * p ≤ 0.05, ** p ≤0.01, *** p ≤ 0.001.

The same tests were then performed for village influent samples. A Chi-square test was used to determine statistical significance in overall resistance frequency across the dates for each of the village influent samples (χ^2 = 20.56, df = 18, p = 0.3023). There was also no statistical significance when comparing each antimicrobial individually across the three sample dates for village influent.

Finally, for the second term time sample date, 21/01/2020, the effluent was compared to the influent samples using a Chi-square test, no statistical significance was identified across any of the sample dates or antimicrobials tested (χ^2 = 16.37, df = 18, p = 0.5665).

The data generated from the replica plating was also reformatted and multiple resistances were noted and displayed in graphical format (Figure 3.18). Most notably

there was consistent linkage across all sample locations of isolates possessing tetracycline and streptomycin resistance. Similarly, resistance to mercury in combination with either of the other four tested antibiotics was identified in each sample population, as was cefotaxime and streptomycin resistance. Interestingly, the effluent sample displayed joint resistances to antimicrobials which were exclusive to either the campus or village influents, demonstrating successful survival of resistances from both campus and village influents to the effluent.
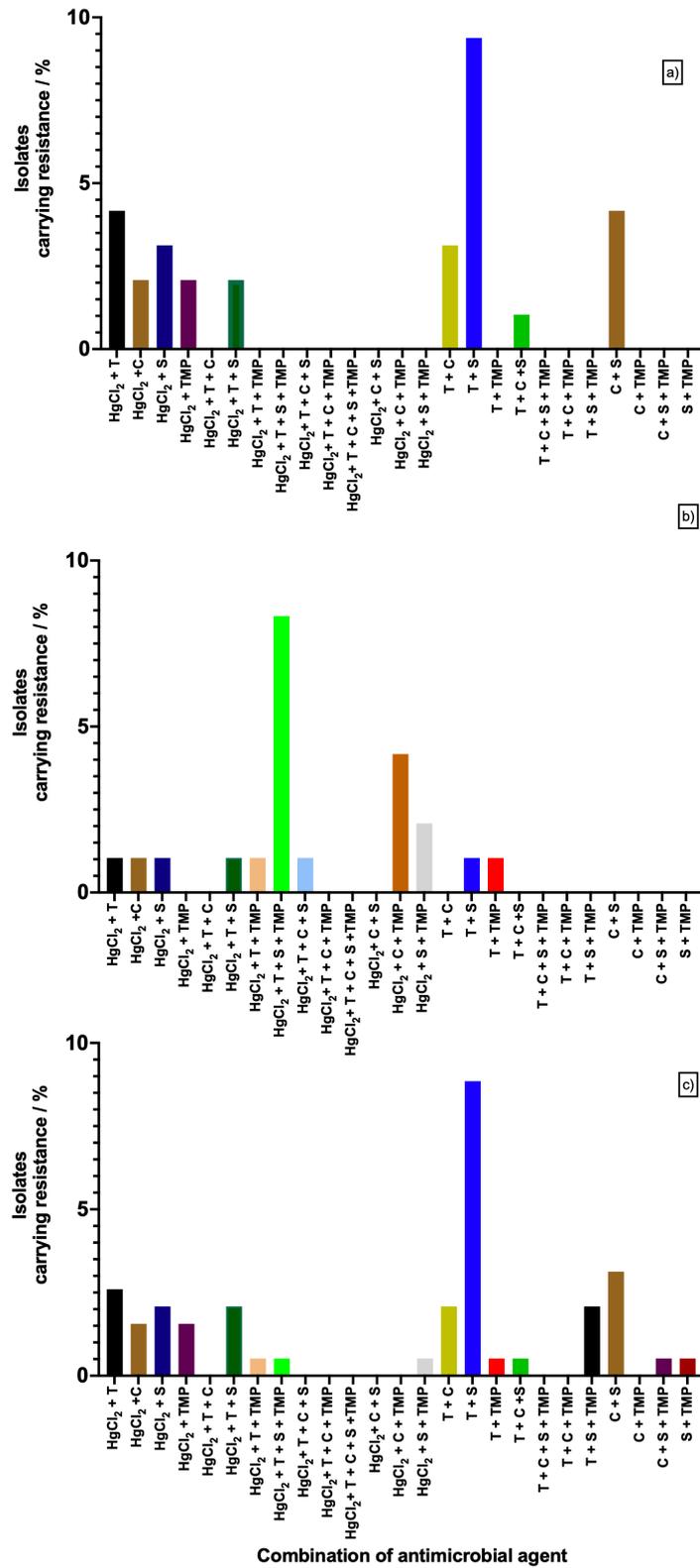
Figure 3.18 Breakdown of resistance to various antimicrobial combinations in campus influent (a), village influent (b) and effluent (c).

HgCl$_2$ refers to mercury (II) chloride, T to tetracycline, C to cefotaxime, S to streptomycin and TMP to trimethoprim.

Further to this, the frequencies of occurrence of individual resistance to the antimicrobials were tested. This was then used to calculate the baseline probability of co-occurrence. This was then compared with the actual frequency to identify if there may be any linkages that may suggest a selective factor. Comparisons were then made across time points and sample locations to identify if there were reservoirs for resistance linkages.

Firstly, the data was analysed as a whole, before breaking down further. Across all influent and effluent samples from all time points, many linkages were identified (Table 3.10). $HgCl_2$ resistance was linked to tetracycline resistance, which also when compounded further with three or more antimicrobials showed linkage to everything except '$HgCl_2$ + TET + CTX + TMP' and '$HgCl_2$ + TET+ CTX + STREP + TMP.' Trimethoprim resistance was also linked to $HgCl_2$ resistance. Similarly, tetracycline resistance demonstrated linkage to the other antimicrobials screened for. The most significant linkages identified were $HgCl_2$ + TET + STREP + TMP, TET + STREP + TMP, $HgCl_2$ + STREP + TMP and TET + CTX + STREP + TMP. Each highly linked combination contains tetracycline and streptomycin.

Combination of Antimicrobials	Occurrence (Actual)/%	Occurrence (Predicted)/%	Fold Difference
$HgCl_2$ + TET	4.82	3.96	**1.22**
$HgCl_2$ + CTX	2.71	4.03	0.67
$HgCl_2$ + STREP	3.61	4.30	0.84
$HgCl_2$ + TMP	2.71	1.54	**1.76**
$HgCl_2$ + TET + CTX	1.51	0.72	**2.10**
$HgCl_2$ + TET + STREP	2.71	0.76	**3.55**
$HgCl_2$ + TET + TMP	0.9	0.27	**3.29**
$HgCl_2$ + TET + STREP + TMP	0.9	0.05	**17.08**
$HgCl_2$ + TET + CTX + STREP	0.6	0.14	**4.37**
$HgCl_2$ + TET + CTX + TMP	0	0.05	0.00
$HgCl_2$ + TET+ CTX + STREP + TMP	0	0.01	0.00
$HgCl_2$ + CTX + STREP	0.6	0.28	**2.16**
$HgCl_2$ + CTX + TMP	0.3	0.28	**1.08**
$HgCl_2$ + STREP + TMP	2.11	0.30	**7.08**
TET+ CTX	3.92	3.21	**1.22**
TET + STREP	8.13	3.43	**2.37**
TET + TMP	2.11	1.23	**1.71**
TET + CTX + STREP	1.2	0.62	**1.95**
TET + CTX + STREP + TMP	0.3	0.04	**7.02**
TET + CTX + TMP	0.3	0.22	**1.35**
TET + STREP + TMP	2.71	0.24	**11.42**
CTX + STREP	3.92	3.48	**1.12**
CTX + TMP	1.81	1.25	**1.44**
CTX + STREP + TMP	0.9	0.24	**3.74**
STREP + TMP	3.01	1.34	**2.26**

Table 3.10 Linkage analysis of total occurrence of resistance to multiple combinations of antimicrobials compared to the predicted occurrence.

Fold difference was calculated by percentage occurrence (actual) divided by percentage occurrence (predicted). Fold difference greater than 1 was considered a linkage and marked in bold (n= 332). This suggests that the occurrence of each of these combinations was not down to chance.

Data for the different influent sites, total influent and effluent were also created (Table 3.11, Table 3.12, Table 3.13, Table 3.14 respectively). The linkage data identified that there were different combinations of antimicrobial resistances linked to different influent sites. The data showed that due to the water treatment process, effluent resistance diversity was reduced but not removed. More complex linkages

containing 3 or more resistances were reduced compared to both influents and the influent as a whole. Similarly, linkage between mercury resistance and any other resistance was also reduced or removed. There was also no linkage in any of the sample sets when comparing sample location to CTX resistance and mercury resistance.

Combination of Antimicrobials	Occurrence (Actual)/%	Occurrence (Predicted)/%	Fold Difference
$HgCl_2$ + TET	6.25	5.04	**1.24**
$HgCl_2$ + CTX	2.08	4.17	0.50
$HgCl_2$ + STREP	4.86	3.99	**1.22**
$HgCl_2$ + TMP	3.47	1.39	**2.50**
$HgCl_2$ + TET + CTX	1.39	0.84	**1.66**
$HgCl_2$ + TET + STREP	2.08	0.80	**2.59**
$HgCl_2$ + TET + TMP	0.69	0.28	**2.48**
$HgCl_2$ + TET + STREP + TMP	0.69	0.05	**15.54**
$HgCl_2$ + TET + CTX + STREP	0.00	0.13	0.00
$HgCl_2$ + TET + CTX + TMP	0.00	0.05	0.00
$HgCl_2$ + TET+ CTX + STREP + TMP	0.00	0.01	0.00
$HgCl_2$ + CTX + STREP	0.00	0.23	0.00
$HgCl_2$ + CTX + TMP	0.00	0.23	0.00
$HgCl_2$ + STREP + TMP	3.47	0.22	**15.65**
TET+ CTX	4.86	3.36	**1.45**
TET + STREP	6.94	3.22	**2.16**
TET + TMP	1.39	1.12	**1.24**
TET + CTX + STREP	0.69	0.54	**1.30**
TET + CTX + STREP + TMP	0.69	0.03	**23.32**
TET + CTX + TMP	0.69	0.19	**3.72**
TET + STREP + TMP	1.39	0.18	**7.77**
CTX + STREP	3.47	2.66	**1.30**
CTX + TMP	1.39	0.93	**1.50**
CTX + STREP + TMP	0.69	0.15	**4.70**
STREP + TMP	3.47	0.89	**3.91**

Table 3.11 Linkage analysis of campus influent occurrence of resistance to multiple combinations of antimicrobials compared to the predicted occurrence.

Fold difference was calculated by percentage occurrence (actual) divided by percentage occurrence (predicted). Fold difference greater than 1 was considered a linkage and marked in bold (n=144). This suggests that the occurrence of each of these combinations was not down to chance.

Combination of Antimicrobials	Occurrence (Actual)/%	Occurrence (Predicted)/%	Fold Difference
HgCl$_2$ + TET	6.25	5.04	**1.24**
HgCl$_2$ + CTX	2.08	4.17	0.50
HgCl$_2$ + STREP	4.86	3.99	**1.22**
HgCl$_2$ + TMP	3.47	1.39	**2.50**
HgCl$_2$ + TET + CTX	3.47	2.26	**1.54**
HgCl$_2$ + TET + STREP	3.47	3.64	0.96
HgCl$_2$ + TET + TMP	2.78	3.76	0.74
HgCl$_2$ + TET + STREP + TMP	2.08	1.25	**1.66**
HgCl$_2$ + TET + CTX + STREP	1.39	0.46	**3.06**
HgCl$_2$ + TET + CTX + TMP	3.47	0.47	**7.39**
HgCl$_2$ + TET+ CTX + STREP + TMP	0.69	0.16	**4.43**
HgCl$_2$ + CTX + STREP	0.69	0.03	**21.27**
HgCl$_2$ + CTX + TMP	1.39	0.10	**14.67**
HgCl$_2$ + STREP + TMP	0.00	0.03	0.00
TET+ CTX	0.00	0.01	0.00
TET + STREP	1.39	0.25	**5.50**
TET + TMP	0.69	0.25	**2.75**
TET + CTX + STREP	0.69	0.26	**2.66**
TET + CTX + STREP + TMP	2.08	2.52	0.83
TET + CTX + TMP	7.64	2.60	**2.93**
TET + STREP + TMP	0.69	0.87	0.80
CTX + STREP	1.39	0.52	**2.65**
CTX + TMP	0.00	0.04	0.00
CTX + STREP + TMP	0.00	0.18	0.00
STREP + TMP	2.78	0.18	**15.36**

Table 3.12 Linkage analysis of Village influent occurrence of resistance to multiple combinations of antimicrobials compared to the predicted occurrence.

Fold difference was calculated by percentage occurrence (actual) divided by percentage occurrence (predicted). Fold difference greater than 1 was considered a linkage and marked in bold (n=144). This suggests that the occurrence of each of these combinations was not down to chance.

Combination of Antimicrobials	Occurrence (Actual)/%	Occurrence (Predicted)/%	Fold Difference
HgCl$_2$ + TET	4.86	3.51	**1.38**
HgCl$_2$ + CTX	2.78	3.96	0.70
HgCl$_2$ + STREP	3.82	3.96	0.96
HgCl$_2$ + TMP	2.78	1.35	**2.07**
HgCl$_2$ + TET + CTX	1.39	0.65	**2.15**
HgCl$_2$ + TET + STREP	2.78	0.65	**4.30**
HgCl$_2$ + TET + TMP	0.69	0.22	**3.16**
HgCl$_2$ + TET + STREP + TMP	0.69	0.04	**17.19**
HgCl$_2$ + TET + CTX + STREP	0.69	0.12	**5.84**
HgCl$_2$ + TET + CTX + TMP	0.00	0.04	0.00
HgCl$_2$ + TET+ CTX + STREP + TMP	0.00	0.01	0.00
HgCl$_2$ + CTX + STREP	0.69	0.25	**2.81**
HgCl$_2$ + CTX + TMP	0.35	0.25	**1.40**
HgCl$_2$ + STREP + TMP	2.08	0.25	**8.41**
TET+ CTX	3.47	3.00	**1.16**
TET + STREP	7.29	3.00	**2.43**
TET + TMP	1.04	1.02	**1.02**
TET + CTX + STREP	1.04	0.55	**1.89**
TET + CTX + STREP + TMP	0.35	0.04	**10.05**
TET + CTX + TMP	0.35	0.19	**1.85**
TET + STREP + TMP	2.08	0.19	**11.10**
CTX + STREP	3.82	3.39	**1.13**
CTX + TMP	1.39	1.15	**1.21**
CTX + STREP + TMP	1.04	0.21	**4.92**
STREP + TMP	2.43	1.04	**2.35**

Table 3.13 Linkage analysis of total influent occurrence of resistance to multiple combinations of antimicrobials compared to the predicted occurrence.

Fold difference was calculated by percentage occurrence (actual) divided by percentage occurrence (predicted). Fold difference greater than 1 was considered a linkage and marked in bold (n=288). This suggests that the occurrence of each of these combinations was not down to chance.

Combination of Antimicrobials	Occurrence (Actual)/%	Occurrence (Predicted)/%	Fold Difference
$HgCl_2$ + TET	4.17	6.25	0.67
$HgCl_2$ + CTX	2.08	3.65	0.57
$HgCl_2$ + STREP	2.08	5.73	0.36
$HgCl_2$ + TMP	2.08	2.60	0.80
$HgCl_2$ + TET + CTX	2.08	0.91	**2.29**
$HgCl_2$ + TET + STREP	2.08	1.43	**1.46**
$HgCl_2$ + TET + TMP	2.08	0.65	**3.20**
$HgCl_2$ + TET + STREP + TMP	2.08	0.15	**13.96**
$HgCl_2$ + TET + CTX + STREP	0	0.21	0.00
$HgCl_2$ + TET + CTX + TMP	0	0.10	0.00
$HgCl_2$ + TET+ CTX + STREP + TMP	0	0.02	0.00
$HgCl_2$ + CTX + STREP	0	0.38	0.00
$HgCl_2$ + CTX + TMP	0	0.38	0.00
$HgCl_2$ + STREP + TMP	2.08	0.60	**3.49**
TET+ CTX	6.25	3.65	**1.71**
TET + STREP	12.5	5.73	**2.18**
TET + TMP	8.33	2.60	**3.20**
TET + CTX + STREP	2.08	0.84	**2.49**
TET + CTX + STREP + TMP	0	0.09	0.00
TET + CTX + TMP	0	0.38	0.00
TET + STREP + TMP	6.25	0.60	**10.47**
CTX + STREP	4.17	3.34	**1.25**
CTX + TMP	4.17	1.52	**2.74**
CTX + STREP + TMP	0	0.35	0.00
STREP + TMP	6.25	2.39	**2.62**

Table 3.14 Linkage analysis of effluent occurrence of resistance to multiple combinations of antimicrobials compared to the predicted occurrence.

Fold difference was calculated by percentage occurrence (actual) divided by percentage occurrence (predicted). Fold difference greater than 1 was considered a linkage and marked in bold (n=48). This suggests that the occurrence of each of these combinations was not down to chance.
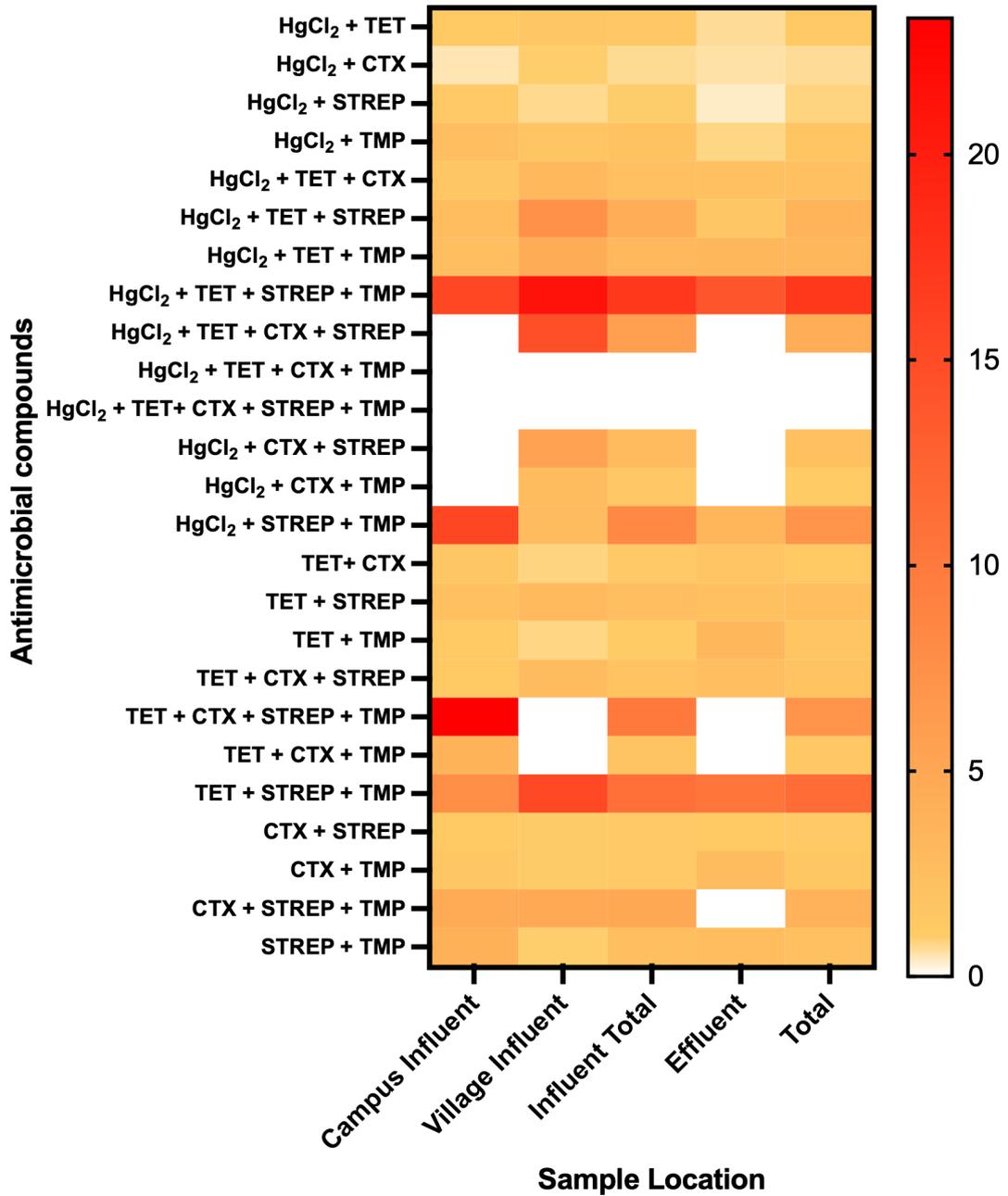
Figure 3.19 Heatmap comparing different sample location fold-difference in resistances.

Fold difference displayed for each combination of resistance. Fold-difference scale is displayed on the right, where less than 1 displayed as white, meaning no linkage between antimicrobials in that sample group. The colours from orange to red showed by how much greater than the predicted occurrence resistance to each antimicrobial combination.

When comparing sample locations side by side, it is possible to identify a reduction

in diversity of multiple resistances to the screened antimicrobials through the

166

wastewater treatment process and reduced occurrence of resistance to some combinations of antimicrobials (Figure 3.19).
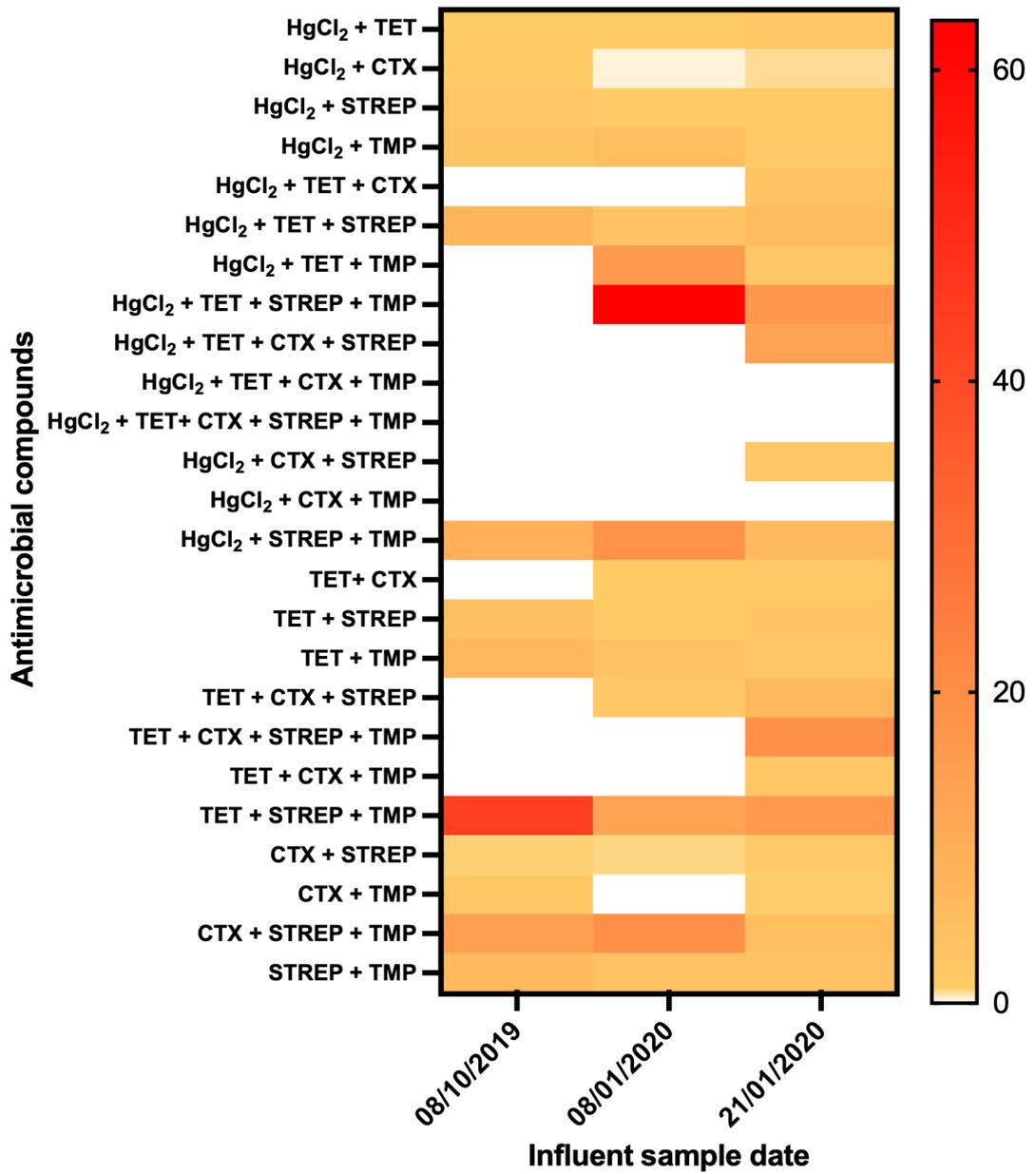
Figure 3.20 Heatmap comparing fold difference of temporal samples from wastewater influent.

Fold difference displayed for each combination of resistance. Fold-difference scale is displayed on the right, where less than 1 displayed as white, meaning no linkage between antimicrobials in that sample group.

167

Across the three time points there is an increase in occurrence of resistance combinations. Influent sample 21/01/2020 possesses the most combinations of resistance greater than the predicted chance of occurrence. Only three combinations of antimicrobial resistance showed no potential linkage out of a possible 25 combinations within that sample. The non-term time influent sample (08/01/2020) possessed a linkage between the four antimicrobials, $HgCl_2$, tetracycline, streptomycin and trimethoprim, 63-fold higher than the predicted chance of occurrence (Figure 3.20). Notably, the term time influent sample 08/10/2019 possessed the least possible linkages of carriage between the three total influent samples screened. However, 85% of all presumptive linkages identified in this sample occurred two-fold more than predicted. The diversity of linkages was reduced but presumptive linkages were more occurrent. The means of linkage for each influent sample were calculated and outliers removed. The result showed a significant decrease in the occurrence of resistance linkage between term time and non-term time influent samples.

3.3   Discussion and Conclusion

3.3.1   qPCR Tn*21* and *rrsA* Validation

In order to determine relative abundance of Tn*21* in the wastewater environment, a house keeping gene, present in all bacterial cells was required to compare abundance of the transposable element against. For that reason, *rrsA*, which encodes the 16S ribosomal RNA, was chosen. However, whilst *rrsA* is one of the most common reference genes used for qPCR techniques, some bacterial species are known to carry more than one copy of the *rrsA* gene for redundancy purposes (Gregory and Dahlberg, 2009). With this caveat in mind, in order to compare relative abundance between samples, *rrsA* could be used as a reference in each sample, and has been widely used (Zhou *et al.*, 2011; Peng *et al.*, 2014).

The Tn*21* primer and probe set were designed to target a conserved region of the Tn*21*. The 38 bp inverted repeat sequence of the Tn*21* transposase gene into the transposase gene itself (Figure 3.21).

Figure 3.21 Diagram showing the target sequences of the qPCR primers to detect Tn*21*-like mobile elements.

Tn*21* IR primer targets the 38 bp IR region of Tn*21* and *tnpA* R targets a section towards the 3' terminal of *tnpA.*

This study was performed using hydrolysis probes (TaqMan). This type of PCR technique allows hydrolysis probes to bind amplified target DNA between a primer set. The probes are made up of 3 components. At the 5' end is a fluorophore, the binding DNA probe which anneals to the target sequence and finally a quencher at the 3' terminal. Upon successful amplification of a target strand of DNA, the Taq polymerase in the reaction mixture uses endonuclease activity to cleave the fluorophore from the probe. This in turn causes fluorescence. The fluorophore does not fluoresce before being cleaved due to being in close proximity to the quencher at the 3' terminal of the probe, which absorbs the emission of fluorescence of the fluorophore. The qPCR thermocycler measures the fluorescence at the end of each thermocycle, enabling quantification of PCR product. Due to the nature of PCR, the product will exponentially increase each cycle. The fluorescence intensity therefore has a direct relationship with the concentration of PCR product. From this a C_q or C_t value can be calculated. This is the number of cycles taken for the fluorescence intensity to exceed an arbitrarily assigned threshold above the background fluorescence.

3.3.2   Tn*21* Carriage in Wastewater

Screening wastewater for Tn*21* presence across the treatment process proved to be a valuable tool to approximate expected presence of Tn*21* and Tn*21-like* transposable elements. The expected yield of mercury resistant and presumptive Tn*21* carrying isolates was lower than the percentage isolated from the replica

plating experiment. This is likely due to TBX growth medium selecting for *Enterobacteriaceae* rather than the total microbial community.

Wastewater treatment does have an impact on abundance of Tn*21*, reducing by two-fold across the time points measured with qPCR (Figure 3.4). A recent study performed in a municipal wastewater treatment plant showed there was between 2.39–3.35 log reduction in copy number from influent to effluent of *merR* and *merD* genes (Yuan *et al.*, 2019).

In the plate counts from the same samples, a ten-fold reduction in colony forming units was observed across the treatment process. This demonstrates that the wastewater treatment process is removing microorganisms despite the lack of legislation to remove the microorganisms. However, not enough to likely reduce the environmental impact of dissemination of resistance genes downstream of the treatment plant. The data suggests that a selection event occurs over the wastewater treatment process as the microbial load decreased and isolation of Tn*21* also reduced. However, there are still 10^4 cfu mL^{-1} being deposited into the environment, which will impact the microbiome downstream and also have wider implications to whole ecosystems. The transposable element Tn*21* is commonly carried on larger self-transmissible plasmids. Stress from the treatment process may result in loss of such plasmids or induce transposition of the Tn*21* leading to this reduction of presence identified.

3.3.2.1 Temporal effects on Tn*21* carriage

The wastewater treatment plant in this study was subjected to varied demand depending on the time of year. A nearby university campus resulted in mass population change between term time and non-term time. Students live and work on the campus and also live in the surrounding villages. The qPCR data showed that between term time and non-term time there was a significant decrease (p = 0.0001) in Tn*21* abundance suggesting that population change impacted the concentration Tn*21* copies within the environment (Figure 3.3 and Figure 3.5). It is possible that in larger wastewater treatment plants a population influx of this proportion, such as a university term time involving international students and students from within the same country, may have a similar impact on carriage of not just Tn*21* transposable elements, but other resistance genes and pathogen transmission into the wider environment. The Tn*21* abundance qPCR assay may be a valuable tool for identifying human polluted environments. The use of integrons has been previously suggested as such a tool, however Tn*21* and Tn*21-like* transposons containing integrons may offer a more refined approach (Gillings *et al.*, 2015). Class I integrons are more prevalent than Tn*21* in the environment so detection could require less sensitivity.

3.3.3 *merA* Phylogeny

Phylogenetic analysis of the *merA* gene allowed the identification of 4 clades (Figure 3.8). Given the specificity of the enzyme to reduce such a toxic ion from mercury (II) to reduced mercury vapour, it was somewhat surprising. Mercuric (II) reductase is a homodimeric protein; its active site is made up of four cysteine residues, two on each

monomer (C136 and C141). Mercury (II) is transferred from a pair of cysteine residues at the C-terminus of one monomer to another pair of cysteine residues in the catalytic site of the other monomer. Mercury (II) ion reduction is mediated by FAD (flavin adenine dinucleotide), which acts as a cofactor, and NADPH (dihydronicotinamide adenine dinucleotide phosphate), a proton carrier (Lian *et al.*, 2014).

Resistance to mercury (II) is thought to have existed for thousands of years, the earliest characterised and sequenced are from the Murray Collection, however it is thought that *merA* may have emerged from thermophilic bacteria in mercury rich hydrothermal environments (Barkay *et al.*, 2010). The collection of mainly *Enterobacteriaceae* from the pre-antibiotic era (1918-1954) from various geographical locations, contained three isolates containing resistance to $HgCl_2$ (Datta and Hughes, 1983).

The generation of four main clades of mercuric (II) reductase gene, *merA*, from the phylogenetic tree may suggest a need for reclassification of the *merA* genes population and possibly the *mer* operon as a whole (Figure 3.8). The plasmid R100 was the first characterised, known to carry mercury resistance at the time. Within its clade is also the *merA* gene from the plasmid pMG101 and the enteroaggregative *E. coli* strain 042 (Chaudhuri *et al.,* 2010). These historic plasmids are widely considered reference points to classic Tn*21*. It can be seen from these data that they may have the most widely conserved *merA* gene. From the tree produced there are 9 isolates containing identical sequence of *merA* to that of plasmid R100, pMG101 and *E. coli* 042. Other isolates in the clade have been isolated from wastewater influent, river

water effluent downstream of a wastewater treatment plant, sheep and a cow shed of a farm local to the wastewater treatment plant from the study. These samples were all identified within a 1 km radius over the time span of four years. Humans in some way have impacted all of the environments in which these mercury resistant organisms were isolated. Further to this, plasmid R100 was isolated from *Shigella flexneri* in Japan in the 1950s. pMG101 was isolated in *Salmonella* Typhimurium from three burns patients in Boston USA in 1973 (McHugh *et al.,* 1975). Both of these isolates were also identified as a result of human impact.

From the amplified region of *merA* that was covered in Sanger sequencing, the alignment of bases 493 to 1,544 of the *merA* gene yielded a strong consensus sequence at >90% similarity (Figure 3.9). The consensus sequence was clearly split by highly conserved regions and variable bases of three to four bases in most cases, with the exception of one region, split by seven variable bases. Due to the target sequence area, it is not possible to determine differences in codon usage across the active site's cysteine residues, as these were located c. 100 bases upstream of the aligned site. This was not considered until after the work in this study was performed. To gain a better understanding into the variability of this region, short-read Illumina technology sequencing of *merRTPCADE* amplicons should be considered to further understand variation and possibly identify evolutionary changes in the *mer* operon.

3.3.4   Isolation of Mercury Resistant *Enterobacteriaceae*

The frequency of mercury resistant *Enterobacteriaceae* occurrence was 16.4%. When examining the replica plating data however, the frequency of mercury resistance was

higher (Table 3.5). 22% of the total of isolates possessed resistance to $HgCl_2$. From the samples collected on 20/01/2020, there was a reduction of mercury resistance frequency noted from influent to effluent; from 32% in influent to 25%. It may be possible that mercury resistant bacteria may be indirectly removed throughout the wastewater treatment process. However, the frequency at which these resistant isolates occurred in this sample set is rather alarming. The population increase of the campus may have impacted the biodiversity of the wastewater bacteria, causing an increase in the presence of mercury resistant *Enterobacteriaceae* observed.

Previous reports from water samples indicate approximately 10-12% of enterobacterial isolates were mercury resistant, and another study using copy number estimated that in Chinese municipal wastewater plants approximately 10^{-3} copies of mer genes occurred per 16S rRNA gene (Maiti and Bhattacharyya, 2013; Yuan *et al.*, 2019). The higher frequency of mercury resistance occurrence reported in this study may be due to selection bias of the TBX media, which was used in order to isolate *Enterobacteriaceae* and exclude potential O157:H7 serovars and other glucuronidase negative *E. coli.* However, Tn*21* and Class I integrons are carried by *Gammaproteobacteria*, so isolation on this media is essential in order to guarantee isolation of the correct phyla (Gillings, 2014).

Over the total time period of sample collection, multiple mercury resistant isolates displayed the same phenotypic resistance patterns and were isolated from different locations. This suggests the possibility of the presence of dominant clones and perhaps the presence of dominant, large self-transmissible plasmids containing mercury resistance and other antibiotic resistance genes. The 2001C27 and 2001E17

strains were all isolated in different locations, campus influent and effluent respectively, on the same date. Each isolate was non-susceptible to the same eight antibiotics (AMP, AMC, S10, OT, CIP, NA, SXT and AZM). This pair of isolates was among the three isolates that contained the most non-susceptibilities to any antibiotics out of those screened. Given the similarities and the presence of the two isolates pre and post wastewater treatment process, it may support the theory. Upon closer inspection, this was more common than thought. Three more isolates, isolated on the same date were also seen to have similar characteristics. CH14, *Enterobacter spp.* isolated from campus influent, VH28, *Enterobacter spp.* and VH32 *Citrobacter freundii* isolated from village influent on the same date. These mercury resistant isolates all possessed resistance to AMC and FOX. Interestingly, these isolates were from two different influent sources, but the presence of these resistances was in two different genera of bacteria. Figure 3.9 shows the *merA* gene of CH14 and VH28 to be very closely related and to exist within the same clade. This further adds to the theory that there may be dominant self-transmissible plasmids allowing transfer of these resistance genes within the environment. However, due to the nature of the isolation of these two strains, from separate influents, it may suggest that there is a dominant human commensal that may carry this DNA.

The mercury resistant isolates that were screened for antibiotic susceptibility using AST under CLSI standards (2.2.17), suggested a link between aminoglycoside presence and mercury resistance. One quarter of all isolates were non-susceptible to aminoglycosides. The model Tn*21* mobile element possesses *aadA1*, conferring resistance to aminoglycosides. The data may suggest presence of the *aadA* gene

176

cassette is present within an integron. It is possible that the presence of aminoglycoside resistance is a fixed gene, or highly relied upon gene in the cassette array of the class I integron.

The well-characterised large transmissible plasmid pMG101, containing Tn*21* shows a Tn*3* family transposable element containing resistance genes to tetracycline (Hooton *et al*. 2021). 15% of isolates found were non-susceptible to tetracycline, despite its current lack of use in human medicine. This may indicate co-occurrence of tetracycline resistance as a result of Tn*21,* likely by carriage on a larger element such as a self-transmissible plasmid. Alternatively, it has a long residence time in the environment, so may maintain minimal selective concentrations (Lundström *et al.*, 2016). What is not known however, is whether the selective driver for the occurrence of Tn*21* is due to mercury resistance, or due to the immobile, but still active class I integron.

3.3.5   Co-Occurrence of ESBL and Mercury Resistance

Previously, mercury resistant isolates containing ESBL genes such as *bla*$_{OXA}$, *bla*$_{CTX-M}$ and *bla*$_{NDM}$ have been reported (Dong *et al.*, 2019; Siddiqui *et al.*, 2020). The ESBL and carbapenemase genes were identified within the class I integron in the transposable element. The previous reports suggested due to the constantly adapting contents of the class I integron that there may be a linkage of co-occurrence identified. From the study targeted on isolating ESBL *Enterobacteriaceae* and then screening for resistance to HgCl$_2$, there was no linkage between the occurrences of the two resistance mechanisms (section 3.2.4, Table 3.7, Table 3.8, Table 3.9, Table

3.10 and Table 3.11). Unfortunately, screening for resistance to the two antimicrobials in the same media was not viable due to the toxicity of each compound to the cells. Similarly, there was also no statistical difference between the carriage of resistances between the two influents from this sample.

AST of the ESBL producing isolates showed 18.5% possessed resistance to quinolones, fluoroquinolones and tetracycline in addition. This combination of resistances is the same as those found in the mobile element IS*Ecp1*. The transposable element has been mobilised by third and fourth generation cephalosporins, which are not only used in veterinary medicine but also human medicine too (Litten *et al.*, Unpublished).

Only four of possible hundreds of known ESBL genes were screened for by PCR. Three isolates of those screened did not possess one of these resistance genes. This may suggest dominance of certain ESBL resistance genes within the wastewater environment. In Europe it is known that there are dominant variants of bla_{CTX-M}, these are $bla_{CTX-M14}$ and $bla_{CTXM-15}$ (Bevan, Jones and Hawkey, 2017).

Some isolates possessed multiple resistance genes to extended spectrum beta-lactams suggesting there may be little fitness cost to carriage of multiple ESBL genes. Secondly, different beta lactamase genes confer resistance to different beta lactams and have slightly different modes of action. For example, the bla_{TEM} gene, like bla_{SHV}, can cleave the lactam ring in penicillin and first generation cephalosporins but is not capable of degrading the oxyamino group of later generation cephalosporins (Shaikh *et al.*, 2015). bla_{CTX-M} acts differently to the bla_{TEM} and bla_{SHV} enzymes and is widespread across humans and animals. The enzyme preferentially hydrolyses

178

cefotaxime and is also active on other third and fourth generation cephalosporins. However, *bla*$_{CTX-M}$ is more readily inhibited by beta lactamase inhibitors such as clavulanic acid (Ma *et al.*, 1998). Therefore, carriage of a combination of both beta lactamase genes is beneficial to survival when subjected to a combination of beta lactamase inhibitors and beta-lactams. It is also worth noting that resistance to beta-lactams may arise from mutation of *ampC* too. This mutation would likely carry a lower fitness cost.

3.3.6 Co-Occurrence of Antimicrobials in Wastewater

First of note, the impact of population change influenced the frequency of resistance in campus influent. Outside of term time, the occurrence of resistance to any of the antimicrobials screened was greatly reduced compared to the other samples screened from term time. Outside of term time, the village influent maintained similar resistance occurrence frequency of the antimicrobials.

Due to limitations on sampling, effluent was only allowed to be sampled by the owners of the wastewater treatment site later on in the study. It is difficult to draw many conclusions as a result. From the one sample screened, during term time, the effluent reduced the prevalence of CTX resistance. Resistance to HgCl$_2$ displayed either no change between village and effluent or a 15% reduction in occurrence. Resistance to TMP was not noted to change from the influent. TET and STREP resistance frequency increased over the course of the wastewater treatment process, suggesting there may be a selection event occurring to either cause dissemination of these resistances or reduce diversity of other strains.

179

Screening each isolate for resistance to the five different antimicrobial compounds by replica plating generated a lot of data, which could be used to identify linkages between different resistances within the wastewater environment. The most notable was the identified potential linkage in almost every sample taken between tetracycline and a second antimicrobial resistance. This further builds on the premise that a selective event for tetracycline resistant bacteria may be occurring. Most surprisingly, the identification of potential linkages between different antimicrobial resistances suggests that a selection event may occur in the gut, before reaching the wastewater treatment plant. Overall, the linkage data suggested that the wastewater treatment process reduced the frequency of compound resistance identified overall. One of the key findings made when identifying potential relationships between occurrence of resistance to the antimicrobials tested was the linkage between $HgCl_2$ and trimethoprim. Trimethoprim competitively interacts with dihydrofolate reductase to prevent the reduction of dihydrofolate to tetrahydrofolate and is often used to treat urinary tract infections (Hitchings, 1973). Given that trimethoprim is used to treat urinary tract infection, it comes as no surprise to identify resistance to the antibiotic in wastewater bacteria. Selection for this resistance may occur in the gut, as the antibiotic is administered orally, or within the affected area. Resistance to trimethoprim is often either intrinsic, formed by a chromosomal mutation of the dihydrofolate reductase gene (*dfr*) or acquired on transmissible plasmids or in gene cassettes of integrons. The class I integron carried by Tn*21* is capable of capturing such gene cassettes as *dfr* and may aid the evolution of Tn*21*-like elements by keeping the mobile element relevant. Equally, without further examining the

integron contents of these bacteria it is hard to know whether the *dfr* genes may be

carried within the class I integron.

3.4 Future Work

3.4.1 Quantification of Mercury (II) in Wastewater

Due to the COVID-19 pandemic, sampling of wastewater was halted due to carriage of the virus in wastewater. Unfortunately, the wastewater taken as samples from this study was completely used for isolation, so it was not possible to screen these samples. The data from this analysis would give an insight as to whether there is likely a co-selection event allowing for more meaningful conclusions to be drawn.

Had it have been possible to collect more wastewater sample, three 2 L samples of water would have been collected weekly, across six weeks over non-term time and term time from each sample location, wastewater influent sites (campus and village), effluent and the brook, up and downstream of the wastewater treatment plant. From each biological sample from each location, three technical repeats would have also been used to quantify the mercury (II) present in each sample. detection would have been performed using a mercury analyser device such as Mercury T2C / T2P machine (Teledyne LeCroy, USA).

3.4.2 Comparative qPCR analysis

To build on the data produced in the qPCR assay, a comparison with the same samples should be made targeting *intI1*, *sul1*, *qacE*, *merA* and *merR* genes. It is likely this would show mercury resistance is more abundant than just occurrence in Tn*21* as Tn*21* is a complex transposable element. Similarly, comparison with *intI1*, *sul1* and

qacE to the abundance of Tn*21* would display similar results to *mer* genes. However, I would hypothesise a strong correlation between *qacEΔ1* abundance and Tn*21*.

3.4.3   Isolation of Mercury Resistant *Enterobacteriaceae*

Due to financial and workload constraints, it was not possible to sequence all the mercury resistant bacteria isolated, so it is not possible to know whether the links trends between non-susceptibility to other antibiotics are true. However, it is still possible to hypothesise which resistances are likely to be co-occurrent on the same mobile element given the literature that has already been published (Pal *et al.*, 2017; Perez-Palacios *et al.*, 2021). Sequencing all the strains collected in this investigation would provide great depth to the AST data and shed light on the presence of dominant sequence types and large self-transmissible plasmids.

3.4.4   *merA* Phylogeny

The phylogenetic tree from Figure 3.8 provided some insight into the variability of what is a highly conserved gene *merA*, the key enzyme in the detoxification of mercury (II). Due to imitations in Sanger sequencing, reliably calling the whole gene in one run would prove difficult to draw a consensus sequence for the whole gene. In future, 'base-walking' of this sequence, using multiple sets of overlapping primers and sequencing the amplicons with Sanger sequencing may allow for full gene alignment of each *merA* gene. Better still, short read sequencing and assembly using Illumina technology of each isolate's whole *mer* operon may provide greater detail

into to potential evolution of the mercury resistance mechanism as a whole, from a genetic standpoint.

3.4.5   ESBL *E. coli*

From the AST data, the resistance profiles of some isolates match that of IS*Ecp1* mobile element. These isolates also possessed *bla*$_{CTX-M}$ further building on this presumption. Sequencing of these isolates would give confirmation and also allow comparison to those IS*Ecp1* isolated on the nearby university dairy farm. These have been shown to be mobilised by low concentrations of cephalosporins, which may be found in wastewater too (Litten *et al.,* unpublished).

3.4.6   Examining Co-occurrence of Antimicrobial Resistance in Mercury Resistant Bacteria

Whilst the replica plating experiments yielded some interesting results, only a small number of antimicrobial classes were used for screening. To get a better understanding of the relationships of other antimicrobials and whether co-occurrence may be identified, screening for other antimicrobial metal resistances such as silver and copper resistance may be beneficial. It is known that silver and copper resistance are often co-occurrent due to the Tn*7/pco/sil* mobile genetic element, meaning they could be used as a reference marker in the future. Whilst cheap and high throughput, there are some drawbacks of this method. The results are generated through determining whether there is growth or not, meaning there is a chance of development of false positives but most importantly, it is therefore not

possible to identify the specific gene allele responsible for the resistance identified. Using targeted metagenomics may be helpful to identify resistance genes present within each wastewater sample, but it would not be possible to identify organism specific resistance profiles.

A link was made between the occurrence of HgCl$_2$ resistance and trimethoprim resistance. This relationship needs further exploring perhaps by examining metagenomic contents of class I integrons from the wastewater environment and sequencing of strains containing both resistances. There are many variants of the *dfr* gene, identifying possible dominant variants from this environment may help identify whether Tn*21* and Tn*21*-like mobile elements are diversifying in order to remain relevant within the environment.

Chapter 4: Whole Genomic Analysis of Bacterial Isolates Containing Tn*21* or Mercury Resistance from Wastewater and the Surrounding Environment.

4.1   Introduction

Since its discovery in the 1950s, Tn*21* has been widely reported globally, in Gram-negative environmental and clinical samples (Nakaya, Nakamura and Murata, 1960; McHugh *et al*., 1975; Zühlsdorf and Wiedemann, 1992; Márquez *et al*., 2008). One study reported Tn*21-tnpA* occurrence to be 1 copy per 10^3 and 10^4 bacteria in a marine environment (Dahlberg and Hermansson, 1995). In another study, approximately 20% of Gram-negative clinical isolates were shown by PCR to possess Tn*21* transposase genes, however many defective transposons have also been reported (Zühlsdorf and Wiedemann, 1992). Tn*21* and mercury resistance transposons have also been isolated from pristine soil environments, suggesting Tn*21* derivatives disseminated before the antibiotic era (Pearson *et al.*, 1996). Three of the isolates in the Murray collection, a collection of isolates from the pre-antibiotic era possessing Tn*21*-like ancestors and a *Pseudomonas* strain isolated from Russian permafrost, identified as a closely related Tn*21* ancestor without the In2 region insertion (Essa *et al.,*2003; Kholodii *et al.,* 2003; Baker *et al.,* 2015). Insertion of the class I integron is likely to have taken place in more recent history and likely since the dawn of the antibiotic era.

Tn*21* and Tn*21*-like family transposons have been shown to be mobilised by their own transposase genes forming cointegrates between self-transmissible plasmids containing accessory genes which may also be beneficial to cell survival (Sherratt, 1989). pMG101 is a well characterised example of a large IncFIB/FII transmissible plasmid containing Tn*21* which was isolated in *Salmonella* Typhimurium from a burns unit in Boston, USA in 1975, is one such example (McHugh *et al.,* 1975). The pMG101

plasmid harbouring Tn*21* also contained Tn*7/pco/sil*, conferring resistance to copper

(*pcoEABCDRS*) and silver (*silESRCFBA*), tellurium resistance (*terZABCDE*), mercury

resistance (*merRTPCADE*), chloramphenicol resistance (*catA*), tetracycline resistance

(*tetRBCD*), β-lactamase resistance (*bla*$_{OXA-1}$), sulphonamide resistance (*sul1*),

aminoglycoside resistance (*aadA1*, *strAB*) and partial resistance to QACs (*qacEΔ1*)

(Randall *et al.,* 2015). Further to this, pMG101 possesses a *Salmonella* genomic island

(SGI) similar to a sequence within *Salmonella* Typhimurium STT240 containing a

pathway for siderophore production; aerobactin synthesis (*iucABCD*) (McHugh *et al.,*

1975). Examining the carriage of Tn*21* and the presence of other accessory genes

such as virulence factors and other acquired antimicrobial resistances may help

identify the driver for the continued retention of Tn*21*.

4.1.1 Aims and Objectives

This chapter aims to explore Tn*21* diversity in a small geographical area (Figure 2.1)

and to characterise isolates carrying Tn*21* and Tn*21*-like elements isolated from

wastewater and surrounding areas. Further to this, a comparison of the mercury

resistance transposons identified will be made to analyse diversity and study

potential evolution of the Tn*21*-like elements. Comparisons of these sequences will

also be made to Tn*21* and Tn*21*-like elements previously published in the literature

to identify any key traits or major differences that may have arisen.

1. Carriage on large plasmids plays a role in Tn*21* mobility. Part of this chapter

   aims to demonstrate how the shortcomings of short read sequencing may be

   overcome using PFGE to resolve plasmid structures. This study also aims to

identify the importance of hybrid DNA assembly as a means for better resolution of genomic architecture, but also for accurate reconstruction of the DNA sequence of mobile genetic elements such as plasmids and transposons including Tn*21*, which can contain repeat sequences that lead to mis-assembly of genomic and mobile genetic element sequences.

2. Alongside Tn*21*, the mercury resistant bacteria isolated carried many other antimicrobial resistance genes, virulence factors, genomic islands and other mobile genetic elements may be carried on large plasmids. In order to identify trends that could be due to co-occurrence or possible co-selection events taking place, identification of these accessory genetic elements as well as the resistance genes present will help to confirm the case.

4.2 Methods and Results

Some of the sequenced mercury resistant organisms were isolated from a small geographical area (Figure 2.1) but from different environmental compartments. Some isolates that were selected for genomic sequencing were collected in the influent of a wastewater treatment plant, water downstream of the same wastewater treatment plant and a nearby dairy unit. The isolates selected for sequencing represented a range of different genera collected across different sample dates and sample locations. Due to the financial and time limitations due the COVID-19 pandemic, only a limited number from the collection were sequenced. 20 of the isolates were sequenced with a mixture of Illumina short read and hybrid sequencing externally at MicrobesNG, and others were sequenced in-house using Oxford Nanopore MinION long read technology. The raw sequencing reads were trimmed, and quality checked (2.2.26 and 2.2.27), assemblies were performed by UniCycler unless stated otherwise (Table 4.1).

Isolate (Isolation Date)	Contigs	Total Length (bp)	Read Length N50 (bp)	Sequencing Method
D1 (22/05/2018)	35	5,053,391	2,574,231	Illumina + Nanopore
BPF1-1 (22/05/2018)	206	5,152,855	67,392	Illumina
BPW2-3 (22/05/2018)	10	5,073,893	4,885,373	Illumina + Nanopore
BPW2-4 (22/05/2018)	13	5,073,874	4,849,817	Illumina + Nanopore
NT7 (18/09/2018)	226	5,438,756	91,620	Illumina
NT50 (18/09/2018)	5	5,306,757	4,935,935	Illumina + Nanopore
NT55 (18/09/2018)	6	5,316,699	4,935,987	Illumina + Nanopore
NT65 (18/09/2018)	20	5,055,767	2,565,242	Illumina + Nanopore
NT67 (18/09/2018)	7	5,355,950	1,349,103	Nanopore
C32 (14/02/2019)	22	5,236,089	923,334	Nanopore
V30 (14/02/2019)	8	6,184,703	5,454,512	Nanopore
11V41 (11/10/2019)	30	6,507,830	3,684,987	Nanopore
A113 (30/01/2020)	3	5,234,291	5,193,748	Nanopore
A114 (30/01/2020)	3	5,215,148	5,197,090	Nanopore
EVAL51 (22/05/2017)	90	5,213,877	244,669	Illumina
EVAL55 (22/05/2017)	138	5,337,455	185,747	Illumina
EVAL56 (22/05/2017)	113	5,066,116	165,136	Illumina
EVAL67 (29/06/2017)	144	5,329,392	183,699	Illumina
EVAL113 (11/07/2017)	88	5,329,392	124,734	Illumina
EVAL397 (16/08/2017)	96	4,956,512	236,608	Illumina

Table 4.1 Assembly statistics from the sequencing of mercury resistant and sensitive bacteria.

4.2.1   Identifying Differences Between Mercury Resistant Isolates with the Same

Antimicrobial Resistance Profiles.

Two *E. coli* isolates: NT50 and NT55 were isolated on the same date from the same

location, downstream of the wastewater effluent release release into a local brook

that runs into the River Soar and displayed identical resistance profiles in AST disc

assay testing. To determine any difference between the two strains sequence types

before sequencing, an XbaI digest of total DNA from the strains was used. The *xbaI*

restriction site (5'-T/CTAGA-3') is not found very often within *E. coli* genomes.

Therefore, the banding pattern of electrophoresed digested genomic DNA fragments

produced when using cleaved by the XbaI restriction enzyme is used to confirm

whether isolates may be different from one another. The XbaI digest of NT50 and

NT55 (Figure 4.1) showed that the strains have the same restriction patterns after

electrophoresis. Strain NT65, which had a different antimicrobial resistance profile,

was also analysed to show variation of sequence types. At the time of performing

these PFGE experiments, it was not known if all of these isolates would be whole

genome sequenced. Mercury resistance operons and Tn*21* are often carried on large

self-transmissible plasmids. Before sequencing, in order to determine any

differences in extra-chromosomal DNA, a PFGE S1 digest was used to identify sizes

of plasmids present within strains as initially these two strains were sequenced with

short-read Illumina and plasmid resolution was not possible. The agarose gel showed

a linear fragment of 150 kb DNA and a second linear fragment approximately 120 kb

which were plasmids (Figure 4.2).

NT50 NT55 NT60 L
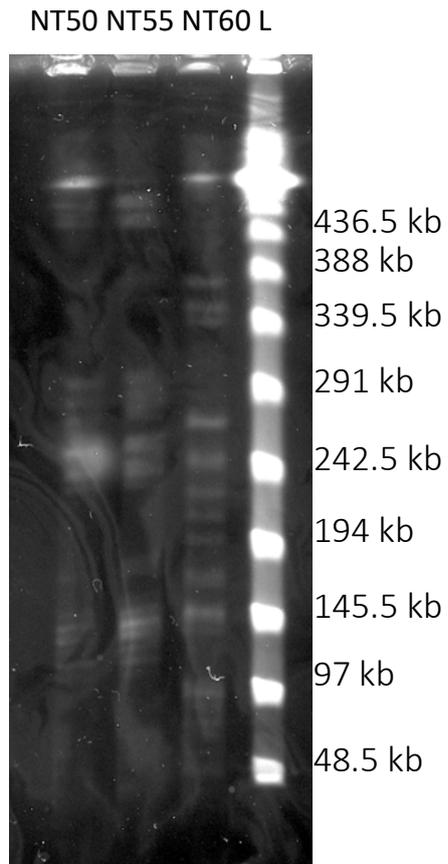
Figure 4.1 Pulse field agarose gel displaying XbaI digest of NT50, NT55 which displayed the same resistance profile, and NT65 with a different resistance profile.

Lane L represents PFG Lambda ladder (Monarch, NEB).
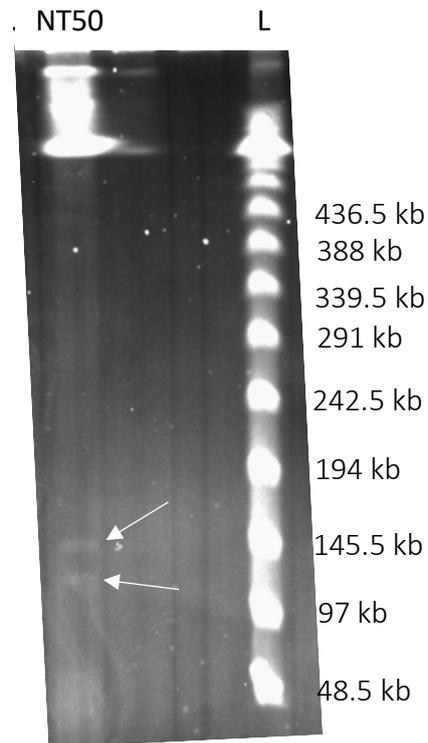
Figure 4.2 Pulse field agarose gel of an XbaI digest of total DNA preparations of strains NT50 and NT55 which displayed the same antimicrobial resistance profile, alongside strain NT65 with a different resistance profile.

Lane L represents PFG Lambda ladder (Monarch, NEB).

4.2.2 Whole Genome Sequence Analysis of Mercury Resistant Isolates

Alongside some of the confirmed mercury resistant bacterial isolates from the wastewater and EVAL-FARMS collection, three isolates which were shown to grow in the presence of mercury and confirmed to possess *merACR* and *intI1* from the EVAL-FARMS collection were also sequenced at MicrobesNG, but sequencing showed they did not possess mercury resistance genes or evidence of Tn*21* presence when sequenced (Iles, 2018). A plasmid containing a Tn*21*-like mobile element isolated from another study from a flock of sheep on the same farm site was also analysed in this study.

4.2.2.1 Sequence Typing Mercury Resistant Bacterial Isolates

Fully sequenced isolates were screened for their species from analysis of their 16S

gene and then sequence type was determined using multiple locus sequence typing

(MLST) (Larsen *et al.*, 2012) (Table 4.2).

Isolate	Isolation Location	Species	Sequence Type
D1	Ditch contaminated with grey Water	*E. coli*	635
BPF1-1	Effluent Pipe Outlet	*E. coli*	635
BPW2-3	Downstream of wastewater treatment site effluent pipe	*E. coli*	635
BPW2-4	Downstream of wastewater treatment site effluent pipe	*E. coli*	635
NT7	Downstream of wastewater treatment site effluent pipe	*E. coli*	1430
NT50	Downstream of wastewater treatment site effluent pipe	*E. coli*	88
NT55	Downstream of wastewater treatment site effluent pipe	*E. coli*	88
NT65	Downstream of wastewater treatment site effluent pipe	*E. coli*	635
NT67	Downstream of wastewater treatment site effluent pipe	*C. freundii*	611
C32	Campus Influent	*E. coli*	1331*
V30	Village Influent	*C. freundii*	496
11V41	Village Influent	*E. coli*	200
A113	Campus Influent	*A. salmonicida*	391
A114	Village Influent	*A. salmonicida*	391
EVAL51	Dairy unit slurry tank	*E. coli*	69
EVAL55	Dairy unit slurry tank	*E. coli*	362
EVAL56 (Hg^s)	Dairy unit slurry tank	*E. coli*	187
EVAL67 (Hg^s)	Dairy unit slurry tank	*E. coli*	69
EVAL113 (Hg^s)	Dairy unit slurry tank	*E. coli*	206
EVAL397	Dairy unit slurry tank	*E. coli*	9504

Table 4.2 Sequence typing, species breakdown and sequencing methods used for each of isolate.

* Most likely sequence type due to read error rate in long read sequencing. Hg^s refers to no identification of *mer* genes.

Most of the sequenced isolates were *E. coli*, due to their initial isolation on TBX agar (2.1.4.5). TBX agar selects for glucuronidase producing bacteria capable of growing in the presence of bile salts. *Citrobacter freundii* and some strains of *Enterobacter cloacae* also produce glucuronidase and survive bile salts. The four samples isolated from the grey water contaminated ditch (D1) and the Black Brook effluent pipe (BPF1-1, NT65) and downstream of the wastewater treatment plant (BPW2-3, BPW2-4) were clonal. These isolates, except NT65, were all sampled on the same date but were collected from 3 different locations. The grey water isolate D1 was isolated further away from BPF1-1, NT65, BPW2-3 and BPW2-4 (Figure 2.1) (see page 78)), but is likely to have leaked from a sewage pipe leading from campus to the WWTP. NT65 was isolated four months later but the same clone was persistent in the Black Brook effluent pipe from the wastewater treatment plant.

Strains NT50 and NT55 were also isolated on the same date as NT65 but were a different sequence type. The difference in sequence types between NT50, NT55 and NT65 confirmed the PFGE data shown in Figure 4.1. The non-wastewater Isolates EVAL51 and EVAL67 were collected one month apart from each other from the slurry tank of a neighbouring dairy unit but were the same sequence types. Overall diversity of the sequence types in the sequenced mercury resistant isolates was limited. The EVAL-FARMS isolate sequence types were more diverse.

To confirm the MLST findings for the isolates with the same sequence type, core genome MLST (cgMLST) was also performed with cgMLSTFinder. cgMLST, compares alleles of genes found within the core parts of the chromosome required for cells to

survive rather than comparing a small number of housekeeping genes (the genes compared varies between species). A core genome (cgST) detected ±10 of another isolate is considered to be closely related enough to be part of the same clonal outbreak. In Table 4.3, some isolates that were shown to have the same sequence type in Table 4.2 are shown to have the same cgST. This strongly suggests that these isolates are clonal. These data meant that the isolates that were the same sequence type which had a hybrid assembly could be used to map the core genomes of those isolates that were sequenced with short-read technology in order to achieve better resolution of differences between isolates.

Isolate	Core Genome Sequence Type
D1	127856
BPF1-1	87180
BPW2-3	127856
BPW2-4	127856
NT65	127856
NT50	1414
NT55	1414
EVAL51	91823
EVAL67 (Hg^s)	100465

Table 4.3 cgMLST results of sequenced isolates that had the same sequence type from MLST searches.

For example: isolates D1, BPF1-1, BPW2-3, BPW2-4 and NT65 had the same ST in MLST but BPF1-1 was shown to be different in cgMLST.

The data from Table 4.3 confirmed that isolates D1, BPW2-3, BPW2-4 and NT65 were clonal. Similarly, isolates NT50 and NT55 are also clonal.

197

4.2.3   Plasmid Carriage of Tn*21*

PlasmidFinder 2.1 was used to identify plasmid incompatibility groups within the sequenced isolates (Camacho *et al.*, 2009; Carattoli *et al.*, 2014) (Table 4.4). All isolates, except the two *Aeromonas salmonicida* isolates (A113 and A114), and one *Citrobacter freundii* (C32) isolate, possessed mercury resistance operons extra-chromosomally. This identification could be made as the annotations of these assemblies identified the *mer* operons on contiguous DNA containing either plasmid transfer genes or plasmid origin of replication sites. The majority of which were carried on large IncF variants or IncI1 plasmids. Further analysis of the assemblies of the mercury resistant *Aeromonas salmonicida* strains and the *Citrobacter freundii* isolate NT67 showed no evidence of plasmid transfer genes, nor origin of replication sites, when using annotation software RASTtk or PlasmidFinder 2.1 (Table 4.4). Similarly, the class I integron was not observed either in these isolates (NT67, A113 and A114). Plasmids that were assembled by either long read technology such as Oxford Nanopore or hybrid techniques such as BPW2-3, NT55 and NT50 produce longer contigs which allowed for resolution of plasmids and gap closing of circular DNA. When the data from Table 4.5 was cross referenced with the data in Table 4.2, it confirms this finding as isolates containing the same cgST as each other were identified to carry the same plasmids and same sized plasmids where there were more than one present. Secondly, when comparing the annotations of RASTtk and the findings of the PlasmidFinder2.1 software, there are discrepancies between the plasmid incompatibility groups documented, meaning one database may possess

inaccuracies or that the replicons of some plasmid incompatibility group genes may

be very similar.

Isolate	Plasmid Incompatibility Groups
D1	IncFIB, IncY
BPF1-1	IncFII
BPW2-3	IncFIB, IncFII
BPW2-4	IncFII, IncFIB
NT7	IncFIB, IncHI1A, IncH1B, IncX1
NT50	IncFIB, IncFII, IncI1, IncL/M, IncQ1, IncX1
NT55	IncFIB, IncFII, IncI1-I(alpha), IncL, IncQ1, IncX1
NT65	IncFIB, IncFII
NT67	-
C32	IncFIA, IncFIB, IncFII, IncQ1
V30	pKPC-CAV1321
11V41	IncB/O/K/Z, IncFII, IncHI1A, IncHI1B(R27), IncR
A113	-
A114	-
EVAL51	IncFIA/FIB, IncFII, IncY
EVAL55	IncFIA, IncFIB, IncFII, IncI2
EVAL56 (Hg^s)	IncFIB, IncFII, IncI1-I(alpha)
EVAL67 (Hg^s)	IncB/O/K/Z, IncFIB, IncFII
EVAL113 (Hg^s)	IncFIB, IncFII, IncY
EVAL397	IncFIB, IncFII

Table 4.4 Plasmid incompatibility groups detected in each isolate.

Detection of these incompatibility group origins of replications in isolates was performed by
PlasmidFinder 2.1.

Limitations of short read sequencing and assembly resulted in difficulties resolving

the plasmid sizes and structures due to the presence of repeat regions within the

strain's chromosomal and plasmid DNA. However, of those that were hybrid and long

read assemblies, plasmid structures were resolved more readily by using the long-

read sequences as a scaffold and then using the short read data to reconstruct the

DNA and close gaps in the assembly. In section 4.2.2.1, cgST showed some isolates

to have the same core genomes. Alongside the data in Table 4.4 this meant it may be possible to also map the contigs for plasmids where the same Inc group was identified within clones to approximate or potentially fully resolve their size and structures (Table 4.5). The mapped assemblies of BPW2-4 and NT65 resolved the same plasmid structure and size as BPW2-3 (Figure 4.3). Isolate D1, isolated from a drainage ditch contaminated with grey water, also had the same cgMLST as these isolates suggesting that the ditch was contaminated with sewage from a leaking pipe that ran to the WWTP from the Campus. Plasmid mapping of isolate D1 against isolate BPW2-3 showed a 26 kb portion missing from the IncFIB plasmid found in the other isolates of the same cgST; identifying a possible horizontal gene transfer event where this IncFIB plasmid gained DNA across the wastewater treatment process (Figure 4.3).

Isolate	Inc Group	Plasmid Size / bp
D1	IncFIB	130,317*
BPF1-1	IncFII	n/a
BPW2-3	IncFIB	156,070*
BPW2-4	IncFIB	156,070*
NT7	IncX1	n/a
NT50	IncFIA /IncFIB/IncFII	148,191*
	IncI1	127,748*
	IncL	58,515
	IncX1	36,305
NT55	IncFIB/IncFII	151,599*
	IncI1-I(alpha)	127,748*
	IncL	58,515
	IncX1	41,257
NT65	IncFIB	156,070*
NT67	IncI1	n/a
C32	IncFIB/IncFII	135,555*
V30	IncFIB [a]	176,654*
11V41	IncHIA/IncHIB/IncFII	360,844*
	IncI1 [b]	97,493
	IncR	39,285
A113	-	-
A114	-	-
EVAL51	IncFIA/FIB, IncFII	n/a*
EVAL55	IncFIB	141,415[c]*
	IncI2	60,022[c]
EVAL56 (Hg[s])	IncFIB/IncFII	n/a
	IncI1-I(alpha)	n/a
EVAL67 (Hg[s])	IncB/O/K/Z, IncFIB, IncFII	n/a
EVAL113 (Hg[s])	IncFIB, IncFII, IncY	n/a
EVAL397	IncFIB, IncFII	n/a*

Table 4.5 Predicted sizes of plasmids found within each isolate and the incompatibility groups identified.

* Indicates Tn*21* or mercury resistance determinants found. [a] PlasmidFinder did not agree with sequence data, annotation identified IncF plasmid over pKPC-CAV on PlasmidFinder. [b] PlasmidFinder identified this as Inc/O/K/Z when in fact annotation identified the region as IncI1. [c] denotes scaffolding to a known reference gave an output of an approximate plasmid size but is not fully corroborated. n/a shows where plasmid size could not be resolved due to short read data being limited for resolving plasmid structure.

Strains BPW2-3, BPW2-4, D1 and NT65 which possessed the same cgST had the

contigs of their plasmids mapped against each other using Circoletto (Figure 4.3). The

two plasmids from NT65 and BPW2-4 that were mapped against each other were the

same size and matched to each other almost perfectly. The reconstruction of the

IncFIB plasmid from isolate D1 however matched less well.
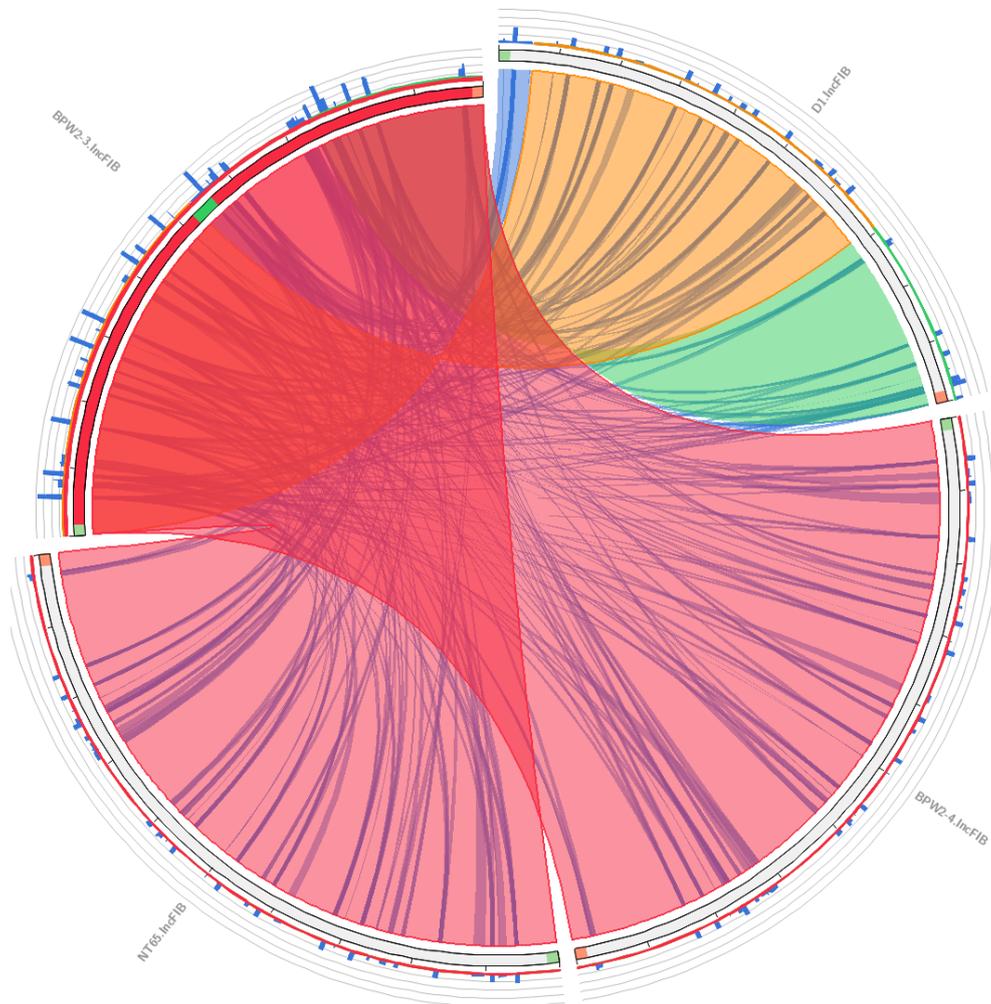
Figure 4.3 Circoletto alignment of contigs containing IncFIB plasmid of identical cgST containing the *mer* operon.
The BPW2-3 IncFIB plasmid was used as the reference contig to the other three plasmid contigs. The green section within the backbone of the BPW2-3 IncFIB plasmid (top left) contains the *mer* operon. The coloured mapping across the figure to other plasmid sequences indicates percentage identity to the reference BPW2-3 IncFIB plasmid. Blue ≤ 25%, green ≤ 50%, orange ≤ 99.9% and red > 99.9% identity. Command line options for Circoletto used was as follows： --e_value 1e-10 --score2colour bit --scoreratio2colour max --maxB1   50 --maxG2 75 --maxO3 99.9999 --out_size 750 --annotation /labs/bat/www/tools/results/circoletto/annotation0008329361 --annocolour scores --ribocolours2allow '(blue|green|orange|red)' --z_by score

In each of these isolates, a *mer* operon was located within the IncFIB plasmid. The structure of the *mer* operon in the IncFIB plasmid was not of Tn*21* but in fact a fusion of the *mer* operon, *urf2* and a Tn*21* resolvase to an IS*26*-like transposase (Figure 4.4). The IS*26*-like transposase in this case forms part of a chain of IS*26*-like elements within the plasmid. Similarly, the assembly of the Tn*21*-like transposon in strain BPF1-1 was similar to isolates BPF1-1, NT65 and BPW2-4. Figure 4.5 shows the two transposons side by side in panels a and b. The structures of the two mobile elements are very similar. However, the transposon in BPF1-1 differed to those within isolates D1 (from a grey water contaminated ditch), BPW2-3, BPW2-4 and NT65 (isolated downstream of the wastewater effluent pipe) as it was not flanked by a second IS*26* transposase or other IS*26* in a chain.
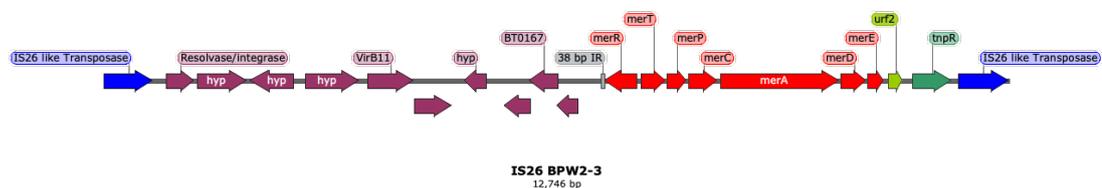
Figure 4.4 Map of BPW2-3 (D1, NT65 and BPW2-4) IS*26*-like transposable element.

The IS*26*-like element resides in an IncFIB plasmid which is 150 kb in size. The *mer* operon contains *merRTPCADE* and *urf2* is located adjacent to it.

All the isolates containing a Tn*21*-like transposon were annotated and are displayed alongside each other in Figure 4.5. The sequences show each isolate contained the full *mer* operon (RTPCADE), including the full 38 bp IR and the gene *urf*2, except for 11V41, which was isolated from village wastewater influent. The sequence annotation suggests that *merE* is missing. However, upon closer inspection of the

203

sequence data, this appears to be an error in the assembly due to base-calling errors in long-read only sequencing. The rest of the genetic structures of these Tn*21*-like transposons varied. Whilst the transposase *tnpA* and *tnpR* was present in most transposon sequences, there was some variation. Isolates A113 and A114 possessed a truncated *tnpA* gene, but still retained the 38 bp IR of the transposase end. Another variation in the transposase region is the truncation of *tnpR* in isolate V30, where an IS*66*-like element has inserted causing deletion of 318 bases of *tnpR* at the 3' terminal. The In2 module which contains the class I integron was absent, as was *tnpM*.

Only six of the sequenced mercury resistant isolates contained a class I integron (NT50, NT55, EVAL51, EVAL55, EVAL397, C32). The contents of which was similar to each other. All of the integrons contained at least one aminoglycoside resistance gene cassette and all except one contained a sulphonamide resistance gene cassette (*sul1* or *sul2*) too. Notably beta lactamase production genes, bla_{TEM} and bla_{OXA-1} were identified within the cassette arrays; bla_{TEM} possessed its own promoter within the cassette, which is uncommon in gene cassettes carried by integrons.
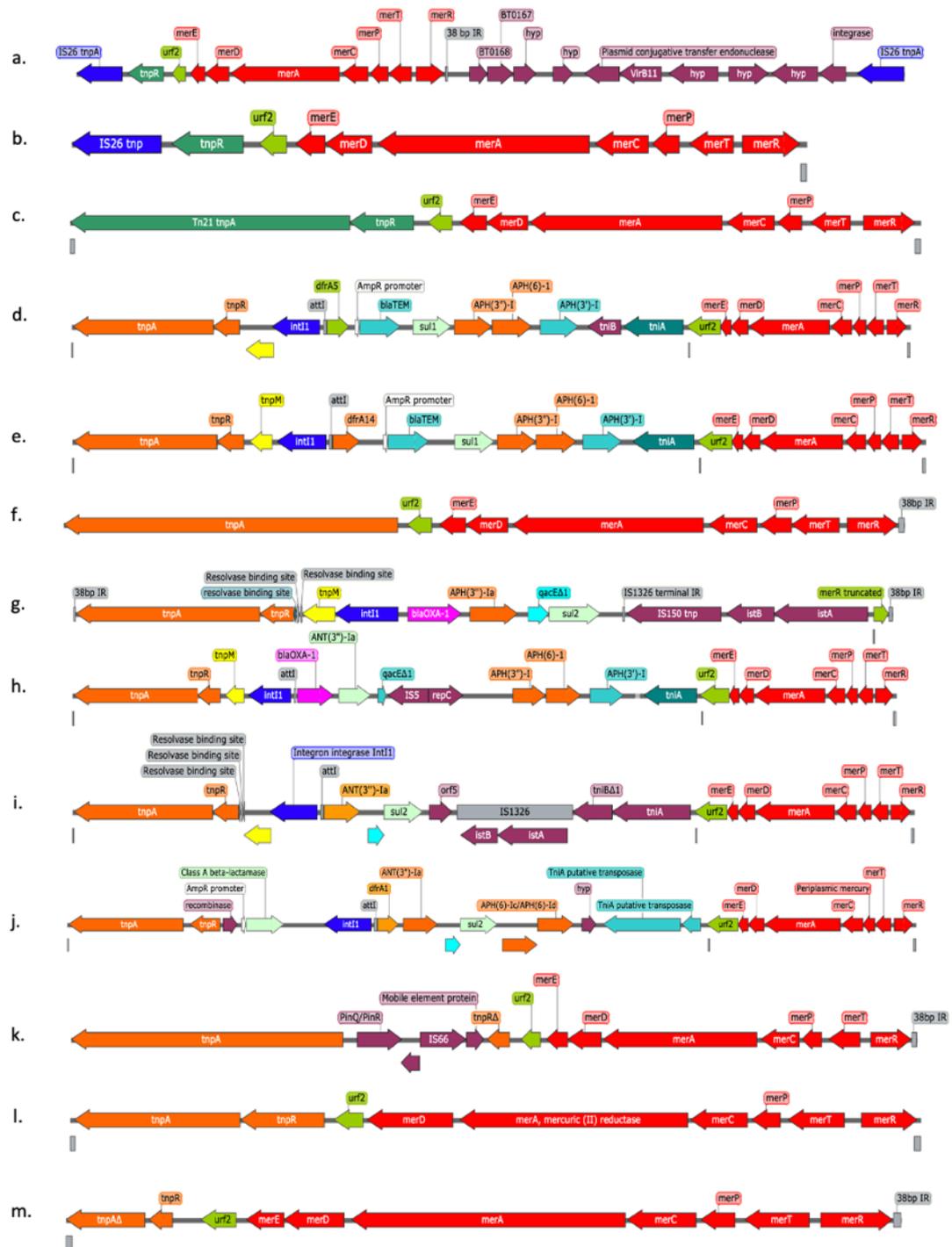
Figure 4.5 Reconstructions of annotated Tn*21*-like genetic elements from wastewater treatment plant and nearby dairy unit.

No scale has been used here due to the largely varying sizes in the genetic elements and difficulties to scale the structures alongside each other. a. IS*26*: D1 BPW2-3 BPW2-4 NT65, 12,746 bp. b. IS*26*: BPF1-1, 5,749 bp. c. Tn*21*-like NT7, 7,465 bp. d. Tn*21*-like NT50, 17,717 bp. e. Tn*21*-like NT55, 17,675 bp. f. Tn*21*-like NT67, 7,299 bp. g. Tn*21* EVAL51, 13,248 bp. h. Tn*21* EVAL55, 19,804 bp. i. Tn*21* EVAL397, 18,056 bp. j. Tn*21* C32, 19,111 bp. k. Tn*21*-like V30, 9,244 bp. l. Tn*21*-like 11V41, 6,262 bp. m. Tn*21*-like A113 and A114, 5,130 bp.

4.2.4 Tn*7/pco/sil* Co-occurrence in Wastewater Isolates

Six of the mercury resistant *E. coli* isolates that were sequenced also carried the copper/silver resistance transposon, Tn*7/pco/sil* (*tnsABCDE, silESRCFBA, pcoEABCDRS*). None of these isolates came from the dairy unit, all of them from wastewater influent and effluent. The Tn*7* mobile genetic element was reported previously to insert chromosomally, preferably to a region downstream of the *glmS* gene. However, in the sequenced isolates possessing the Tn*7/pco/sil* transposable element, the element was inserted between *yhiM* and *yhiN* (Figure 4.6). *yhiM* encodes an inner membrane protein, and *yhiN* encodes a predicted oxidoreductase. The use of alternate insertion sites in the *E. coli* chromosome for the Tn7/*sil*/*pco* element that are not the canonical *att*Tn*7* site (5'- TCATTTGACGCCGAAGTCACTGGCTTACGCTCCCG -3') has also been observed in pig *E. coli* isolates and the *E. coli* strain J53 with plasmid pMG101-B (Chalmers *et al.*, 2018; Hooton *et al.*, 2021). Also contained within this Tn*7* mobile genetic element are other ORFs. Between the *tns* and *sil* operons exist ORFs of an antitoxin encoded protein, AAA family ATPase and a hypothetical protein. One final ORF is located between the *sil* and *pco* operons (Figure 4.6). This ORF encodes a peptidoglycan DD-metalloendopeptidase (MepM), which in the presence of a divalent metal ion promotes cell elongation. This configuration of the Tn*7* was identified in 8 of the mercury resistant strains that were sequenced (Figure 4.5). One strain, BPF1-1 possessed Tn*7/pco/sil* between *yhiMN*; however, *tnsABCDE* (the Tn*7* transposase module) was also identified within the 3' region of *glmS*.

Figure 4.6 Chromosomal insertion Tn7/pco/sil between *yhiM* and *N*. This transposable element was found in this configuration in all of the wastewater isolates.

Tn*7* inserts between *yhiM* and *yhiN* at an *att*Tn*7* recognition site causing a 5 bp duplication at the 5' terminal of the *tnsABCDE* transposase genes and at the 3' end of the Tn*7*, in this case downstream of *pcoEABCDRSE*. The Tn*7* across all the isolates sequenced in this study possessed Tn*7/sil/pco* in this configuration, including the *hipB, recF, mepM* and two hypothetical protein coding regions.

4.2.4.1   Co-Occurrence of Other Metal Resistance

Two *Aeromonas salmonicida* strains possessed the chromosomal carbapenemase producing gene, *cphA5*, as well as the Tn*21 mer* operon. They also carried a zinc resistance operon, *czcABCD*, and an arsenic resistance operon, *arsABCD*. These resistance genes have not been identified in any of the *E. coli* or *Citrobacter* spp. that were sequenced. All but two of the sequenced mercury resistant isolates from the EVAL-FARMS collection possessed the arsenic resistance operon (*arsABCR*) (Gladysheva, Oden and Rosen, 2002). One of these isolates also possessed *arsD* within the operon. The *ars* operon and tellurite resistance operon *terABCD* was also detected in isolate 11V41.

4.2.5   Evidence of Further Tn*21* Evolution

Bacterial evolution takes place constantly and dynamically. Mobile elements play a role in this process but are also evolving simultaneously. In one such bacteria isolated from a nearby dairy unit, an *E. coli* isolate, EVAL51, contained a Tn*21*-like element which had lost the *mer* operon apart from 249 bases of the *merR* gene. The 3' region

207

of the *merR* gene forms part of the 38 bp IR (L) of Tn*21*, which is required for the

mobility of the mobile genetic element and has a strong evolutionary pressure to

retain it. Further to this, the Tn*21*-like element in EVAL-Farms isolate 51 is flanked by

an IS*91* and florphenicol/chloramphenicol resistance gene (Figure 4.7). Also of note,

the cassette array of the class I integron possesses the classically characterised

combination of *sul*, *qacEΔ1* and *aadA1*; *bla$_{OXA-1}$* has inserted also, which was first

identified in pMG101 (Mchugh *et al.*, 1975). *tni* transposition module and IS*1353*

have also been lost from this transposon.
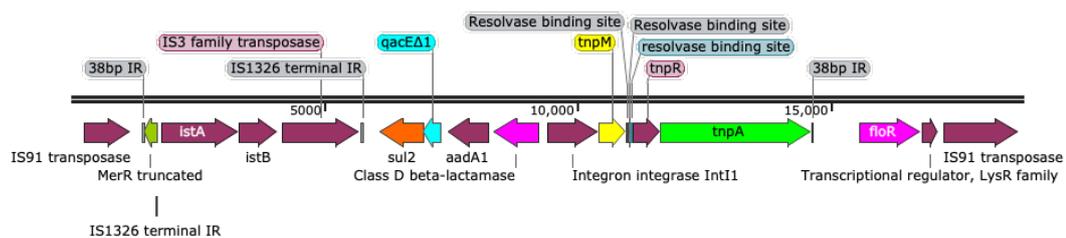
**Figure 4.7 Tn*21*-like element from EVAL-Farms isolate 51 with *merTPCADE* deletion and
lack of *tni* module.**

This Tn*21* variant contains a deletion of *merTPCADE* and contains the addition of the class D
beta lactamase, *bla$_{OXA-1}$* to the class I integron. The Tn*21*-like element is also flanked by IS*91*
insertion sequences, with the florfenicol resistance gene *floR* inserted downstream of *tnpA*.

EVAL51 is not the only isolate characterised which shows evidence of Tn*21* evolution,

likely through mutation. Another isolate from the EVAL collection, EVAL55, possessed

a similar Tn*21* structure to that found in EVAL51. However, the *mer* operon was

intact. The *tnpA* repeat was flanked by the same *floR* IS*91* element. However, IS*91*

was missing from the *mer* side of the Tn*21* mobile genetic element. Both Tn*21*-like

elements were identified on IncF plasmids. The gene arrangement of this Tn*21*-like

208

mobile element was similar to the one found in EVAL51, which was isolated from the same location three years previously (Figure 4.5).

4.2.5.1 Co-occurrence of Other Acquired Antimicrobial Resistance Genes

Whilst Tn*21* may possess gene cassettes conferring resistance to antimicrobials, not all of the resistance genes acquired by that organism may be present within the integron. In the sequenced isolates from this study, multiple isolates possessed other acquired antibiotic resistance genes. In both strains (EVAL51 and EVAL55), IS*91* contained *floR*, providing resistance to chloramphenicol and florfenicols, close to the Tn*21*-like mobile elements (Figure 4.7). Both strains also possessed another resistance gene conferring trimethoprim resistance (*dfrA36*), and macrolide resistance (*macAB-tolC*). However, the two dairy *E. coli* isolates could be differentiated by their resistance genes: EVAL55 also contained bla_{TEM-1B} in addition to the bla_{OXA-1} gene identified in the cassette arrays of the Tn*21* class I integron found in both EVAL51 and EVAL55.

macAB-tolC was identified in strains harbouring IncFIB plasmids such as those in NT50 and NT55 (Figure 4.8). This tripartite protein effluxer complex provides resistance to macrolides and has been shown to offer some protection against polymyxins and aminoglycosides. MacAB-TolC is an efflux pump, more than antimicrobials can be exported, the tripartite cassette is also linked to potential virulence by exporting heat-stable toxins (Yamanaka *et al.*, 2008).
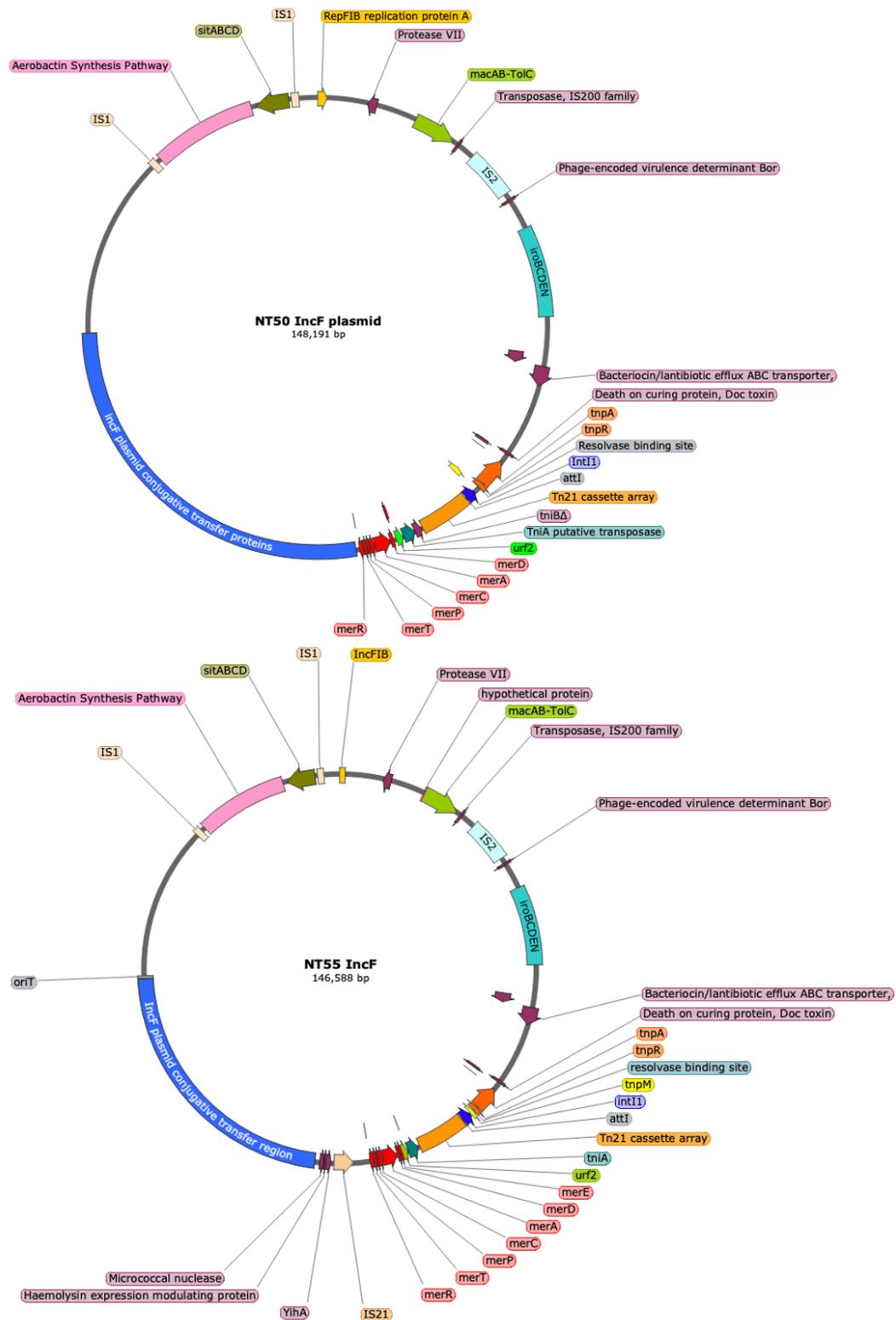
Figure 4.8 Plasmid Maps derived from sequence assemblies of the IncF plasmids from wastewater *E. coli* isolates NT50 and NT55.

Both plasmids contain a Tn*21*-like mobile element providing resistance to mercury, aminoglycosides, beta-lactams and trimethoprim. Macrolide resistance provided by *macAB-tolC* is present. Both plasmids contain IncF conjugative transfer regions for mobilisation. The

two plasmids also possess genes for the biosynthesis of aerobactin and enterobactin iron chelators. The two plasmids are also capable of chelating manganese with *sitABCD*, normally found within *Salmonella enterica* serovar Typhimurium.

4.2.6   Modelling the Development of Tn*21*

The Tn*21* and Tn*21*-like transposable elements that were isolated and sequenced were aligned against published Tn*21*, Tn*21*-like elements (Table 4.6) and other Tn*3* family transposon sequences from Tn*501*, Tn*1696* and Tn*1721*. Figure 4.9 shows a phylogenetic tree of the aligned transposable elements. Whilst Tn*501* possesses *tnpAR* and a *mer* operon, it differs from Tn*21* as *tnpM*, In2 and *merC* are absent (Figure 1.5). Figure 4.9 is a phylogenetic tree that identifies the relationships between the NCBI sequences and isolates sequenced within this study (2.2.28.1) containing Tn*21* and Tn*21*-like transposable elements to the commonly used reference transposons for Tn*21* (found in pMG101, R100 and Enteroaggregative *E. coli* strain 042) and similar Tn*3* family transposons Tn*501*, Tn*1696* and Tn*1721*. The phylogenetic tree suggested that the Tn*21*-like elements in strains EVAL51 and EVAL 55 were ancestors of the classic Tn*21* described by Liebert *et al.* (1999) rather than possible descendants of it as they were located closer to the root on the phylogram. One of which (EVAL51) contains a deletion of the *mer* operon. These two Tn*21*-like sequences are likely to have come from the same source, where incorrect transposition or replication occurred causing a deletion of the *mer* operon to result in EVAL51.

Tn*1696* is very similar in structure to Tn*21*; they both contain *merRTPCADE* and *tnpAR* (Stokes and Hall, 1991). However, there is no *tnpM* or *urf2 in Tn1696*. Tn*1696* does

not possess a *tni* module but instead contains In4, an integron with a slightly expanded cassette array with *aacC1*, *orfE*, *aadA2* and *cmlA1* inserted 5' of the *qacEΔ1* gene. Isolates NT7, BPF1-1, D1 and A114 did not possess an integron in their sequences but were more closely related to Tn*1696*. All of the isolates that were sequenced contained *merC* in the *mer* operon so were not likely to be similar to that of Tn*501*; this is shown in   Figure 4.9 and Figure 4.10.

The Tn*21*-like transposable element was almost identical to the Tn*21* of the pMG101 plasmid and R100. The most significant difference between this transposable element and the reference sequences was the absence of IS*1353*, which contains *orfAB*, in the In2 region.
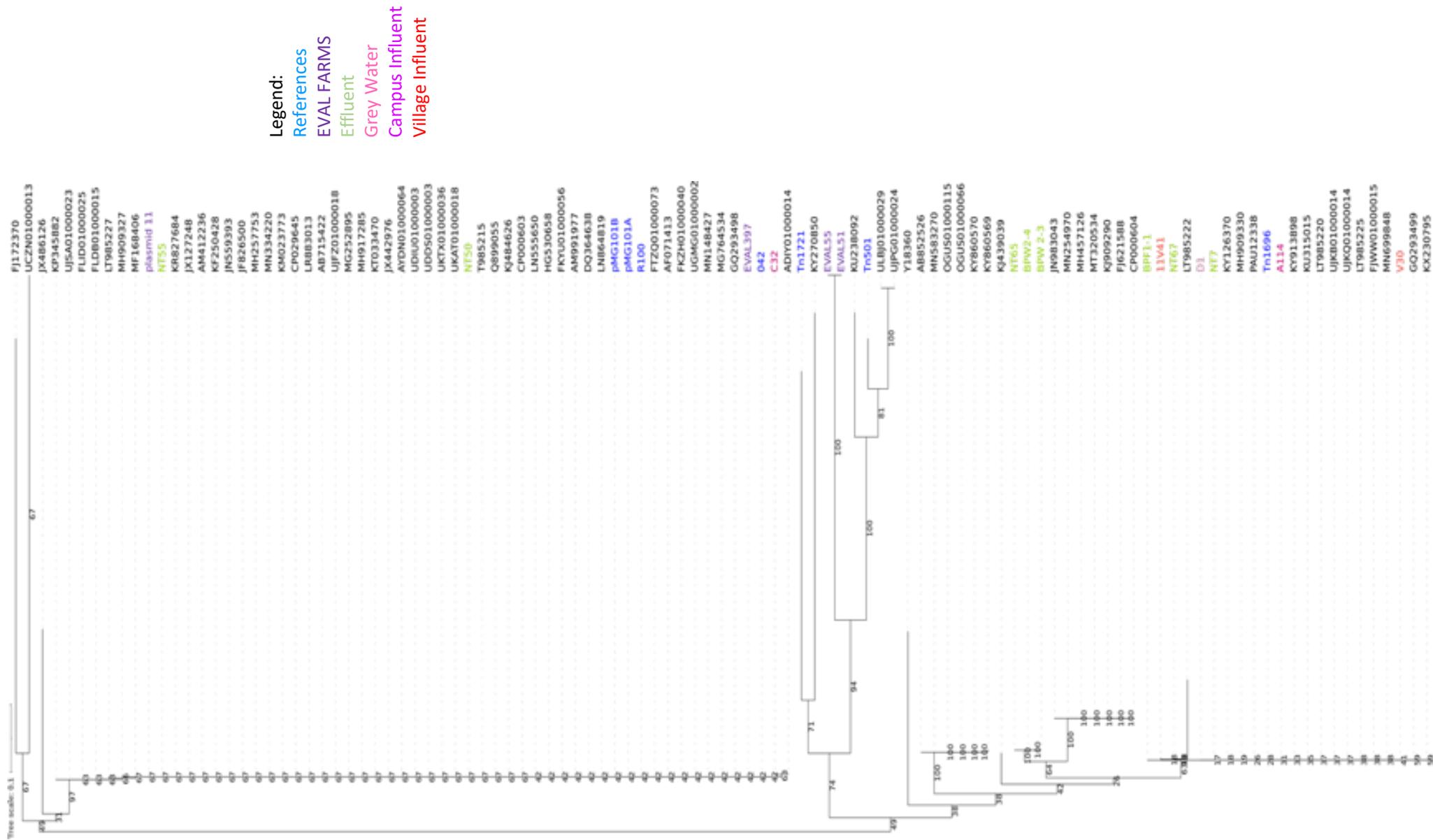
Figure 4.9 Phylogenetic tree comparing reference Tn*21*s and Tn*3* family transposons with NCBI published environmental Tn*21*s and sequenced Tn*21* from wastewater en

Sequences were aligned using MAFFT and the phylogenetic tree was created from the alignment. Bootstrapping was performed to the value of 100 and displayed on each bra

8.1).
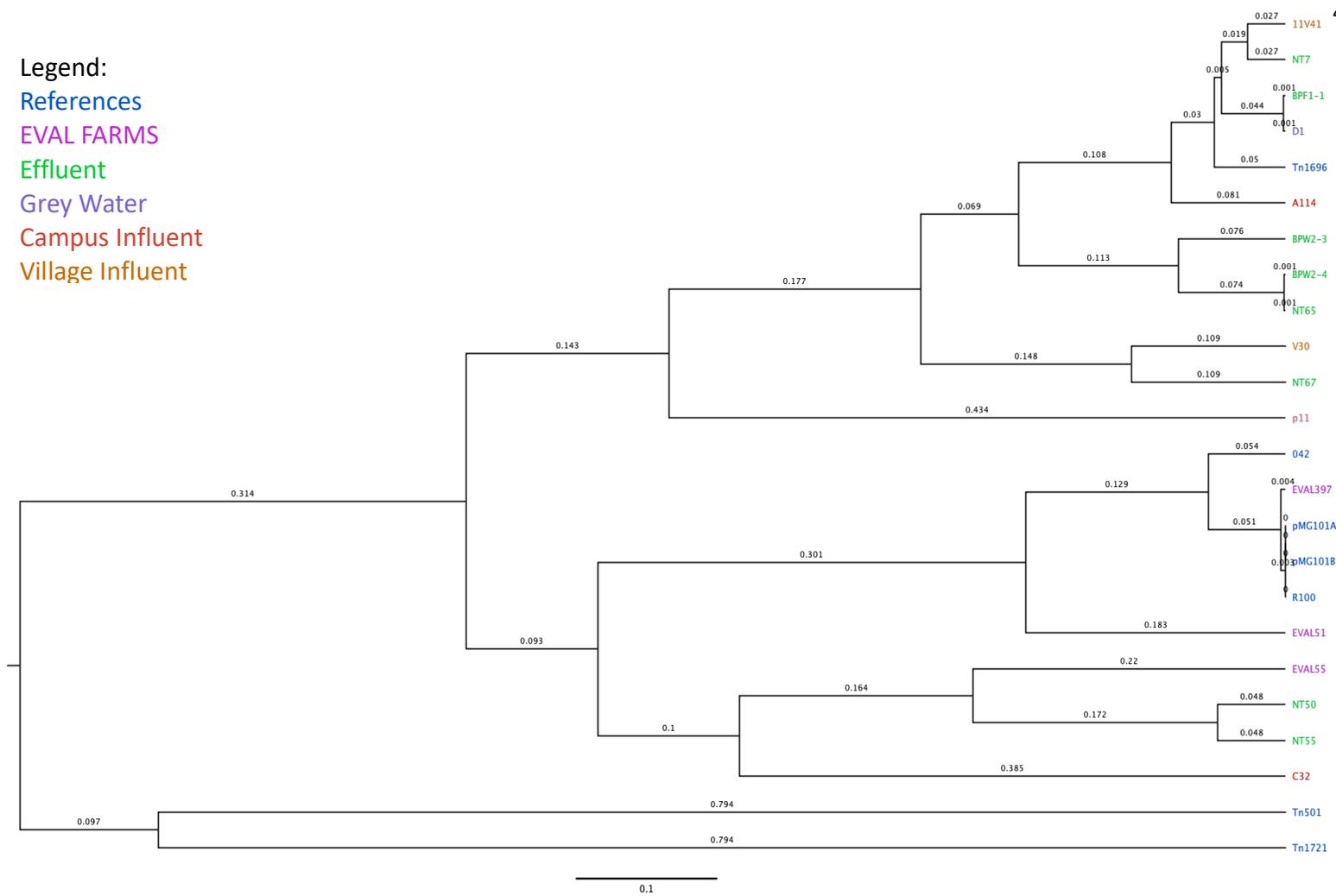
Figure 4.10 Phylogenetic tree comparing mercury resistance transposons sequenced in this study against reference Tn*21*, Tn*21*-like transposons and other Tn*3* family transposons.

Sequences were aligned using MAFFT and the phylogenetic tree was produced from the alignment.

4.2.7   Carriage of Virulence Genes

The sequenced isolates were also screened for carriage of virulence genes using VirulenceFinder 2.0 to identify any other forms of co-occurrence (Table 4.6). Unfortunately this tool cannot be used to determine virulence genes in *Aeromonads* or *Citrobacter spp.*, meaning it was not possible to identify possible genes which may cause virulence within isolates NT67, V30, A113 and A114.

However, upon searching manually, isolates V30 and NT67 (both *Citrobacter freundii*) contained the enterobactin synthesis pathway *fep* genes, a siderophore which chelates iron (Buchanan *et al.*, 1999). *iucA*, encoding aerobactin was also found (Russo and Gulick, 2019). NT67 also contained *virK*, which encodes a protein allowing the aggregation of cells in order to colonise host cells, such as in the intestines (Tapia-Pastrana *et al.*, 2012). Isolates A113 and A114, did not possess many virulence factors either, however they did possess type 6 secretory system (T6SS) and virulence factor *vasK*. *vasK* has been previously identified with T6SS in *Vibrio cholerae* and is used to kill other Gram-negative bacteria (Miyata *et al.*, 2013).

Whilst the majority of the isolates (11) contained a limited set of virulence genes (two or three genes), there were equally others harbouring a great variety of such virulence genes including genes such as *astA*, which encodes heat stable toxin EAST-1, *aggR*, encoding a protein that allows agglutination, and multiple iron sequestering mechanisms (*iuc, fep, iro* and *iut*). The main iron sequestering mechanism detected was *iroBCDEN*, however *iutA*, aerobactin was also detected in seven isolates (Table 4.6).

215

There were also some limitations to the use of VirulenceFinder. Whilst carrying an extensive list of virulence genes, isolate C32 also harboured the full yersiniabactin synthesis (*ybt*) pathway (Miller *et al.*, 2010). This pathway also produces a siderophore which causes the chelation of iron. This isolate along with three others (EVAL51, EVAL55 and EVAL67) also possessed *sitABCDE*, a pathway capable of providing resistance to peroxides.

Isolate	Virulence Genes Detected
D1	*celb, gad, lpfA*
BPF1-1	*gad, lpfA*
BPW2-3	*gad, lpfA*
BPW2-4	*gad, lpfA*
NT7	*gad, lpfA*
NT50	*gad, lpfA, iss, mchF, iroBCDEN, iutA*, ompT**
NT55	*gad, lpfA, iss, mchF, iroBCDEN, iutA*, ompT**
NT65	*gad, lpfA*
NT67*	*fep, iutA, virK*
C32	*cea, chuA, clbB, focCG, fyuA, gad, hra, iha, iroBCDEN, irp2, iss, iucABCD, iutA, kpsE, kpsMII_K5, mchBCF, mcmA, ompT, papA_F48, papC, pic, sat, sfaD, sitABCDE, tcpC, traT, usp, vat, yfcV*
V30*	*fep, iutA*
11V41	ORF3, ORF4, *aafABCD, aap, aar, aatA, aggR, astA, capU, fyuA, gad, irp2, kpsM, lpfA, pet, pic, traT*
A113*	T6SS, v*asX*
A114*	T6SS, v*asX*
EVAL51	*afaABCD, astA, chuA, eilA, espP, hra, iss, iucABCD, iutA, kpsE, kpsMIII_K98, lpfA, ompT, sitABCDE, traT*
EVAL55	*astA, chuA, espP, gad, hra, iss, iucABCD, iutA, kpsE, kpsM_M15, mcmA, papA_F48, papC, sitABCDE, traT*
EVAL56 (Hg[s])	*cib, gad, lpfA, ompT, traT*
EVAL67 (Hg[s])	*astA, cba, chuA, cia, cma, eilA, hlyF, hra, iss, kpsE, kpsMII_K5, lpfA, ompT, papC, sitABCDE, traT*
EVAL113 (Hg[s])	*astA, gad*
EVAL397	*gad, hra, lpfA, ompT, traT*

Table 4.6 Virulence associated genes detected using VirulenceFinder 2.0.

* denotes the genes listed for particular isolate were identified manually due to limitations of VirulenceFinder 2.0 (Joensen *et al.*, 2014; Tetzschner *et al.*, 2020).

4.3 Discussion

4.3.1 Assembly and Annotation of Isolates Using Different Sequencing Technologies

Unfortunately, due to financial limitations, only a few of the isolates could be sequenced. Of those sequenced, six isolates were sequenced using both short and long-read sequencing technologies for hybrid assembly with UniCycler (Wick *et al.*, 2017). Seven isolates were sequenced with only Illumina short-read sequencing technology and the remaining six isolates were sequenced using Oxford Nanopore long-read technology. The usefulness of hybrid assemblies has been highlighted, allowing greater resolution of chromosomes and plasmid DNA alike, resulting in longer contiguous assembled sections of DNA, and more accurate assembly (Miller *et al.*, 2017). Long-read assembly from Oxford Nanopore yielded some read errors, which caused difficulty in annotation of genomes. Despite this problem, the benefits of better assembly may have outweighed the drawbacks. Bacteria sequenced using Oxford Nanopore technologies yielded 30 or less contigs compared to between 90 and 138 using Illumina short-read.

4.3.2 Sequence Typing and Plasmid Carriage of the Mercury Resistance Transposons

Across the isolates sequenced in this study, it is clear that there are dominant clones which are carrying mercury resistance in the wastewater and effluent but are different to those found within dairy slurry within the same small geographical range of 1 km.

217

Further to this, the importance of performing core genome sequence typing has been highlighted as, it not only compares the six genes normally used in MLST studies, but also the small nucleotide polymorphisms of genes found within the core genome of the species. Whilst the MLST results showed the five wastewater isolates, D1, BPF1-1, BPW2-3, BPW2-4 and NT65 to be the same sequence type (Table 4.2), comparison of the core genomes revealed in fact that isolate BPF1-1 was not related to the other four isolates, differing by over 40,000 sequence changes (Table 4.3). The use of cgMLST allows high resolution identification of sequences by comparing over 1,000 genes, but cgMLST can identify if it is likely two isolates are part of a clonal outbreak (Zhou *et al.*, 2017, 2020). In epidemiological uses of this software, isolates with less than ten allele differences identified when using cgMLST are likely from the same clonal outbreak. It was possible to test this by using the long-read sequence data from isolate BPW2-3 and compare the hybrid assemblies using the short-read data of the other related isolates. Before the comparisons were made, the short-read assemblies were used in the generation of the phylograms (Figure 4.9 and Figure 4.10). The phylogram (Figure 4.9) shows isolates D1, BPW2-3 and BPW2-4 to have identical genetic structures, confirming that the Tn*21*-like mobile element is likely to be carried on the same plasmid. The genome assemblies showed the isolates BPW2-4 and NT65 to possess the same IncFIB 156 kb plasmid as BPW2-3.

The IncFIB plasmid itself contained a Tn*21*-like mobile element within an IS*26* insertion sequence. The contigs from the Illumina short-read only data of the isolates D1, BPW2-4 and NT65 showed the Tn*21*-like transposon to be present. Isolate NT65 was isolated from the same location downstream of the wastewater treatment plant

as BPW2-3, BPW2-4 and BPF1-1 had been, but four months later, suggesting the presence of a persistent clone within the water environment just downstream of the wastewater effluent pipe or within the wastewater treatment plant. Another possibility could be a biofilm may have been present in the effluent pipe of the wastewater treatment plant or further upstream of the sample site, feeding into the brook itself. BPF1-1 was found in the exit site of effluent pipe to the brook in which BPW2-3, BPW2-4 and NT65 were collected. It was also possible to confirm that these isolates had the same plasmid structure. IS*26* is known to form chains of IS*26* elements by causing duplications of the IS*26* transposase at each end of the element. This gives it the potential to reorganise the structure of plasmids by mobilisation and insertion of the IS*26* element in a different locus of the plasmid (He *et al.*, 2015). IS*26* has also recently been noted as clinically relevant due to its role in the evolution and transmission of *bla$_{NDM}$* genes throughout various environments (Zhao *et al.*, 2021). Isolate D1 however did not have the same size IncFIB plasmid as isolates BPW2-3, BPW2-4 and NT65, but it was similar. This was partly expected as it was isolated from a ditch contaminated with grey water (Figure 2.1, location D) nearly 400 m from the sample point downstream of the wastewater treatment plant (location F). The mapping of the plasmids in Figure 4.4 showed that not only was the plasmid shorter, but it had regions with between 50 and 99.9% identity to the IncFIB plasmid of BPW2-3.

4.3.3   Genetic Structure of Sequenced Tn*21*-Like Elements

The biggest difference between the mercury resistance transposons sequenced in this study was in the carriage of the Tn*402*-like module containing a class I integron (Figure 1.5). The model for evolution of Tn*21* is that Tn*402* inserted between two ORFs (*tnpM* and *urf2*) in an ancestor transposon to give the core structure of Tn*21* (Figure 1.12) (Liebert *et al.,* 1999). Structures similar to that of an ancestor were identified in this study across multiple sample locations and dates and in different genera. Two ancestral transposons to Tn*21* stand out. The first of which is a Tn*5060*, isolated from a *Pseudomonas* spp. strain in Siberian permafrost and the other is Tn*5075* isolated from cerebrospinal fluid in 1931 before the antibiotic era (Figure 4.11) (Essa *et al.*, 2003; Kholodii *et al.*, 2003). The latter of the two transposons was shown to have >99.6% identity to Tn*21* without the In2 region. The prevalence of these structures may suggest that the class I integron within the Tn*21* is either still readily mobile or that other mercury resistance transposons may be widely disseminated. However, the annotated genetic structures of these HgR Tn*21* variants/ancestors lack any evidence of the *tnpM* ORF, which was initially thought to modulate Tn*21* transposition. The lack of this ORF but the presence of *urf2* within the identified structures challenges the previous model of Tn*21* evolution (Figure 1.12). A study by (Essa *et al.*, 2003) hypothesized from evidence from sequences of transposable elements related to Tn*21* that the integron region (In2) within Tn*21* inserted to an unknown gene *urf2M* forming two ORFs, *urf2* and *tnpM*, suggesting it was unlikely that the protein encoding *tnpM* to be a modulator of Tn*21* transposition.
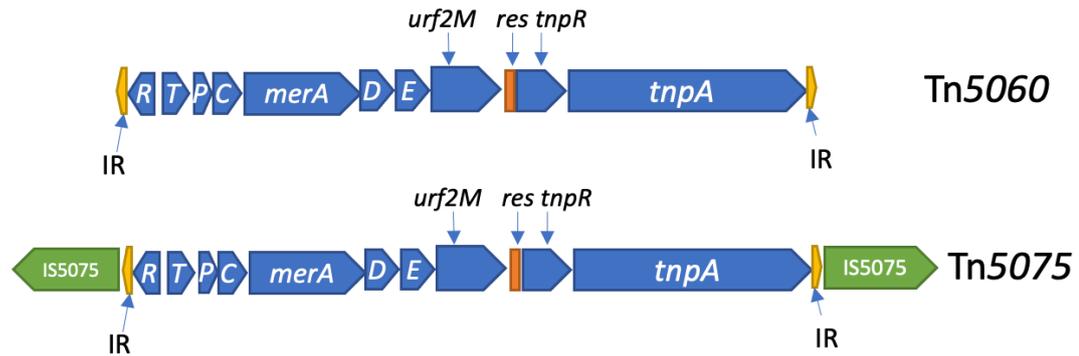
Figure 4.11 Evolutionary ancestors to Tn*21* Tn*5060* and Tn*5075*.

Both transposable elements are examples of possible ancestors to the Tn*21* transposable element displaying a fusion of the *urf2* and *tnpM* ORFs before the insertion of In2. Not to scale.

4.3.3.1 Transposase Gene Configuration and Variation

Seven of the sequenced Tn*21*-like transposon-carrying isolates had either different or truncated transposase regions. Five *mer* transposons contained an IS*26* transposase in the place of the *tnpA* gene normally seen in Tn*3/21* family transposable elements. In the case of four of these isolates (D1, BPW2-3, BPW2-4 and NT65), the IS*26* transposase had inserted to replace *tnpA* and had formed a chain of IS*26* composite transposons, suggesting that the IS*26* element had been present in this structure for some time (Iida *et al.*, 1984). In contrast, BPF1-1, which also had a similar structure of Tn*21*-like element to those of D1, BPW2-3, BPW2-4 and NT65 was not contained in a chain of IS*26* suggesting this had more recently integrated to the IncFIB plasmid than the other four isolates described. The truncation or loss of *tnpA* suggests that it may be redundant due to low transposition frequency in comparison to plasmid transfer frequency and stability of large plasmids such as IncF plasmids. Both the Tn*21* and the IS*26* transposition mechanisms are RecA–

independent which form cointegrates which may explain the production of this structure.

Further to this, the mercury resistance found in the two *Aeromonas* isolates A113 and A114 possessed truncated *tnpA* genes. It is hypothesised that mobility of this element would therefore have been lost, further work to attempt to mobilise this transposon is needed to confirm this. This is particularly interesting in terms of Tn*21* evolution as the mobile elements in which these were identified, were located on the chromosome and are now stuck there.

4.3.3.2   Cassette Array Variation

As previously mentioned, not all the isolates sequenced in this study possessed a class I integron within the mercury resistance transposons they carried (Figure 4.5). All of the isolates containing a class I integron (6) had similarities. Isolates NT50 and NT55 possessed the same cgST and very similar class I integron cassette array structures. Both contained gene cassettes for three different aminoglycoside genes (APH6-I, APH3'-I and APH3"-1) sulphonamide resistance (*sul1*) and beta-lactamase *bla*$_{TEM-1B}$. Where the integron content differed between NT50 and NT55 was the presence of trimethoprim resistance (folate reductase pathway). Both possessed a gene encoding an alternate folate reductase (*dfrA*) providing resistance to trimethoprim. However, NT50 possessed *dfrA5* and NT55 possessed *dfrA14.* Both ORFs of the *dfrA* genes are 474 base pairs in length. However, there are 52 bp differences in the dfrA gene sequences between *dfrA5* and *dfrA14*. In this case it is likely that given both isolates have identical cgMLST sequences, the *dfrA* gene

cassettes were recently captured. However, this may have taken place in the gut or urinary tract, in the wastewater treatment process or in the Black Brook, downstream of the effluent outlet before isolation (Table 4.3, Figure 4.5).

The other isolate from wastewater possessing a class I integron was C32. This differed from the isolates NT50 and NT55 in cgST but also in gene cassette array contents (Figure 4.5). Most notably however was that bla_{TEM-1} had inserted outside of the integron itself, causing deletion of *tnpM* and any resolvase *res* sites, which are used for cointegrate resolution. This suggests that the bla_{TEM-1} gene was inserted by an IS element or another mobile element. The cassette array in the C32 Tn*21*-like element possessed partial resistance to quaternary ammonium compounds (*qacEΔ1*), trimethoprim resistance (*dfrA1*), sulphonamide resistance (*sul2*) and three aminoglycoside resistance genes (ANT3"-Ia, APH6'-Ic and APH3"-I). Together, the three isolates (NT50, NT55 and C32) are particularly interesting as they each possess multiple aminoglycoside resistance genes. Some of these aminoglycoside resistance genes overlap in their function; for example, in C32, ANT3"-1 and APH3"-I, both provide resistance to streptomycin. However, the presence of the three genes provides additional resistance to kanamycin, neomycin, paramycin and spectinomycin (Shaw *et al.*, 1993). The fact that multiple isolates have captured multiple aminoglycoside resistance genes suggests there may be a selective advantage to possess multiple aminoglycoside resistance genes.

The Tn*21*-like elements in the EVAL51 and EVAL55 strains also carried multiple aminoglycoside resistance genes within their integron cassette arrays (Figure 4.5). These two isolates also possessed genes capable of providing resistance to extended

spectrum beta-lactams such as oxacillinase (*bla$_{OXA-1}$*), first identified in the Tn*21*-like transposon of pMG101 (McHugh *et al.*, 1975). Similar to the *bla$_{TEM-1}$* gene cassettes in the integrons of NT50, NT55 and C32, the *bla$_{OXA-1}$* gene cassettes in EVAL51 and EVAL 55 possessed their own individual promoter regions for the ORF. This is particularly interesting as gene cassettes often rely on expression from the Pc promoter upstream of the *attI* site of the class I integron which is modulated heavily by the SOS response. However, it is also possible to see gene cassettes with their own promoters too (Bissonnette *et al.*, 1991).

Bacterial isolate EVAL397 was the only Tn*21* carrying isolate containing a class I integron with a cassette array markedly similar to that of the 'classic Tn*21*' described by Liebert, *et al*. (1999) (Figure 4.12). This isolate was also the only sequenced isolate to possess orf5, thought to be a puromycin resistance gene, and IS*1326*, an insertion sequence. The cassette array described in the 'classic Tn*21*' may still be of relevance today. Given that the isolate was detected in a dairy slurry tank, it is likely that a selection event to keep this mobile element may still be occurring in the bovine gut. This is supported by the EVAL-FARMS study, which showed that the selection events from 3rd and 4th generation cephalosporins, tetracycline and aminoglycosides to disseminate multi-drug resistance genes within a dairy slurry tank are likely in the gut of the cow rather than the slurry tank itself (Baker *et al.*, 2021).
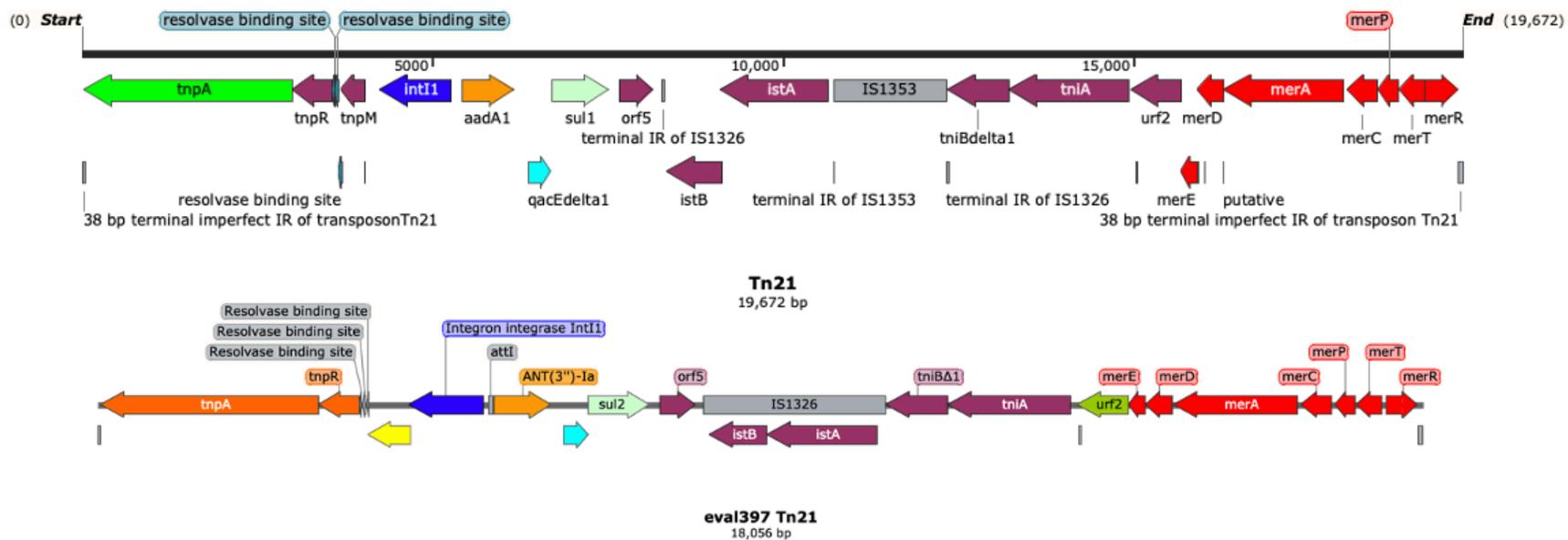
Figure 4.12 Schematics of Tn*21* from R100 (Top) and EVAL397 (Bottom).

The two mobile elements are almost identical despite isolation almost 60 years apart. EVAL397 differs with the lack of IS*1353*.

Although the sample size is small, the variation of gene cassettes within the cassette arrays between sequenced isolates does not appear to be particularly diverse, but this may alternatively suggest that successful variants of Tn*21* have disseminated between humans and animals. Further to this, within both of these environments it is possible that the main stressors are constantly feeding into the slurry tanks, muck heaps or wastewater treatment plants or points of isolation resulting in the selection of dominant gene cassettes. However, quinolones are not used on this dairy unit and beta lactams are known to degrade quickly (Braschi *et al.*, 2013).

Bacterial isolates from the wastewater treatment plant and the effluent may have been selected for at sub-lethal concentrations of sulphonamides. Unpublished data (Garduno Jimenez) analysing the wastewater influents at the campus influent, village influent and effluent for antimicrobial contents showed there to be 1,590 ng L^{-1} tetracycline and 1,166 ng L^{-1} sulphapyridine (a sulphonamide derivative) present. Both concentrations of antimicrobials were reduced by ten-fold across the wastewater treatment process to the effluent. It is possible therefore that whilst there is limited diversity in the cassette arrays noted that sulphonamides could therefore co-select for the whole Tn*21*.

One study into class I integrons in soil showed that cassette dissemination was a slow process, inter sample diversity of gene cassettes between areas in close proximity is reduced however across larger distances inter sample cassette diversity increased, whilst maintaining some common gene cassettes despite the larger distance (Ghaly *et al.*, 2019). Similarly, a study into the wastewater treatment plants feeding into the

226

Thames basin showed a dominance of certain resistance gene cassettes throughout multiple sample sites. All of the diverse cassettes were then being fed into the wider environment of the entire Thames basin, which in turn would have great impact on biological diversity due to humans (Amos *et al.*, 2015). In a smaller study such as this one where the faecal material in the wastewater treatment plant is from a mobile, and international human population, the impact on the antimicrobial resistance gene diversity may still be large as the techniques used to isolate resistance harbouring bacteria are very specific. Using metagenomic sequencing on water samples collected from the same sample locations, sampling sites further downstream and other larger municipal wastewater sites around the same area fed by larger populations may also help identify trends in persistent gene cassettes carried by bacteria.

Whilst the diversity of the cassette arrays in this study may be limited, comparing the sample set to the wider cassette array population from published Tn*21s* (Table 8.1), shows that whilst Tn*21* has disseminated worldwide very successfully, it still retains its ability to capture antibiotic resistance gene cassettes. This may be what drives its success rather than the carriage of mercury resistance genes.

4.3.4 Co-occurrence of Tn*7/pco/sil* in Mercury Resistant Strains

Expansion of this element into chromosomes may have arisen due to a selective advantage, or due to low cost of retaining such a large mobile element (33 kb). Interestingly, the discovery of more and more isolates containing this Tn*7/pco/sil* located between *yhiM* and *yhiN* calls into question whether previous work

227

determining the Tn*7* resolution mechanism and TnsD recognition site are truly understood (Peters and Craig, 2001; Hooton *et al.*, 2021). The transposition of Tn7/*sil*/*pco* into the genomic region between *yhiM* and *yhiN* is consistent with other reports that this Tn*7* element can insert into the chromosome of other *E. coli* strains at non-canonical *att*Tn*7* sites (Chalmers *et al.,* 2018; Hooton *et al.,* 2021). Transposition of the Tn*7* element onto the chromosome resulted in the host strain gaining resistance genes to silver and copper in the form of *silESRCFBAGP* and *pcoEABCDRS*, which form with the Tn*7* transposition machinery - a recognised, likely ancient, and widely distributed, mobile genetic element (Hao *et al*., 2015; Hobman and Crossman, 2015; Staehlin *et al*., 2015). The use of this *att*Tn*7* site for transposition target sequence recognition suggests that chromosomal insertion occurred via the TnsABC+D rather than the TnsABC+E system. Whilst both mechanisms are able to induce transposition to the chromosome, TnsABC+D promotes insertion by 'cut and paste' mechanism to the specific attTn*7* sites (5'-CCCGC-3') which is duplicated and normally 23 bases from the C-terminal end of the *glmS* gene (Waddell and Craig, 1989; Bainton *et al.,* 1993; Mitra *et al*, 2010). In the case of pMG101-B the same *att*Tn*7* site is observed between the *yhiN* and *M* genes in the *E. coli* J53 chromosome (Hooton *et al.*, 2021)*.* Tn*7* and some other Tn*7*-like elements undergo chromosomal insertion by inserting downstream (3') of *glmS* within the host chromosome, as it is considered a safe site that does not interfere with host cell function and is therefore more likely to propagate to the daughter cells (Peters and Craig, 2001).

From the genome assemblies 6 isolates contain Tn*7/pco/sil*, the mobile element has inserted using this mechanism between *yhiN* and *yhiM*. Chromosomal insertion of Tn*7* is reported to be less frequent than transposition to conjugal plasmids. Experimental work performed on two plasmid archetypes, pMG101-A, a 380 kb metal and multidrug resistance plasmid, and pMG101-B, a smaller version at 156 kb, suggests there may be a reason for chromosomal insertion of such large transposons. pMG101A has lost mobilisation ability due to *orf009*, part of the *tra* genes needed for conjugative plasmid transfer, having an integrative and conjugative element (ICE) inserting in the middle of it and separating it into two parts (Figure 4.13) (Hooton, *et al.*, 2021). Over time it is possible that the loss of conjugative ability of the plasmid may have increased the likelihood of transposition induction by the TnsABC+D mechanism. As a result, to ensure that the Tn*7* element is passed onto daughter cells, spontaneous chromosomal insertion of the Tn*7* occurred (Craig, 1996). This hypothesis agrees with the reported function of TnsC, which may evaluate insertion target sites for *attTn7* acting as a regulator for the TnsD and E proteins. TnsD then binds to a consensus sequence allowing for insertion of the Tn*7* (Hauer and Shapiro, 1984). A similar chain of events has allowed pMG101-B/Tn*7* to integrate at an alternative chromosomal locus. This chromosomal insertion by TnsABC+D may take place more frequently in the environment than previously thought, this is supported by the proportion of sequenced isolates that possessed this Tn*7* element from the wastewater treatment plant.
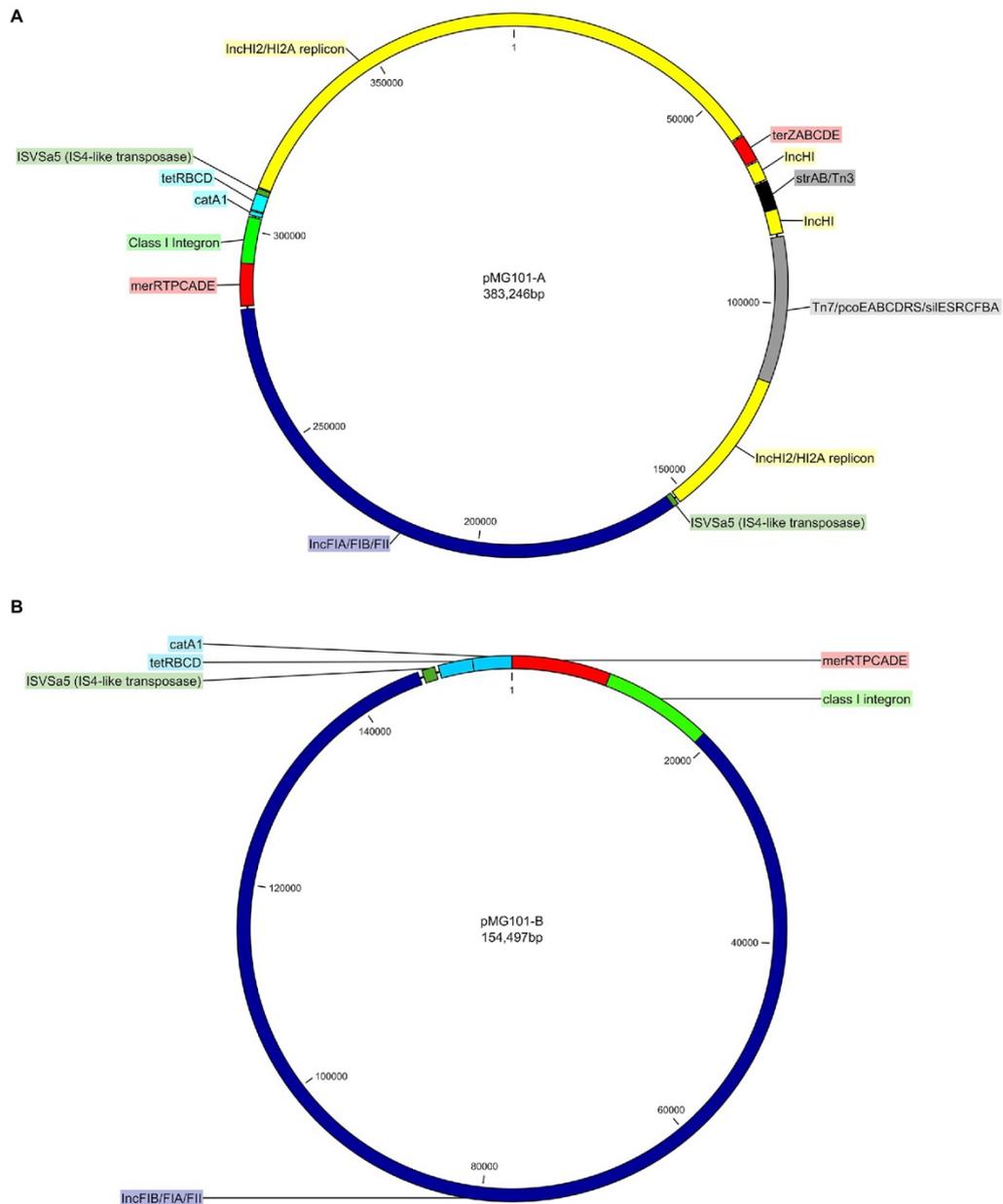
Figure 4.13 Circular genetic map of IncH/IncF plasmid pMG101-A (Top) and pMG101-B (Bottom).

Circular genetic map of IncH/IncF plasmid pMG101-A – the IncHI2/HI2A replicon (yellow), IncFIA/FIB/FII component (blue), Tn7/pco/sil mobile unit (grey), ter and mer metal resistance operons (red), Tn3/strAB antimicrobial resistance gene (ARGs; black), tetBCD/catA1 (light blue), class I integron (light green), and two copies (dark green) of ISVSa5 (IS4-like transposase) are all highlighted. (B) Circular genetic map of IncF plasmid pMG101-B – the IncFIA/FIB/FII backbone of pMG101-B (blue), class I integron (light green), tetBCD/catA1 (light blue), mer operon (red), and a single copy (dark green) of ISVSa5 (IS4-like transposase). (Hooton, *et al.,* 2021). Reproduction of this figure was permitted by Frontiers in Microbiology.

4.3.5   Virulence Factor Carriage in Mercury Resistant Bacteria

All of the isolates in this study were isolated from sites where there was contamination from human or animal waste and were isolated from a small geographic area. As a result, the bacteria collected are likely to carry antimicrobial resistance genes, and genes involved in survival or colonisation of the gut (Table 4.6). Virulence proteins encoded by genes such as *aggR*, *lpfA*, ORF3 and ORF 4 give bacterial cells the ability to aggregate and agglutinate to help colonise gut surfaces (Jordan *et al.*, 2004; Morin *et al.*, 2013). These proteins are advantageous outside of the gut environment too, they help in the production of biofilms to adhere to surfaces such as pipework in a sewage system (Sheikh *et al.*, 2001; Mohamed *et al.*, 2007). This could therefore allow biofilms of such commensal bacteria to form in wastewater treatment sites, dairy slurry tanks and lagoons (Schramm, Schroeder and Jonas, 2020).

For bacteria such as pathogenic *E. coli* or commensal gut bacteria, that are either invasive or that may colonise host organisms respectively, iron chelation and storage is extremely important. 6 of the organisms sequenced possessed at least one mechanism to chelate iron (*fep, ybf*, *iro*, *iut* or *iuc*) often by the production of siderophores (Andrews, Robinson and Rodríguez-Quiñones, 2003). Iron chelators are required for the cells to survive in competitive environments such as a host organism's gut. Their role is to have a higher affinity for iron than the host organism's intrinsic iron chelators. Ferrous and ferric ions are essential for enzymatic processes within both host cells and bacterial cells, like the TCA cycle (Andrews, Robinson and Rodríguez-Quiñones, 2003). Multiple siderophores have been detected in different

genera of microorganisms but have the same overall function. The bacteria isolated in this study were no different. Aerobactin, enterobactin and yersiniabactin were detected alongside *iroBCDEN*, which was first described in *Salmonella enterica* (Sorsa *et al.*, 2003). All of which are classified as pathogenicity factors. These virulence factors have since been found in genomic islands or islets on transmissible plasmids within other strains. This is nothing new, the mobility of virulence genes and genomic islands has been well documented, they have been identified in urinary tract infection causing *E. coli* similarly (Sorsa *et al.*, 2003).

With plasmid located genomic islands containing antimicrobial resistance genes, such as seen in pMG101A, perhaps selection for acquired antimicrobial resistance genes which are often carried on transmissible plasmids may inadvertently co-select for more virulent strains to arise (Hooton, *et al.,* 2021). This is also seen in this study, for example, the large IncF plasmids carried in isolates NT50 and NT55 harboured the Tn*21*-like mobile genetic elements and siderophore *iroBCDEN* and aerobactin synthesis pathways (Figure 4.8).

Alongside the carriage of these aggregative genes and iron chelating systems are also extracellular proteases (such as *espP*) (Table 4.6). Such extracellular proteases are often associated with UTI causing pathogens (Foxman *et al.*, 1995). Finally, genes encoding proteins which are secreted extracellularly have also been shown to cause virulence. In this study, *kpsE* and *kpsM* are good examples, both been identified to allow the production of group II capsules (Pavelka, Wright and Silver, 1991). The identification of all these proteases, capsular proteins and virulence factors described may suggest that mercury resistance transposons from both wastewater

treatment plants and dairy slurry could be linked to the dissemination of pathogens into the environment beyond where these organisms were isolated from.

4.3.6   Evolution of Tn*21*

Since the discovery of the Tn*21* mobile element, several hypotheses have been postulated on how it evolved. The most commonly accepted of which is that a Tn*402*-like element, containing a class I integron, inserted into a genetic element 'Tn*21Δ*' containing transposase genes *tnpMRA* and *merRTPCADE* in a hypothetical protein *urf2M* (Figure 1.12). Two other IS elements, IS*1326* and IS*1353* then inserted into the Tn*402* module of the Tn*21* element forming the model Tn*21*. From this structure, Tn*21* may go on to then capture and excise more gene cassettes using IntI1 integrase. Amongst the isolates sequenced from the wastewater and dairy slurry, a range of different Tn*21* structures have been identified. In some isolates, expansion of the cassette array has been shown with carriage of multiple aminoglycoside resistance cassettes (such as ANT3' and APH6-I). Similar to the cassette array identified within pMG101, *bla*$_{OXA-1}$ has been documented within the gene cassette array of bacteria isolated from dairy slurry. Despite the loss of the 3' *attI* site in the Tn*21*, the mobile element can still capture new gene cassettes, providing a selective benefit for keeping the Tn*21*.

Tn*21* may also be evolving to lose obsolete DNA, by spontaneous deletion from replication error or the insertion and excision of insertion sequences. As previously described, EVAL51 carries a Tn*21* but *merTPCADE* was lost. The 3' terminal of the *merR* gene forms part of the 38 base IR of the Tn*21* and is essential for mobilisation.

233

Loss of the *mer* operon within this Tn21-like genetic element suggests that the slurry tank environment in which this strain was isolated from may not drive selection to keep the *mer* operon as there is not enough bioavailable mercury (II) to require the expression of the *mer* operon that would otherwise kill cells. A similar structure to the Tn*21* in EVAL51 was previously isolated in the slurry tank three years before, suggesting the mobile element may be mobile within the slurry or the gut of dairy cattle (Ibrahim *et al.*, 2016).

Isolates EVAL51 and EVAL55 were collected from the same sample. Tn21 sequences from these isolates were flanked by IS*91* carrying *floR* (Figure 4.7). IS*91*-like elements are known to be associated with the formation of complex class I integrons and may mobilise DNA upstream of the IS*91*-like sequence causing 'genetic slippage' (Tavakoli *et al.*, 2000). Expansion of Tn*21* and/ or the fusion of Tn*21* with a mobile element such as IS*91* may allow for easier dissemination of the mobile element (Mendiola *et al.*, 1994).

The *mer* operons in all except three isolates (A113, A114 and C32) were borne on plasmids (Table 4.5). 14 of which were carried on IncFIB, IncHI and IncI1 plasmids. Whilst the *mer* operons and transposase genes were identified on these plasmids (Figure 4.5), there were many isolates that did not carry the In2 section of the transposon, but possessed no *tnpM*, thought to modulate Tn*21* transposition or resolvase binding sites which allow for resolution of Tn*21* cointegrates upon mobilisation (Hyde and Tu, 1985). Isolate NT7 may have lost the In2 region causing the *tnpM* and resolvase sites to be deleted too. Further to this, the isolates containing a similar sequence to NT7 such as: D1, BPW2-3, BPW2-4 and NT65 have an IS*26*-like

transposase in place of the *tnpA* gene of Tn*21*. This may be another evolutionary break away from the Tn*21* discussed by Liebert, *et al* (1999). Diversification in transposable elements is expected as they disseminated into the environment, the possibilities of interactions with other genetic elements is therefore endless.

Tn*21* is a large enough mobile element where there are many ways in which evolution may cause diversification of its descendants, such as insertions to the *mer* operon from insertion sequences, alongside deletion of the *mer* operon all together. Similarly, isolates described in this study have shown that the transposase genes are no longer necessary to dissemination due to their carriage on large self-transmissible plasmids. However, the most likely mechanism in which diversification of Tn*21* and Tn*21*-like elements will be through the action of IntI1-integrase and the gene cassette array. Diversity in the cassette arrays within the class I integron has been identified across the strains published on the NCBI database (Table 8.1), but also within the isolates sampled and sequenced in this study. The diversity in cassette arrays across the Tn*21* and Tn*21*-like transposons suggests that the presence of the integron is likely the reason for the retention of the *mer* operon in the absence of mercury selection. This is supported by the structure of Tn*21* in EVAL51 (Figure 4.7) where the *mer* operon has been lost.

Isolates from wastewater not only possessed a varied integron cassette array but there was also co-occurrence of another Tn*3*-like element which carried resistance genes for tetracycline (*tetAR*) flanking the Tn*21*-like transposons, and the Tn*21*-like element was found within a plasmid which is supported by the co-occurrence data (3.2.5). The presence of the two genetic elements adjacent to each other on a self-

transmissible plasmid may allow for the co-selection of the whole plasmid rather than co-selection causing the survival of certain resistance genes. If the dairy slurry tank contents caused tetracycline presence to become a selective pressure, the plasmid containing the tetracycline resistance genes would likely be retained by a cell rather than just the genes encoding tetracycline efflux. Data showed that the wastewater influent had relatively high concentrations of tetracycline in it. Up to 1,560 ng L^{-1} was detected in the influent and a ten-fold reduction was noted in the effluent (Garduno Jimenez, unpublished). As Tn*21* is located adjacent tetracycline resistance genes, tetracycline may be a selective pressure in the wastewater environment and may co-select for the carriage of Tn*21* large transmissible IncF, IncI or IncH plasmids. In aquatic biofilms, the minimum selective concentration of tetracycline is known to be approximately 10 μg L^{-1} (Lundström *et al.*, 2016).

The phylogeny work in Figure 4.9 shows not only the diversity of the Tn*21*-like elements, but also how different variants of the transposon have diverged and disseminated into the environment. Tn21-variants that are included in the phylogram have been identified in nearly every continent. In each geographical location different cassette arrays have been identified within these Tn*21*-like elements, perhaps there may be local bias on gene cassettes within cassette arrays depending on the selective demands of each environment.

4.4   Future work

Sequencing work in this study was limited due to financial constraints. As a result, only 20 of the mercury resistant isolates were sequenced leaving many other equally interesting isolates to be characterised. Therefore, sequencing of more of the isolates from the collection may also help with understanding more about the evolution of Tn*21* in the wastewater environment and the surrounding area. More sequencing data would also allow a better insight into the differences seen across the treatment process itself. As mentioned in the Discussion, short-read and long-read sequencing platforms have their drawbacks. Hybrid assembly of those isolates already sequenced may allow for better assembly and annotation of accessory genes and mobile elements such as Tn*21*. A large amount of data was collected in this chapter, hybrid sequencing of some of these isolates could be used to resolve plasmid sizes and structures. The data could then be used to perform larger comparisons to plasmids containing Tn*21* from the NCBI database.

Metagenome sequencing would greatly help the understanding of changes occurring throughout the treatment process and may help identify resistance genes and other mobile elements that were not investigated in this study. Perhaps identification of selection events may also be observed. By sampling upstream, downstream and at the point of wastewater effluent dispersion into the Black Brook it would be possible to identify whether the wastewater affects the diversity of antimicrobial resistance genes within the brook.

Given that Tn7/*sil*/*pco* appears quite frequently in bacterial sequences of many of the mercury resistant isolates collected, it may be beneficial to screen the collection

of isolates for resistance to Ag (I) and Cu (II). Screening phenotypically and by PCR for the presence of either the resistance genes or Tn*7* transposase genes may help identify a link between the carriage of Tn*7* and Tn*21*. It is possible that large IncF and IncHI plasmids may be responsible for the carriage of both mobile genetic elements? The work performed in Hooton *et al*. (2021) may support this idea.

Targeted metagenomics identifying class I integrons and class I integrons of Tn*21*-like mobile genetic elements may help gain greater understanding of cassette array diversity within the wastewater environment. Potentially this may also help identify differences between general gene cassette carriage and cassette carriage in Tn*21*.

Chapter 5: Assessment of Integron Gene Cassette Carriage in Tn*21*, Tn*21*-like Mobile Elements and the Organisms Which May Carry Them

5.1 Introduction

Integrons have been documented globally, in bacteria from soil, water, wastewater and humans (Gillings, 2014; Amos *et al*., 2015; Ghaly *et al.,* 2019). Whilst integrons readily capture gene cassettes, they may capture and excise them within the life cycle of a single cell. Analyses of whole genome sequences of bacteria and metagenomes that contain integrons has identified that approximately 65% of gene cassettes found within integrons encode polypeptides with no known functions or homologues in DNA or protein databases, a further 15% of the ORFs have identity to hypothetical proteins (Stokes *et al.*, 2001; Boucher *et al.*, 2007). Within class I integrons carried by Tn*21* and Tn*21*-like elements, the gene cassettes identified in previously published works often confer resistance to antibiotics (Table 8. 8.1). Similarly, in Chapter 4, the integrons carried in the Tn*21*-like elements in sequenced isolates contained antibiotic resistance gene cassettes but no gene cassettes encoding hypothetical proteins. In this way, the class I integron found within Tn*21* and Tn*21*-like elements has a rather unique characteristic compared to other class I integron gene cassette arrays (1.8). The loss of the 3' region of the *attC* site due to a *sul1* cassette insertion and subsequent 66 bp deletion of the 3' end of the *qacE* gene means this gene is no longer mobile (Liebert *et al*., 1999). However, downstream of this *sul1* insertion, the end of the cassette array can also be variable. Downstream of this, an ORF known as orf5 is typically present which encodes a protein thought to confer puromycin resistance; but in some cases, a second ORF, known as orf6 has also been documented downstream of orf5 (Partridge *et al.*, 2001).

Gene cassettes carried by class I integrons in Tn*21* or Tn*21*-like mobile genetic elements have been identified within wastewater environments in both influent and effluent of a wastewater treatment plant (Amos *et al.,* 2015). Previously, Gillings *et al*. (2015) suggested that identification of cassettes and class I integrons could be used as a proxy to detect anthropogenic pollution, in environments such as wastewater effluent or industrial waste effluent. By analysing cassette contents in heavily polluted areas, it can help also identify any selective pressures the resultant pollution may be placing on the microbial population.

In addition to identifying the contents of class I integron gene cassette arrays of Tn*21* and Tn*21*-like organisms it would be beneficial to identify which taxa of microorganisms from environmental samples are harbouring these mobile elements. Novel techniques have recently been developed to aid identification of the taxa to which genes or accessory elements, such as plasmids belong to, within microbial environmental samples. One such technique is epicPCR, which utilises entrapment of single bacterial cells from an environmental sample in poly-acrylamide beads and emulsifying the beads, before PCR to amplify a gene of interest and fuse it to a housekeeping gene from the same cell (Spencer *et al*., 2016) (Figure 5.9). The amplicons from the environmental cells within the emulsion are then extracted and sequenced, taxonomy of cells carrying the genes of interest can then be assessed. Larger scale techniques for metagenomics also exist such as Hi-C (High-throughput chromosomal confirmation capture) (Belton *et al.*, 2012). This method is a high through-put technique where chromatin is crosslinked with biotin-labelled nucleotides, in a digestion and re-ligation of DNA. This results in DNA such as plasmid

DNA being covalently bonded to host cell chromosomal DNA. Upon sequencing, ligated products will allow the identification of which bacterial host possesses the crosslinked DNA. This is something traditional shotgun metagenomic approaches are unable to do.

For Tn*21*, the use of epicPCR could uncover more information than isolation and culturing techniques to help identify which bacteria may be carrying such an element within a particular environment.

5.1.1 Aims and Objectives

This chapter describes metagenomic experiments aimed at gaining more information of what variable DNA is carried by Tn*21*-like elements, and which organisms may carry Tn*21* and Tn*21*-like elements, within the wastewater environment.

1. Firstly, amplification and sequencing of class I integron gene cassette arrays of Tn*21* and Tn*21*-like mobile elements from metagenomic DNA of wastewater will be used to aid in the understanding of gene cassette turnover across the wastewater treatment process.

2. The information gathered from the targeted amplicon sequencing of wastewater metagenomic DNA samples in this study may also help identify whether the presence of an immobilised integron within Tn*21* may be the reason for the persistence of Tn*21* within the wastewater environment.

The development of epicPCR has been identified as a means to target the amplification of DNA in cells containing a functional gene of interest in certain environments whilst retaining which taxonomic group the cell with the gene belongs

242

to (Spencer *et al.*, 2016). epicPCR has been used predominantly alongside short-read sequencing techniques to identify functional genes and their taxa (Spencer *et al.*, 2016; Hultman *et al.*, 2018; Hall *et al.*, 2020).

3. In this chapter, epicPCR using the Oxford Nanopore long-read sequencing platform was explored, in order to gain better resolution of 16S genes to give more specific details regarding which organisms may be carrying Tn*21* and Tn*21*-like mobile elements.

4. This chapter also discusses the limitations of using techniques such as epicPCR, as well as other novel metagenomic techniques currently in use, which may be viable for wider use with further refinement.

5.2 Methods and Results

Firstly, class I integrons associated with Tn*21* and Tn*21*-like mobile elements from wastewater influents and effluents were amplified. These amplicons were sequenced and assembled to identify gene cassettes carried by these class I integrons. Secondly, epicPCR was performed on cells extracted from the same wastewater samples.

5.2.1 Sequencing and Analysis of Tn*21*-like Integrons and Their Gene Cassette Arrays in Wastewater Metagenomes

Metagenomic DNA samples from wastewater influents (campus influent and village influent) and effluent, were sampled across two time points and gene cassette contents of class I integrons belonging to Tn*21* and Tn*21*-like transposons were specifically amplified. The primers used were *tnpM*_F as a forward primer, to target Tn*21* specific sequences, and pooled reverse primers *qacE*_R, targeting the 3' CS region of Tn*21* class I integron, and HS286_R to target the 3' *attC* element of any other gene cassettes to capture non-*qac* containing integrons (Table 2.3 and Figure 5.1). The amplicons were then purified and sequenced using Illumina Nextera XT (by DeepSeq) (2.2.24, Figure 5.2). Assembly of the reads produced was performed using MEGAHIT, then SPAdes (2.2.29). The number of amplicon reads and the number of bases generated in the metagenomics across all six samples was evenly distributed. However, there was large variation between the assembled outputs for each sample (Table 5.1). On first examination, these data would suggest that there may be more

diversity in the influent sample from the village and effluent term time sample because of the increased number of contigs per number of amplicon reads.
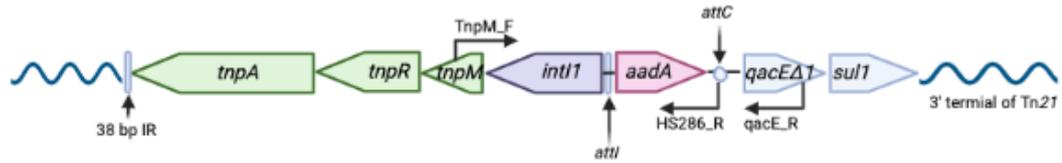
Figure 5.1 The transposase terminal of Tn*21* and the primer binding sites for amplification of class I integrons from Tn*21* and Tn*21*-like transposons.

*TnpM*_F amplifies target DNA from midway through *tnpM*. HS286_R amplifies the 3' terminal of gene cassette *attC* sites in the direction towards *tnpM*. *qacE*_R amplifies the 3' terminal of *qacEΔ1* in the same direction. If the gene cassette array contains more gene cassettes than shown in the diagram, then these cassettes will also be amplified.
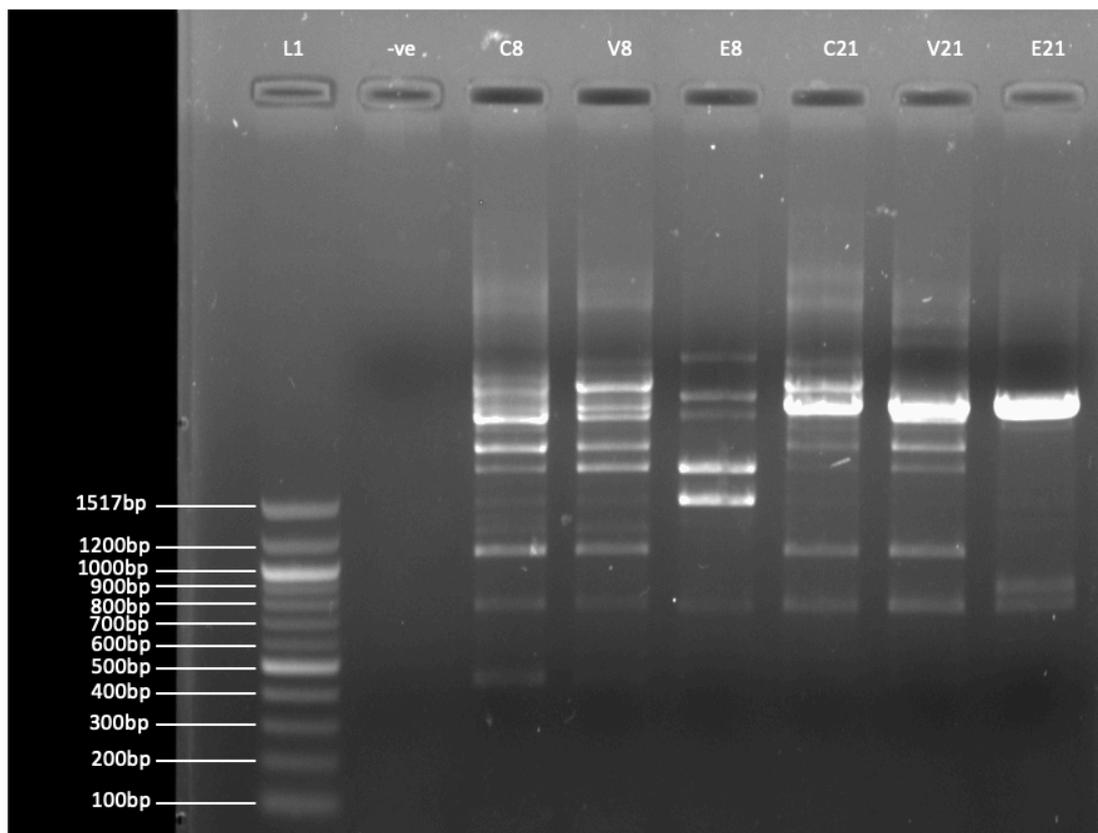
Figure 5.2 Ethidium Bromide 2.0% agarose gel image of amplified Tn*21* and *Tn21*-like class I integrons.

Lane L1 contains a 100 bp DNA ladder (New England BioLabs), -ve contains nuclease-free water. Lanes C8, V8, E8, C21, V21 and E21 contain amplified integron DNA of wastewater samples taken from campus influent pre-term time, village influent pre-term time, effluent pre-term time, campus influent during term time, village influent during term time and effluent during term time, respectively.

Sample	Number of Reads	Number of Bases	Contigs
Campus influent pre-term time	337,658	169,504,316	605
Village influent pre-term time	275,846	138,474,692	544
Effluent pre-term time	285,683	143,412,866	514
Campus influent term time	316,589	158,927,678	833
Village influent term time	340,025	170,692,550	2,338
Effluent term time	313,979	157,617,458	1,250

Table 5.1 Sequencing yields from Illumina MiSeq and number of assembled contigs from the assembly method described.

5.2.1.1   Identification of *attC* Sites of Tn*21*-like Class I Integrons in Wastewater

Firstly, contigs from the assembled class I integron amplicons were annotated using RASTtk and then further analysed manually to identify gene cassettes and *intI1*. RASTtk is a fully automated process which identifies protein-encoding genes within assembled DNA and assigns their function (Aziz *et al.*, 2008). The annotations allowed quick and easy identification of resistance gene cassettes. However, many gene cassettes present may contain hypothetical or unknown ORFs. HattCI was also used to identify gene cassettes by analysing the 3' *attC* site (Table 5.2) (Pereira *et al.*, 2016). The HattCI output was varied across the samples, some of the proposed 3' *attC* sites were part of open reading frames and *attI* sites are not detected using this software. There were very few *attC* sites detected in all samples, but this is to be expected given that previously isolated sequences of gene cassette arrays from Tn*21* and Tn*21*-like mobile elements are not particularly variable.

Sample	HattCI number of cassettes identified
Campus influent pre-term time	20
Village influent pre-term time	20
Effluent pre-term time	8
Campus influent term time	31
Village influent term time	92
Effluent term time	11

Table 5.2 HattCI output detailing the number of *attC* sites identified in each wastewater sample.

5.2.1.2   Gene Cassette Origins: Using *attC* Sites to Estimate Cassette Taxon

The custom script attC-taxa.sh, uses covariance models to predict the bacterial taxa from which the detected gene cassettes may have originated, based on the folding patterns of the *attC* sites (Ghaly, Tetu and Gillings, 2021). The output of this script listed each gene cassette location and proposed taxa of origin (Table 5.3). However, due to the currently limited dataset, some *attC* sites which were identified were not able to be classified. In turn, some of the *attC* sites may be false positives where their DNA sequence has similarities to an *attC* structure but is not an *attC*, further narrowing down the number of true cassettes identified. Similarly, some *attC* sites present in different contigs may be duplications of each other due to assembly from thousands of reads.

Sample	Number of hits HattCI	Taxon	Number of *attC*s
Campus influent pre-term time	20	*Xanthomonadales*	3
Village influent pre-term time	20	*Xanthomonadales*	3
Effluent pre-term time	8	*Xanthomonadales*	1
Campus influent term time	31	*Xanthomonadales*	5
Village influent term time	92	*Xanthomonadales*	35
Effluent term time	11	*Xanthomonadales*	2

Table 5.3 Taxa of identified *attC*s in each metagenomic sample when running attC-taxa.

Total number of *attC*s is less than number identified from HattCI as not present in the database.

5.2.1.3 Integron Cassette Array Structure Variation Across Tn*21*-Like Mobile Elements in Wastewater

In all of the six sequenced metagenomic amplicon samples, there was a general trend which matched the classic class I integron structure of Tn*21*. The order of the cassettes from *intI1* to 3' terminal of the gene cassette array was *aadA*, *qacEΔ1* and *sul1* (Figures 5.3 – 5.8). However, the presence of additional folate reduction pathway genes in the cassette array was also found in all but one sample: effluent term time (Table 5.4). Notably, the uptake or diversity of folate resistance cassettes within influent samples was greater than in the effluent samples. *dfrA12* was found more often than any of the other *dfrA* genes. In contrast to the diversity in the presence of different *dfr* gene alleles, there was very little variation in the *aadA* genes (Table 5.4).

Sample	*dfrA* Genes	*aadA* Genes	*sul* Genes	Other Resistance Genes	Other Genes
Campus influent pre-term time	*dfrA1* *dfrA7* *dfrA12* *dfrA17*	*ant(3'')-Ia*	*sul2*	*qacEΔ1* *bla$_{OXA-1}$* *arr-3*	DUF1010 HAD-hydrolase
Village influent pre-term time	*dfrA* *dfrA1* *dfrA12* *dfrA17*	*ant(3'')-Ia* *aac(6')-Ib-cr* *aadA2*	*sul2*	*qacEΔ1* *bla$_{OXA-1}$* *catB* *arr-3*	DUF1010
Effluent pre-term time	*dfrA5*	*ant(3'')-Ia*	-	*qacEΔ1*	IS*26 tnpA*
Campus influent term time	*dfrA* *dfrA12* *dfrA17* *dfrB3*	*ant(3'')-Ia*	*sul1*	*qacEΔ1*	DUF1010 *estX* *serB*
Village influent term time	*dfrA* *dfrA1* *dfrA5* *dfrA12*	*ant(3'')-Ia* *aadA2*	*sul2*	*qacEΔ1*	DUF1010
Effluent term time	-	*ant(3'')-Ia* *aadA2*	*sul2*	*qacEΔ1*	-

Table 5.4 Gene cassettes isolated from class I integrons of Tn*21*-like transposable elements from wastewater.

Breakdown of any genes detected in verified gene cassettes within a gene cassette or in assembled integrons.

Owing to the assembly pathway described in the methods (2.2.29.2), it was possible to reconstruct some full and some partial integron gene cassette arrays for each sample from the short-read sequence data. Coverage of assembled gene cassette arrays varied within and between samples. Whilst some of the coverage may have been due to frequency of integron occurrence, it is also likely that PCR efficiency may also have impacted amplification of some Tn*21* cassette array regions over others (Figure 5.3 – 5.8). The HS286_R primer, is a universal primer which amplifies *attC* sites, however, it may still cause bias towards amplification of some gene cassettes over others.

250
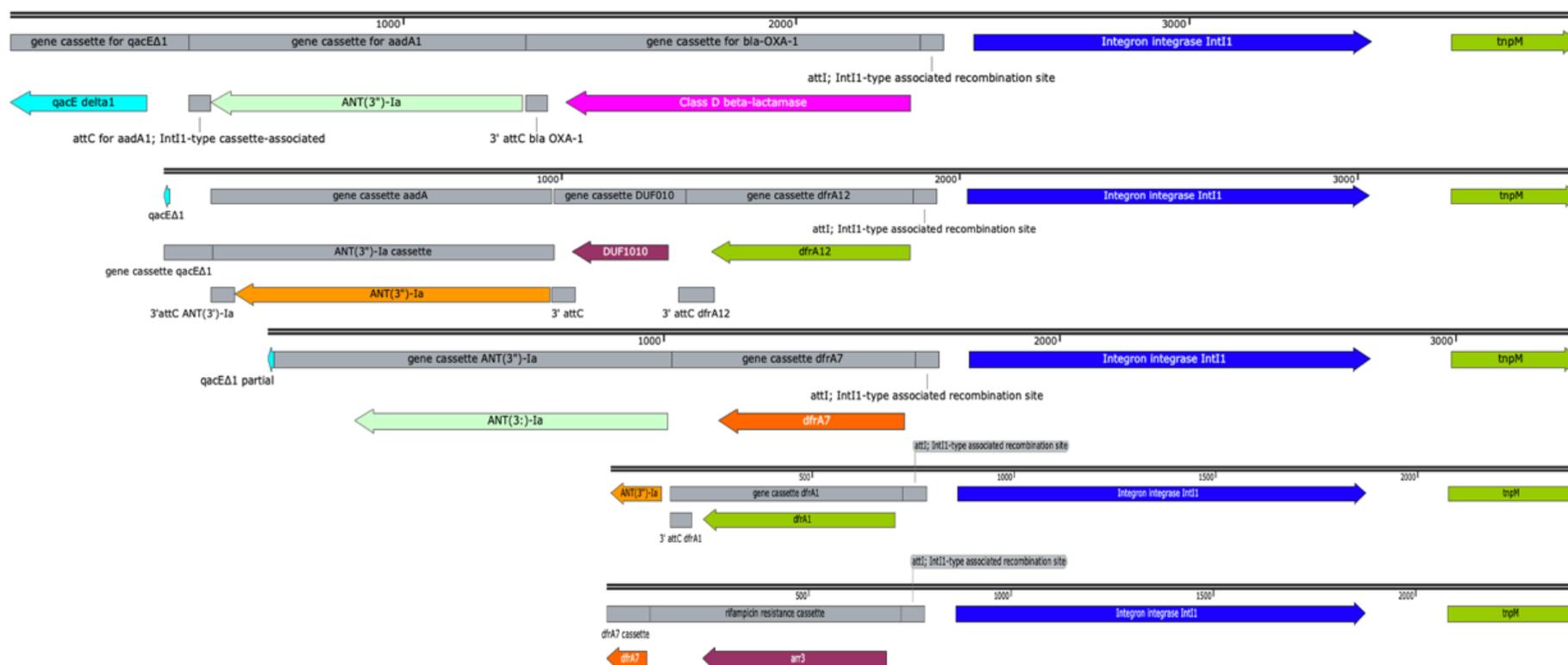
Figure 5.3 Assembled integron cassette arrays from wastewater sample campus influent pre-term time.

Annotations for each integron cassette are displayed below the DNA line to show size. Where possible, integron assemblies have been scaled against each other.
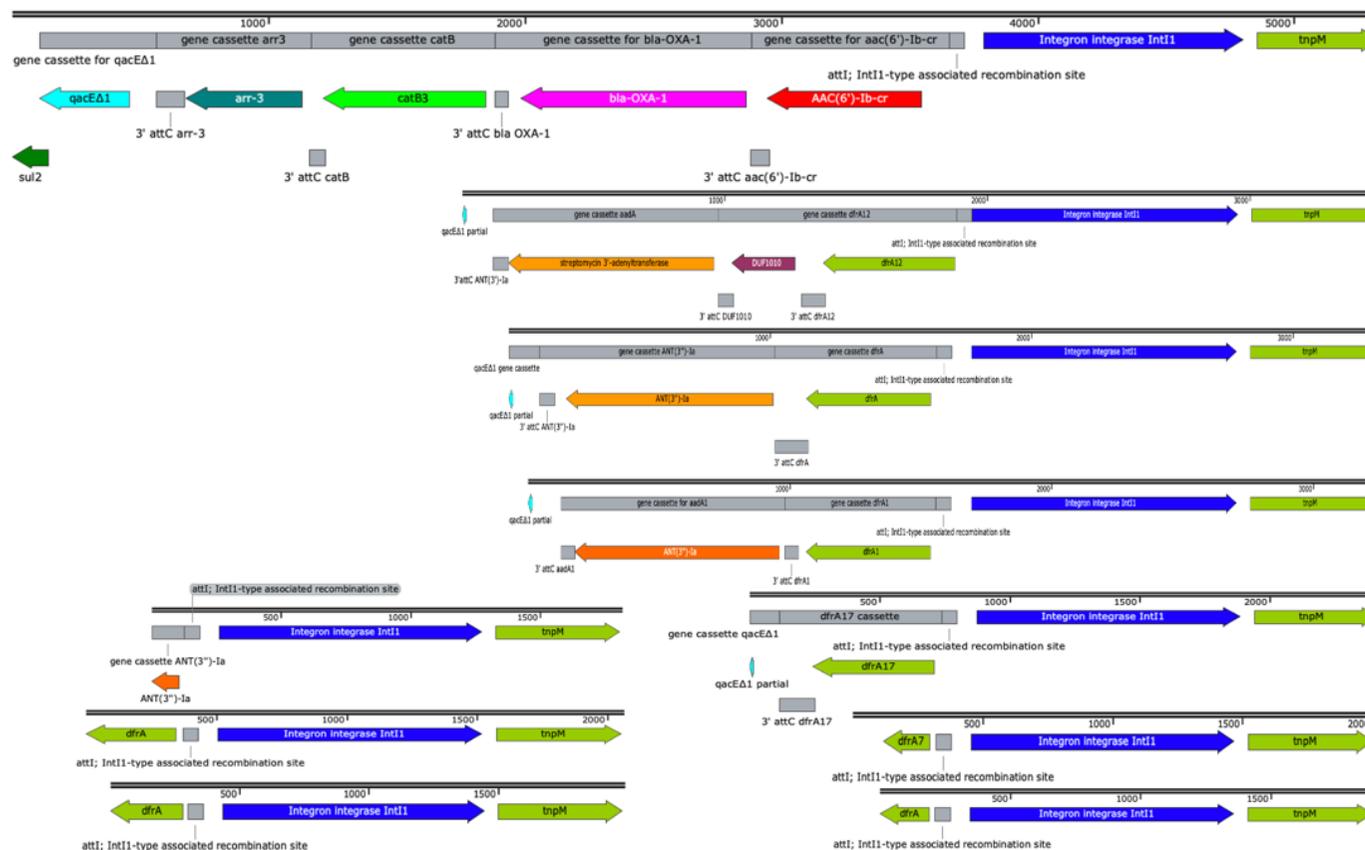
Figure 5.4 Assembled integron cassette arrays from wastewater sample village influent pre-term time.

Annotations for each integron cassette are displayed below the DNA line to show size. Where possible, integron assemblies have been scaled against each other.
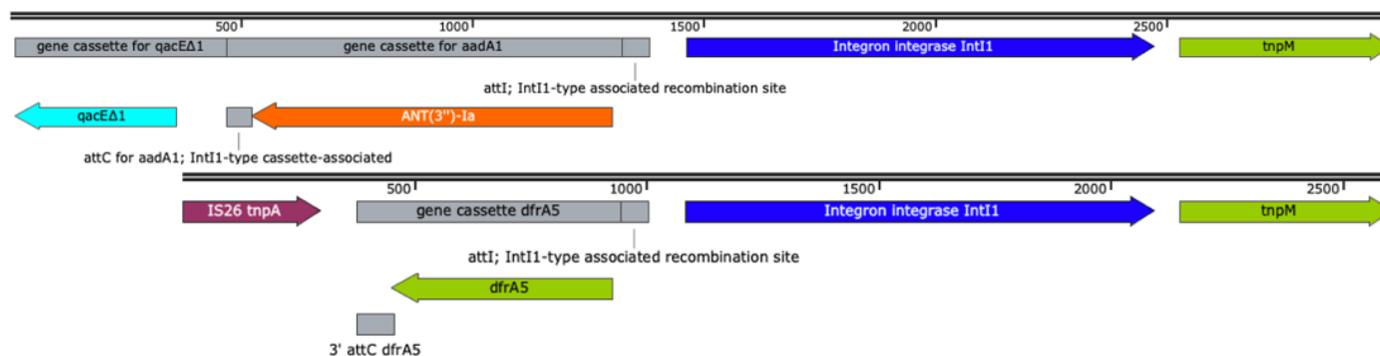
Figure 5.5 Assembled integron cassette arrays from wastewater sample effluent pre-term time.

Annotations for each integron cassette are displayed below the DNA line to show size. Where possible, integron assemblies have been scaled against each other.
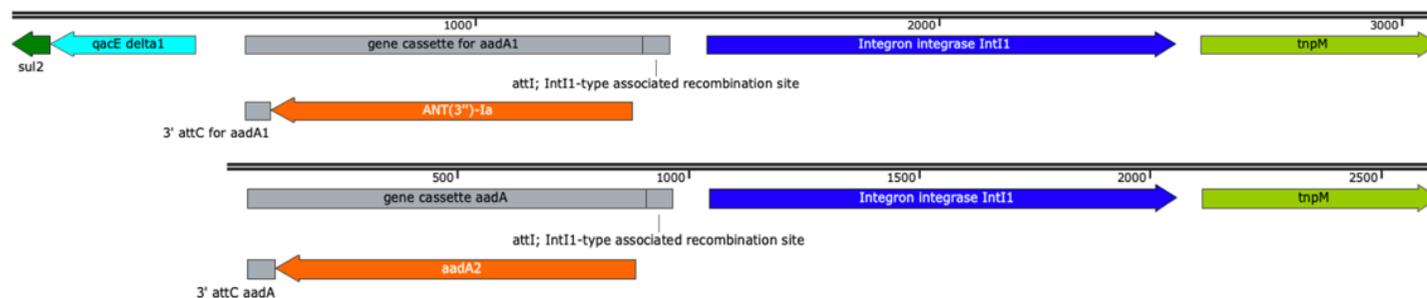
Figure 5.6 Assembled integron cassette arrays from wastewater sample effluent during term time.

Annotations for each integron cassette are displayed below the DNA line to show size. Where possible, integron assemblies have been scaled against each other.
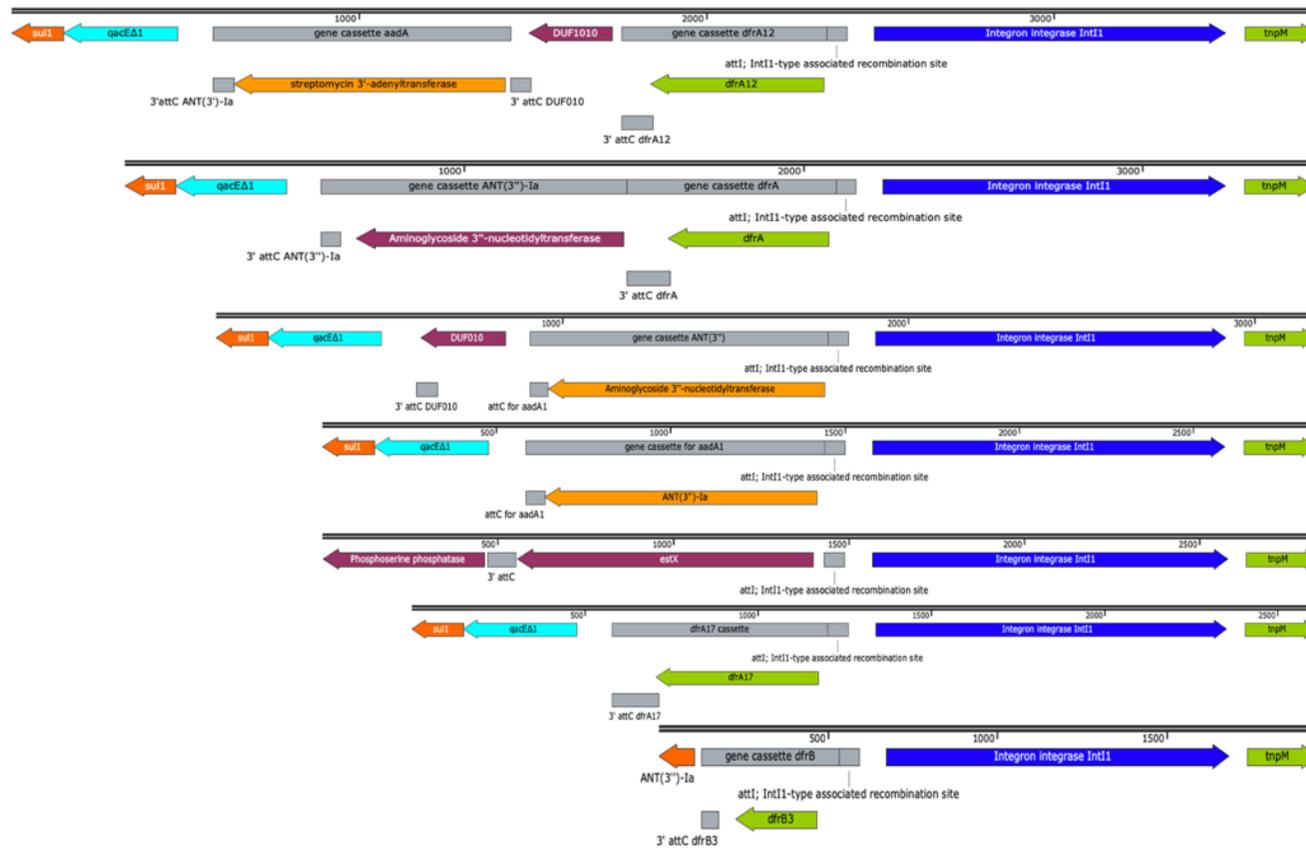
Figure 5.7 Assembled integron cassette arrays from wastewater sample campus influent during term time.

Annotations for each integron cassette are displayed below the DNA line to show size. Where possible, integron assemblies have been scaled against each other.

254
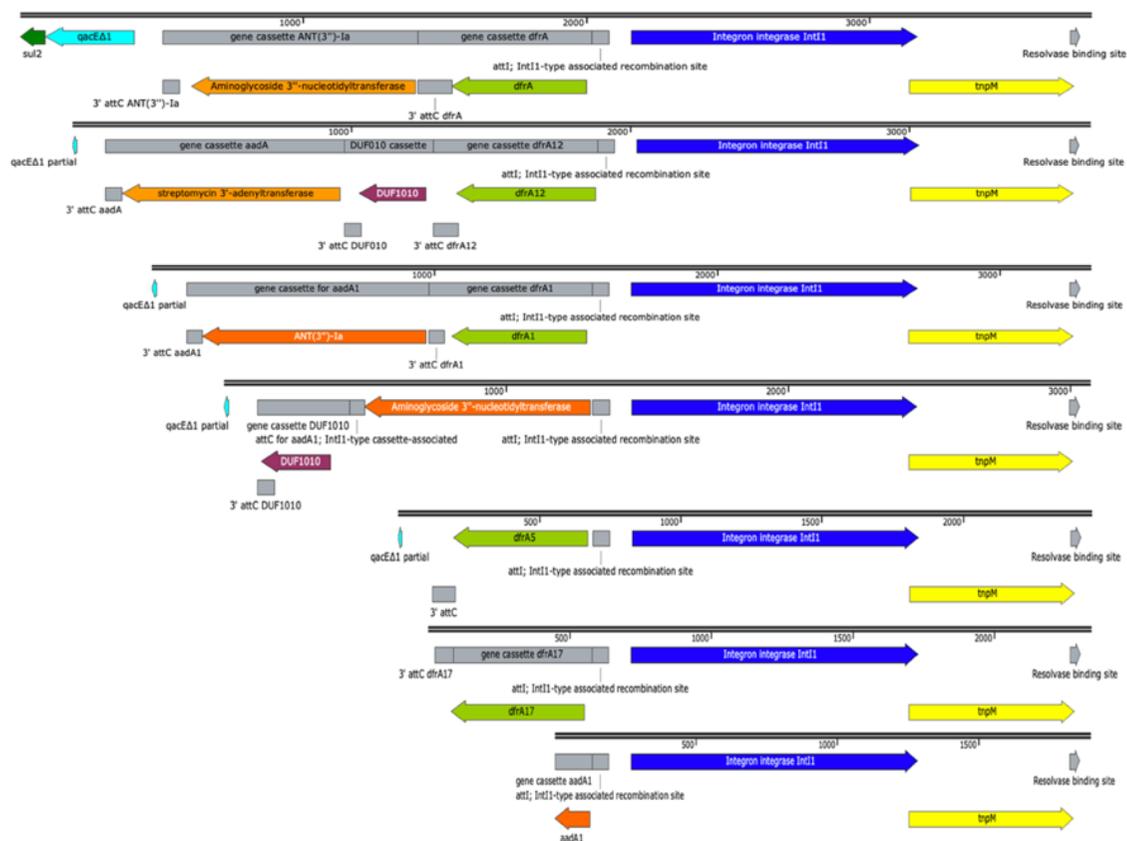
Figure 5.8 Assembled integron cassette arrays from wastewater sample village influent during term time.

Annotations for each integron cassette are displayed below the DNA line to show size. Where possible, integron assemblies have been scaled against each other.

5.2.2   epicPCR

epicPCR or emulsion paired isolation concatenation PCR is a relatively new and ambitious molecular technique, first published in 2016 (Figure 5.9). The technique aims to facilitate mass screening of entrapped single environmental bacterial cells in acrylamide beads for a functional gene and fusing it to a housekeeping gene, the 16S rDNA gene, to allow identification of the genera harbouring the target genes within a sample (Spencer *et al.*, 2016). The short, fused amplified products are then sequenced using short read Illumina Nextera sequencing technology, and the reads are then processed in a bioinformatics pipeline (2.2.24.5) to determine the family to which the organisms carrying such genes belong. Due to the use of Illumina sequencing, finer resolution is not possible as each sequence in epicPCR covers only 300 bp of the 16S rDNA gene.

In this section, epicPCR could represent an opportunity in two ways. The first of which was to identify bacteria from wastewater environments which may harbour Tn*21* and Tn*21*-like transposable elements, by targeting the conserved 38-base-pair IR region of Tn*21* and a section of the Tn*21* transposase gene, *tnpA*. The second was to adapt epicPCR for use with long-read sequencing technology in order to achieve greater host resolution of Tn*21* carrying bacteria by covering a larger region of the 16S rDNA gene.
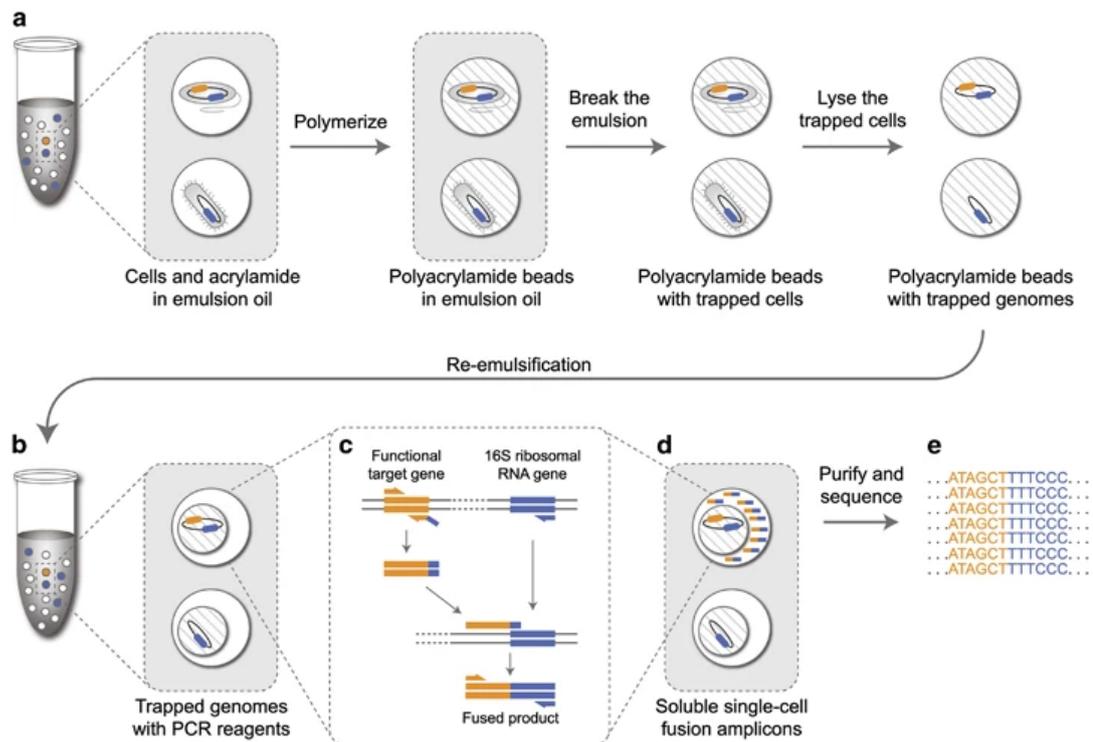
Figure 5.9 epicPCR workflow.

(a) Microbial cells in acrylamide suspension are mixed into emulsion oil. The emulsion droplets are polymerized into polyacrylamide beads containing single cells. The emulsion is broken and the cells in the polyacrylamide beads are treated enzymatically to destroy cell walls, membranes and protein components, and expose the genomic DNA. (b) Polyacrylamide-trapped, permeabilized microbial cells are encapsulated into an emulsion with fusion PCR reagents. (c) Fusion PCR first amplifies a target gene with an overhang of 16S rRNA gene homology. With a limiting concentration of overhang primer, the target gene amplicon will anneal and extend into the 16S rRNA gene, forming a fusion product that continues to amplify from a reverse 16S rRNA gene primer. (d) The fused amplicons only form in the emulsion compartments where a given microbial cell has the target functional gene. (e) After breaking the emulsion, the fused amplicons are prepared for next-generation sequencing. The resulting DNA sequences are concatemers of the target functional gene and the 16S rRNA gene of the same cell. (Reproduced from Spencer *et al.* (2016), permission to reproduce this figure was granted by Springer Nature).

5.2.2.1   Refinement of epicPCR for MinION

epicPCR is a highly complex, multistep process in which there are many steps where

errors can be introduced. Therefore, this can increase the risk of aberrant results

which impairs the depth and quality of results. The original published method

(Spencer *et al.*, 2016) utilised Phusion polymerase (New England BioLabs, USA) for amplification of fused products, blocking and nested PCRs. However, for use with the Oxford Nanopore MinION sequencing system, Q5 and Q5 Hot-Start polymerases (New England BioLabs, USA) were favoured over Phusion polymerase in order to reduce the polymerase error rate due to Q5's higher proof-reading ability compared to Phusion polymerase. Phusion polymerase is at least 50X less error prone than Taq polymerase, but Q5 is at least 100X less error prone than Taq polymerase (New England BioLabs, no date). The use of Q5 polymerase should yield higher quality, more accurate amplicons across the epicPCR workflow before sequencing, in order to combat the slightly increased error rate of long-read sequencing.

The generation of the amplicon fusion products was first tested on metagenomic DNA preps of the wastewater samples that would be used for epicPCR (Figure 5.10 and Figure 5.11). Despite checking primers, the metagenomic DNA samples displayed a large amount of non-specific amplification within the round of fusion PCR without nesting (1542R), whereas fusion PCR using the nested primer 1492R yielded little off targeting. The PCR would then be attempted in the actual acrylamide beads.

Figure 5.10 Expected PCR product from fusion PCR step in the epicPCR process.

During the fusion PCR step of epicPCR, ten-fold less bridging primer is used than the end primers. The unfused amplified product with the bridging primer R1-F2' tail acts as a primer to amplify the housekeeping gene (16S rDNA).
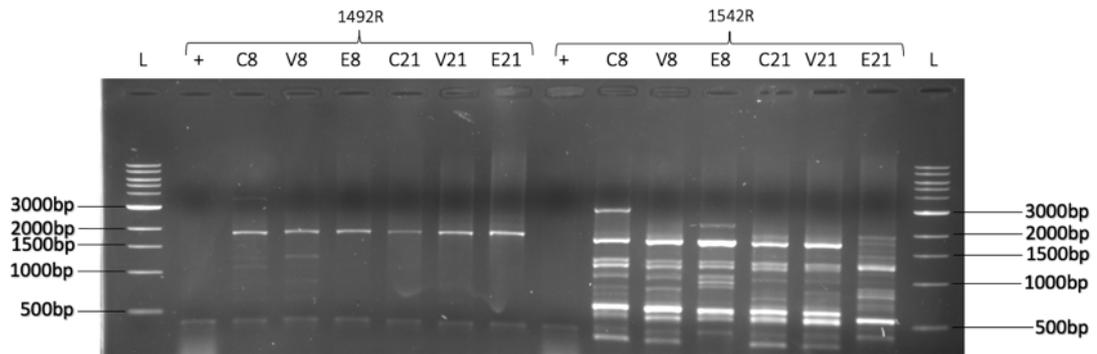
Figure 5.11 epicPCR fusion PCR tests on nested (1492R) and non-nested (1542R).

The desired product sizes in both PCR reactions are 1,898 bp and 1,926 bp for 1492R and 1542R, respectively. The unfused product (Tn*21* marker only) is 392 bp. Both positive controls contained unfused product, fused product was not achieved in these (1,898 and 1,926 bp). The reason is likely due the addition of too little DNA. The unfused product acts as a primer to amplify the 16S rRNA gene, if there is not enough target gene amplified, then the 16S gene is not amplified. The size marker used was 1 kb ladder (New England BioLabs). Lanes C8, V8, E8, C21, V21 and E21 contain amplified integron DNA of wastewater samples taken from campus influent pre-term time, village influent pre-term time, effluent pre-term time, campus influent during term time, village influent during term time and effluent during term time, respectively.

When polyacrylamide beads were formed, fluorescence microscopy was used to ensure cells had been encapsulated by the beads correctly. 1-2 cells per 100 polyacrylamide beads formed is the desirable range. This decreased the chances of multiple cells being captured within one polyacrylamide bead. This in turn would have caused the formation of chimeras of fused products of cells which may not carry the target Tn*21 tnpA* gene (Figure 5.12).
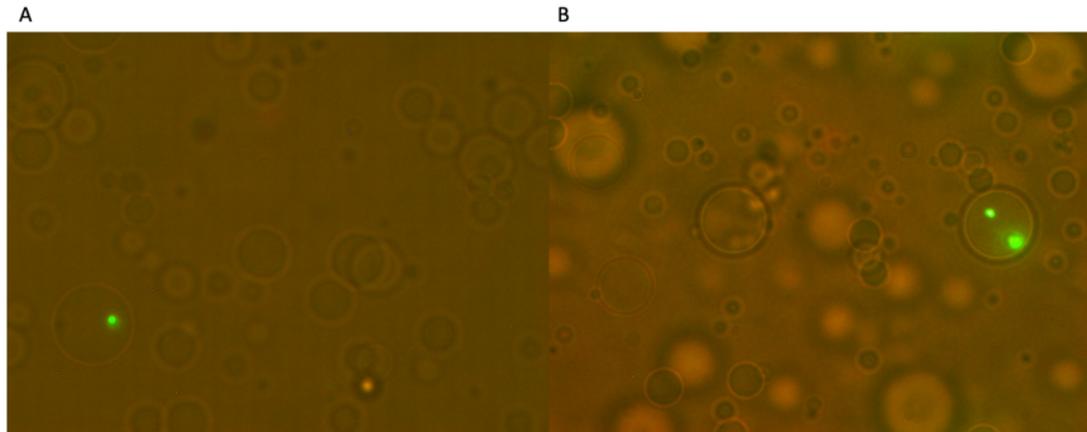
Figure 5.12 SYBR Green Fluorescence Imaging of polyacrylamide beads.

Panel A shows one cell fluorescing within the large number of polyacrylamide beads. Panel B shows multiple cells within one polyacrylamide bead, which is to be avoided and therefore the sample pictured could not be used for epicPCR. Images were taken at 100X focus using oil immersion lens (Axiovert 135 TV, Zeiss).

After the formation of cell encapsulated beads the epicPCR procedure was followed as described in section 2.2.23.4**.** After much time spent refining and repeating steps of the epicPCR process, only one environmental sample, term time effluent (E21) produced fused products and was taken forwards for sequencing (Figure 5.13). Due to financial and time constraints (partly due to the COVID-19 pandemic), the other samples that were meant to be amplified with epicPCR were not sequenced because no fusion products could be generated. Within the sample E21, there were two extra amplicons and non-specific bands of approximately 50-100 bp. To avoid carrying these non-target amplicons forwards to the sequencing run, the quadruplicates were all electrophoresed on a 1% 1X TAE agarose gel and the correct sized band was purified with a Monarch™ DNA gel extraction kit (New England BioLabs). The DNA extracted from E21 was ready to be sequenced alongside the nested PCRs of metagenomic DNA which were performed in parallel.

260
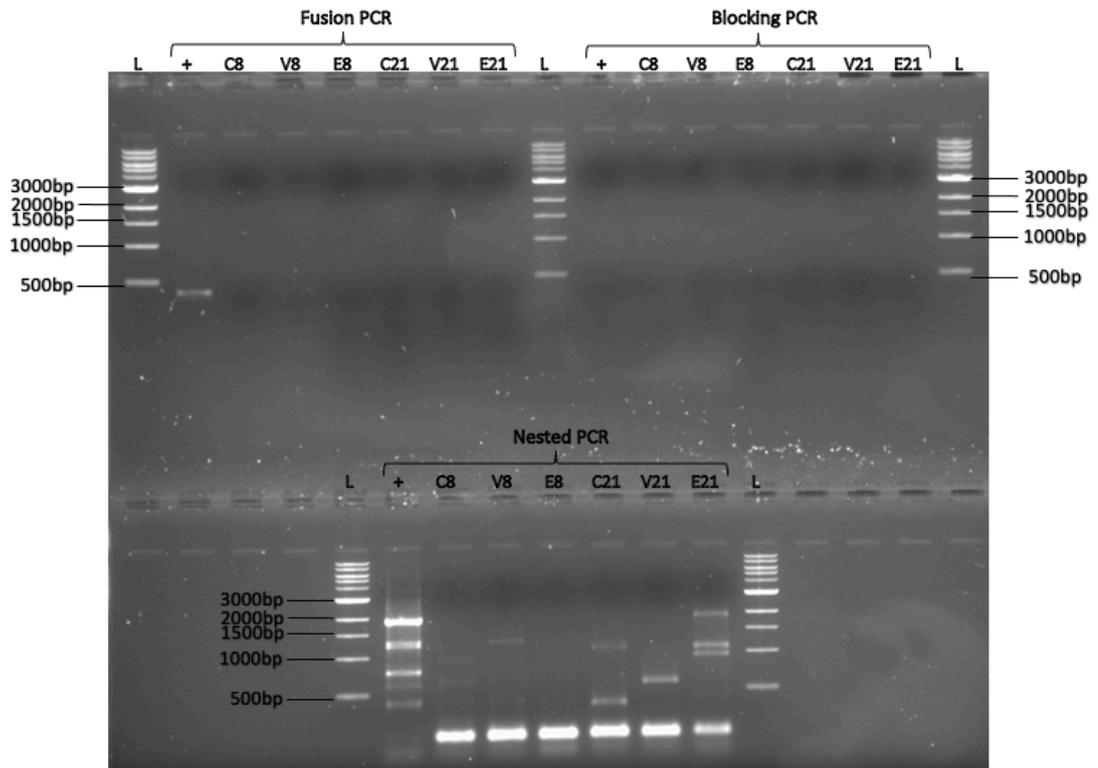
Figure 5.13 Image of a 1% 1X TAE agarose gel containing DNA from each PCR step within the epicPCR procedure.

+ represents a DNA only positive control (*E. coli* strain J53 pMG101) which was run in parallel throughout the epicPCR process. The size marker used was 1 kb ladder (New England BioLabs). Of all the environmental samples only E21 (term time effluent) produced a usable result for long read sequencing. Lanes C8, V8, E8, C21, V21 and E21 contain amplified integron DNA of wastewater samples taken from campus influent pre-term time, village influent pre-term time, effluent pre-term time, campus influent during term time, village influent during term time and effluent during term time, respectively.

For sequencing preparation, barcoding PCR using the four primer PCR protocol (Oxford Nanopore Technologies, 2018), which amplifies the PCR product and attaches sequence barcodes was trialled. However, it caused the DNA libraries to be unusable due to shearing and generation of artefacts (Figure 5.14). Barcoding by ligation was used instead to yield higher quality DNA. The four primer PCR barcoding protocol uses LongAMP polymerase, which has a higher error rate than Q5, so it was in line with the aims to use ligation over PCR for barcoding to prevent unnecessary

introduction of error to the fused amplicons. Thus maintaining the highest base accuracy as possible.
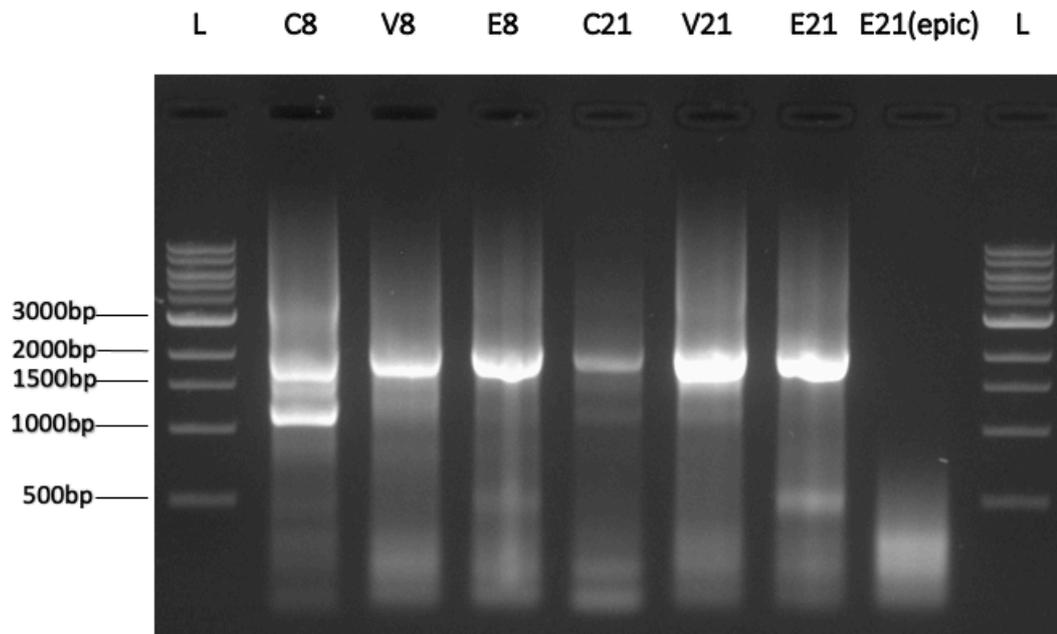
Figure 5.14 Image of a 1% 1X TAE agarose gel containing DNA from each barcoding PCR sample.

The size marker used was 1 kb ladder (New England BioLabs). DNA in each lane was caused to shear, which is shown by the blurring of bands imaged. This rendered the DNA libraries from the PCR barcoding process to be unusable. Lanes C8, V8, E8, C21, V21 and E21 contain amplified integron DNA of wastewater samples taken from campus influent pre-term time, village influent pre-term time, effluent pre-term time, campus influent during term time, village influent during term time and effluent during term time, respectively. E21(epic) refers to the epicPCR fused amplicon sample.

5.2.2.2   epicPCR: Sequencing and Comparison

The prepared DNA was then sequenced using a MinION (Oxford Nanopore Techologies) according to the manufacturer's instructions, and sequenced for 4.5 hrs. This yielded 742.7 Mb across 1,063,718 reads with an average quality score of 10.53 and average read length of 702 bases (Figure 5.15). The mean average read

score was 10.53 and the modal read quality was 8.95. This meant that the sequence

quality met minimum requirements for the experiment to proceed. Despite pooling

each barcoded DNA sample in equimolar concentrations for final loading for DNA

sequencing, the number of reads produced did not reflect this (Table 5.5).
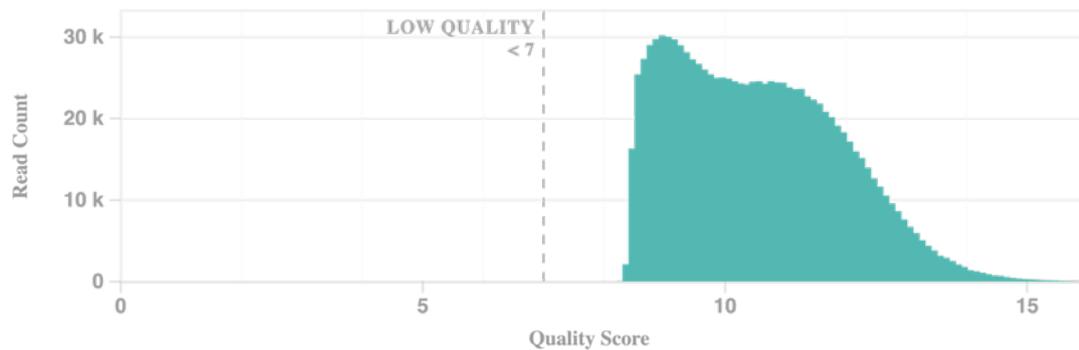
Figure 5.15 Sequencing breakdown of the quality score of each read versus the number of reads.

Read quality scores less than 7 in Nanopore sequencing should be filtered out and not used.

Sample	Number of reads	Mean Quality Score
C8	264,414	10.5
V8	355,231	10.4
E8	84,413	10.6
C21	64,069	10.5
V21	96,733	10.6
E21	53,516	10.6
E21_epicPCR	100,132	11

Table 5.5 Breakdown of read number and quality.

A breakdown of the sequencing quality scores and the number of reads generated by each sample from the sequencing run of the epicPCR sample E21_epicPCR and the non epicPCR parallel samples. These statistics were the raw outputs from the MinION before filtering and processing using Fastq 16S.

The reads were then filtered using cutadapt v3.4 to remove the Tn*21 tnpA* end of the amplicon so that 16S rRNA amplicons could be processed further (Martin, 2011). The trimmed reads were then processed using Oxford Nanopore's cloud-based software EPI2ME in a program called Fastq 16S (Table 5.6). Notably, the epicPCR had the highest percentage of reads that could not be processed by the BLASTN program from the filtering process.

Sample	Number of viable reads post filtering	Percentage of reads lost from filtering
C8	71,754	72.9%
V8	97,941	72.4%
E8	14,685	82.6%
C21	24,844	61.2%
V21	28,836	70.2%
E21	15,407	71.2%
E21_epicPCR	16,651	83.4%

Table 5.6 Breakdown of viable reads after filtering for running through the EPI2ME pipeline Fastq 16S.

The parameters used for filtering were as follows: Minimum read length 700 bases, minimum identity 80%, maximum error value 10^{-10} (conservative).

The trimmed reads from each sample were now viable for taxonomic classification, the epicPCR sample E21_epicPCR had the highest success in determining taxa of the reads, 66.6% of the reads were able to be assigned to a taxonomic group. In comparison, 31.3% of the metagenomic DNA PCR parallel sample E21 was able to be assigned to a taxonomic group, over two-fold difference.

The trimmed reads from the epicPCR were then used to construct phylogenetic trees for each sample. Figure 5.16 displays the genera classifications of reads from the epicPCR sample. Most 16S reads displayed were from Gram-negative bacterial

groups (97.4%), however the presence of the Gram-positive bacterial taxa was also noted; the generation of the phylogenetic trees allowed for quick visual comparison of the taxonomic groups represented to the metagenomic PCR prep of the parallel sample E21 (Figure 5.16 and 5.17). Similarly, to the epicPCR sample of E21, the metagenomic PCR parallel also contained reads from Gram-positive genera. In fact, across all of the other metagenomic PCR samples the same was true (data not shown). Also, of interest across the two parallel samples, epicE21 and E21, was the variation of genera between them. This was particularly noticeable in the Gram-positive taxa in the metagenomic PCR of E21 where nine different genera of Gram-positive organisms were detected, opposed to six in the epicPCR sample (E21epicPCR) (Figure 5.16 and 5.17). In both figures, however, there is rather great variation in the genera present within the samples.
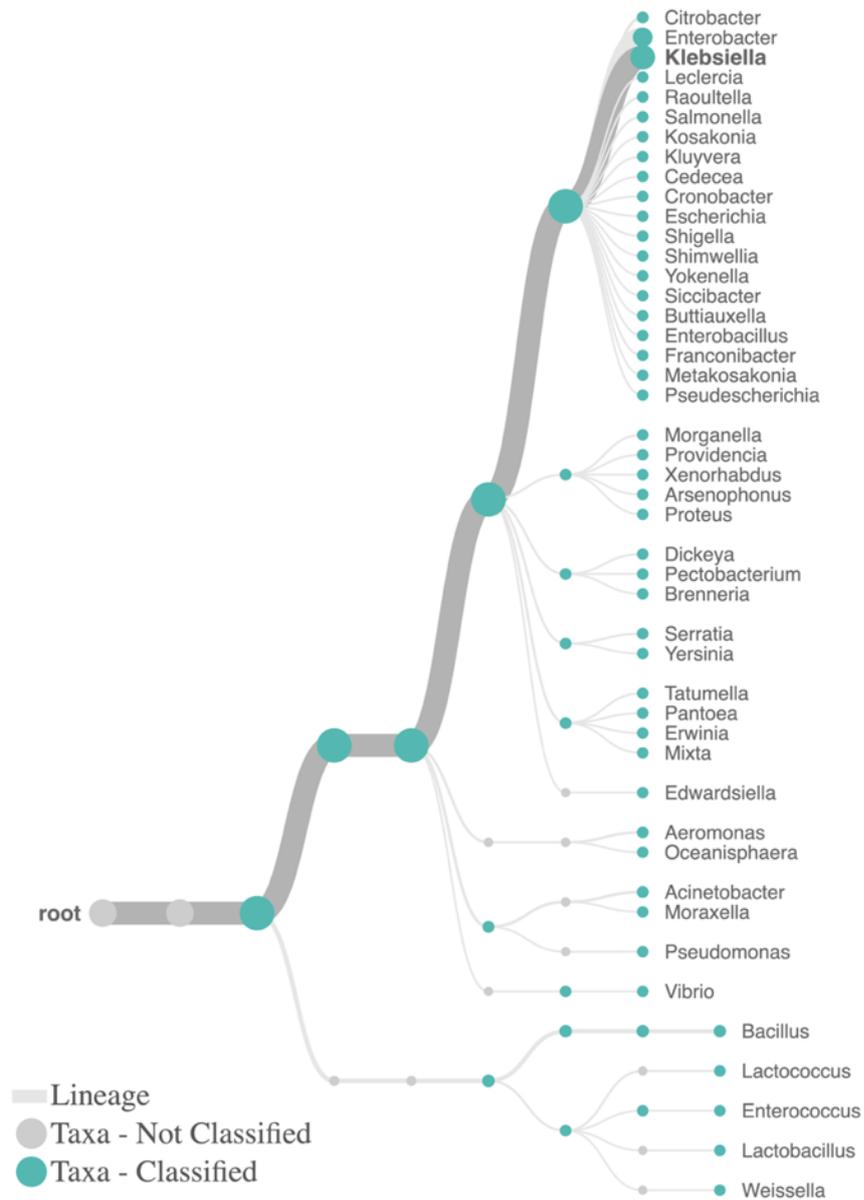
Figure 5.16 Taxonomic breakdown of the most abundant genera in sample E21_epicPCR identified using EPI2ME Fastq 16S.

The line weighting in the phylogenetic tree represents abundance of reads within the lineage. Parameters for BLAST were minimum identity 80%, maximum error value 10^{-10} (conservative).
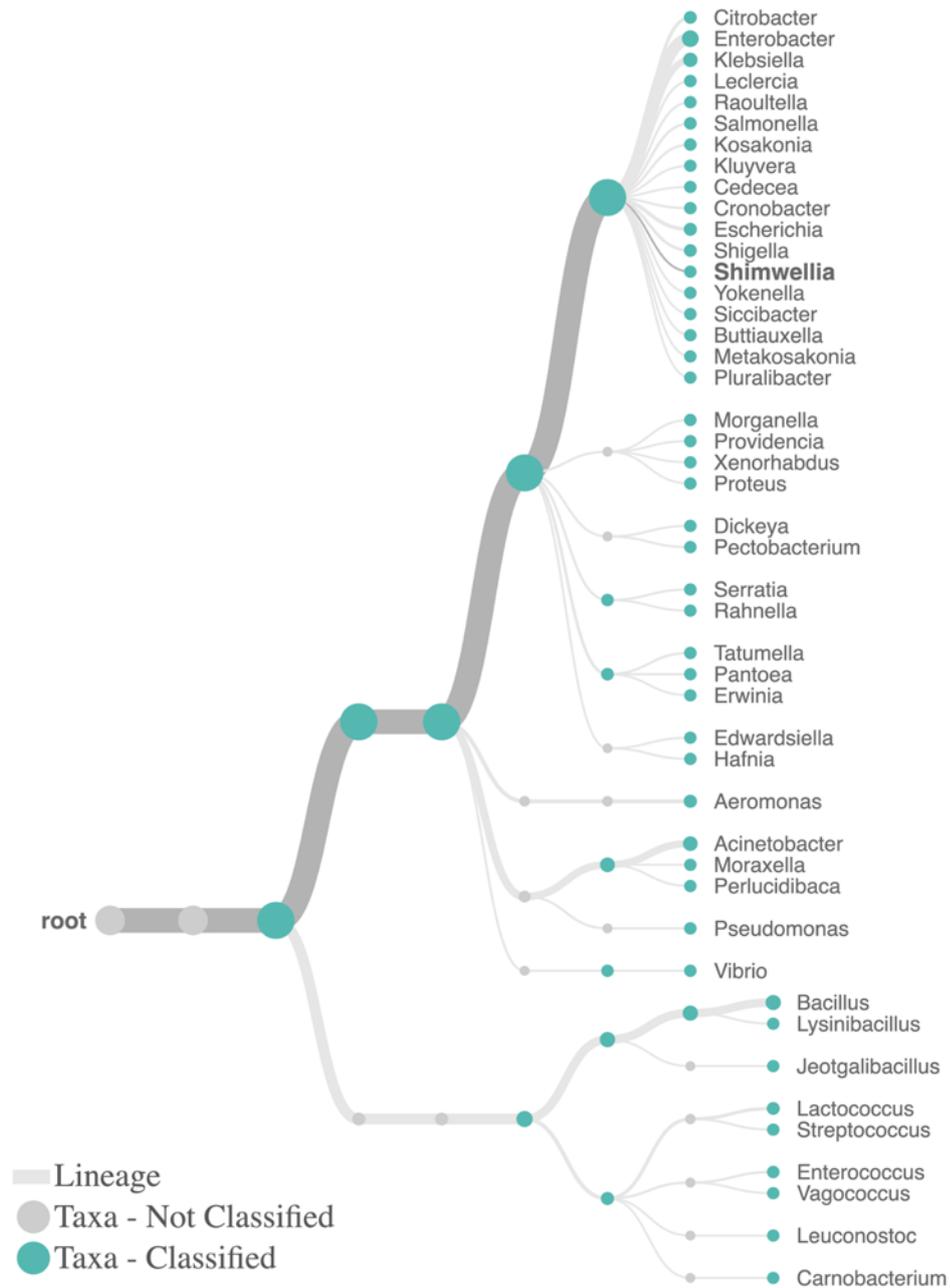
Figure 5.17 Taxonomic breakdown of the most abundant genera in sample E21 identified using EPI2ME Fastq 16S.

The line weighting in the phylogenetic tree represents abundance of reads within the lineage. Parameters for BLAST were minimum identity 80%, maximum error value 10^{-10} (conservative).

Each sequenced sample (both metagenomic PCR samples and the single epicPCR sample) contained similar proportions of bacterial families detected (Figure 5.18). The data in this figure helps display the large variation of bacterial families detected. The proportion of each bacterial family detected varied between samples. Between the epicPCR sample E21 and the metagenomic PCR sample of E21, there was again difference between the two, which was to be expected as epicPCR is discriminate in the production of fusion products from individual cells whereas the parallel metagenomic DNA amplicons will amplify and fuse any Tn*21* product with any 16S rDNA gene within the DNA added to the PCR reaction mixture.
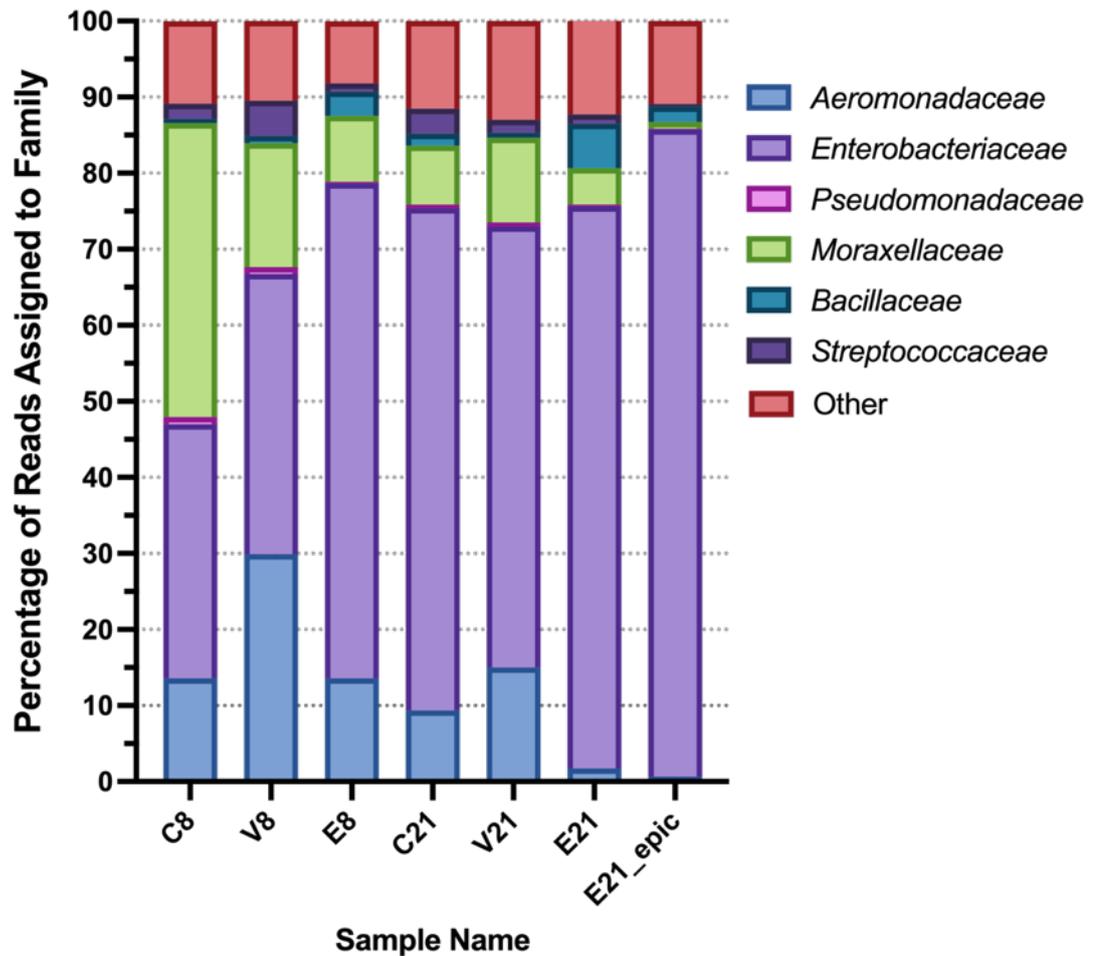
Figure 5.18 Percentage breakdown of each major isolated family from Fastq 16S.

For ease of display, the data was broken down to familial taxonomic level.

5.3 Discussion

5.3.1 Cassette Array Variation

In all six samples trimethoprim and aminoglycoside resistance genes were identified-commonly carried by Tn21 integrons. Similarly, there were also some gene cassettes present that have been less widely reported in the literature. One such cassette identified encoded a protein from the DUF1010 family. This protein of unknown function is part of a family of plasmid encoded proteins from Gram-negative bacteria.

The gene cassette may have been acquired and retained at random, or more likely may have a function that confers selective advantage, because in some instances, the DUF1010 gene cassette was not the most recently captured gene by the class I integron (Figure 5.3, 5.4, 5.7 and 5.8).

5.3.2   Using *attC*-taxa to Estimate Origins of Gene Cassettes

The identification of *attC* sites assembled contigs from the targeted metagenomics samples were performed with HattCI, which was able to identify most of the *attC* sites (Pereira *et al.*, 2016). *attC* sites are short fragments of DNA found at the ends of gene cassettes, normally 59 bp in length that are targeted by integron associated integrases to capture gene cassettes (Section 1.8.2, Figures 1.7 and 1.8). Partial sites were not detected by the software. However, attC-taxa was able to identify that all of the *attC* sites detected in each sample belonged to the family *Xanthomonadales* from their structure (Figure 5.19) (Ghaly, Tetu and Gillings, 2021)*.* Several studies have shown that integron associated integrases may cluster into clades. The two major clades of which are marine and soil/freshwater (Mazel, 2006; Boucher *et al.*, 2007; Cambray *et al.*, 2011). However, the third major clade in which gene cassettes may originate from is the *Xanthomonadales*, which is responsible for the carriage of the greatest range of resistance gene cassettes (Ghaly, Tetu and Gillings, 2021). However, it is not yet understood as to why these gene cassettes originate from *Xanthomonadales*-like bacteria. Whilst the samples come from water, the water in question is not fresh water as the bacteria in it come from the gut which would explain these findings. The *Xanthomonadales* covariance model is more sensitive

270

than the other clades attC-taxa as there are more gene cassettes used to build its model. The *Xanthomonadales* family is believed to be one of the larger contributors to the spread of resistance cassettes in mobile integrons (Ghaly, Tetu and Gillings, 2021).
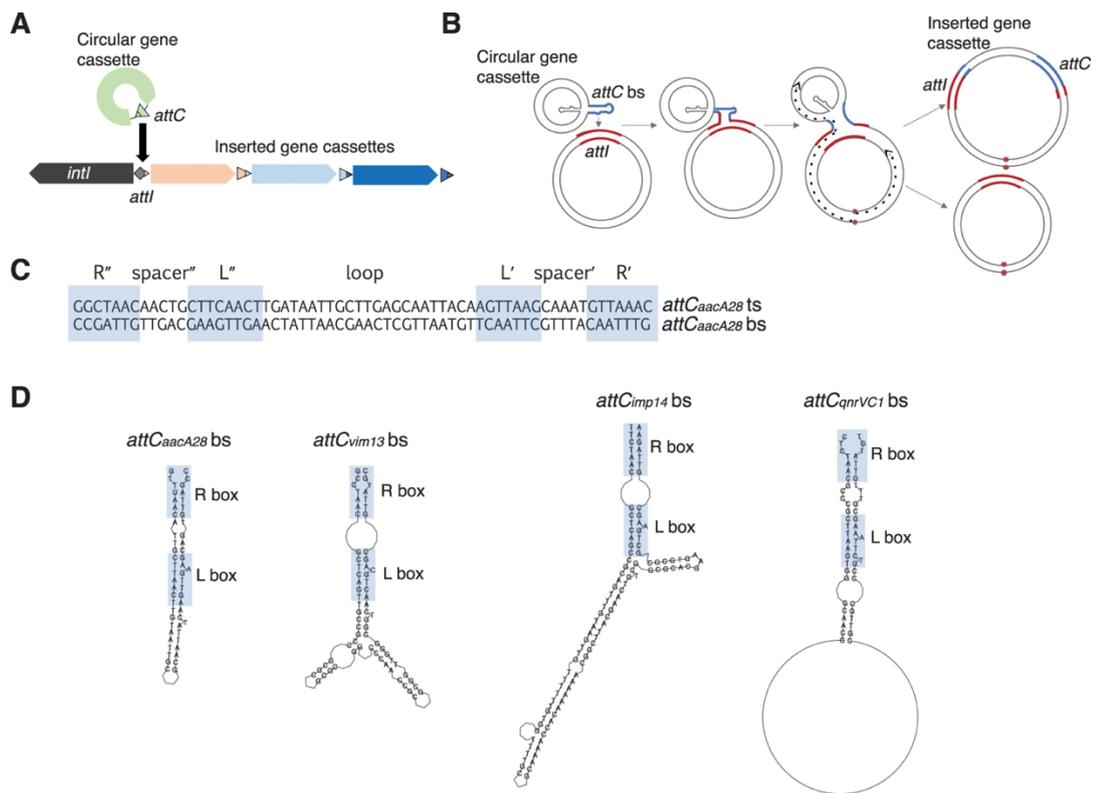
Figure 5.19 The role of *attC* folding structure in gene cassette insertion.

A) Integrons carry an integron integrase gene (*intI*) that encodes a tyrosine recombinase (IntI). IntI facilitates the insertion of circular gene cassettes by mediating recombination between the cassette-associated recombination (*attC*) and integron-associated recombination (*attI*) sites. IntI activity can result in arrays of gene cassettes that vary considerably in size (1 to +300). B) Cassette insertion involves the recombination between *attI* and only the bottom strand of *attC* (*attC* bs). This results in an atypical Holliday junction, which can only be resolved by replication (dotted black arrows; lagging strand not shown). Replication of the recombinogenic strand produces a daughter molecule with the inserted cassette at the *attI* site, while replication of the alternate strand generates the integron without the inserting cassette. C) The palindromic nature of *attC*s gives rise to their single-stranded folding structure. All *attC* sites have two sets of inverted repeats (R'/R" and L'/L"), which allow the folding of single-stranded *attC*s. Two spacers, spacer" and spacer', separate R" from L" and L' from R', respectively. The middle region of the *attC* is known as the loop and is highly variable in sequence and size. D) Shown are the predicted bottom strand folding structures of *attC*s from four antibiotic resistance gene cassettes (*aacA28*, *vim13*, *imp14,* and *qnrVC1*). The variable degree of base-pairing beyond the R and L boxes generates considerable structural diversity among different *attC*s, which in turn impacts their recombination efficiency by different IntIs. (Reproduced from Ghaly, Tetu and Gillings (2021). Permission was granted to reproduce this figure by Springer Nature).

Samples taken from the term time sample point possessed the most *attC* sites from

the HattCI analysis and as a result had the most *attC*s classified by the attC-taxa

272

program. However, the number of *attCs* detected does not mean they are all from different gene cassettes, but in fact due to assembly limitations it is possible to have duplications of contiguous DNA with overlaps that may contain the identical *attC* sites. In the case of the term time village influent sample, there were 18 contigs which contained *attC* sites identical to the *attC* site which is associated with *qacEΔ1*.

The data displayed in Table 5.3 shows that there are a very large number of *attC*s identified by HattCI, but their taxonomic origins could not be established using attC-taxa. This is due to a limited dataset being used to develop the covariance models used in this recently developed program (Ghaly, Tetu and Gillings, 2021). The other reason for lack of identification of some cassette taxa may be due to inability to determine the folding structure of the *attC* sites. attC-taxa uses the package 'infernal,' which is used to predict the shapes and folding structures of RNA molecules (Nawrocki and Eddy, 2013). A covariance model from attC-taxa was used from a database of known *attC* sites to predict the taxon of the cassette in question. *attC* sites may work in a similar manner to RNA when forming stem-loop structures when undergoing recombination events for capture or circularisation of gene cassettes, which is why this package is used (Ghaly, Tetu and Gillings, 2021). As with all newly developed scripts and models, such as this covariance model for predicting cassette taxon, models will need further refinement and updates.

5.3.3 Resistance Gene Cassette Carriage

In each of the six samples, the archetypal Tn*21* cassette array (*aadA1-qacEΔ1-sul1*) was detected, with little variation. However, the addition of folate reductase gene

cassettes, conferring resistance to trimethoprim was detected, and multiple variants of this *dfrA* gene were noted in many of the samples. In some cases, due to limited sequence length, definition of the subtype was not possible. Across the samples there was strong presence of the *dfrA1*, *dfrA12* and *dfrA17* subtypes of the gene. The presence of these subtypes was seen in the influent samples at both locations and time points (08/01/2020 and 21/01/2020). Due to the COVID-19 pandemic only two sample time points were collected. The original sampling plan was to further sample during the March-April 2020 time period weekly for six weeks and repeat this work to identify any notable temporal change.

Folate pathway inhibitors such as trimethoprim and sulfamethoxazole are potent antibiotics which are frequently prescribed in order to treat bacterial infections. In the UK the combination of sulfamethoxazole, a sulphonamide, and trimethoprim (co-trimoxazole) are prescribed most commonly to treat urinary tract infections but are also used to treat acne and chest infections (NHS, 2018). Patients take courses of these medications from as little as three days to four to six weeks, depending on the severity of the infection (NHS, 2018). In 2017, a report by the National Institute for Health and Care Excellence (NICE) determined that 34% of urinary tract infection causing bacteria were found to be resistant to trimethoprim. It is unsurprising to identify a wide range of trimethoprim resistance gene cassettes in wastewater. The common use of the co-trimoxazole antibiotic cocktail may therefore be responsible for driving the uptake and carriage of such cassettes in Tn*21*-like mobile elements, especially as sulphonamide resistance from *sul1* or *sul2* is immobilised at the 3' end of the class I integron in many Tn*21* elements. Class I integrons and Tn*21*-like genetic

elements have been previously identified in clinical urinary tract infections caused by Gram-negative bacterial isolates (Márquez *et al.*, 2008). Alongside these clinical findings and the common use of co-trimoxazole to treat community-acquired infections, co-selection events within the patients are likely the cause of the dissemination of these class I integrons and resistance cassettes such as *dfrA* into the wastewater environments. It is difficult to identify the proportion of a population with ongoing infections such as urinary tract infections, but it is likely a small percentage. Infectious isolates such as a urinary tract infection causing bacteria are perhaps likely to require virulence factors that allow the production of biofilms via agglutination or the production of extracellular polypeptides and bundle forming pilli (Sheikh *et al.*, 2001; Mohamed *et al.*, 2007). Expression of these factors outside of a host could perhaps contribute to biofilm formation within wastewater and sewage systems. Shedding of such bacteria from hosts would explain the isolation of these cassette arrays in the wastewater itself. A recent study investigating the carriage of ARGs in sewer pipes identified a large variation in biofilms, containing many taxonomic groups, including some potential human pathogens (Morales Medina *et al.*, 2020). However, instances of *Gammaproteobacteria* such as those which may be commensal bacteria and often carry Tn*21*-like mobile elements, progressively decline as sewage biofilms mature (Auguet *et al.*, 2015)

Interestingly, in all instances where the *dfrA* cassettes were identified, it was the most recently inserted gene cassette. Compared to the NCBI database of Tn*21*-like mobile elements, the frequency of occurrence of *dfrA* genes is higher in the wastewater samples from this study. Of the 139 true Tn*21*-like mobile elements

(containing transposase, *mer* genes and integron), only five contained *dfr* genes within the integron cassette array (Table 8.1). These sequences in NCBI belonged to *Salmonella, Klebsiella* and *Vibrio,* originating from the USA and China. This population of integron cassettes arrays from Tn*21*-like elements suggests there may be a more dominant local population due to the antimicrobial selective pressures within the host environment. However, downstream of the treatment plant it was seen in Chapter 4 that the cassette arrays within Tn*21* integrons were different again, containing *bla*$_{TEM-1}$ and their own promoters. Although these were isolated on a different sample date to the metagenomic samples, it suggests that diversity of cassette array may change further downstream or that by the time of the second sample, bacteria carrying these gene cassettes may have passed downstream. In order to further understand this, more samples would need to be screened over more locations and also over multiple time points. Similar work was performed on class I integrons and demonstrated the change in diversity between gene cassettes in soils over both small and large geographical distances (Ghaly *et al.*, 2019). However, in water systems where bacteria may be more mobile due to the impacts of currents and flow of water, perhaps the dynamics of this model would differ (Section 1.8.4). One study examining occurrence of class I integrons in a river, upstream and downstream of a treatment plant serving a large human population suggested that intervention on a water environment, such as the release of effluent would affect the diversity of gene cassettes, the most notable was the increase in abundance of *qacEΔ1* gene cassettes (Amos *et al.*, 2018).

The review of Liebert *et al.* (1999), suggested how diversification of Tn*21* has taken place within Tn*21*-like element cassette arrays, however, many of those identified in these UK wastewater samples are novel in comparison to those described. Within the two decades since this review, it is likely that the evolution of the Tn*21* class I integron is the most dynamic region of the mobile element. A more recent review suggested that gene cassette variation of Tn*21*-like gene cassette arrays was limited, however, the data in Table 8.1 suggests otherwise (Mindlin and Petrova, 2018). From this current study, it may be possible to suggest that the gene cassette array of Tn*21* class I integrons may differ depending on the environment they were isolated in. For example, in wastewater samples from this study Tn*21* and previous information from clinical urinary tract infection isolates would suggest the typical cassette array would contain *sul1*, *qacEΔ1*, *aadA1* and *dfrA*, although more data is required to confirm this (NHS, 2018).

It is also important to highlight *dfrA* was not the only folate inhibitor resistance gene detected. In the sample from the campus influent, during term time, *dfrB3* was detected (Figure 5.7). This folate inhibitor resistance gene belongs to a completely separate family of folate resistance genes. The main difference between *dfrB* and *dfrA* is sequence length, *dfrB* is approximately half the length of *dfrA* (237 nt and 474 nt, respectively). *dfrB* acts as a homotrimeric enzyme and may provide resistance to antibiotics such as trimethoprim at least three times the concentration that *dfrA* can (Pattishall *et al.*, 1977).

The rifampicin resistance gene cassette *arr-3*, which allows ribosomal protection to rifampicin (Table 1.2), was also detected in two of the six samples (Figure 5.3 and

5.4). *arr-3* has previously been detected in wastewater isolated integrons in China followed by *dfrA27*. Here the similar layout of gene cassettes within the integron was identified in campus influent pre-term time but with *dfrA7* instead (Figure 5.3). Whilst rifampicin resistance gene cassettes have been described in integrons in Europe and Asia, *arr-3* has been more commonly reported in Asia than Europe. The difference between the two alleles is a single amino acid substitution, K98 to R98. Previous reports also suggest that *arr-2* is found most frequently within wastewater influent and is integron associated (Arlet *et al.*, 2001; Gadou *et al.*, 2018; Surleac *et al.*, 2020). However, the cassette array collected in this study suggests that *arr-3* may be more commonplace in the influent of this UK wastewater treatment plant. Interestingly, the rifampicin resistance cassettes were only identified within the pre-term time influents. Temporal change may have meant these gene cassettes were not detected within the term time influent samples or perhaps cassette frequency of other gene cassettes may have meant amplification of smaller integron cassette arrays was favoured over shorter gene cassette arrays from PCR bias. To gain more certainty on these findings, repeating this across a continuous time series is required. Although, extension times of the integron cassette array PCRs were intentionally longer to try to recover longer cassette arrays, which may have otherwise not been detected. Similarly, to allow more even distribution of each cassette array each sample was split into quadruplicates for PCR amplification before being pooled for purification and sequencing.

It is also important to note that the sample area in question in this study is fed by a mainly rural population, compared to other studies on tertiary wastewater

treatment plants focusing on large urban treatment sites. However, the campus influent is fed by a large proportion of international students, as well as students from across the UK which could potentially explain the results found here. Further investigation could uncover more information about the prevalence of these resistance gene cassettes.

The *estX* gene cassette found in the term time campus influent sample, which encodes an unknown esterase, has previously been associated with the carriage of a resistance gene cassette *satA*, encoding streptothricin resistance (Jones-Dias *et al.*, 2016). Moreover, this is the first time *estX* has been identified without the neighbouring *satA* cassette. The assembled gene cassette array from campus influent at term time, instead displayed a second novel gene cassette following the *estX* cassette in its array. This was a partial open reading frame matching that of *serB*, a phosphoserine phosphatase, part of the family of halo-acid-like hydrolases. Neither of these two proteins have any known antimicrobial resistance functions, but both were identified within a Tn*21*-like class I integron found in wastewater influent from a university campus during term time. Whilst Tn*21* is widely linked with the carriage of resistance genes, it is perhaps also surprising that most integron gene cassettes outside of the Tn*21* captured from the environment are not resistance genes or have no known function (Partridge *et al.*, 2009).

Further to these non-resistance gene cassettes identified in cassette arrays was a domain of unknown function, *duf1010*, sometimes referred to as *gcuD, gcuF* or *orfD, orfF*. This has been previously reported in a cassette array containing the carbapenem resistance gene bla_{IMP-18}, *gcuD* and bla_{OXA-2} and an In706 containing

bla$_{IMP-18}$-aadA43Δ-bla$_{OXA-2}$-gcuD (Martínez *et al.*, 2012; Molina-Mora, Batán and García, 2020). It still is not clear what the function of this *duf1010* gene product is. However, it appears to have been detected in a large variety of bacterial gene cassette arrays globally. The Duf1010 gene cassette has been detected in wastewater in mainland Europe (Mesquita *et al.*, 2021). In this case the gene cassette was found on a cassette array also harbouring *aadA2* and *dfrA12* with the structure *dfrA12-duf1010-aadA2*. This cassette array structure was identified across village influent samples in this study and has also been detected in *E. coli* in Taiwan (Chang, Chang and Chang, 2007). In the campus influent however, the streptomycin 3'-o-adentyltransferase (*aadA2*) gene was not present, but *ant(3'')-Ia* (*aadA1*) was. However, what is most interesting about this *duf1010* cassette is that between entering the wastewater treatment plant to leaving in the effluent, the cassette is lost or undetected (Table 5.4, Figure 5.4 and Figure 5.8). Inadvertently, not only was the *duf1010* cassette undetected, but so was the *dfrA12* which preceded it in the cassette array. However, the limitations of this study mean that it is not possible to determine whether the cassette arrays identified in the effluent necessarily come from the same bacteria as the effluent.

Along with the already noted loss of diversity in cassette arrays across the wastewater treatment process, perhaps a selective event may occur within this process. Since such selective events tend to invoke an SOS-response within cells, it is more than possible cassette turnover of the *duf1010,* and other cassettes may have occurred when shuffling the gene cassette array to provide stronger transcription of

280

a gene cassette. The role of this dynamic movement and constant turnover of cassettes allows the potential for transmission within the environment.

During the identification of the Tn*21* class I integrons in the targeted metagenome amplicons, one particular gene cassette array stood out. The 5,341 bp integron sequence in the pre-termtime village influent sample was the longest sequence by almost 1,000 bp (Figure 5.4). The cassette array itself is rather complex containing resistance cassettes to multiple classes of antimicrobials *aac(6')-Ib-cr-bla$_{OXA-1}$-catB3-arr-3-qacEΔ1-sul2*. The gene *aac(6')-Ib-cr*, fluoroquinolone-acetylating aminoglycoside 6'-N-acetyltransferase, a plasmid mediated quinolone resistance gene (PMQR) simultaneously provides aminoglycoside and quinolone resistance and has previously been associated with ESBL producing *E. coli* and *K. pneumoniae* (Shin *et al.*, 2009)*. A similar cassette array, containing *arr-2* instead of *arr-3* was identified in two strains providing resistance to eight classes of antibiotics from a collection of both clinical and non-clinical *E. coli* isolates in Kenya (Kiiru *et al.*, 2013). The study from Kiiru *et al.* (2013) also further supports Shin *et al.* (2009) as all *aac(6')-Ib-cr* positive strains contained resistance determinants to beta lactams. However, without being able to isolate the microorganism a cassette array came from, it may prove difficult to identify whether corresponding resistance determinants to beta-lactams would be present within the host bacterial strain. It is also possible that hybridisation of aborted PCR reactions within a reaction mixture may occur causing the generation of artefacts to create new cassette arrays.

This large cassette array complex provides resistance to seven different classes of antimicrobials (including the 3'CS). The detection of *bla$_{OXA-1}$* within the cassette array

is also surprising, with this gene having first been identified within a class I integron in the 1970s in pMG101 (Mchugh et al., 1975; Hooton et al., 2021). When comparing the cassette arrays identified in these wastewater samples to some of the cassette arrays of the whole genomes of isolates found downstream of the wastewater effluent release, bla_{OXA-1} was not found, instead bla_{TEM-1B} and its promoter was present within the cassette arrays (Figure 4.4). However, the bla_{OXA-1} gene cassette was identified in 2017 in the dairy slurry isolates EVAL51 and EVAL 55 (Figure 4.5). The other major resistance determinant identified within the Tn*21* cassette arrays was the chloramphenicol resistance gene *catB3*. This chloramphenicol acyltransferase is characteristic of Gram-negative isolates such as *Salmonella spp.* Typhimurium*, Acinetobacter baumannii* and *E. coli* and is often found both chromosomally and on plasmids, typically on integron cassette arrays (Bunny, Hall and Stokes, 1995; Tosini *et al.*, 1998; Houang *et al.*, 2003; Wang *et al.*, 2003). Dissemination of resistance gene cassettes has been documented clearly across multiple studies and the data from this study. No novel gene cassettes or gene cassette arrays to Tn*21*-like mobile elements were identified, however, this study consolidated findings of other gene cassette arrays identified elsewhere in the world. Although primers were not added or used to detect the entirety of the 3' CS, including the *sul* gene located at the 3' terminal of *qacEΔ1*, the sequence assemblies in five of the six samples were assembled. Small fragments of the *sul* genes downstream of *qacEΔ1* were also identified. This is because the *qacEΔ1*_R primer formed part of the open reading frame for *sul* which explains this.

282

The information collected by screening the wastewater DNA samples for Tn*21* integron cassette arrays helped uncover what resistance gene cassettes are abundant in the environment and allowed for the prediction from where these cassettes may have once originated. Furthermore, it is not possible to identify from this information what may be organisms carrying these mobile elements and their resistance gene cassette cargo.

5.3.4 epicPCR: A Solution and a Problem

The epicPCR technique was initially designed for high-throughput screening of marine and wastewater environmental samples for antimicrobial resistance genes of interest whilst maintaining information of the organisms they were found in (Spencer *et al.*, 2016; Hultman *et al.*, 2018). Unlike other high-throughput techniques or metagenomic techniques, it is actually possible to be very targeted within epicPCR approach in order to gather this information. Targeted amplicon sequencing of metagenomic DNA samples allows identification of genes present within an environment, but it does not inform the investigator as to which organisms they belong to. This is the case for the work performed on screening wastewater samples for Tn*21*-like integron cassette arrays (5.2.1). Whilst in this case it was possible to determine the origin of the gene cassettes thanks to the development of covariance models based on the folding of *attC* sites, determination of the full mobile element was not possible without further work.

Understanding more about mobile genetic elements, their bacterial hosts and their possible dissemination routes is of significant interest in the effort to combat the

283

spread of acquired antimicrobial resistance genes. Shotgun metagenomics was the first widely used approach to identify this particular topic in a high throughput manner; however, it was still extremely difficult to identify bacterial hosts and rare mobile elements or resistance genes due to the lack of depth in coverage depth. Other methods such as Hi-C have offered the potential to link accessory genetic information such as plasmids to the chromosomal DNA of the host cell in a metagenomic DNA sample (Stalder *et al.*, 2019; McInnes *et al.*, 2020). Upon metagenomic sequencing and assembly, it would be possible to determine which organism a gene found belongs to and may allow for partial reconstruction of important environmental plasmids from the crosslinked genomes. This method would aid identification of genetic elements that have a role in the dissemination of resistance genes, mobile elements and virulence factors throughout the environment by lateral gene transfer. One of the major problems with bacterial shotgun metagenomics is the inability of the method to link host cells with plasmid replicons. A recent seminar given by (Van Shaik, 2021) suggested that whilst Hi-C was a promising technique, it too was unreliable. Cross-linking of accessory DNA elements such as plasmids to their host cells was found to have poor efficiency. This meant it was extremely difficult to establish with certainty a true depiction of plasmid carriage within the microbiome. In this sense, epicPCR should be more advantageous over Hi-C if the focus of investigation is identifying specific elements or resistance genes within the chosen environment. The two procedures, and other similar high throughput techniques are by no means cheap, but perhaps they should be used in conjunction with each other.

Another recently developed method, is OIL-PCR (Diebold *et al.*, 2021). The process is effectively the same as epicPCR but with less steps, meaning less complex methodology, which should reduce the chances of failure in the protocol. The key difference in this process to the original epicPCR method is the introduction of lysozyme to the emulsified, acrylamide encapsulated cells in the PCR master mix. This also helps with the capturing of Gram-positive bacteria which are often harder to lyse. In the case of Tn*21*, the latter may not be particularly useful.

5.3.4.1   Why epicPCR?

It is known that Tn21 and class I integrons are carried by Gammaproteobacteria. However, there has been less study into diverse carriage of Tn*21* and Tn*21*-like elements within this large taxonomic group. epicPCR was favourable, offering a clearer insight into the organisms which may carry Tn*21* within the wastewater environment. OIL-PCR, later named epicPCR2.0, would have been considered a promising alternative technique instead, had the method been available or known about at the time of experimental planning and data collection. However, as previously stated, OIL-PCR uses lysozyme which particularly helps lysis of Gram-positive cells (Roman *et al*, 2021). Given that Tn*21* is found in Gram-negative bacteria, it is likely this would not have been as advantageous to this study. Theoretically, the use of Oxford Nanopore's MinION would have an advantage over Illumina as the capability of producing longer reads of non-fragmented DNA, as it is capable of covering a larger portion of the 16S rDNA gene in one sequencing read. The Illumina sequencing platform requires the fragmentation and tagging of shorter

lengths of DNA to generate libraries of DNA sequences to generate reads. The original epicPCR methods use Illumina Nextera sequencing methods and demonstrate that whilst using short sections of the 16S rRNA gene it can still supply good information regarding the carriage of functional genes (Spencer *et al.*, 2016; Hultman *et al.*, 2018). Since the initial work performed on optimisation for epicPCR in this study, and the development of a sequencing method, another research group has since developed a method for epicPCR using the long-read PacBio sequencing platform in order to identify phage and host cell interactions in estuarine environments (Sakowski *et al.*, 2021). Whilst this study is the first documented use of long-read sequencing for epicPCR, the data here is believed to be the first use of Oxford Nanopore MinION to perform epicPCR. Since the production of this data, a revised method has been published, which streamlined the epicPCR process suggesting increased reliability (Roman *et al.*, 2021). Since its initial release of epicPCR, there have been multiple modifications in order to achieve greater reliability. The refinement of such methods for different purposes will allow for greater uptake of such valuable methods by the wider scientific community in metagenomic sequencing methods.

For Tn*21*, epicPCR appeared to be an attractive method to use. It was the cheaper of the options available and could complement the targeted amplicon sequencing being used in analysing the same wastewater samples. Having learned epicPCR on the BBSRC Professional Internship for PhD Students, it was clear that the technique could be used to identify organisms carrying Tn*21* and Tn*21*-like elements. However, due

to the lack of reliability in the process, a very large period of time was spent refining the technique for use on this mobile element.

The first difficulty was which region of the Tn*21* transposon to target. There are several regions, which are specific to Tn*21*-like elements. The first one targeted the *intI1* and the *tnpM* genes, a stretch of DNA of similar size to the 27F and 1542R and 1492R (nested primer). However, successful formation of fused products proved very difficult. Targeting the mercury resistance genes was considered, but this is not specific to Tn*21*. In addition to this, epicPCR has previously been performed to identify mercury resistance in both Gram-positive and Gram-negative organisms (Hall *et al.*, 2020). As a result, the primers tnpA_F, targeting the 38 bp IR of IR$_{tnp}$, and *tnpA_R*, a section of the Tn*21* transposase gene *tnpA*, yielding a product of 392 bp were designed (Figure 5.10).

5.3.4.2 Long-Read epicPCR for Greater Host Organism Resolution

The sequencing process and Fastq 16S program allowed easy throughput of approximately one million reads for taxonomic identification. However, as can be seen from the phylogenetic tree, 2.6% of reads were assigned to Gram-positive taxonomic orders within the E21 epicPCR sample, namely *Lactobacillales* and *Bacillales* (Figure 5.16 and Figure 5.18). This is interesting for a number of reasons. Firstly, Tn*21* is synonymous with Gram-negative gammaproteobacteria, and secondly this result is of interest due to the efforts made to reduce errors where possible within the amplification by the usage of high fidelity Q5 DNA polymerase and native ligation of barcodes to prevent unnecessary amplification or introduction of errors

to the DNA which may then cause misclassification of organisms carrying Tn*21*-like elements. To try and combat this, many precautions were taken. The BLASTN parameters used for the trimmed reads was to reduce the chance of falsely assigning reads to taxa. Minimum read length and read quality of the trimmed reads was set to 700 bp and 7, respectively to avoid the use of short, fragmented and lower quality reads which may be misclassified to taxonomic groups. To ensure every cell containing Tn*21* was amplified, the fusion PCR reaction mixture was divided into 16 50 µL reactions per sample and similarly, blocking and nested PCR were in quadruplicate. Although the 16S rDNA primers used: 27F, 1492R and 1542R, are universal primers it is still possible that some bacterial 16S rDNA genes may be favoured in amplification over others. The single celled approach to amplification of the Tn*21* carrying organisms should have helped reduce any bias.

Strict parameters were used for BLASTN to reduce the probability of taxonomic misclassification: firstly the 'conservative' function was set to less than 1 x10^{-10} and the minimum percentage identity was set to 80% were used to achieve the phylogeny data displayed in Figure 5.16 to 5.19. The conservative approach to the taxonomic classification of Tn*21* harbouring environmental bacterial cells may help overcome potential read errors from the basecalling of the Oxford Nanopore software. In conjunction with these parameters, this specificity is also useful to fulfil the original rationale of using long read sequencing technologies in epicPCR to further define species which may carry Tn*21* but could be used on longer more variable targets such as integron gene cassette arrays. The cost of this stringent filtering was the loss of some reads which may not have contained basecalling errors. However, the loss of

some reads outweighs the gross misclassification of taxa. As with any filtering, some errors or outliers may still be identified within the results.

There are two main advantages of epicPCR using long read technologies. The first of which, is that despite the high error rate of Oxford Nanopore basecalling, the method can still be useful when targeting longer sections of housekeeping genes when aiming to achieve higher resolution of organisms carrying a desired functional gene. The second advantage is linked to the first: long read data acquired from epicPCR requires very little processing afterwards thanks to Oxford Nanopore's proprietary cloud-based platform EPI2ME. After basecalling, using guppy, cutadapt was used to trim attached functional genes from reads and processing may resume through the EPI2ME platform, including quality analysis, using a plug and play method for defining program parameters. Overall, this approach is considered more user-friendly, rather than relying on developing or adapting custom scripts from original methods (Spencer *et al.*, 2016). When used alongside epicPCR 2.0, the experimental procedure may also be more reliable.

5.3.4.3 The Potential Disadvantages of Using Oxford Nanopore Sequencing in epicPCR

Long-read technology such as Nanopore has its disadvantages compared to the use of short-read technologies. The long-read platform works by attaching a motor protein to the 5' terminal of target DNA in a library which then when bound to a sequencing pore in the presence of ATP (an ATPase) causes the translocation of DNA through sequencing pores. As the strands of DNA translocate through the membrane

bound pores, the membrane depolarises, causing a voltage change. Each base type, adenine, cytosine, guanine and thiamine all depolarise the membrane at measurably different voltages, allowing for definition between bases (Branton *et al.*, 2008). The chemistry of the sequencing process requires high stringency. An optimal DNA translocation rate through the nanopores is between 300 and 450 b s^{-1}. This can be affected by the concentration of DNA added, or by the volume of priming buffer (which contains ATP) added. Alongside this stringency, lengths of DNA containing a repeat of a single base can sometimes be misread causing the introduction of small insertions or deletions to sequences. However, since the initial release of Oxford Nanopore sequencing technology, basecalling software has moved a long way, with error rates for raw reads reportedly down at 1.7% for microbial genomes under the 'super accuracy' base calling model (Oxford Nanopore Technologies, 2021). However, realistically, error rate is still often reported at approximately 10%.

5.3.4.4   epicPCR vs. Metagenomic DNA

Fused amplicons of Tn*21 tnpA* and 16SrDNA from metagenomic DNA extracted from the same wastewater samples were analysed in parallel with the epicPCR procedure. One of these sets of amplicons was in parallel to the successful epicPCR sample, E21**.** The data in Figure 5.18 shows the variation between performing epicPCR and the DNA only PCR of sample E21. epicPCR is shown to have reduced familial diversity than its targeted metagenomic counterpart. This is of little surprise due to the single celled amplification approach to avoid off targeting and generation of chimeras when amplifying Tn*21* carrying organisms. The single celled approach of epicPCR highlights

the need to aim for 1-2 cells per 100 acrylamide beads, which aids reduce the chance of this chimera formation (Figure 5.12). This therefore shows that to some extent that epicPCR can be used successfully on long-read platforms. Notably, carriage of Tn*21* was predominantly by *Enterobacteriaceae*, given the sample location of this sample, it was expected.

Further to this, some Gram-positive bacteria were detected in the metagenomic fusion PCRs due to the use of free DNA within the metagenomic samples. This allowed *tnpA* to fuse with any 16S rDNA gene, including genes from Gram-positive organisms, for detection within the taxonomic classification within these samples. Surprising was the lack of any *Bifidobacteriaceae*, another Gram-positive organism which is commonly found within the gut. Whilst most organisms within this genus are anaerobes and would not be expected to survive within wastewater, it is still somewhat surprising that remnants of their 16S rRNA gene was not detected.

Unfortunately, due to the circumstances, vital comparisons across epicPCR samples from each wastewater sample location and timepoint cannot be made. However, the metagenomic DNA PCRs still give some insight into how the overall taxonomic distribution of the samples varies. The most important changes identified were the loss of *Aeromonad* abundance across the wastewater treatment process. This is more notable in the term time samples: influent (C21, V21) to the effluent (E21). Similarly, across both sample sets, the proportion of *Enterobacteriaceae* reads increased over the wastewater treatment process, this may be due to the losses of other bacterial families from the wastewater treatment process. It was also noted that there was loss of *Aeromonads*. *Moraxellaceae* were also seen to be lost over the

wastewater treatment process. *Moraxellaceae* exist as commensal gut bacteria, they are also water and soil borne. It is of little surprise to identify this family within wastewater. However, across all of the metagenomic PCR samples, *Acinetobacter* made up almost all of the carriage of this family. In comparison, in the epicPCR sample, all but one read within this family was associated to an *Acinetobacter* species (Figure 5.20). *Acinetobacter* has been detected harbouring class I integrons and Tn*21* in a wide range of environmental settings; from nosocomial and clinical isolates in Europe and South America, to wastewater in India (Gonzalez *et al.*, 1998; Ploy *et al.*, 2000; Sultan *et al.*, 2020). What is particularly interesting from the epicPCR sample however is that 50% of the reads assigned to the *Moraxellaceae* are assigned to the species *A. johnsonii* (Figure 5.20). This species is commonly isolated on human skin and whilst often non-pathogenic, has occasionally been reported to cause human infections along with other non-*baumannii* species of *Acinetobacter* (Karah *et al.*, 2011). *A. johnsonii* has also been shown to be particularly capable of capturing accessory DNA such as insertion sequences, transposons and gene cassettes (Montaña *et al.*, 2016). Tn*21* indicator genes such as the class I integron, have been previously identified within many *Acinetobacter* species (Gonzalez *et al.*, 1998; Liebert, Hall and Summers, 1999).

Figure 5.20 Lineages of reads detected in epicPCR E21 of *Moraxallaceae*.

The lineage of this family is predominantly from *Acinetobacter*, only one read came from *Moraxella* genus. The thickness of the line on the phylogram indicates the proportion of reads belonging to each lineage. Green nodes indicate species which were detected by BLASTN and grey nodes indicate taxa that were not identified. For example: *A. baumanii* was detected with a green node, but ancestor DNA which could only allow for distinguishing of the genera was not detected. (Note, due to EPI2ME software, it was not possible to unmerge the overlapping species at the top of the figure, containing *A. baumannii* and the lower one with *A. pittii*).

The reads which were assigned to *Enterobacteriaceae* taxa were also explored

further within the epicPCR sample. As shown in Figure 5.18, the majority of epicPCR

reads were assigned to this family. Upon further investigation, approximately 4,000

of the 9,500 reads did not have enough information to determine which genus or species they belonged to. However, of the remaining reads assigned to this family able to define genus and or species, 3,000 were assigned to the *Klebsiella* genus and 1,500 to *Enterobacter* genus (Figure 5.21). This again is interesting as the data suggests that in the future, when selecting for Tn*21* harbouring bacterial isolates, selective isolation for *Klebsiella* and *Enterobacter* on media such as Eosin Methylene Blue Agar and MacConkey agar, respectively or the use of CHROMagar may yield more of these organisms. *Klebsiella,* particularly the *pneumoniae* species is clinically relevant, having been placed in the WHO priority pathogens list in 2017 due to reports of carbapenem resistance (WHO, 2017). Carriage of Tn*21* in such isolates by proxy means carriage of a class I integron, in which carbapenem resistance genes may be captured. In one such report, a Tn*21*-like genetic element had fused with a backbone of Tn*1722* causing loss of the integron but in its place existed a mobile element containing *bla*$_{KPC-3}$, encoding carbapenem resistance (Zhang *et al.*, 2019). This is not the only report of carbapenemase activity discovered within Tn*21*s class I integrons (Chowdhury *et al.*, 2011).
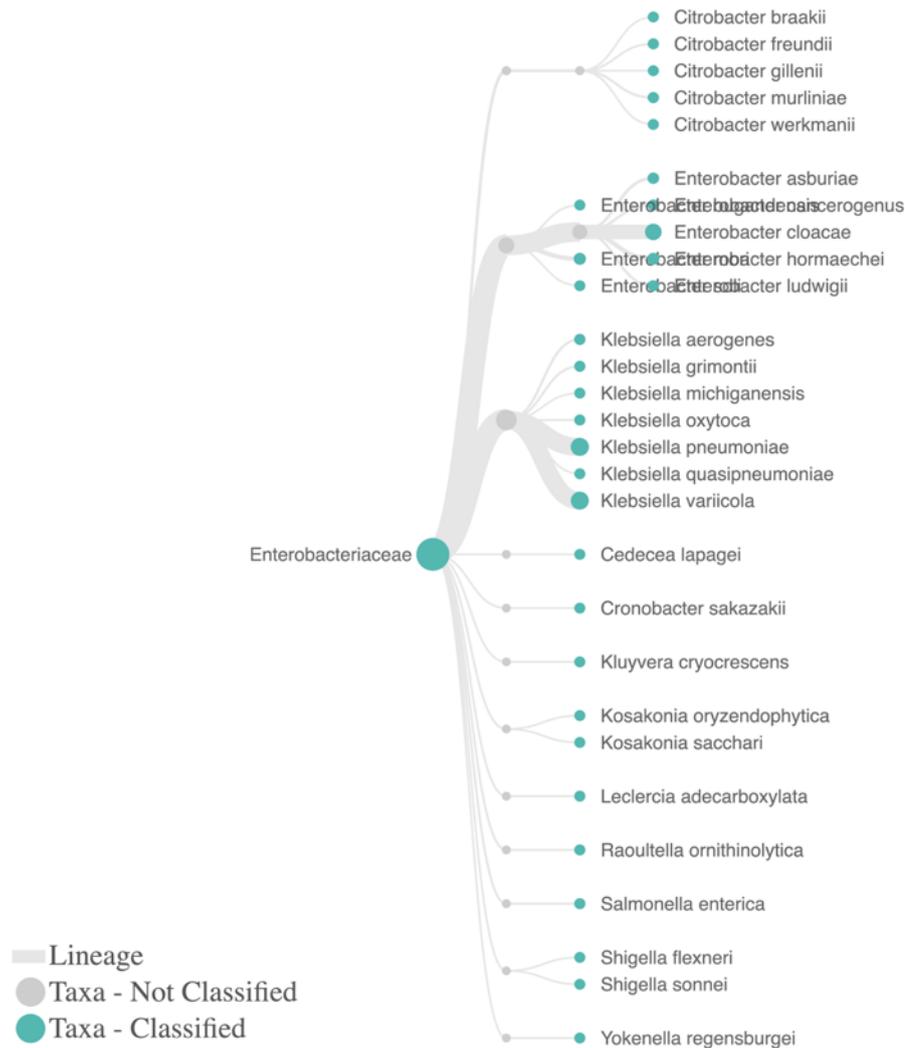
Figure 5.21 Lineages of reads detected in epicPCR E21 of the *Enterobacteriaceae* family.

The lineage of this family is predominantly from two genera *Klebsiella* and *Enterobacter*. Only the top 30 most abundant clades are shown in this image. The thickness of the line on the phylogram indicates the proportion of reads belonging to each lineage. Green nodes indicate species which were detected by BLASTN and grey nodes indicate taxa that were not identified. For example: *K. pneumoniae* was detected with a green node, but ancestor DNA which could only allow for distinguishing of the genus *Klebsiella* was not detected. (Note, due to EPI2ME software, it was not possible to unmerge the overlapping species at the top of the figure, containing some *Enterobacter* species).

Further comparison across epicPCR samples would have been performed to help

validate the results from other metagenomic DNA samples; however, due to financial

and time limitations, the planned 16S PCR of each sample was unable to take place.

Similarly, running epicPCR using the Illumina sequencing method, as per Hultman, *et al* (2018), alongside this work with Oxford Nanopore samples was considered, however due to lost time from the COVID-19 pandemic this was not possible.

5.4 Conclusions

5.4.1 Tn*21* cassette PCR

The targeted amplicon sequencing of the variable region, containing the Tn*21* cassette arrays across the wastewater treatment process identified the large variety of resistance and non-resistance gene cassettes carried within the Tn*21.* The targeted amplicon sequencing also helped identify changes to the cassette arrays across the wastewater treatment process. The Tn*21* class I integron was shown to be active in the acquisition of new genes such as trimethoprim resistance genes (*dfrA*) and the loss of these genes downstream in wastewater treatment process.

5.4.2 Long-Read epicPCR

Whilst epicPCR was originally developed for short-read Illumina sequencing, it is clear that the use of long-read sequencing has its place in this technique. If the focus is to further interrogate sequenced epicPCR amplicons, to determine genus or species-specific classification of the organisms carrying a functional gene or secondly to target functional genes or genetic elements which may be variable in length, the using long-read sequencing technology such as Oxford Nanopore would be greatly beneficial compared to the original method. As long-read technologies become

refined and error-rate decreases, this technique could become more widely used. The work performed in this chapter to identify the organisms which may carry Tn*21* and Tn*21*-like elements acts to show the differences in Tn*21* between the organisms detected through metagenomic techniques isolation of strains, but also the wider impact a mobile element such as this may have on the wider community and its dissemination into the wider environment post wastewater processing.

5.4.3 Overall Conclusion

The work performed in the two experiments described in this chapter together highlight the need for greater biological surveillance of wastewater treatment. Wastewater influent effectively acts as a non-invasive proxy of a whole population's gut microbiome. It may also help identify potential outbreaks of infection but can clearly be seen to survey particular genetic elements. In the attempts to further understand and overcome antimicrobial resistance, monitoring of wastewater and other waste sources, such as farm manure for mobile elements associated with resistance gene dissemination such as class I integrons may prove to be a particularly useful tool.

5.5   Future Work

5.5.1   Resistance Gene Cassette Carriage in Wastewater

The targeted amplification and sequencing methods described in this Chapter have provided novel insight as to the identity of the resistance genes carried by Tn*21* and Tn*21*-like genetic elements. This data is qualitative not quantitative. To further the impact of the information gained here, qPCR to determine abundance of these gene cassettes should be used to determine the relative impacts these genes may have on dissemination to the environment beyond wastewater treatment. Multiple samples across a large timescale, influents and effluent may help identify temporal change within the resistance genes shed to the wastewater treatment site and what is removed by the treatment process. Repeating this experimental process alongside performing targeted amplicon metagenomic work on an urban tertiary wastewater treatment plant may also allow for valuable comparisons between smaller changing international populations and the constant flow of large populations in an urban environment.

Whilst the cassette diversity of the wastewater effluent appears to be reduced, resistance gene cassettes are still dispelled into open bodies of water, disseminating into the wider environment. The anthropogenic impact of the current state of wastewater treatment leaves a lot to be desired, however until there is legislation introduced, it is hard to foresee any efforts to reduce this impact.

5.5.2 Using epicPCR to Identify Tn*21* Carrying Bacteria

epicPCR requires refinement for each different target functional gene and also for use with Oxford Nanopore and multiple issues along the way. Due to the prolonged process, only one sample was able to be used in the sequencing and analysis. To build on the work performed in this chapter, further refinement, or the use of the recently published OIL-PCR (epicPCR 2.0) (Roman *et al.*, 2021) may help yield results more reliably than adaptation of the original methods described (Spencer *et al.*, 2016; Hultman *et al.*, 2018).

Alongside the work performed, direct comparison to the original epicPCR method using Illumina sequencing to detect Tn*21* carriage within wastewater should also be performed to help further validate this study. Similarly, investigations should be performed to identify whether it would be possible to perform hybrid technology (short and long-read) sequencing on longer DNA targets.

As previously stated, as long-read technology improves and the reliability of such an innovative technique such as epicPCR is refined, the use of long-read epicPCR may prove more useful. For example, in the surveillance of antimicrobial organisms, it may one day be possible to not only identify multidrug resistance organisms carrying genes of interest, but it may be possible to extract whole class I integron gene cassette arrays from organisms too to give more context across the sampled bacterial communities.

Chapter 6: Mobilisation of Tn*21,* Tn*21-like* and IS*26*-like Sequences Through Chemical Stress

6.1   Introduction

Since its discovery, *Tn21* and Tn*21*-like transposons have been documented globally. This family of self-transmissible genetic elements are capable of mobilising between plasmids by the formation and resolution of cointegrates (Shapiro, 1979). Tn*21* transposition, (alongside other related Tn*3* family transposable elements) is RecA-independent - its mobility is self-regulated by its own transposase, encoded by *tnpA*. What is less well understood however, is what factors affect expression of these transposase genes. Previous work has shown that Tn*3* family transposition may be stimulated by environmental factors such as temperature, UV light exposure and restricted access to carbon sources (Kretschmer and Cohen, 1979; Pfeifer and Blaseio, 1990; Eichenbaum and Livneh, 1998; Lamrani *et al.,* 1999). A Tn*3* family mercury resistance transposon, Tn*501*, possessing a *mer* operon similar to that of Tn*21*, was shown to be mobilised by sublethal concentrations of Hg (II) stress. In fact, it was thought that Tn*501* required the presence of Hg (II) in order for mobilisation to occur (Kitts *et al.,* 1982). The hypothesis for this transposition induction was that transcription of the *mer* operon continued from the *mer* operon into the transposase genes causing expression of the transposase genes, which, in turn induced formation and resolution of cointegrates.

Tn*21,* however, has a Tn*402*-like element inserted into what is thought to be a transposase modulator gene *urf2M,* which is located between the *mer* operon and the transposase genes. Therefore, transcription of the *mer* operon from P*merT* is unlikely to cause transcription of the transposase genes, because there is an 11 kb Tn*402*-like element between the *mer* operon and the transposition genes. Studies by Hyde and Tu (1985) suggested that despite the splitting of *urf2* by insertion, TnpM still played a role in transposition by upregulating

transposition but down regulating resolution of cointegrates (1.9). Equally of note, the modulator is not essential for transposition of the Tn*21* but may reduce transposon resolution of cointegrates (Hyde and Tu, 1985).

6.1.1   Aims and Objectives

1. This chapter aims to identify whether Tn*21* and Tn*21*-like transposons may be mobilised when exposed to stress in the form of Hg (II) and if so what concentrations of Hg (II) affect the transposition frequency.

2. The study aims to also build on research performed by other groups that demonstrated temperature, UV and access to carbon sources also modulate transposition frequencies of Tn*3* family transposons.

3. Sequence analysis of the Tn*21* transposases and their promoter regions in Tn*21* variants that were isolated from wastewater and dairy slurry allows the construction of a possible model of how the expression of the RecA-independent transposase *tnpA* may be regulated.

6.2   Methods and Results

A series of conjugation events were used to transfer the plasmid RP4-8 into environmental bacterial isolates, from wastewater effluent and dairy slurry (2.2.20). RP4-8 was then used as a template to trap induced mercury resistance transposons when exposed to different concentrations of mercury (II) chloride (section 2.2.20, Figure 6.1). The transconjugant plasmid was then mated out of the environmental bacterial isolate in question to a chromogenic recipient *E. coli* K-12 J53 strain (*E. coli* K-12 J53 RFP kan[r]) (section 2.2.19, Figure 6.2).
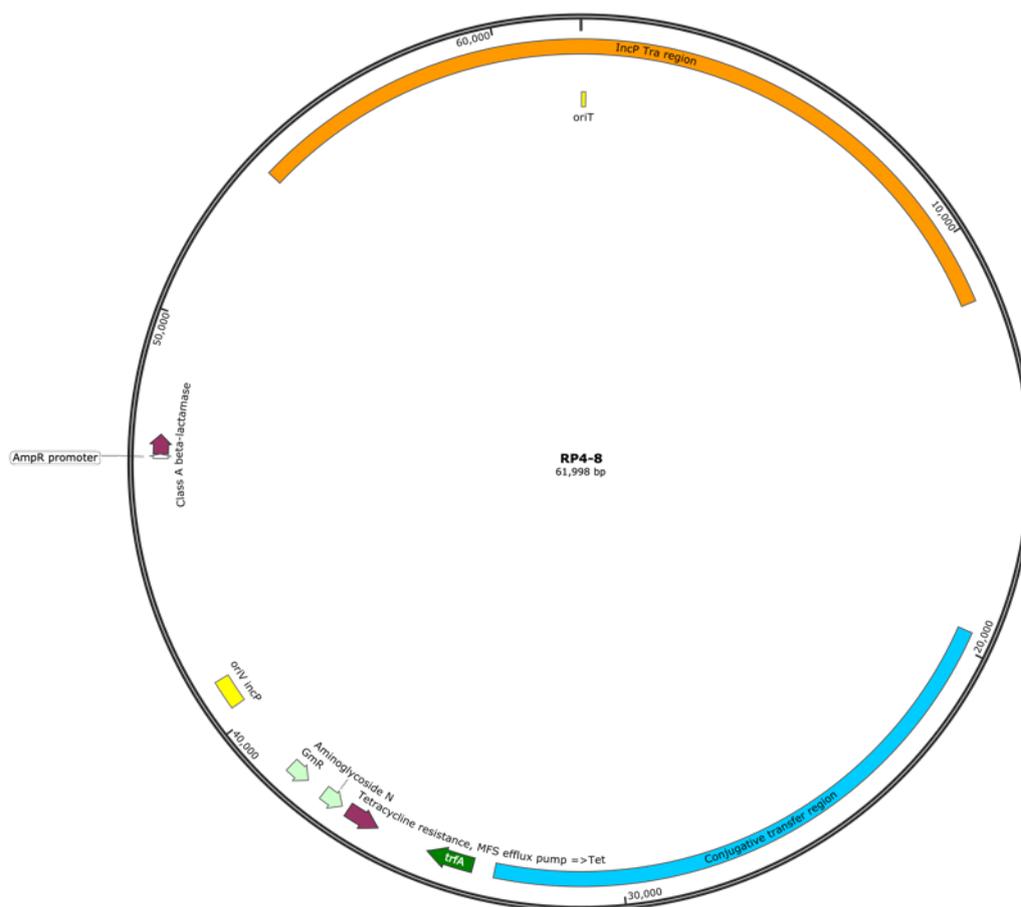

Figure 6.1 Plasmid map of RP4-8.

Plasmid map of RP4-8, showing the relevant resistance genes, origin sites and transfer regions of the IncP plasmid.

Figure 6.2 Reconstruction of the chromosomal insertion of *mrfp1* and kan[r] to the 3' region downstream of *glmS*.

mrfp1 and kan[r] were inserted to the 3' intergenic region of *glmS* using the gene doctoring method described in section 2.2.19 (Lee *et al*., 2009). *glmS* homology regions were used to guide insertion of the DNA to this region. The kan[r] gene is contained within FLP recombinase sites for future removal of the resistance gene to generate an antibiotic sensitive recipient strain.

6.2.1 Identification of Mercury (II) Stress Induced Transposition of EVAL 397 Tn*21*

The RP4-8 mediated transposition experiments conducted on Tn*21* (section 2.2.20) identified that with the addition of mercury (II) chloride ($HgCl_2$) in low concentrations (up to 2 μg mL^{-1}), transposition could be induced more readily than with no added $HgCl_2$ (Table 6.1). When there was no added $HgCl_2$, the transposition frequency was recorded as 2.44×10^{-7}. At 1 μg mL^{-1} of $HgCl_2$, a transposition frequency of 1.06×10^{-6} was observed. At 2 μg mL^{-1} of $HgCl_2$, the mean recorded transposition frequency was calculated as 2.91×10^{-5}. At concentrations greater than or equal to 4 μg mL^{-1} of $HgCl_2$ transposition was undetectable.

[$HgCl_2$] / μg mL^{-1}	Mean transposition fold-difference
1	4.25
2	27.5
4	N/A

Table 6.1 Fold difference of transposition of Tn*21* from EVAL397 compared to no added $HgCl_2$.

Three biological repeats were performed for each concentration of HgCl$_2$ and spread plated in triplicate, giving three technical repeats for each biological replicate.

When between 1 and 2 μg mL^{-1} of HgCl$_2$ was added to the donor cells, a 6.5-fold increase in transposition was observed. By increasing the concentration of HgCl$_2$, an increase of mobility of Tn*21* was recorded. Between 2 and 4 μg mL^{-1} of HgCl$_2$ was added to the donor cells, the Tn*21* transposase expressions appears to be repressed, and transposition no longer occurs. The transconjugant contained within the recipient strain *E. coli* J53 RFP kan^r was then sequenced using long-read MinION technology (2.2.25). This was to ensure resolution of the large transconjugant RP4-8 Tn*21* plasmid to confirm insertion, an example can be seen in Figure 6.3.
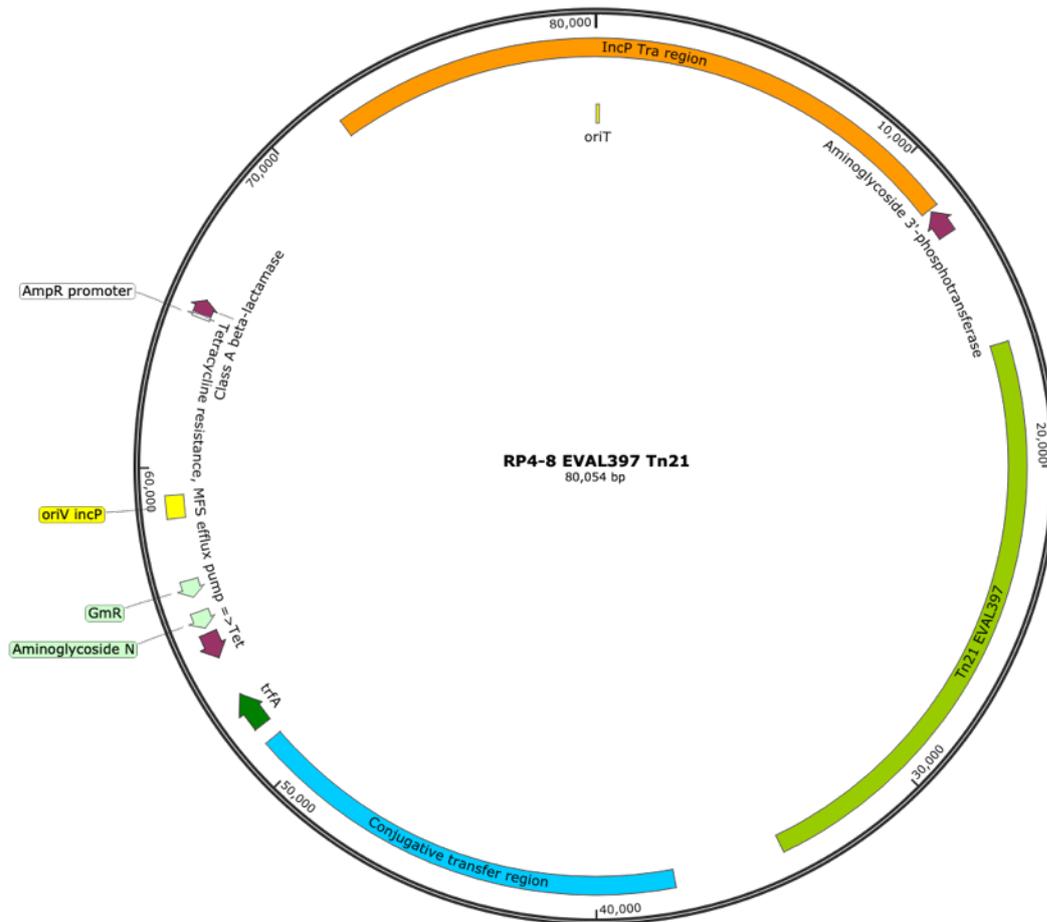
Figure 6.3 Plasmid map of RP4-8 Tn*21* transconjugant.

Tn*21* from EVAL397 inserted into RP4-8 and was sequenced using long-read MinION technology after conjugation of the transconjugant in to J53 RFP Kan^r.

6.2.2 Identification of Mercury (II) Stress Induced Transposition of BPW2-4, IS*26*-Like Mobile Element

An environmental wastewater isolate possessing a Tn*21*-like transposable element whereby the class I integron, *tnpM* and insertion sequences were missing was also screened. This transposable element produced similar results. The frequency of each transposition event for different $HgCl_2$ concentration were calculated (Table 6.2).

At 1 µg mL^{-1} of HgCl$_2$, the mean transposition frequency was 12-fold higher than when there

was no added HgCl$_2$ stress. At 2 µg mL^{-1} of HgCl$_2$, transposition frequency was 34-fold higher

than when there was no added HgCl$_2$ stress on the cells. In turn this is a 2.83-fold difference

in transposition frequency between the two concentrations of HgCl$_2$. Above 2 µg mL^{-1} of HgCl$_2$

however, transposition events were undetected.

[HgCl$_2$] / µg mL^{-1}	Mean transposition frequency	Mean transposition fold-difference from no added HgCl$_2$
None added	1.30x10^{-7}	-
1	1.56x10^{-6}	12
2	4.41x10^{-6}	34
4	Not detected	N/A

Table 6.2 Mean transposition frequency and fold difference of transposition of IS*26*-like element in BPW2-4 compared to no added HgCl$_2$.

Three biological repeats were performed for each concentration of HgCl$_2$ and spread plated in triplicate, giving three technical repeats for each biological replicate.

6.3   Discussion

Horizontal gene transfer (HGT) is a phenomenon which helps drive the evolution of bacterial diversity and adaptation to survival in constantly changing environments (Gogarten and Townsend, 2005). The process of HGT is dynamic and many bacterial genomes are affected by such events. The aim of this study was to determine whether environmental pressures such as stress from Hg(II) may influence the frequency of transposition events. Interestingly, mobilisation of a MGE from one plasmid to another was identified for Tn*21* by the formation of co-integrates and resolution to produce copies of the MGE within both resultant plasmids. Previously, natural transformation of Tn*21*-like elements was identified in *Acinetobacter spp.* however, it is not known what drives the transposition process (Domingues *et al.*, 2012). Further bioinformatic analysis beyond the transposition experiments in this study resulted in a suggested model for which regulation of the transposase gene, *tnpA*, may be modulated or mediated.

Tn*21* is a Tn*3* family mobile element, meaning it contains its own transposase genes to form co-integrates and resolution of these co-integrates, although formation of cointegrates is not obligatory (Bennett, de la Cruz and Grinsted, 1983). The transposition is therefore thought to be RecA-independent; however, on rare occasions, RecA may also allow homologous recombination without forming a cointegrate as an intermediate for such an event. RecA independent transposable elements have a lower frequency of transposition due to tighter regulation on the transposase genes.

6.3.1   RP4 plasmid series mediated conjugation

The RP4 series of transmissible plasmids provided a backbone to mediate the uptake of large, resolved transposons. RP4-8 was used as an intermediate for transfer of the Tn*21* from the donor strains to the recipient strain as both transposable elements were both plasmid-borne initially. The rationale, therefore, was to standardise the type of mobility of the transposable elements detected. Direct conjugation of transferrable IncF plasmids would not reflect true transposition of the mobile elements, but instead conjugation of a plasmid. It was key for the *E. coli* J53 RFP kan^r recipient strain not to be subjected to unnecessary mercury stress during the conjugation of the modified RP4-8 plasmid to prevent damage or death of the recipient cells and ensure optimal transfer of the plasmid. A wash step of the donor cells to remove free Hg (II) was therefore essential before mixing with the recipient cells.

6.3.2   Oxidative stress enhances the formation of Tn*21* and Tn*21*-like elements

Both strains carrying these mercury resistance transposons mobilised the genetic elements at sub-lethal concentrations of mercury (II) chloride. At more toxic concentrations Tn*21* transposition was not detected, suggesting that increased stress to potentially toxic levels may inhibit transposition due to transcription activity of the *mer* operon. The homo-dimeric structure of MerR allows strict control of expression of the *mer* operon by winding and underwinding a 19 bp region. A recent study by LaVoie and Summers (2018) described growth of bacterial strain *E. coli* MG1655 sensitive to mercury (II) chloride was consistently reduced but not completely inhibited, at 3 µM HgCl$_2$, this concentration was sublethal for HgCl$_2$ sensitive strains (LaVoie and Summers, 2018). Yu, *et al*. (1996) showed transcription of the

mer operon was operating at 10% of maxima at 50 nM of HgCl$_2$ but activity increased to 90% of maximum transcription at 200 nM. From this study, at concentrations of 1 and 2 μg mL^{-1} (3.7 and 7.4 μM, respectively), *mer* operon expression is predicted to be at a maximal rate and starting to decline (Yu, Chu and Misra, 1996). Above these concentrations, transcription rate is reduced, as is the mobility of the transposable elements. Given the differences of these two transposable elements, it is somewhat surprising to report transposition at a similar frequency.

Tn*21* and Tn*501* were previously identified to have putative promoter regions for their transposase genes *tnpA* and *tnpR* within the *res* sites; however, transposase expression is not necessarily solely controlled by the transcription from these sites. Tn*501* demonstrated that transposition could be enhanced when stimulated by growth in the presence of Hg (II) (Kitts *et al.*, 1982). Greater than 90% of cointegrate formation was also noted, meaning a 90% increase in transposition rate. The reasoning for this was thought to be due to the transcription of the *mer* operon, upstream of the *res* sites of *tnpA* and *tnpR*. Secondary transcription control is therefore likely regulated by the MerR homodimer. In Tn*21* and *Tn21*-like elements however, the *res* sites, and the *tnpA* and *tnpR* genes are separated from the *mer* operon by the class I integron (Figure 1.11 Figure 1.13). Instead, *tnpM* was thought to modulate *tnpA* and *tnpR* expression (Hyde and Tu, 1985), although this could not be substantiated. From the data collected, it may be possible to suggest physiological change in the host cell environment may cause induction of transposition. Further to this, subjecting organisms to other oxidising stressors may display similar changes to transposition frequencies.

Similarly, the understanding of antibiotics modes of action which result in lethality is evolving (Dwyer *et al.,* 2014). It has been theorised that upon the action of antibiotics, they may cause physiological effects resulting in production of reactive oxygen species which may result in cell death (Dwyer *et al.*, 2014). Hg (II) is a redox-inactive metal which depletes cells of their antioxidant compounds, particularly thiol-containing compounds (Ercal, Gurer-Orhan and Aykin-Burns, 2001). This imbalance caused by Hg (II) toxicity may therefore have a consequent effect which induces stress responses to mobilise such insertion sequences and mobile elements. Other Tn*3* family elements mobility have been shown to be modulated by the actions of proteins involved in general stress response. These stressors include deprivation of carbon sources, UV light and temperature (Kretschmer and Cohen, 1979; Blaseio and Pfeifer, 1990; Eichenbaum and Livneh, 1998; Lamrani *et al.*, 1999). The results in this study suggest that generation of oxidative stress such as Hg (II) may have a similar effect. The mechanisms for this modulation are less well characterised. The *tnpR* and *tnpA* genes are thought to be transcribed from the same promoter region, where the resolvase binding site is located (Sherratt, 1989).

6.3.2.1   Analysis of Transposase Promoter Region

Upon bioinformatic analysis of Tn*21* and the Tn*21*-like elements in this study, a homologous region with 76% identity to the OxyR binding consensus sequence 5'-GATAGBYHWDRVCTATC-3' was identified within the second resolvase binding site upstream of the *tnpR* and *tnpA* genes, between -62 and -79 from the start codon of *tnpR* (Salgado *et al.,* 2013). We propose that the identification of this may further support the hypothesis that perhaps sub-lethal

oxidative stress may enhance the formation of co-integrates in Tn*21* and Tn*21*-like transposases.

Tn*21*, like many other transposable elements has been widely disseminated globally, likely due to human carriage, but dissemination can occur horizontally as well as vertically. The mobilisation and resolution mechanisms of Tn*3* transposons have been well characterised as a family, but there is more knowledge required to fully understand how the transposition mechanism is induced in Tn*21.*

6.4   Conclusions

The short study showed that a Tn*21*-like element and an IS*26*-like element were capable of mobilising under sub-lethal stress from Hg (II). In fact, both elements were not only able to mobilise, but transposition frequency was increased at low concentrations of Hg (II), up to 2 µg mL^{-1}. These results build on the work of other groups that previously determined that mobility of transposons such as Tn*501* is enhanced by sub-lethal concentrations of Hg (II). This work was performed in the late 1980s and early 1990s where a solid theory by Kitts *et al.* (1982) was constructed with the knowledge at the time. Literature on Tn*21* and other Tn*3* family transposases suggested that transposase regulation was not well characterised at that time. However, with time, the understanding of other pathways such as oxidative stress response and the roles of such proteins within this signal cascade helps to expand these hypotheses. The model that could be suggested is that Tn*3* family transposases, in particular Tn*21* may be regulated by the oxidative stress signal response; in this case, OxyR.

6.5 Future work

The data collected by inducing transposition with Hg (II) stress and the analysis of the promoter region 5' of *tnpR* and *tnpA* in Tn*21* suggests OxyR may play a role in the mobilisation of this transposon by regulating transposase *tnpA* transcription regulation. In order to test this hypothesis, the following experiments should be considered. Firstly, the promoter region of the transposase genes should be cloned upstream of a reporter element such as *luxCDABE* in a reporter plasmid like pLUX (Burton *et al.*, 2010). *luxCDABE* reporter or other luciferase reporters may be used over other reporter elements such as fluorescence or beta-galactosidase as they are more sensitive than other reporters and also have lower background signal compared to fluorescent reporters (Hakkila *et al.*, 2002). This reporter plasmid could then be transformed into a lab *E. coli* strain and in a strain which has *ΔoxyR*. Next, the two transformed strains should be exposed to sub lethal amounts of different oxidative stressors, UV, H_2O_2, Hg (II) and other antimicrobials over a one-hour duration and record luminescence change against the wild-type. Secondly, RT-qPCR to measure change in expression of the reporter gene in wild type and *ΔoxyR* strains between no stress and facing different amounts and different stressors causing oxidative stress. Finally, a bacterial strain containing Tn*21* could be manipulated to generate *ΔoxyR* to measure the transposition frequency compared to wild-type host. An *ΔoxyR* strain should result in the reduction or prevention of Tn*21* mobility when oxidative stress is applied compared to the wild-type host. Within the same knock-out strain however, no change to the transposition rate would be expected. Without oxidative stress, it is hard to predict the expression rate and transposition frequency of the *tnp* genes and Tn*21*, respectively; however, Tn*21* should not differ from the wildtype.

Finally, as noted in the discussion, transposition rate changes seemed to mirror that of the *in vivo* study on *mer* operon expression performed by Yu, Chu and Misra (1996). In order to further determine the concentration of $HgCl_2$ induces transposition of Tn*21*, repeating the transposition induction experiments using a finer scale of concentrations of $HgCl_2$ between no added $HgCl_2$ and 4 µg mL^{-1} may show where transposition rate begins to increase and also further identify where transposition stops. To confirm this, RNA should be extracted from each concentration of mercury stress for real-time qPCR to examine the expression rate of Tn*21 tnpA*. Furthermore, using the same refined scale of $HgCl_2$ concentration increments may be used to identify expression of the *mer* operon. The data alongside each other from the same strains may identify whether *mer* operon transcription may start as transposition frequency reduces.

Chapter 7: General Discussion and Concluding Remarks

7.1   Discussion

The aims of this research, described in Chapter 1, were:

1) To further understand the prevalence of Tn*21* in wastewater and what drives carriage persistence, and evolution, of this large mobile element.

2) To detect which organisms within the wastewater environment, carry Tn*21* and Tn*21*-like transposons and assess the variation of class I integron gene cassette arrays within these mobile elements.

3) Finally, work was aimed at further understanding how the mobile element may be mobilised and disseminates to the wider environment.

7.1.1   Prevalence of Tn*21* in Wastewater

The qPCR data in Chapter 3 suggested that Tn*21* could be used as a biomarker to detect the spread of antimicrobial resistance in wastewater. The data collected also showed that Tn*21* is lost over the wastewater treatment process, but the overall loss of bacterial DNA over the process is not substantial. This suggests that the abundance of different bacterial families changes over the wastewater treatment process. Recent studies like Auguet *et al* (2015) demonstrated this change across biofilm communities isolated from sewer systems. Seasonal change in wastewater bacterial communities have also been observed (Zhang *et al.*, 2018). In addition to these data, epicPCR from the wastewater effluent sample suggested that Tn*21* may be more prevalent in *Klebsiella* and *Enterobacter* genera, compared to other *Enterobacteriaceae* such as *E. coli*, which was the initial focus in Chapter 3. The overall survival of *Klebsiella* and *Enterobacter* through the wastewater treatment process suggests a selective

316

advantage for these genera to disseminate into the environment from the effluent. These data further suggested that to isolate more Tn*21* from the environment and gain a better understanding of the carriage of this mobile element, enterobacteria rather than *E. coli* should be selectively isolated. The use of a less selective media such as MacConkey or CHROMagar rather than TBX may be more favourable. The WHO suggest that genera such as *Escherichia* and *Enterococcus* should be used as indicators of faecal pollution, resulting in large investment into their monitoring (including antimicrobial resistance) by researchers and governments around the world (Santo Domingo and Edge, 2010). However, surveillance of resistomes of these genera only represents a small portion of the bacterial species which are capable of lateral gene transfer. Further monitoring using increasingly affordable metagenomic techniques may allow for better understanding.

The prevalence of Tn*21* has long been noted globally since its discovery in the mid 1950s (Nakaya, Nakamura and Murata, 1960), and the identification in this suburban wastewater treatment plant (and geographically close farm) is just one such example. One interesting result was the isolation of *E. coli* downstream of the wastewater treatment plant four months apart from each other of identical clones, possessing the same plasmids, which carried Tn*21*. Whether these particular isolates originated from the wastewater treatment plant or not, the persistence of particular *E. coli* clones containing such resistance genes is therefore somewhat troubling. Another study identified identical plasmids from different species containing carbapenemase producing genes in East Anglian wastewater treatment plants confirming that these incidents are not geographically localised but likely to be taking place globally (Ludden *et al.*, 2017). This thesis supports the claims made by Ludden *et al* (2017) that the

wastewater treatment process does not at present stop the transfer of antimicrobial resistance genes from influent to effluent.

7.1.1.1 Drivers for the Occurrence and Diversity of Tn*21* and Tn*21*-like Elements in Wastewater

Selection for resistance to antimicrobials and the impact of anthropogenic pollution was unknown when antimicrobial metals and antibiotics were first introduced and is now one of the most pressing challenges for the human population to tackle. Clinical mercury resistance determinants were first found linked to both Gram-positive and Gram-negative isolates in *Staphylococcus aureus* and *Shigella flexneri* in the 1950s and 1960s (Moore, 1960; Nakaya, Nakamura and Murata, 1960). Mercury is no longer used in medicine, or as disinfectants, and is being eliminated in industrial uses; however, resistance to the antimicrobial metal is still present in clinical bacterial strains, and in human and animal isolates (Mchugh *et al.*, 1975; Hanczvikkel *et al.*, 2018; Argudín, Hoefer and Butaye, 2019). As Gram-negative bacteria containing mercury resistance transposons, including those containing class I integrons, have since been identified frequently around the globe, understanding the drivers for its continued transmission is key for the development of solutions to remove, or more realistically, to reduce the number of metal and antibiotic resistant microorganisms and their DNA from wastewater.

It is difficult to suggest from the data collected whether the occurrence of Tn*21* and Tn*21*-like transposable elements may be due to selection from co-occurrence of antimicrobial resistance genes, or whether co-selection drives the carriage of mercury resistance

transposons. This is because due to COVID-19, wastewater samples could not be collected over a prolonged timeframe or to determine the concentration of Hg (II) within the wastewater environment. However, from the results, it is possible to speculate to some extent about factors which may cause the retention of these mobile elements.

Firstly, the presence of active class I integrons in these mobile elements in wastewater samples gives some insight into the dynamics of these Tn*21*-like mobile elements as change in cassette diversity was identified over the wastewater treatment process. This suggests that the wastewater treatment possibly selects for particular gene cassette arrays or the MGEs or organisms that carry them. Moreover, the higher variation of cassette arrays across the wastewater influents suggests that the integron is likely the driver of Tn*21* retention amongst the Gram-negative bacterial population; where a resistance gene cassette carried within the gene cassette array may provide a selective advantage in the survival of the host organism. However, without taking more regular wastewater samples it is hard to be certain.

Mercury containing compounds are now very rarely used, if at all, in modern medicine. It is somewhat surprising that mercury resistance is still identified in human clinical and wastewater isolates. However, according to current theory on microbial evolution, genetic turnover and lack of selection should result in the loss of such resistance genes (Vos *et al.*, 2015). This idea is supported by the whole genome sequences of isolates EVAL51 and EVAL55. These two strains were not isolated in wastewater, but they had lost the *mer* operon, except for *merR*. The 3' terminal of the ORF of *merR* forms part of the 38 bp IR required for transposition of the Tn*21*-like element, and therefore there is a selective pressure for retention of this DNA sequence even if the rest of the *mer* genes are lost. Evolutionary traits

319

such as the loss of a complex resistance operon may suggest its redundancy in some environments. Similar incidences of phenotypic reversion have been identified in environments where antimicrobial stressors are removed, but in antimicrobial resistances acquired by mutations (Dunai *et al.*, 2019). Current medical and veterinary antibiotic use could co-select for antibiotic resistance gene cassettes in integrons, leading to an increase in the frequency of Tn*21*-like elements without the *mer* operon in the future (Pal *et* al, 2015). At present mercury resistant bacteria and *mer* operons are still isolated frequently.

The other reason for the retained presence for Tn*21* carriage in wastewater is likely due to the carriage of the mobile element on large transmissible plasmids. This is perhaps the most important factor in horizontal gene transfer for Tn*21* (Liebert, Hall and Summers, 1999). The transposition mechanism of Tn*21* is RecA-independent (Martinez and de la Cruz, 1988). This means the resulting transposition frequency is much lower than RecA-dependent mechanisms because RecA-independent transposons rely on their own transposases. In turn it is possible that the transposition of Tn*21* to these large mobile plasmids may be where these mobile elements remain the most stable. Data from Chapter 6 suggested that oxidative stress may drive the induction of Tn*21* transposition, but the frequency increase is still negligible when compared to RecA-dependent transposons.

Moreover, data from Chapter 3, which examined links between the co-occurrence of different antimicrobial resistances identified that tetracycline resistance was individually linked with the occurrence of resistance to cefotaxime, trimethoprim, streptomycin and mercury (II) chloride in *E. coli* isolated from wastewater influents and effluent. Resistance to mercury (II) chloride was also linked to the presence of trimethoprim resistance. Similar

instances of Tn*21*-like transposons carrying other transposons were shown in the Tn*21* evolutionary pathway by Liebert *et al* (1999) (Figure 1.12). The same wastewater samples were frozen and then prepared for targeted amplicon metagenomic sequencing to identify gene cassettes within Tn*21*-like class I integrons, which is described in Chapter 5. Resistance determinants to trimethoprim (*dfrA*, *dfrB*) were found in all but one sample (Effluent term time–21/01/2020), suggesting that this co-occurrence of both mercury and trimethoprim resistance is located on the same mobile genetic element.

Studies on mercury resistance in commensal *E. coli* from the human gut are rare, but one study found that 27% of the human population were carriers of mercury resistant *E. coli* strains; and in comparison, human populations that were known to have been exposed to mercury had a 6% higher occurrence of mercury resistant *E. coli*; which suggested mercury did drive selection in such environments (Skurnik *et al.*, 2010). Similarly, it was suggested that dental amalgam may contribute to the presence of mercury resistance (Summers *et al.*, 1993; Bender, 2008). This study suggests that the sample environment shapes the selective pressures that the microbiome is exposed to. For example, quaternary ammonium compound resistance in class I integrons has previously been linked with the persistence of class I integrons due to the gain of a selective advantage in the human environment due to the widespread use of QACs (Gillings, Xuejun, *et al.*, 2008). However, it is likely that biocides such as QACs have driven persistence of Tn*21* as their use has not been regulated in the same manner that antibiotics are. Previous work on co-occurrence using inverse PCR techniques, found that Tn*21* and other resistance determinants associated with this mobile element (such as QAC resistance) were found on large plasmids rather than carried on the chromosome (Pal

et al., 2015). The meta-analysis also suggested that the potential for co-selection of resistance to biocides, metals, antibiotics and other antimicrobials is higher in clinically important bacterial strains, possibly due to the abundance of large plasmids which were found to be associated with toxin anti-toxin systems which also helps explain their persistence (Pal *et al.*, 2015). This may explain why the same Tn*21*-carrying *E. coli* clone was isolated four months apart. The toxin anti-toxin system found on the large IncF plasmids in these isolates likely drives the persistence of this bacterial clone (Page and Peti, 2016). For an organism containing such a system on a plasmid, it can no longer survive without the plasmid. The toxin produced causes cell death and has a long half-life whereas the anti-toxin's half-life is much shorter, ensuring the propagation of the plasmid.

The carriage of Tn*21* through the wastewater treatment process has been explored across this thesis, but the reasons for its prevalence are not so well understood. Previous studies have used samples collected from wastewater treatment plant influent as a proxy to determine the diversity of human gut microbiomes (Gillings *et al.*, 2015; Newton *et al.*, 2015; An *et al.*, 2018). However, it is important to note that microbial cells isolated from wastewater are not just from the collective human population gut. Part of the microbiome in the influent will come from other reservoirs, such as UTI causing bacteria, rainwater run-off, hospital, food manufacturing, field run-off rodents and wild animals (Department of Health, 2014). Chapter 5 explored cassette variability of gene cassette arrays, where resistance to trimethoprim was abundant across all samples. Trimethoprim is predominantly used in the treatment of urinary tract infections and a number of clinical UTI isolates have been found to carry Tn*21*-like mobile genetic elements suggesting this may be one of many reservoirs of Tn*21* dissemination

into the wastewater in the first place (Márquez *et al.*, 2008; National Institute for Health and Care Excellence, 2017).

It is important to remember that selection for acquisition of resistance genes is multi-faceted. Wastewater treatment plants have high volumes of water constantly flowing through containing pharmaceuticals, illegal drugs, hormones, antibiotics and biocides which may elicit a selective response (Castiglioni *et al.*, 2008; Blair *et al.*, 2013; Rodriguez-Mozaz *et al.*, 2020). Tn*21* itself is a complex multi-antimicrobial resistance carrying element, whereby multiple selection factors from compounds to wash laundry, mouthwashes, clinical antibiotic usage and exposure to mercury or tetracycline may all drive the selection of this mobile element (Östman *et al.*, 2017). The areas explored within this thesis explores but a small part of this wider selective process.

7.1.2   The Future of Wastewater and Agricultural Waste Management

The multi-factorial problem of resistance gene dissemination into the environment after wastewater treatment processing therefore needs careful consideration to address it. The widely adopted system for wastewater treatment in the UK uses activated sludge to treat raw influent, an old but widely used method (the sampled site from this study uses trickling filter, an older system) (Hyde, 1937). This particular method of treatment is rather energy intensive, using electrical energy to generate oxygen bubbles in the sludge for other microorganisms to break down organic compounds within the sewage liquor. The rural wastewater treatment plant which water was sampled from in this thesis is not the typical treatment plant. First, influent is passed through settling tanks to remove larger particulates from the water. The

323

clarified liquid goes through a trickling filter. The settled solids are then collected and taken off site for anaerobic digestion. Next, the clarified liquor is passed through sand filtration before discharge to remove fine particulates.

It is estimated that 1% of Europe's energy usage goes into wastewater treatment, with the UK spending £15 million yearly to run these activated sludge sites (Ruijgers, 2021). This process has been shown to be inefficient in removing bacterial load, rather causing selection for *Betaproteobacteria* which changes the diversity of the effluent microbiota (Płaza *et al.*, 2021). Co-selection has been demonstrated at population and species levels and confirmed by examining the relationship of sub-inhibitory antimicrobial agents to track resistance gene expression (Pal et al., 2015b; Ju et al., 2018). From these studies and the work performed here, it is clear that the current state of wastewater treatment is becoming more unsustainable in both energy use and in recycling water before release so that it may be potable. As the global human population continues to rise, the strain on inefficiencies of wastewater treatment will be exacerbated further.

To combat this, the development of new technologies to replace or be used in synergy with what is a rather archaic treatment process are being investigated. There are multiple steps within the wastewater treatment process which each play different roles in the sanitisation of wastewater before release back into larger bodies of water. The processes of relevance to this thesis however are the implementation of disinfection steps. The first main technology of interest is the implementation of UV treatment of clarified liquor. A step which should be additional to current treatment processes before release of wastewater as effluent.

The number of publications examining the use of this UV and photocatalysis methods have been widely investigated but are still prohibitively expensive (Iervolino *et al.*, 2019). However, ultra-violet radiation (UV-C) is known to be capable of inactivating organic active compounds such as pharmaceuticals, one of the other main contaminants found within wastewater, as well as being able to inactivate DNA and RNA (Cutler and Zimmerman, 2011; Barrett *et al.*, 2016). A problem with the sole use of UV-C is that DNA damage may be repairable, using the cells' own DNA repair mechanisms, however, in comparison addition of hydrogen peroxide, (H_2O_2) generates free radicals which prevent the repair of DNA (Hassen *et al.*, 2000; Pablos *et al.*, 2013). The UV irradiation process combined with hydrogen peroxide allows for fast disinfection but is still rather expensive and energy intensive. The same study did however suggest that UV-A treatment, by harnessing the sun's energy in conjunction with fixed TiO_2 could be used as an alternative to combat some of these disadvantages (Pablos *et al.*, 2013). This technology has a longer kill step but is also more efficient in removing pharmaceuticals and requires less monitoring to ensure disinfection.

The use of anaerobic digestion is already widely used. Whilst this would not remove bacterial load but change the diversity of it; it may be more sustainable and more easily implemented across both wastewater and agricultural environments. Digestate and gas collected may be used to offset the energy usage (Holm-Nielsen, Al Seadi and Oleskowicz-Popiel, 2009). In Denmark, this has already been performed successfully, where digestate is used to power the treatment plant and produce excess electricity to power the entire local energy grid (TheCivilEngineer.org, 2017). It is important to note that there are factors to consider which may affect the technologies used in a geographical area, such as the type of water coming as

325

influent to a treatment plant, whether industrial or domestic. Within an agricultural setting, the uptake of anaerobic digestion could be advantageous for similar reasons. Gasses collected may be used to generate electricity to aid powering of the farm and digestate may provide richer materials for use on the land with a lower risk of antimicrobial resistance gene dissemination to soils, which would normally risk leaching into water courses (Holm-Nielsen, Al Seadi and Oleskowicz-Popiel, 2009). These are but two of many techniques being investigated. The introduction of new technologies to reduce the dissemination of antimicrobial resistance genes to the wider environment is still likely a few years away, but research into solutions is promising. The main limiter is cost of investment into these technologies, whatever they may be in the future (Iervolino *et al.*, 2019).

7.2   Conclusion

This study has highlighted that despite the literature reviews on Tn*21*, there is still a lot to be understood about this mobile element. It still appears to play a vital role in providing a pathway to spread resistance genes, which are carried by the class I integron, and it is likely to be important to its continued persistence within the environment. The Tn*21* transposon has a lower transposition rate compared to Tn*1* and Tn*7* transposable elements with transposition rates normally between 10^{-5} and 10^{-7}. This is in part due to its size, but more due to being a RecA independent transposition mechanism (Sherratt, 1989). It is likely disseminated further due to its carriage on larger transmissible plasmids. It is unclear whether the persistence of Tn*21* within the environment is due to co-selection or co-occurrence. However, changes in gene cassette array variation were documented across the wastewater

treatment plant. This suggests that co-selection may be likely the driver for the persistence of mercury resistance around the globe, however, the dataset in this study is limited partially due to COVID-19 and co-selection needs to be confirmed by quantifying antimicrobial stressors present in each particular sample time point and location. Class I integrons have been found to be persistent in multiple environments other than wastewater treatment plants and are efficient at disseminating resistance genes between species (Gillings, 2014). Data in this thesis suggests that the mercury resistance operon may now be accessory to the transmission and dominance of class I integrons. Evidence from whole genome sequences from a neighbouring dairy unit supported this conclusion, where almost the entirety of the *mer* operon was lost in isolates EVAL51 and EVAL55. Conversely, isolates containing large IncF and IncI plasmids were detected within wastewater samples which carried a *mer* operon with Tn*21* and no In2 region or were part of IS*26*-like mobile elements (Figure 4.5) which may also indicate that the evolution of Tn*21* is divergent.

Evidence from this study suggests that the carriage of Tn*21* within the wastewater environment may be due to the carriage of class I integrons on self-transmissible plasmids. To reduce the further spread of antimicrobial resistance genes into the environment, more action needs to be taken in the sanitisation of wastewater and agricultural faecal waste before either releasing to rivers or spreading onto soils. It is equally important to note that Tn*21* is but one of many mobile elements responsible for the horizontal transfer of resistance genes within the microbiome. Similarly, wastewater is only one reservoir for the dissemination of resistance genes into the environment. Understanding the way that the spread of disease, antimicrobial resistance, and other pollutants are released by humans into

the wider environment affects the biodiversity has been highlighted within scientific literature. The development of measures to control these pollutants are necessary to prevent further ecological impact.

7.3   Future Direction of Work

To further the understand the persistence of Tn*21* within wastewater co-selection and co-occurrence needs more investigation. Screening wastewater samples for antimicrobial compounds such as metals, biocides and antibiotics in conjunction with bacterial replica plating experiments like the one performed in Chapter 3 could help to identify whether Tn*21* is co-selected by another antimicrobial. The presence of Tn*21* could then be confirmed using PCR on isolates that possess mercury resistance. In addition, performing the replica plating on a greater number of bacterial isolates per sample over a longer time period from each sample location would enable the development of a model to identify the antimicrobial selection pressures that microorganisms found within wastewater face. Bacterial isolates should also be screened for resistance to more classes of antimicrobials including other toxic metals and biocides, such as silver, copper or chlorhexidine. qPCR analysis suggested that Tn*21* abundance is reduced over the wastewater treatment process. This should similarly be performed over more time points across the year to identify whether seasonal change may affect the abundance of Tn*21* within wastewater. Silver resistance (*sil*) and copper resistance (*pco*) operons were identified in a number of isolates that were whole genome sequenced. To add to the work on co-selection and co-occurrence, qPCR could be used to identify whether similar trends in the carriage of these resistance operons occurs. In the whole

genome sequences of isolates sequenced in Chapter 4, the *sil* and *pco* operons were identified chromosomally.

Both replica plating and qPCR methods could also be used to identify the presence of Tn*21* and antimicrobial selective processes in other larger wastewater treatment plants, upstream and downstream of effluent release and non-water environments. Similar investigations could be performed on agricultural slurry amended fields and animal husbandry sites to identify any differences between systems.

Data from the epicPCR showed that using selective media such as TBX only helped identify a limited number of bacterial isolates that may carry Tn*21*. In the future, isolation of mercury resistant bacterial isolates should take place on MacConkey agar supplemented with mercury (II) chloride. AST using disc diffusion assays is useful to determine which antimicrobials each isolated organism was resistant to. However, to understand more about the resistances these isolates carried, minimum inhibitory concentration (MIC) testing should be performed.

The use of epicPCR proved to be a good tool to identify which organisms carry Tn*21* within the effluent. In the future, the PCR reagents from this thesis could be used in the newer epicPCR 2.0 method to increase the reliability of the epicPCR technique (Roman, *et al.*, 2021). This method could then be used on more wastewater samples, collected from various stages across wastewater treatment to identify how the diversity of Tn*21* carrying organisms varies. Further validation of the long-read epicPCR is also required. Comparison of the long-read epicPCR against the short-read technique should be performed.

Finally, Chapter 6 focussed on Tn*21* transposition induction using mercury (II) chloride. The data suggested that both Tn*21* and the IS*26*-like element which were tested were mobilised by low concentrations of mercury (II) chloride. To further understand the mechanisms of Tn*21* and Tn*3* family transposition, promoter analysis and expression analysis with reporters in wild type and *ΔoxyR* knock out strains are needed. Tn*21* is a complex transposon, a lot of information is still required to fully understand its continued prevalence, either by co-selection or co-evolution, and its mechanism for induction of mobility is still not fully understood.

References

Aertsen, A. and Michiels, C. W. (2006) "Upstream of the SOS response: figure out the trigger," *Trends in Microbiology*, 14(10), pp. 421–423.

Amos, G., Gozzard, E., Carter, C. E., Mead, A., Bowes, M. J., Hawkey, P. M., Zhang, L., Singer, A. C., Gaze, W. H. and Wellington, E. M. H. (2015) "Validated predictive modelling of the environmental resistome," *The ISME Journal*, 9(6), pp. 1467–1476.

Amos, G., Ploumakis, S., Zhang, L., Hawkey, P. M., Gaze, W. H. and Wellington, E. M. H. (2018) "The widespread dissemination of integrons throughout bacterial communities in a riverine system," *The ISME Journal*, 12(3), pp. 681–691.

An, X.-L., Chen, Q.-L., Zhu, D., Zhu, Y.-G., Gillings, M. R., Su, J.-Q. and Elliot, M. A. (2018) "Impact of Wastewater Treatment on the Prevalence of Integrons and the Genetic Diversity of Integron Gene Cassettes," *Applied and Environmental Microbiology*, 84, pp. 2766–2783.

Andrews, S. (2010) *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed: September 29, 2021).

Andrews, S. C., Robinson, A. K. and Rodríguez-Quiñones, F. (2003) "Bacterial iron homeostasis," *FEMS Microbiology Reviews*, 27(2–3), pp. 215–237.

Ansari, Aseem. Z., Chael, Mark. L. and O'Halloran, Thomas. V. (1992) "Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR," *Nature*, 355, pp. 87–89.

Arakawa, Y., Murakami, M., Suzuki, K., Ito, H., Wacharotayankun, R., Ohsuka, S., Kato, N. and Ohta, M. (1995) "A novel integron-like element carrying the metallo-β-lactamase gene bla(IMP)," *Antimicrobial Agents and Chemotherapy*, 39(7), pp. 1612–1615.

Arber, W. (2000) "Genetic variation: molecular mechanisms and impact on microbial evolution," *FEMS Microbiology Reviews*, 24(1), pp. 1–7.

Argudín, M. A., Hoefer, A. and Butaye, P. (2019) "Heavy metal resistance in bacteria from animals," *Research in Veterinary Science*, 122(November 2018), pp. 132–147.

Arlet, G., Nadjar, D., Herrmann, J., Donay, J., Rouveau, M., Lagrange, P. and Philippon, A. (2001) "Plasmid-mediated rifampin resistance encoded by an arr-2-like gene cassette in Klebsiella pneumoniae producing an aCC-1 class c β-lactamase," *Antimicrobial Agents and Chemotherapy*, 45(10), pp. 2971–2972.

Arthur, A. and Sherratt, D. (1979) "Dissection of the transposition process: A transposon-encoded site-specific recombination system," *Molecular and General Genetics MGG 1979 175:3*, 175(3), pp. 267–274.

Auguet, O., Pijuan, M., Batista, J., Borrego, C. M. and Gutierrez, O. (2015) "Changes in microbial biofilm communities during colonization of sewer systems," *Applied and Environmental Microbiology*, 81(20), pp. 7271–7280.

Ayliffe, G. A., Collins, B. J., Brightwell, K. M. and Lowbury, E. J. (1969) "Varieties of Aseptic Practice in Hospital Wards," *The Lancet*, 294(7630), pp. 1117–1120.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., *et al.* (2008) "The RAST Server: Rapid Annotations using Subsystems Technology," *BMC Genomics*, 9(1), pp. 1–15.

Baker, K. S., Burnett, E., McGregor, H., Deheer-Graham, A., Boinett, C., Langridge, G. C., Wailan, A. M., Cain, A. K., Thomson, N. R., Russell, J. E. and Parkhill, J. (2015) "The Murray collection of pre-antibiotic era Enterobacteriacae: a unique research resource," *Genome Medicine*, 7(1), p. 97.

Baker, M., Williams, A. D., Hooton, S. P. T., Helliwell, R., King, E., Dodsworth, T., Baena-Nogueras, R. M., Warry, A., Ortori, C. A., Todman, H., Gray-Hammerton, C. J., Pritchard, A. C. W., Iles, E., Emes, R. D., Cook, R., Jones, M. A., Barrett, D. A., Kypraios, T., Ramsden, S. J., *et al.* (2021) *Antimicrobial resistance in dairy slurry tanks: a critical point for measurement and control* .

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. and Pevzner, P. A. (2012) "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing," *Journal of Computational Biology*, 19(5), p. 455.

Barkay, T., Kritee, K., Boyd, E. and Geesey, G. (2010) "A thermophilic bacterial origin and subsequent constraints by redox, light and salinity on the evolution of the microbial mercuric reductase," *Environmental Microbiology*, 12(11), pp. 2904–2917.

Barkay, T., Miller, S. M. and Summers, A. O. (2003) "Bacterial mercury resistance from atoms to ecosystems." *FEMS microbiology reviews*, 27(2–3), pp. 355–84.

Baronti, C., Curini, R., D'Ascenzo, G., Di Corcia, A., Gentili, A. and Samperi, R. (2000) "Monitoring natural and synthetic estrogens at activated sludge sewage treatment plants and in a receiving river water," *Environmental Science and Technology*, 34(24), pp. 5059–5066.

Barrett, M., Fitzhenry, K., O'Flaherty, V., Dore, W., Keaveney, S., Cormican, M., Rowan, N. and Clifford, E. (2016) "Detection, fate and inactivation of pathogenic norovirus employing settlement and UV treatment in wastewater treatment facilities," *Science of The Total Environment*, 568, pp. 1026–1036.

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y. and Dekker, J. (2012) "Hi–C: A comprehensive technique to capture the conformation of genomes," *Methods*, 58(3), pp. 268–276.

Bender, M. (2008) *A Report to US House of Representatives Government Oversight Subcommittee on Domestic Policy Assessing State and Local Regulations to Reduce Dental*

Mercury Emissions Facing Up to the Hazards of Facing Up to the Hazards of Mercury Tooth Fillings Mercury To.

Bennett, P. M., de la Cruz, F. and Grinsted, J. (1983) "Cointegrates are not obligatory intermediates in transposition of Tn3 and Tn21," *Nature 1983 305:5936*, 305(5936), pp. 743–744.

Berglund, B. (2015) "Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics," *Infection Ecology & Epidemiology*, 5(1), p. 28564.

Bevan, E. R., Jones, A. M. and Hawkey, P. M. (2017) "Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype," *Journal of Antimicrobial Chemotherapy*, 72(8), pp. 2145–2155.

Bissonnette, L., Champetier, S., Buisson, J. P. and Roy, P. H. (1991) "Characterization of the nonenzymatic chloramphenicol resistance (cmlA) gene of the In4 integron of Tn1696: similarity of the product to transmembrane transport proteins," *Journal of bacteriology*, 173(14), pp. 4493–4502.

Blair, B. D., Crago, J. P., Hedman, C. J., Treguer, R. J. F., Magruder, C., Royer, L. S. and Klaper, R. D. (2013) "Evaluation of a model for the removal of pharmaceuticals, personal care products, and hormones from wastewater," *Science of The Total Environment*, 444, pp. 515–521.

Blaseio, U. and Pfeifer, F. (1990) "Transformation of Halobacterium halobium: development of vectors and investigation of gas vesicle synthesis," *Proceedings of the National Academy of Sciences*, 87(17).

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. (1997) "The complete genome sequence of Escherichia coli K-12.," *Science*, 277(5331), pp. 1453–62.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) "Trimmomatic: a flexible trimmer for Illumina sequence data," *Bioinformatics*, 30(15), pp. 2114–2120.

Bortolaia, V., Kaas, R., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R., Rebelo, A., Florensa, A., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., *et al.* (2020) "ResFinder 4.0 for predictions of phenotypes from genotypes," *The Journal of antimicrobial chemotherapy*, 75(12), pp. 3491–3500.

Boucher, Y., Cordero, O. X., Takemura, A., Hunt, D. E., Schliep, K., Bapteste, E., Lopez, P., Tarr, C. L. and Polz, M. F. (2011) "Local mobile gene pools rapidly cross species boundaries to create endemicity within global Vibrio cholerae populations," *mBio*, 2(2).

Boucher, Y., Labbate, M., Koenig, J. E. and Stokes, H. W. (2007) "Integrons: mobilizable platforms that promote genetic diversity in bacteria," *Trend in Microbiology*, 15(7), pp. 301–309.

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Ventra, M. di, Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. v, Ramsey, J. M., *et al.* (2008) "The potential and challenges of nanopore sequencing," *Nature biotechnology*, 26(10), p. 1146.

Braschi, I., Blasioli, S., Fellet, C., Lorenzini, R., Garelli, A., Pori, M. and Giacomini, D. (2013) "Persistence and degradation of new β-lactam antibiotics in the soil and water environment," *Chemosphere*, 93(1), pp. 152–159.

Broussard, L. A., Hammett-Stabler, C. A., Winecker, R. E. and Ropero-Miller, J. D. (2002) "The Toxicology of Mercury," *Labratory Medicine*, 8(33), pp. 614–625.

Brown, H. J., Stokes, H. W. and Hall, R. M. (1996) "The integrons In0, In2, and In5 are defective transposon derivatives.," *Journal of bacteriology*, 178(15), pp. 4429–37.

Buchanan, S. K., Smith, B. S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., van der Helm, D. and Deisenhofer, J. (1999) "Crystal structure of the outer membrane active transporter FepA from Escherichia coli," *Nature Structural Biology*, 6(1), pp. 56–63.

Bunny, K., Hall, R. M. and Stokes, H. W. (1995) "New mobile gene cassettes containing an aminoglycoside resistance gene, aacA7, and a chloramphenicol resistance gene, catB3, in an integron in pBWH301," *Antimicrobial agents and chemotherapy*, 39(3), pp. 686–693.

Burmeister, A. R. (2015) "Horizontal Gene Transfer," *Evolution, Medicine, and Public Health*, 2015(1), p. 193.

Burrell, R. E. (2003) "A scientific perspective on the use of topical silver preparations.," *Ostomy/wound management*, pp. 19–24.

Burton, N. A., Johnson, M. D., Antczak, P., Robinson, A. and Lund, P. A. (2010) "Novel Aspects of the Acid Response Network of E. coli K-12 Are Revealed by a Study of Transcriptional Dynamics," *Journal of Molecular Biology*, 401(5), pp. 726–742.

Cacace, D., Fatta-Kassinos, D., Manaia, C. M., Cytryn, E., Kreuzinger, N., Rizzo, L., Karaolia, P., Schwartz, T., Alexander, J., Merlin, C., Garelick, H., Schmitt, H., de Vries, D., Schwermer, C. U., Meric, S., Ozkal, C. B., Pons, M. N., Kneis, D. and Berendonk, T. U. (2019) "Antibiotic resistance genes in treated wastewater and in the receiving water bodies: A pan-European survey of urban settings," *Water Research*, 162, pp. 320–330.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009) "BLAST+: architecture and applications," *BMC Bioinformatics*, 10(1), pp. 1–9.

Cambray, G., Guerout, A.-M. and Mazel, D. (2010) "Integrons," *Annual Review of Genetics*, 44(1), pp. 141–166.

Cambray, G., Sanchez-Alberola, N., Campoy, S., Guerin, É., Da Re, S., González-Zorn, B., Ploy, M. C., Barbé, J., Mazel, D. and Erill, I. (2011) "Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons," *Mobile DNA*, 2(1), p. 6.

Carattoli, A., Zankari, E., García-Fernández, A., Larsen, M., Lund, O., Villa, L., Aarestrup, F. M. and Hasman, H. (2014) "In silico detection and typing of plasmids using PlasmidFinder and

plasmid multilocus sequence typing," *Antimicrobial agents and chemotherapy*, 58(7), pp. 3895–3903.

Casadaban, M. J., Chou, J. and Cohen, S. N. (1982) "Overproduction of the Tn3 transposition protein and its role in DNA transposition," *Cell*, 28(2), pp. 345–354.

Castiglioni, S., Zuccato, E., Chiabrando, C., Fanelli, R. and Bagnatl, R. (2008) "Mass spectrometric analysis of illicit drugs in wastewater and surface water," *Mass Spectrometry Reviews*, 27(4), pp. 378–394.

De Cazes, M., Belleville, M.-P., Petit, E., Salomo, M., Bayer, S., Czaja, R., De Gunzburg, J. and Sanchez-Marcano, J. (2016) "Erythromycin degradation by esterase (EreB) in enzymatic membrane reactors," *Biochemical Engineering Journal*, 114, pp. 70–78.

Chainier, D., Barraud, O., Masson, G., Couve-Deacon, E., François, B., Couquet, C. Y. and Ploy, M. C. (2017) "Integron digestive carriage in human and cattle: A 'one health' cultivation-independent approach," *Frontiers in Microbiology*, 8(SEP).

Chalmers, G., Rozas, K., Amachawadi, R., Scott, H., Norman, K., Nagaraja, T., Tokach, M. and Boerlin, P. (2018) "Distribution of the pco Gene Cluster and Associated Genetic Determinants among Swine Escherichia coli from a Controlled Feeding Trial," *Genes*, 9(10).

Chandler, M. and Mahillon, J. (2002) "Insertion Sequences Revisited," in *Mobile DNA II*, pp. 305–366.

Chang, L. L., Chang, T. M. and Chang, C. Y. (2007) "Variable gene cassette patterns of class 1 integron-associated drug-resistant Escherichia coli in Taiwan," *Kaohsiung Journal of Medical Sciences*, 23(6), pp. 273–280.

Chaudhuri, R. R., Sebaihia, M., Hobman, J. L., Webber, M. A., Leyton, D. L., Goldberg, M. D., Cunningham, A. F., Scott-Tucker, A., Ferguson, P. R., Thomas, C. M., Frankel, G., Tang, C. M., Dudley, E. G., Roberts, I. S., Rasko, D. A., Pallen, M. J., Parkhill, J., Nataro, J. P., Thomson, N. R., *et al.* (2010) "Complete Genome Sequence and Comparative Metabolic Profiling of the Prototypical Enteroaggregative Escherichia coli Strain 042," *PLoS ONE*. Edited by N. Ahmed, 5(1), p. e8801.

Chen, C. Y., Wu, K. M., Chang, Y. C., Chang, C. H., Tsai, H. C., Liao, T. L., Liu, Y. M., Chen, H. J., Shen, A. B. T., Li, J. C., Su, T. L., Shao, C. P., Lee, C. Te, Hor, L. I. and Tsai, S. F. (2003) "Comparative genome analysis of Vibrio vulnificus, a marine pathogen," *Genome Research*, 13(12), pp. 2577–2587.

Chen, K.-T. and Lu, C. L. (2018) "CSAR-web: a web server of contig scaffolding using algebraic rearrangements," *Nucleic Acids Research*, 46(W1), pp. W55–W59.

Chiang, Y. N., Penadés, J. R. and Chen, J. (2019) "Genetic transduction by phages and chromosomal islands: The new and noncanonical," *PLOS Pathogens*, 15(8), p. e1007878.

Chowdhury, P. R., Ingold, A., Vanegas, N., Martínez, E., Merlino, J., Merkier, A. K., Castro, M., Rocha, G. G., Borthagaray, G., Centrón, D., Toledo, H. B., Márquez, C. M. and Stokes, H. W. (2011) "Dissemination of Multiple Drug Resistance Genes by Class 1 Integrons in Klebsiella

pneumoniae Isolates from Four Countries: a Comparative Study," *Antimicrobial Agents and Chemotherapy*, 55(7), p. 3140.

Christopher, M. (2013) *Class 1 Integrons and Their Impact on the Mobility of Antibiotic Resistance in Clinical Environments*.

Clausen, P. T. L. C., Aarestrup, F. M. and Lund, O. (2018) "Rapid and precise alignment of raw reads against redundant databases with KMA," *BMC Bioinformatics*, 19(1), pp. 1–8.

Clinical and Laboratory Standards Institute (2018) *M100 Performance Standards for Antimicrobial Susceptibility Testing*.

Collis, C. M., Grammaticopoulos, G., Briton, J., Stokes, H. W. and Hall, R. M. (1993) "Site-specific insertion of gene cassettes into integrons," *Molecular Microbiology*, 9(1), pp. 41–52.

Cook, M., Molto, E. and Anderson, C. (1989) "Fluorochrome labelling in roman period skeletons from dakhleh oasis, Egypt," *American Journal of Physical Anthropology*, 80(2), pp. 137–143.

Craig, P. J. (1982) "Environmental Aspects of Organometallic Chemistry," *Comprehensive Organometallic Chemistry*, 2, pp. 979–1020.

Cruz-Morató, C., Ferrando-Climent, L., Rodriguez-Mozaz, S., Barceló, D., Marco-Urrea, E., Vicent, T. and Sarrà, M. (2013) "Degradation of pharmaceuticals in non-sterile urban wastewater by Trametes versicolor in a fluidized bed bioreactor," *Water Research*, 47, pp. 5200–5210.

Cutler, T. D. and Zimmerman, J. J. (2011) "Ultraviolet irradiation and the mechanisms underlying its inactivation of infectious agents," *Animal Health Research Reviews*, 12(1), pp. 15–23.

Dahlberg, C. and Hermansson, M. (1995) "Abundance of Tn3, Tn21, and Tn501 transposase (tnpA) sequences in bacterial community DNA from marine environments," *Applied and environmental microbiology*, 61(8), pp. 3051–3056.

Darzentas, N. (2010) "Circoletto: visualizing sequence similarity with Circos," *Bioinformatics*, 26(20), pp. 2620–2621.

Daschner, F. and Schuster, A. (2004) "Disinfection and the prevention of infectious disease: No adverse effects?," *American Journal of Infection Control*, 32(4), pp. 224–225.

Datta, N. and Hughes, V. M. (1983) "Plasmids of the same Inc groups in Enterobacteria before and after the medical use of antibiotics," *Nature*, 306(5943), pp. 616–617.

Davies, J. (2006) "Where have All the Antibiotics Gone?," *The Canadian Journal of Infectious Diseases & Medical Microbiology*, 17(5), p. 287.

Department of Health (2014) *Antimicrobial Resistance (AMR) Systems Map Overview of the factors influencing the development of AMR and the interactions between them*.

Diebold, P. J., New, F. N., Hovan, M., Satlin, M. J. and Brito, I. L. (2021) "Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion pcr," *eLife*, 10.

Domingues, S., Harms, K., Fricke, W. F., Johnsen, P. J., Silva, G. J. da and Nielsen, K. M. (2012) "Natural Transformation Facilitates Transfer of Transposons, Integrons and Gene Cassettes between Bacterial Species," *PLOS Pathogens*, 8(8), p. e1002837.

Domingues, S., da Silva, G. J. and Nielsen, K. M. (2012) "Integrons," *Mobile Genetic Elements*, 2(5), pp. 211–223.

Dong, D., Li, M., Liu, Z., Feng, J., Jia, N., Zhao, H., Zhao, B., Zhou, T., Zhang, X., Tong, Y. and Zhu, Y. (2019) "Characterization of a NDM-1- Encoding Plasmid pHFK418-NDM From a Clinical Proteus mirabilis Isolate Harboring Two Novel Transposons, Tn6624 and Tn6625," *Frontiers in Microbiology*, 10, p. 2030.

Dubnau, D. (2003) "DNA Uptake in Bacteria," *Annual Reviews*, 53, pp. 217–244.

Duffus, J. H. (2002) "'Heavy Metals'-A Meaningless Term?," *Pure Applied Chemistry*, 74(5), pp. 1999–2001.

Dunai, A., Spohn, R., Farkas, Z., Lázár, V., Györkei, Apjok, G., Boross, G., Szappanos, B., Grézal, G., Faragó, A., Bodai, L., Papp, B. and Pál, C. (2019) "Rapid decline of bacterial drug-resistance in an antibiotic-free environment through phenotypic reversion," *eLife*, 8.

Durán, N., Durán, M., de Jesus, M. B., Seabra, A. B., Fávaro, W. J. and Nakazato, G. (2016) "Silver nanoparticles: A new view on mechanistic aspects on antimicrobial activity," *Nanomedicine: Nanotechnology, Biology, and Medicine*, 12(3), pp. 789–799.

Dwyer, D. J., Belenky, P. A., Yang, J. H., MacDonald, I. C., Martell, J. D., Takahashi, N., Chan, C. T. Y., Lobritz, M. A., Braff, D., Schwarz, E. G., Ye, J. D., Pati, M., Vercruysse, M., Ralifo, P. S., Allison, K. R., Khalil, A. S., Ting, A. Y., Walker, G. C. and Collins, J. J. (2014) "Antibiotics induce redox-related physiological alterations as part of their lethality," *Proceedings of the National Academy of Sciences*, 111(20), pp. E2100–E2109.

Dziewit, L., Baj, J., Szuplewska, M., Maj, A., Tabin, M., Czyzkowska, A., Skrzypczyk, G., Adamczuk, M., Sitarek, T., Stawinski, P., Tudek, A., Wanasz, K., Wardal, E., Piechucka, E. and Bartosik, D. (2012) "Insights into the Transposable Mobilome of Paracoccus spp. (Alphaproteobacteria)," *PLoS ONE*.

Ehrlich, P. (1908) "Experimental researches on specific therapy. On immunity with special relationship between distribution and action of antigens. Harben Lecture."

Eichenbaum, Z. and Livneh, Z. (1998) "UV light induces IS10 transposition in Escherichia coli.," *Genetics*, 149(3), p. 1173.

Ercal, N., Gurer-Orhan, H. and Aykin-Burns, N. (2001) "Toxic metals and oxidative stress part I: mechanisms involved in metal-induced oxidative damage," *Current topics in medicinal chemistry*, 1(6), pp. 529–539.

Escudero, J. A., Loot, C., Nivina, A. and Mazel, D. (2015) "The Integron: Adaptation On Demand," *Microbiology Spectrum*, 3(2).

Essa, A. M. M., Julian, D. J., Kidd, S. P., Brown, N. L. and Hobman, J. L. (2003) "Mercury resistance determinants related to Tn21, Tn1696, and Tn5053 in enterobacteria from the preantibiotic era.," *Antimicrobial agents and chemotherapy*, 47(3), pp. 1115–9.

Eterpi, M., McDonnell, G. and Thomas, V. (2009) "Disinfection efficacy against parvoviruses compared with reference viruses," *Journal of Hospital Infection*, 73(1), pp. 64–70.

European Commission (2017) *Science for Environment PolicyTackling Mercury Pollution in the EU and Worldwide*.

Fazlara, A. and Ekhtelat, M. (2012) "The Disinfectant Effects of Benzalkonium Chloride on Some Important Foodborne Pathogens," *Journal for Agricultural & Environmental Science*, 12(1), pp. 23–29.

Fleming, A. (2001) "On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929.," *Bulletin of the World Health Organization*, 79(8), p. 780.

Foguel, S. R., Dunford, N. and Schwartz, J. T. (1963) "Bacteriophage-Induced Mutation In Escherichia coli," *Proceedings of the National Academy of Sciences of the United States of America*, 50(2), pp. 1043–1051.

Foxman, B., Zhang, L., Palin, K., Tallman, P. and Marrs, C. F. (1995) "Bacterial Virulence Characteristics of Escherichia coli Isolates from First-Time Urinary Tract Infection."

Freedman, Z., Zhu, C. and Barkay, T. (2012) "Mercury resistance and mercuric reductase activities and expression among chemotrophic thermophilic Aquificae.," *Applied and environmental microbiology*, 78(18), pp. 6568–75.

Gadou, V., Guessennd, N. K., A.Toty, A., Konan, F., Ouattara, M. B., Dosso, M., Diène, S. M., Djaman, J. A. and Rolain, J.-M. (2018) "Molecular Detection of the Arr-2 Gene in Escherichia coli and Klebsiella pneumoniae Resistant to Rifampicin in Abidjan, Côte D'Ivoire," *Microbiology Research Journal International*, 23(4), pp. 1–8.

Gardner, M., Comber, S., Scrimshaw, M. D., Cartmell, E., Lester, J. and Ellor, B. (2012) "The significance of hazardous chemicals in wastewater treatment works effluents," *Science of the Total Environment*, 437, pp. 363–372.

Gaze, W. H., Zhang, L., Abdouslam, N. A., Hawkey, P. M., Calvo-Bado, L., Royle, J., Brown, H., Davis, S., Kay, P., Ba Boxall, A. and Wellington, E. M. (2011) "Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment," *The ISME Journal*, 5, pp. 1253–1261.

Gerba, C. P. (2015) "Quaternary Ammonium Biocides: Efficacy in Application," *Applied and Environmental Microbiology*. Edited by V. Müller, 81(2), pp. 464–469.

German, N., Doyscher, D. and Rensing, C. (2013) "Bacterial killing in macrophages and amoeba: Do they all use a brass dagger?," *Future Microbiology*, 8(10), pp. 1257–1264.

Ghaly, T. M., Chow, L., Asher, A. J., Waldron, L. S. and Gillings, M. R. (2017) "Evolution of class 1 integrons: Mobilization and dispersal via food-borne bacteria," *PLoS ONE*, 12(6), pp. 1–11.

Ghaly, T. M., Geoghegan, J. L., Alroy, J. and Gillings, M. R. (2019) "High diversity and rapid spatial turnover of integron gene cassettes in soil," *Environmental Microbiology*, 21(5), pp. 1567–1574.

Ghaly, T. M., Penesyan, A., Pritchard, A., Qi, Q., Rajabal, V., Tetu, S. G. and Gillings, M. R. (2021) "Methods for the targeted sequencing and analysis of integrons and their gene cassettes from complex microbial communities," *bioRxiv*, p. 2021.09.08.459516.

Ghaly, T. M., Tetu, S. G. and Gillings, M. R. (2021) "Predicting the taxonomic and environmental sources of integron gene cassettes using structural and sequence homology of attC sites," *Communications Biology*, 4(1), pp. 1–8.

Giachi, G., Pallecchi, P., Romualdi, A., Ribechini, E., Lucejko, J. J., Colombini, M. P. and Lippic, M. M. (2013) "Ingredients of a 2,000-y-old medicine revealed by chemical, mineralogical, and botanical investigations," *Proceedings of the National Academy of Sciences of the United States of America*, 110(4), pp. 1193–1196.

Gillings, M. R. (2014) "Integrons: past, present, and future.," *Microbiology and molecular biology reviews: MMBR*, 78(2), pp. 257–77.

Gillings, M. R. (2017) "Class 1 integrons as invasive species," *Current Opinion in Microbiology*, pp. 10–15.

Gillings, M. R., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M. and Stokes, H. W. (2008) "The evolution of class 1 integrons and the rise of antibiotic resistance.," *Journal of bacteriology*, 190(14), pp. 5095–100.

Gillings, M. R., Gaze, W. H., Pruden, A., Smalla, K., Tiedje, J. M. and Zhu, Y.-G. (2015) "Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution.," *The ISME journal*, 9(6), pp. 1269–79.

Gillings, M. R., Xuejun, D., Hardwick, S. A., Holley, M. P. and Stokes, H. W. (2008) "Gene cassettes encoding resistance to quaternary ammonium compounds: a role in the origin of clinical class 1 integrons?," *The ISME Journal*, 3(2), pp. 209–215.

Gladysheva, T. B., Oden, K. L. and Rosen, B. P. (2002) "Properties of the Arsenate Reductase of Plasmid R773," *Biochemistry*, 33(23), pp. 7288–7293.

Van Goethem, M. W., Pierneef, R., Bezuidt, O. K. I., Van De Peer, Y., Cowan, D. A. and Makhalanyane, T. P. (2018) "A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils," *Microbiome*, 6(1), p. 40.

Gogarten, J. P. and Townsend, J. P. (2005) "Horizontal gene transfer, genome innovation and evolution," *Nature Reviews Microbiology*, 3(9), pp. 679–687.

Goldstein, C., Lee, M. D., Sanchez, S., Hudson, C., Phillips, B., Register, B., Grady, M., Liebert, C., Summers, A. O., White, D. G. and Maurer, J. J. (2001) "Incidence of Class 1 and 2 Integrases in Clinical and Commensal Bacteria from Livestock, Companion Animals, and Exotics," *Antimicrobial Agents and Chemotherapy*, 45(3), p. 723.

Gonzalez, G., Sossa, K., Bello, H., Dominguez, M., Mella, S. and Zemelman, R. (1998) "Presence of integrons in isolates of different biotypes of Acinetobacter baumannii from Chilean hospitals," *FEMS Microbiology Letters*, 161, pp. 125–128.

Gordon, O., Slenters, T. V., Brunetto, P. S., Villaruz, A. E., Sturdevant, D. E., Otto, M., Landmann, R. and Fromm, K. M. (2010) "Silver Coordination Polymers for Prevention of

Implant Infection: Thiol Interaction, Impact on Respiratory Chain Enzymes, and Hydroxyl Radical Induction," *Antimicrobial Agents and Chemotherapy*, 54(10), pp. 4208–4218.

Götz, A., Pukal, R., Smit, E., Tietze, E., Prager, R., Tschäpe, H., van Elsas, J. and Smalla, K. (1996) "Detection and characterization of broad-host-range plasmids in environmental bacteria by PCR," *Applied and environmental microbiology*, 62(7), pp. 2621–2628.

Grass, G., Rensing, C. and Solioz, M. (2011) "Metallic copper as an antimicrobial surface," *Applied and Environmental Microbiology*, pp. 1541–1547.

Gregory, S. T. and Dahlberg, A. E. (2009) "Genetic and structural analysis of base substitutions in the central pseudoknot of Thermus thermophilus 16S ribosomal RNA," *RNA*, 15(2), pp. 215–223.

Griffiths, A. J., Miller, J. H., Suzuki, D. T., Lewontin, R. C. and Gelbart, W. M. (2000) *Prokaryotic transposons*.

Grinsted, J. and Brown, N. L. (1984a) "A Tn21 terminal sequence within Tn501: complementation of tnpA gene function and transposon evolution," *MGG Molecular & General Genetics*, 197(3), pp. 497–502.

Grinsted, J. and Brown, N. L. (1984b) "A Tn21 terminal sequence within Tn501: complementation of tnpA gene function and transposon evolution," *MGG Molecular & General Genetics*, 197(3), pp. 497–502.

Grinsted, J., De La Cruz, F. and Schmitt, R. (1990) "The Tn21 subgroup of bacterial transposable elements," *Plasmid*, 24(3), pp. 163–189.

Guerin, É., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Re, S. Da, Gonzalez-Zorn, B., Barbé, J., Ploy, M. C. and Mazel, D. (2009) "The SOS response controls integron recombination," *Science*, 324(5930), p. 1034.

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. (2013) "QUAST: quality assessment tool for genome assemblies," *Bioinformatics*, 29(8), pp. 1072–1075.

Haas, L. (1999) "Papyrus of Ebers and Smith," *Journal of neurology, neurosurgery, and psychiatry*, 67(5), p. 578.

Hakkila, K., Maksimow, M., Karp, M. and Virta, M. (2002) "Reporter Genes lucFF, luxCDABE, gfp, and dsred Have Different Characteristics in Whole-Cell Bacterial Sensors," *Analytical Biochemistry*, 301(2), pp. 235–242.

Hall, J. P. J., Harrison, E., Pärnänen, K., Virta, M. and Brockhurst, M. A. (2020) "The Impact of Mercury Selection and Conjugative Genetic Elements on Community Structure and Resistance Gene Transfer," *Frontiers in Microbiology*, 0, p. 1846.

Hall, V., O'Neill, G. L., Magee, J. T. and Duerden, B. I. (1999) "Development of Amplified 16S Ribosomal DNA Restriction Analysis for Identification of Actinomyces Species and Comparison with Pyrolysis-Mass Spectrometry and Conventional Biochemical Tests," *Journal of Clinical Microbiology*, 37(7), p. 2255.

Hanczvikkel, A., Füzi, M., Ungvári, E. and Tóth, Á. (2018) "Transmissible silver resistance readily evolves in high-risk clone isolates of Klebsiella pneumoniae," *Acta Microbiologica et Immunologica Hungarica*, 65(3), pp. 387–403.

Hardwick, S. A., Stokes, H. W., Findlay, S., Taylor, M. and Gillings, M. R. (2008) "Quanti¢cation of class1integron abundance in natural environments using real-time quantitative PCR," *FEMS Microbiol Lett*, 278, pp. 207–212.

Harshey, R. M. (2014) "Transposable phage Mu," *Microbiology Spectrum*, 2(5).

Hasman, H., Mevius, D., Veldman, K., Olesen, I. and Aarestrup, F. M. (2005) "β-Lactamases among extended-spectrum β-lactamase (ESBL)-resistant Salmonella from poultry, poultry products and human patients in The Netherlands," *Journal of Antimicrobial Chemotherapy*, 56(1), pp. 115–121.

Hassen, A., Mahrouk, M., Ouzari, H., Cherif, M., Boudabous, A. and Damelincourt, J. J. (2000) "UV disinfection of treated wastewater in a large-scale pilot plant and inactivation of selected bacteria in a laboratory UV device," *Bioresource Technology*, 74(2), pp. 141–150.

Hauer, B. and Shapiro, J. A. (1984) "Control of Tn7 transposition," *MGG Molecular & General Genetics*, 194(1–2), pp. 149–158.

He, S., Hickman, A. B., Varani, A. M., Siguier, P., Chandler, M., Dekker, J. P. and Dyda, F. (2015) "Insertion Sequence IS 26 Reorganizes Plasmids in Clinically Isolated Multidrug-Resistant Bacteria by Replicative Transposition," *mBio*. Edited by J. E. Davies, 6(3).

Hedges, R. W. and Jacob, A. E. (1974) "Transposition of ampicillin resistance from RP4 to other replicons," *Molecular & general genetics : MGG*, 132(1), pp. 31–40.

Heffron, F. (1983) *Tn3 and Its Relatives*, *Mobile Genetic Elements*.

Hobman, J. L. (2017) "Antimicrobial metal ion resistance and its impact on co-selection of antibiotic resistance," *Culture*, 37(1).

Hobman, J. L. and Crossman, L. C. (2015) "Bacterial antimicrobial metal ion resistance," *Journal of Medical Microbiology*, 64, pp. 471–497.

Hochhut, B., Lotfi, Y., Mazel, D., Faruque, S. M., Woodgate, R. and Waldor, M. K. (2001) "Molecular analysis of antibiotic resistance gene clusters in Vibrio cholerae O139 and O1 SXT constins," *Antimicrobial Agents and Chemotherapy*, 45(11), pp. 2991–3000.

Hodson, M. E. (2004) "Heavy metals - Geochemical bogey men?," *Environmental Pollution*, 129(3), pp. 341–343.

Holmes, A. J., Gillings, M. R., Nield, B. S., Mabbutt, B. C., Nevalainen, K. M. H. and Stokes, H. W. (2003) "The gene cassette metagenome is a basic resource for bacterial genome evolution," *Environmental Microbiology*, 5(5), pp. 383–394.

Holm-Nielsen, J. B., Al Seadi, T. and Oleskowicz-Popiel, P. (2009) "The future of anaerobic digestion and biogas utilization," *Bioresource Technology*, 100(22), pp. 5478–5484.

Holt, K. B. and Bard, A. J. (2005) "Interaction of Silver(I) Ions with the Respiratory Chain of Escherichia coli: An Electrochemical and Scanning Electrochemical Microscopy Study of the Antimicrobial Mechanism of Micromolar Ag+," *Biochemistry*, 44, pp. 13214–13223.

Hooton, S. P. T., Pritchard, A. C. W., Asiani, K., Gray-Hammerton, C. J., Stekel, D. J., Crossman, L. C., Millard, A. D. and Hobman, J. L. (2021) "Laboratory Stock Variants of the Archetype Silver Resistance Plasmid pMG101 Demonstrate Plasmid Fusion, Loss of Transmissibility, and Transposition of Tn7/pco/sil Into the Host Chromosome," *Frontiers in Microbiology*, 0, p. 2339.

Houang, E., Chu, Y., Lo, W., Chu, K. and Cheng, A. (2003) "Epidemiology of rifampin ADP-ribosyltransferase (arr-2) and metallo-beta-lactamase (blaIMP-4) gene cassettes in class 1 integrons in Acinetobacter strains isolated from blood cultures in 1997 to 2000," *Antimicrobial agents and chemotherapy*, 47(4), pp. 1382–1390.

Huaizhi, Z. and Yuantao, N. (2001) "China's Ancient Gold Drugs," *Gold Bulletin*, 34(1), pp. 24–29.

Hultman, J., Tamminen, M., Pärnänen, K., Cairns, J., Karkman, A. and Virta, M. (2018) "Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent," *FEMS Microbiology Ecology*, 94(4).

Hutchings, M., Truman, A. and Wilkinson, B. (2019) "Antibiotics: past, present and future," *Current Opinion in Microbiology*, pp. 72–80.

Hyde, C. G. (1937) "The Beautification and Irrigation of Golden Gate Park with Activated Sludge Effluent on JSTOR," *Sewage Works Journal*, 9(6), pp. 929–941.

Hyde, D. R. and Tu, C.-P. D. (1985) *tnpM: A Novel Regulatory Gene That Enhances Tn21 Transposition and Suppresses Cointegrate Resolution*, Cell.

Ibrahim, D. R., Dodd, C. E. R., Stekel, D. J., Ramsden, S. J. and Hobman, J. L. (2016) "Multidrug resistant, extended spectrum β-lactamase (ESBL)-producing Escherichia coli isolated from a dairy farm," *FEMS Microbiology Ecology*. Edited by P. Simonet, 92(4), p. fiw013.

Ichikawa, H., Ikeda, K., Amemura, J. and Ohtsubo, E. (1990) "Two domains in the terminal inverted-repeat sequence of transposon Tn3," *Gene*, 86(1), pp. 11–17.

Iervolino, G., Zammit, I., Vaiano, V. and Rizzo, L. (2019) "Limitations and Prospects for Wastewater Treatment by UV and Visible-Light-Active Heterogeneous Photocatalysis: A Critical Review," *Topics in Current Chemistry*, 378(1), pp. 1–40.

Iida, S., Mollet, B., Meyer, J. and Arber, W. (1984) "Functional characterization of the prokaryotic mobile genetic element IS26," *Molecular and General Genetics*, 198(1), pp. 84–89.

Iles, E. A. (2018) *The prevalence and potential mobility of Tn21-like mercury resistance transposons in E. coli isolated from dairy cattle slurry*.

Ioannou, C. J., Hanlon, G. W. and Denyer, S. P. (2007) "Action of disinfectant quaternary ammonium compounds against Staphylococcus aureus," *Antimicrobial Agents and Chemotherapy*, 51(1), pp. 296–306.

Joensen, K., Scheutz, F., Lund, O., Hasman, H., Kaas, R., Nielsen, E. M. and Aarestrup, F. M. (2014) "Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli," *Journal of clinical microbiology*, 52(5), pp. 1501–1510.

Jones, I. A. and Joshi, L. T. (2021) "Biocide use in the antimicrobial era: A review," *Molecules*.

Jones, S. E. and Elliot, M. A. (2017) "Streptomyces Exploration: Competition, Volatile Communication and New Bacterial Behaviours," *Trends in Microbiology*, 25(7), pp. 522–531.

Jones-Dias, D., Manageiro, V., Ferreira, E., Barreiro, P., Vieira, L., Moura, I. B. and Caniça, M. (2016) "Architecture of class 1, 2, and 3 integrons from gram negative bacteria recovered among fruits and vegetables," *Frontiers in Microbiology*, 7(SEP), p. 1400.

Jordan, D. M., Cornick, N., Torres, A. G., Dean-Nystrom, E. A., Kaper, J. B. and Moon, H. W. (2004) "Long Polar Fimbriae Contribute to Colonization by Escherichia coli O157:H7 In Vivo," *Infection and Immunity*, 72(10), p. 6168.

Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C. S., Singer, H. P., Johnson, D. R., Zhang, T. and Bürgmann, H. (2018) "Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes," *The ISME Journal 2018 13:2*, 13(2), pp. 346–360.

Karah, N., Haldorsen, B., Hegstad, K., Simonsen, G. S., Sundsfjord, A. and Samuelsen, Ø. (2011) "Species identification and molecular characterization of Acinetobacter spp. blood culture isolates from Norway," *The Journal of antimicrobial chemotherapy*, 66(4), pp. 738–744.

Karkman, A., Do, T. T., Walsh, F. and Virta, M. P. J. (2018) "Antibiotic-Resistance Genes in Waste Water," *Trends in Microbiology*, 26(3), pp. 220–228.

Katoh, K. and Standley, D. M. (2013) "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability," *Molecular Biology and Evolution*, 30(4), p. 772.

Kholodii, G., Mindlin, S., Petrova, M. and Minakhina, S. (2003) "Tn *5060* from the Siberian permafrost is most closely related to the ancestor of Tn *21* prior to integron acquisition," *FEMS Microbiology Letters*, 226(2), pp. 251–255.

Kholodii, G. Ya., Mindlin, S. Z., Bass, I. A., Yurieva, O. V., Minakhina, S. V. and Nikiforov, V. G. (1995) "Four genes, two ends, and a res region are involved in transposition of Tn5053: a paradigm for a novel family of transposons carrying either a mer operon or an integron," *Molecular Microbiology*, 17(6), pp. 1189–1200.

Kiiru, J., Butaye, P., Goddeeris, B. M. and Kariuki, S. (2013) "Analysis for prevalence and physical linkages amongst integrons, ISEcp1, ISCR1, Tn21 and Tn7 encountered in Escherichia coli strains from hospitalized and non-hospitalized patients in Kenya during a 19-year period (1992-2011)," *BMC Microbiology*, 13(1), pp. 1–14.

Kitts, P. A., Lamond, A. and Sherratt, D. J. (1982) "Inter-replicon transposition of Tn1/3 occurs in two sequential genetically separable steps," *Nature 1982 295:5850*, 295(5850), pp. 626–628.

Kitts, P., Symington, L., Burke, M., Reedt, R. and Sherratt, D. (1982) "Transposon-specified site-specific recombination (Tnl/Tn3/y8/Tn501/transposition mechanism)," *Proc. NatL Acad. Sci. USA*, 79, pp. 46–50.

Klappenbach, J. A., Saxman, P. R., Cole, J. R. and Schmidt, T. M. (2001) "Rrndb: The ribosomal RNA operon copy number database," *Nucleic Acids Research*, 29(1), pp. 181–184.

Knapp, C. W., Dolfing, J., Ehlert, P. A. I. and Graham, D. W. (2010) "Evidence of Increasing Antibiotic Resistance Gene Abundances in Archived Soils since 1940," *Environmental Science & Technology*, 44(2), pp. 580–587.

Kostriken, R., Morita, C. and Heffron, F. (1981) "Transposon Tn3 encodes a site-specific recombination system: Identification of essential sequences, genes, and actual site of recombination," *Proceedings of the National Academy of Sciences of the United States of America*, 78(7 I), pp. 4041–4045.

Kretschmer, P. and Cohen, S. N. (1979) "Effect of temperature on translocation frequency of the Tn3 element," *Journal of bacteriology*, 139(2), pp. 515–519.

Kumar, V. A. and Khan, S. (2015) "Defining multidrug resistance in Gram-negative bacilli," *Indian Journal of Medical Research*, pp. 491–493.

Kurenbach, B., Gibson, P. S., Hill, A. M., Bitzer, A. S., Silby, M. W., Godsoe, W. and Heinemann, J. A. (2017) "Herbicide ingredients change Salmonella enterica sv. Typhimurium and Escherichia coli antibiotic responses," *Microbiology*, 163(12), pp. 1791–1801.

De la Cruz, F. and Davies, J. (2000) "Horizontal gene transfer and the origin of species: Lessons from bacteria," *Trends in Microbiology*, pp. 128–133.

De La Cruz, F., Frost, L. S., Meyer, R. J. and Zechner, E. L. (2010) "Conjugative DNA metabolism in Gram-negative bacteria," *FEMS Microbiology Reviews*, 34(1), pp. 18–40.

De La Cruz, F. and Grinsted, J. (1982) "Genetic and Molecular Characterization of Tn2J, a Multiple Resistance Transposon from R100.1," *Journal of Bacteriology*, pp. 222–228.

Labbate, M., Case, R. J. and Stokes, H. W. (2009) "The integron/gene cassette system: an active player in bacterial adaptation.," *Methods in molecular biology (Clifton, N.J.)*, 532, pp. 103–125.

Lamrani, S., Ranquet, C., Gama, M.-J., Nakai, H., Shapiro, J. A., Toussaint, A. and Maenhaut-Michel, G. (1999) "Starvation-induced Mucts62-mediated coding sequence fusion: a role for ClpXP, Lon, RpoS and Crp," *Molecular Microbiology*, 32(2), pp. 327–343.

Larsen, M., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R., Jelsbak, L., Sicheritz-Pontén, T., Ussery, D., Aarestrup, F. M. and Lund, O. (2012) "Multilocus sequence typing of total-genome-sequenced bacteria," *Journal of clinical microbiology*, 50(4), pp. 1355–1361.

LaVoie, S. P. and Summers, A. O. (2018) "Transcriptional responses of Escherichia coli during recovery from inorganic or organic mercury exposure," *BMC Genomics*, 19(1).

Lawrence, J. G. and Ochman, H. (1998) "Molecular archaeology of the Escherichia coli genome," *Proceedings of the National Academy of Sciences of the United States of America*, 95(16), pp. 9413–9417.

Lee, A. M. and Singleton, S. F. (2004) "Inhibition of the Escherichia coli RecA protein: zinc(II), copper(II) and mercury(II) trap RecA as inactive aggregates," *Journal of Inorganic Biochemistry*, 98(11), pp. 1981–1986.

Lee, D. J., Bingle, L. E., Heurlier, K., Pallen, M. J., Penn, C. W., Busby, S. J. and Hobman, J. L. (2009) "Gene doctoring: A method for recombineering in laboratory and pathogenic Escherichia coli strains," *BMC Microbiology*, 9, pp. 1–14.

Lefort, V., Longueville, J.-E. and Gascuel, O. (2017) "SMS: Smart Model Selection in PhyML," *Molecular Biology and Evolution*, 34(9), pp. 2422–2424.

Lévesque, C., Brassard, S., Lapointe, J. and Roy, P. H. (1994) "Diversity and relative strength of tandem promoters for the antibiotic-resistance genes of several integron," *Gene*, 142(1), pp. 49–54.

Levy, S. B. (2001) "Antibacterial household products: cause for concern.," *Emerging Infectious Diseases*, 7(3 Suppl), p. 512.

Li, H. (2016) "Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences," *Bioinformatics*, 32(14), pp. 2103–2110.

Li, H., Zhu, X. and Ni, J. (2011) "Comparison of electrochemical method with ozonation, chlorination and monochloramination in drinking water disinfection," *Electrochimica Acta*, 56(27), pp. 9789–9796.

Lian, P., Guo, H.-B., Riccardi, D., Dong, A., Parks, J. M., Xu, Q., Pai, E. F., Miller, S. M., Wei, D.-Q., Smith, J. C. and Guo, H. (2014) "X-ray Structure of a Hg2+ Complex of Mercuric Reductase (MerA) and Quantum Mechanical/Molecular Mechanical Study of Hg2+ Transfer between the C-Terminal and Buried Catalytic Site Cysteine Pairs," *Biochemistry*, 53(46), p. 7211.

Liebert, C. A., Hall, R. M. and Summers, A. O. (1999) "Transposon Tn21, flagship of the floating genome." *Microbiology and molecular biology reviews*, 63(3), pp. 507–22.

Liebert, C. A., Wireman, J., Smith, T. and Summers, A. O. (1997) "Phylogeny of mercury resistance (mer) operons of gram-negative bacteria isolated from the fecal flora of primates.," *Applied and Environmental Microbiology*, 63(3), p. 1066.

Lok, C.-N., Ho, C.-M., Chen, R., He, Q.-Y., Yu, W.-Y., Sun, H., Tam, P. K.-H., Chiu, J.-F. and Che, C.-M. (2006) "Proteomic Analysis of the Mode of Antibacterial Action of Silver Nanoparticles," *Journal of Proteome Research*, 5(4), pp. 916–924.

Ludden, C., Reuter, S., Judge, K., Gouliouris, T., Blane, B., Coll, F., Naydenova, P., Hunt, M., Tracey, A., Hopkins, K. L., Brown, N. M., Woodford, N., Parkhill, J. and Peacock, S. J. (2017) "Sharing of carbapenemase-encoding plasmids between Enterobacteriaceae in UK sewage uncovered by MinION sequencing," *Microbial genomics*, 3(7), p. e000114.

Lundström, S. v., Östman, M., Bengtsson-Palme, J., Rutgersson, C., Thoudal, M., Sircar, T., Blanck, H., Eriksson, K. M., Tysklind, M., Flach, C. F. and Larsson, D. G. J. (2016) "Minimal selective concentrations of tetracycline in complex aquatic bacterial biofilms," *Science of The Total Environment*, 553, pp. 587–595.

Ma, L., Ishii, Y., Ishiguro, M., Matsuzawa, H. and Yamaguchi, K. (1998) "Cloning and Sequencing of the Gene Encoding Toho-2, a Class A β-Lactamase Preferentially Inhibited by Tazobactam," *Antimicrobial Agents and Chemotherapy*, 42(5), p. 1181.

MacDonald, D., Demarre, G., Bouvier, M., Mazel, D. and Gopaul, D. N. (2006) "Structural basis for broad DNA-specificity in integron recombination," *Nature*, 440(7088), pp. 1157–1162.

Madigan, M. T., Martinko, J. M., Bender, K. S., Buckley, D. H. and Stahl, D. A. (2015) *Brock Biology of Microorganisms*.

Maekawa, T., Yanagihara, K. and Ohtsubo, E. (1996) "A cell-free system of Tn3 transposition and transposition immunity," *Genes to Cells*, 1(11), pp. 1007–1016.

Magiorakos, A. P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., Harbarth, S., Hindler, J. F., Kahlmeter, G., Olsson-Liljequist, B., Paterson, D. L., Rice, L. B., Stelling, J., Struelens, M. J., Vatopoulos, A., Weber, J. T. and Monnet, D. L. (2012) "Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance," *Clinical Microbiology and Infection*, 18(3), pp. 268–281.

Maillard, J.-Y. (2005) "Antimicrobial biocides in the healthcare environment: efficacy, usage, policies, and perceived problems," *Therapeutics and Clinical Risk Management*, 1(4), p. 307.

Maiti, A. and Bhattacharyya, S. (2013) "Isolation and Characterization of Mercury Resistant Bacteria from Haldia river sediments," *IOSR Journal Of Environmental Science*, 5(3), pp. 23–28.

Marchesi, J. R., Sato, T., Weightman, A. J., Martin, T. A., Fry, J. C., Hiom, S. J., Dymock, D. and Wade, W. G. (1998) "Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA," *Applied and Environmental Microbiology*, 64(6), p. 2333.

Márquez, C., Labbate, M., Raymondo, C., Fernández, J., Gestal, A. M., Holley, M., Borthagaray, G. and Stokes, H. W. (2008) "Urinary Tract Infections in a South American Population: Dynamic Spread of Class 1 Integrons and Multidrug Resistance by Homologous and Site-Specific Recombination," *Journal of Clinical Microbiology*, 46(10), p. 3417.

Martin, M. (2011) "Cutadapt removes adapter sequences from high-throughput sequencing reads," *EMBnet.journal*, 17(1), pp. 10–12.

Martinez, E. and de la Cruz, F. (1988) "Transposon Tn21 encodes a RecA-independent site-specific integration system," *Molecular & General Genetics*, 211(2), pp. 320–325.

Martínez, T., Vazquez, G. J., Aquino, E. E., Goering, R. V. and Robledo, I. E. (2012) "Two Novel Class I Integron Arrays Containing IMP-18 Metallo-β-Lactamase Gene in Pseudomonas aeruginosa Clinical Isolates from Puerto Rico," *Antimicrobial Agents and Chemotherapy*, 56(4), p. 2119.

Mavridou, D. A. I., Gonzalez, D., Clements, A. and Foster, K. R. (2016) "The pUltra plasmid series: A robust and flexible tool for fluorescent labeling of Enterobacteria," *Plasmid*, 87–88, pp. 65–71.

Mazel, D. (2006) "Integrons: Agents of bacterial evolution," *Nature Reviews Microbiology*, 4(8), pp. 608–620.

Mchugh, G. L., Moellering, RobertC., Hopkins, CyrusC. and Swartz, MortonN. (1975) "Salmonella Typhimurium Resistat to Silver Nitrate, Chloramphenicol, And Ampicillin: A New Threat in Burn Units?," *The Lancet*, 305(7901), pp. 235–240.

McInnes, R. S., McCallum, G. E., Lamberte, L. E. and van Schaik, W. (2020) "Horizontal transfer of antibiotic resistance genes in the human gut microbiome," *Current Opinion in Microbiology*, 53, pp. 35–43.

Melnyk, A. H., Wong, A. and Kassen, R. (2015) "The fitness costs of antibiotic resistance mutations," *Evolutionary Applications*, 8(3), pp. 273–283.

Mendiola, M. V., Bernales, I., De, F. and Cruz, L. A. (1994) "Differential roles of the transposon termini in IS91 transposition," *Proceedings of the National Academy of Science USA*, 91, pp. 1922–1926.

Mesquita, E., Ribeiro, R., Silva, C. J. C., Alves, R., Baptista, R., Condinho, S., Rosa, M. J., Perdigão, J., Caneiras, C. and Duarte, A. (2021) "An Update on Wastewater Multi-Resistant Bacteria: Identification of Clinical Pathogens Such as Escherichia coli O25b:H4-B2-ST131-Producing CTX-M-15 ESBL and KPC-3 Carbapenemase-Producing Klebsiella oxytoca," *Microorganisms*, 9(3), p. 576.

Miller, J. R., Zhou, P., Mudge, J., Gurtowski, J., Lee, H., Ramaraj, T., Walenz, B. P., Liu, J., Stupar, R. M., Denny, R., Song, L., Singh, N., Maron, L. G., McCouch, S. R., McCombie, W. R., Schatz, M. C., Tiffin, P., Young, N. D. and Silverstein, K. A. T. (2017) "Hybrid assembly with long and short reads improves discovery of gene family expansions," *BMC Genomics*, 18(1), pp. 1–12.

Miller, M. C., Fetherston, J. D., Pickett, C. L., Bobrov, A. G., Weaver, R. H., DeMoll, E. and Perry, R. D. (2010) "Reduced synthesis of the Ybt siderophore or production of aberrant Ybt-like molecules activates transcription of yersiniabactin genes in Yersinia pestis," *Microbiology*, 156(7), pp. 2226–2238.

Mindlin, S. Z. and Petrova, M. A. (2018) "On the Origin and Distribution of Antibiotic Resistance: Permafrost Bacteria Studies," *Molecular Genetics, Microbiology and Virology 2017 32:4*, 32(4), pp. 169–179.

Mirzaei, N., Kafilzadeh, F. and Kargar, M. (2008) "isolation and identification of Mercury Resistant Bacteria from Kor River, Iran," *Journal of Biological Sciences*, 8(5), pp. 935–939.

Miyata, S. T., Unterweger, D., Rudko, S. P. and Pukatzki, S. (2013) "Dual Expression Profile of Type VI Secretion System Immunity Genes Protects Pandemic Vibrio cholerae," *PLOS Pathogens*, 9(12), p. e1003752.

Mohamed, J. A., Huang, D. B., Jiang, Z. D., DuPont, H. L., Nataro, J. P., Belkind-Gerson, J. and Okhuysen, P. C. (2007) "Association of putative enteroaggregative Escherichia coli virulence genes and biofilm production in isolates from travelers to developing countries," *Journal of Clinical Microbiology*, 45(1), pp. 121–126.

Molina-Mora, J. A., Batán, R. G. and García, F. (2020) "From Pan-genome to the Genomic Context of the Two Integrons of ST-111 Pseudomonas Aeruginosa AG1: A VIM-2-carrying Old-acquaintance and a Novel Imp-18-carrying Integron."

Montaña, S., Schramm, S. T. J., Traglia, G. M., Chiem, K., Noto, G. P. Di, Almuzara, M., Barberis, C., Vay, C., Quiroga, C., Tolmasky, M. E., Iriarte, A. and Ramírez, M. S. (2016) "The Genetic Analysis of an Acinetobacter johnsonii Clinical Strain Evidenced the Presence of Horizontal Genetic Transfer," *PLoS ONE*, 11(8).

Moore, B. (1960) "A NEW SCREEN TEST AND SELECTIVE MEDIUM FOR THE RAPID DETECTION OF EPIDEMIC STRAINS OF STAPH. AUREUS," *The Lancet*, 276(7148), pp. 453–458.

Morales Medina, W. R., Eramo, A., Tu, M. and Fahrenfeld, N. L. (2020) "Sewer biofilm microbiome and antibiotic resistance genes as function of pipe material, source of microbes, and disinfection: field and laboratory studies," *Environmental Science: Water Research & Technology*, 6(8), pp. 2122–2137.

Morby, A. P., Hobman, J. L. and Brown, N. L. (1995) "The role of cysteine residues in the transport of mercuric ions by the Tn501 MerT and MerP mercury-resistance proteins," *Molecular Microbiology*, 17(1), pp. 25–35.

Morin, N., Santiago, A. E., Ernst, R. K., Guillot, S. J. and Nataro, J. P. (2013) "Characterization of the AggR Regulon in Enteroaggregative Escherichia coli," *Infection and Immunity*, 81(1), p. 122.

Moura, A., Pereira, C., Henriques, I. and Correia, A. (2012) "Novel gene cassettes and integrons in antibiotic-resistant bacteria isolated from urban wastewaters," *Research in Microbiology*, 163(2), pp. 92–100.

Mugnier, P. D., Poirel, L. and Nordmann, P. (2009) "Functional Analysis of Insertion Sequence ISAba1, Responsible for Genomic Plasticity of Acinetobacter baumannii," *Journal of Bacteriology*, 191(7), pp. 2414–2418.

Murrell, J. C., McDonald, I. R. and Bourne, D. G. (1998) "Molecular methods for the study of methanotroph ecology," *FEMS Microbiology Ecology*, 27(2), pp. 103–114.

Nakaya, R., Nakamura, A. and Murata, Y. (1960) "Resistance transfer agents in Shigella," *Biochemical and Biophysical Research Communications*, 3(6), pp. 654–659.

National Institute for Health and Care Excellence (2017) *Antibiotic resistance is now "common" in urinary tract infections.*

Nawrocki, E. P. and Eddy, S. R. (2013) "Infernal 1.1: 100-fold faster RNA homology searches," *Bioinformatics (Oxford, England)*, 29(22), pp. 2933–2935.

Nelson, M. L. and Levy, S. B. (2011) "The history of the tetracyclines," *Annals of the New York Academy of Sciences*, 1241(1), pp. 17–32.

Nevers, P. and Saedler, H. (1977) *Transposable genetic elements as agents of gene instability and chromosomal rearrangements, Nature.*

New England BioLabs (no date) *High-Fidelity PCR | NEB*. Available at: https://international.neb.com/applications/dna-amplification-pcr-and-qpcr/pcr/high-fidelity-pcr (Accessed: September 21, 2021).

Newton, R. J., McLellan, S. L., Dila, D. K., Vineis, J. H., Morrison, H. G., Eren, A. M. and Sogin, M. L. (2015) "Sewage Reflects the Microbiomes of Human Populations," *mBio*, 6(2).

NHS (2018) *Trimethoprim*. Available at: https://www.nhs.uk/medicines/trimethoprim/ (Accessed: September 14, 2021).

Nies, D. H. (1999) "Microbial heavy-metal resistance," *Applied Microbiology and Biotechnology*, 51(6), pp. 730–750.

Nikaido, H. (2009) "Multidrug Resistance in Bacteria," *Annual Review of Biochemistry*, 78(1), pp. 119–146.

Nöllmann, M., Byron, O. and Stark, W. M. (2005) "Behavior of Tn3 resolvase in solution and its interaction with res," *Biophysical Journal*, 89(3), pp. 1920–1931.

Ohtsubo, H., Ohmori, H. and Ohtsubo, E. (1979) "Nucleotide-sequence analysis of Tn3 (Ap): Implications for insertion and deletion," *Cold Spring Harbor Symposia on Quantitative Biology*, 43(2), pp. 1269–1277.

Ojo, K. K., Sapkota, A. R., Ojo, T. B. and Pottinger, P. S. (2008) "Antimicrobial resistance gene distribution: a socioeconomic and sociocultural perspective.," *GMS Krankenhaushygiene interdisziplinar*, 3(3), p. Doc26.

Östman, M., Lindberg, R. H., Fick, J., Björn, E. and Tysklind, M. (2017) "Screening of biocides, metals and antibiotics in Swedish sewage sludge and wastewater," *Water Research*, 115, pp. 318–328.

Oxford Nanopore Technologies (2017a) "Genomic DNA by Ligation," 24(05386273), pp. 2–17.

Oxford Nanopore Technologies (2017b) "Native barcoding genomic DNA," 24(05386273), pp. 2–17.

Oxford Nanopore Technologies (2018) "Four-primer PCR Before start checklist Four-primer PCR," pp. 1–8.

Oxford Nanopore Technologies (2019) "Native barcoding genomic DNA (with EXP-NBD104 , EXP- Before start checklist Native barcoding genomic DNA (with EXP-NBD104 , EXP-," pp. 1–6.

Oxford Nanopore Technologies (2021) *Accuracy*. Available at: https://nanoporetech.com/accuracy (Accessed: September 20, 2021).

Pablos, C., Marugán, J., van Grieken, R. and Serrano, E. (2013) "Emerging micropollutant oxidation during disinfection processes using UV-C, UV-C/H2O2, UV-A/TiO2 and UV-A/TiO2/H2O2," *Water Research*, 47(3), pp. 1237–1245.

Page, R. and Peti, W. (2016) "Toxin-antitoxin systems in bacterial growth arrest and persistence," *Nature Chemical Biology*, 12(4), pp. 208–214.

Pal, C., Asiani, K., Arya, S., Rensing, C., Stekel, D. J., Larsson, D. G. J. and Hobman, J. L. (2017) *Chapter Seven - Metal Resistance and Its Association With Antibiotic Resistance*, *Microbiology of Metal Ions*.

Pal, C., Bengtsson-Palme, J., Kristiansson, E. and Larsson, D. G. J. (2015) "Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential," *BMC Genomics*, 16(1), p. 964.

Park, H. J., Kim, J. Y., Kim, J., Lee, J. H., Hahn, J. S., Gu, M. B. and Yoon, J. (2009) "Silver-ion-mediated reactive oxygen species generation affecting bactericidal activity," *Water Research*, 43(4), pp. 1027–1032.

Partridge, S. R. (2011) "Analysis of antibiotic resistance regions in Gram-negative bacteria," *FEMS Microbiology Reviews*, pp. 820–855.

Partridge, S. R., Brown, H. J., Stokes, H. W. and Hall, R. M. (2001) "Transposons Tn1696 and Tn21 and Their Integrons In4 and In2 Have Independent Origins," *Antimicrobial Agents and Chemotherapy*, 45(4), p. 1263.

Partridge, S. R., Kwong, S. M., Firth, N. and Jensen, S. O. (2018) "Mobile Genetic Elements Associated with Antimicrobial Resistance," *Clinical Microbiology Reviews*, 31(4).

Partridge, S. R., Tsafnat, G., Coiera, E. and Iredell, J. R. (2009) "Gene cassettes and cassette arrays in mobile resistance integrons: Review article," *FEMS Microbiology Reviews*, 33(4), pp. 757–784.

Pattishall, K. H., Acar, J., Burchall, J. J., Goldstein, F. W. and Harvey, R. J. (1977) "Two distinct types of trimethoprim-resistant dihydrofolate reductase specified by R-plasmids of different compatibility groups.," *Journal of Biological Chemistry*, 252(7), pp. 2319–2323.

Pavelka, M. S., Wright, L. F. and Silver, R. P. (1991) "Identification of Two Genes, kpsM and kpsT, in Region 3 of the Polysialic Acid Gene Cluster of Escherichia coli K1," *Journal of Bacteriology*, 173(15), pp. 4603–4610.

Pearson, A. J., Bruce, K. D., Osborn, A. M., Ritchie, D. A. and Strike, P. (1996) "Distribution of class II transposase and resolvase genes in soil bacteria and their association with mer genes," *Applied and environmental microbiology*, 62(8), pp. 2961–2965.

Pei, R., Kim, S. C., Carlson, K. H. and Pruden, A. (2006) "Effect of River Landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG)," *Water Research*, 40(12), pp. 2427–2435.

Pellegrini, C., Celenza, G., Segatore, B., Bellio, P., Setacci, D., Amicosante, G. and Perilli, M. (2011) "Occurrence of Class 1 and 2 Integrons in Resistant Enterobacteriaceae Collected from a Urban Wastewater Treatment Plant: First Report from Central Italy," *Microbial Drug Resistance*, 17(2), pp. 229–234.

Peng, S., Stephan, R., Hummerjohann, J. and Tasara, T. (2014) "Evaluation of three reference genes of Escherichia coli for mRNA expression level normalization in view of salt and organic acid stress exposure in food," *FEMS Microbiology Letters*, 355(1), pp. 78–82.

Pereira, M. B., Wallroth, M., Kristiansson, E. and Axelson-Fisk, M. (2016) "HattCI: Fast and Accurate attC site Identification Using Hidden Markov Models," *Journal of Computational Biology*, 23(11), pp. 891–902.

Perez-Palacios, P., Delgado-Valverde, M., Gual-de-Torrella, A., Oteo-Iglesias, J., Pascual, Á. and Fernández-Cuenca, F. (2021) "Co-transfer of plasmid-encoded bla carbapenemases genes and mercury resistance operon in high-risk clones of Klebsiella pneumoniae," *Applied Microbiology and Biotechnology 2021 105:24*, 105(24), pp. 9231–9242.

Pérez-Valdespino, A., Celestino-Mancera, M., Villegas-Rodríguez, V. L. and Curiel-Quesada, E. (2013) "Characterization of mercury-resistant clinical Aeromonas species," *Brazilian Journal of Microbiology*, 44(4), pp. 1279–1283.

Peters, J. E. and Craig, N. L. (2001) "Tn7: Smarter than we thought," *Nature Reviews Molecular Cell Biology*, 2(11), pp. 806–814.

Pignato, S., Coniglio, M. A., Faro, G., Weill, F. X. and Giammanco, G. (2009) "Plasmid-mediated multiple antibiotic resistance of Escherichia coli in crude and treated wastewater used in agriculture," *Journal of Water and Health*, 07(2), p. 251.

Płaza, G., Jałowiecki, Ł., Głowacka, D., Hubeny, J., Harnisz, M. and Korzeniewska, E. (2021) "Insights into the microbial diversity and structure in a full-scale municipal wastewater treatment plant with particular regard to Archaea," *PLOS ONE*. Edited by L. Brusetti, 16(4), p. e0250514.

Ploy, M. C., Denis, F., Courvalin, P. and Lambert, T. (2000) "Molecular characterization of integrons in Acinetobacter baumannii: Description of a hybrid class 2 integron," *Antimicrobial Agents and Chemotherapy*, 44(10), pp. 2684–2688.

Poirel, L., Naas, T., Thomas, I. Le, Karim, A., Bingen, E. and Nordmann, P. (2001) "CTX-M-Type Extended-Spectrum β-Lactamase That Hydrolyzes Ceftazidime through a Single Amino Acid Substitution in the Omega Loop," *Antimicrobial Agents and Chemotherapy*, 45(12), p. 3355.

Quandt, J., Clark, R. G., Venter, A. P., Clark, S. R. D., Twelker, S. and Hynes, M. F. (2004) "Modified RP4 and Tn5-Mob derivatives for facilitated manipulation of large plasmids in Gram-negative bacteria," *Plasmid*, 52(1), pp. 1–12.

Raleigh, E. A. and Low, K. B. (2013) *Conjugation*, *Brenner's Encyclopedia of Genetics: Second Edition*.

Ralph Pearson, B. G. (1963) "Hard and Soft Acids and Bases," *Physical and Inorganic Chemistry* , 12, p. 13.

Recchia, G. D. and Hall, R. M. (1995) *Gene cassettes: a new class of mobile element*, *Microbiology*.

Rimmer, C. C., Miller, E. K., McFarland, K. P., Taylor, R. J. and Faccio, S. D. (2010) "Mercury bioaccumulation and trophic transfer in the terrestrial food web of a montane forest," *Ecotoxicology*, 19(4), pp. 697–709.

Roberts, M. C., Leroux, B. G., Sampson, J., Luis, H. S., Bernardo, M. and Leitão, J. (2008) "Dental Amalgam and Antibiotic- and/or Mercury-resistant Bacteria," *Journal of Dental Research*, 87(5), pp. 475–479.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., De Wit, C. A., Hughes, T., Van Der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., *et al.* (2009) "A safe operating space for humanity," *Nature*, 461(7263), pp. 472–475.

Rodriguez-Mozaz, S., Vaz-Moreira, I., Varela Della Giustina, S., Llorca, M., Barceló, D., Schubert, S., Berendonk, T. U., Michael-Kordatou, I., Fatta-Kassinos, D., Martinez, J. L., Elpers, C., Henriques, I., Jaeger, T., Schwartz, T., Paulshus, E., O'Sullivan, K., Pärnänen, K. M. M., Virta, M., Do, T. T., *et al.* (2020) "Antibiotic residues in final effluents of European wastewater treatment plants and their impact on the aquatic environment," *Environment International*, 140, p. 105733.

Roman, V. L., Merlin, C., Virta, M. P. J. and Bellanger, X. (2021) "EpicPCR 2.0: Technical and Methodological Improvement of a Cutting-Edge Single-Cell Genomic Approach," *Microorganisms*, 9(8), p. 1649.

Rowe-Magnus, D. A., Guerout, A. M., Ploncard, P., Dychinco, B., Davies, J. and Mazel, D. (2001) "The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons," *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), pp. 652–657.

Ruijgers, H. (2021) "The Future of Wastewater Treatment Plant Energy Management | WWD," *Water & Wastes Digest*, 30 March.

Russo, T. A. and Gulick, A. M. (2019) "Aerobactin Synthesis Proteins as Antivirulence Targets in Hypervirulent Klebsiella pneumoniae," *ACS Infectious Diseases*, 5(7), pp. 1052–1054.

Rye, R. and Holland, H. (1998) "Paleosols and the evolution of atmospheric oxygen: a critical review," *American journal of science*, 298(8), pp. 621–672.

Sakowski, E. G., Arora-Williams, K., Tian, F., Zayed, A. A., Zablocki, O., Sullivan, M. B. and Preheim, S. P. (2021) "Interaction dynamics and virus-host range for estuarine actinophages captured by epicPCR," *Nature microbiology*, 6(5), pp. 630–642.

Santo Domingo, J. and Edge, T. (2010) *Identification of primary sources of faecal pollution*.

Schramm, F. D., Schroeder, K. and Jonas, K. (2020) "Protein aggregation in bacteria," *FEMS Microbiology Reviews*, 44(1), pp. 54–72.

Severino, P. and Magalhães, V. D. (2002) "The role of integrons in the dissemination of antibiotic resistance among clinical isolates of Pseudomonas aeruginosa from an intensive care unit in Brazil," *Research in Microbiology*, 153(4), pp. 221–226.

Van Shaik, W. (2021) "Antimicrobial resistance in microbiomes."

Shaikh, S., Fatima, J., Shakil, S., Rizvi, S. M. D. and Kamal, M. A. (2015) "Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment.," *Saudi journal of biological sciences*, 22(1), pp. 90–101.

Shaw, K. J., Rather, P. N., Hare, R. S. and Miller, G. H. (1993) "Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes," *Microbiological reviews*, 57(1), pp. 138–163.

Sheikh, J., Hicks, S., Dall'Agnol, M., Phillips, A. D. and Nataro, J. P. (2001) "Roles for Fis and YafK in biofilm formation by enteroaggregative Escherichia coli," *Molecular microbiology*, 41(5), pp. 983–997.

Sherratt, D. (1989) "Tn3 and Related Transposable Elements: Site-Specific Recombination and Transposition," in Berg, D. E. and Howe, M. M. (eds) *Mobile DNA*, pp. 163–184.

Shibata, N., Doi, Y., Yamane, K., Yagi, T., Kurokawa, H., Shibayama, K., Kato, H., Kai, K. and Arakawa, Y. (2003) "PCR Typing of Genetic Determinants for Metallo-β-Lactamases and Integrases Carried by Gram-Negative Bacteria Isolated in Japan, with Focus on the Class 3 Integron," *Journal of Clinical Microbiology*, 41(12), pp. 5407–5413.

Shin, S. Y., Kwon, K. C., Park, J. W., Song, J. H., Ko, Y. H., Sung, J. Y., Shin, H. W. and Koo, S. H. (2009) "Characteristics of aac(6')-Ib-cr Gene in Extended-Spectrum β-Lactamase-Producing Escherichia coli and Klebsiella pneumoniae Isolated from Chungnam Area," *The Korean Journal of Laboratory Medicine*, 29(6), pp. 541–550.

Siddiqui, M. T., Mondal, A. H., Gogry, F. A., Husain, F. M., Alsalme, A. and Haq, Q. Mohd. R. (2020) "Plasmid-Mediated Ampicillin, Quinolone, and Heavy Metal Co-Resistance among ESBL-Producing Isolates from the Yamuna River, New Delhi, India," *Antibiotics*, 9(11), pp. 1–17.

Silver, S., Phung, L. T. and Silver, G. (2006) "Silver as biocides in burn and wound dressings and bacterial resistance to silver compounds," in *Journal of Industrial Microbiology and Biotechnology*, pp. 627–634.

Skurnik, D., Ruimy, R., Ready, D., Ruppe, E., Bernède-Bauduin, C., Djossou, F., Guillemot, D., Pier, G. B. and Andremont, A. (2010) "Is exposure to mercury a driving force for the carriage of antibiotic resistance genes?," *Journal of Medical Microbiology*, 59(7), pp. 804–807.

Solioz, M. (2016) "Copper Oxidation State and Mycobacterial Infection," *Mycobacterial Diseases*, 6(2).

Sone, Y., Nakamura, R., Pan-Hou, H., Itoh, T. and Kiyono, M. (2013) *Role of MerC, MerE, MerF, MerT, and/or MerP in Resistance to Mercurials and the Transport of Mercurials in Escherichia coli*, *Biol. Pharm. Bull*.

Sorsa, L. J., Dufke, S., Heesemann, J. and Schubert, S. (2003) "Characterization of an iroBCDEN Gene Cluster on a Transmissible Plasmid of Uropathogenic Escherichia coli: Evidence for Horizontal Transfer of a Chromosomal Virulence Factor," *Infection and Immunity*, 71(6), p. 3285.

Sorum, H., Roberts, M. C. and Crosa, J. H. (1992) "Identification and cloning of a tetracycline resistance gene from the fish pathogen Vibrio salmonicida," *Antimicrobial Agents and Chemotherapy*, 36(3), pp. 611–615.

Spencer, S. J., Tamminen, M. v., Preheim, S. P., Guo, M. T., Briggs, A. W., Brito, I. L., A Weitz, D., Pitkänen, L. K., Vigneault, F., Virta, M. P. and Alm, E. J. (2016) "Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers," *ISME Journal*, 10(2), pp. 427–436.

Stalder, T., Press, M. O., Sullivan, S., Liachko, I. and Top, E. M. (2019) "Linking the resistome and plasmidome to the microbiome," *The ISME Journal*, 13(10), pp. 2437–2446.

Stanisich, V. A., Bennett, P. M. and Richmond, M. H. (1977) "Characterization of a translocation unit encoding resistance to mercuric ions that occurs on a nonconjugative plasmid in Pseudomonas aeruginosa.," *Journal of Bacteriology*, 129(3), p. 1227.

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K. and Schmidt, T. M. (2015) "rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development," *Nucleic Acids Research*, 43(D1), pp. D593–D598.

Stokes, H. W. and Gillings, M. R. (2011) "Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens," *FEMS Microbiology Reviews*, 35(5), pp. 790–819.

Stokes, H. W. and Hall, R. M. (1991) "Sequence analysis of the inducible chloramphenicol resistance determinant in the TN1696 integron suggests regulation by translational attenuation," *Plasmid*, 26(1), pp. 10–19.

Stokes, H. W., Holmes, A. J., Nield, B. S., Holley, M. P., Helena Nevalainen, K. M., Mabbutt, B. C. and Gillings, M. R. (2001) "Gene Cassette PCR: Sequence-Independent Recovery of Entire Genes from Environmental DNA," *Applied and Environmental Microbiology*, 67(11), pp. 5240–5246.

Stokes, H. W., O'Gorman, D. B., Recchia, G. D., Parsekhian, M. and Hall, R. M. (1997) "Structure and function of 59-base element recombination sites associated with mobile gene cassettes," *Molecular Microbiology*, 26(4), pp. 731–745.

Sultan, I., Ali, A., Gogry, F. A., Rather, I. A., Sabir, J. S. M. and Haq, Q. M. R. (2020) "Bacterial isolates harboring antibiotics and heavy-metal resistance genes co-existing with mobile genetic elements in natural aquatic water bodies," *Saudi Journal of Biological Sciences*, 27(10), pp. 2660–2668.

Summers, A. O., Wireman, J., Vimy, M. J., Lorscheider, F. L., Marshall, B., Levy, S. B., Bennett, S. and Billard, L. (1993) "Mercury released from dental 'silver' fillings provokes an increase in mercury- and antibiotic-resistant bacteria in oral and intestinal floras of primates," *Antimicrobial Agents and Chemotherapy*, 37(4), pp. 825–834.

Surleac, M., Barbu, I. C., Paraschiv, S., Popa, L. I., Gheorghe, I., Marutescu, L., Popa, M., Sarbu, I., Talapan, D., Nita, M., Iancu, A. V., Arbune, M., Manole, A., Nicolescu, S., Sandulescu, O., Streinu-Cercel, A., Otelea, D. and Chifiriuc, M. C. (2020) "Whole genome sequencing snapshot of multi-drug resistant Klebsiella pneumoniae strains from hospitals and receiving wastewater treatment plants in Southern Romania," *PLOS ONE*, 15(1), p. e0228079.

Szuplewska, M., Czarnecki, J. and Bartosik, D. (2014) "Autonomous and non-autonomous Tn3 -family transposons and their role in the evolution of mobile genetic elements," *Mobile Genetic Elements*, 4(6), pp. 1–4.

Tapia-Pastrana, G., Chavez-Dueñas, L., Lanz-Mendoza, H., Teter, K. and Navarro-García, F. (2012) "VirK Is a Periplasmic Protein Required for Efficient Secretion of Plasmid-Encoded Toxin from Enteroaggregative Escherichia coli," *Infection and Immunity*, 80(7), p. 2276.

Tavakoli, N., Comanducci, A., Dodd, H. M., Lett, M. C., Albiger, B. and Bennett, P. (2000) "IS1294, a DNA Element That Transposes by RC Transposition," *Plasmid*, 44(1), pp. 66–84.

Tetzschner, A. M. M., Johnson, J. R., Johnston, B. D., Lund, O. and Scheutz, F. (2020) "In Silico genotyping of Escherichia coli isolates for extraintestinal virulence genes by use of whole-genome sequencing data," *Journal of Clinical Microbiology*, 58(10).

TheCivilEngineer.org (2017) *A Danish wastewater treatment plant produces more energy than it consumes*, *TheCivilEngineer.org*. Available at: https://www.thecivilengineer.org/news-center/latest-news/item/1171-a-danish-wastewater-treatment-plant-produces-more-energy-than-it-consumes (Accessed: September 28, 2021).

Thisted, T. and Gerdes, K. (1992) "Mechanism of post-segregational killing by the hok/sok system of plasmid R1. Sok antisense RNA regulates hok gene expression indirectly through the overlapping mok gene," *Journal of Molecular Biology*, 223(1), pp. 41–54.

Tosini, F., Visca, P., Luzzi, I., Dionisi, A. M., Pezzella, C., Petrucca, A. and Carattoli, A. (1998) "Class 1 Integron-Borne Multiple-Antibiotic Resistance Carried by IncFI and IncL/M Plasmids in Salmonella enterica Serotype Typhimurium," *Antimicrobial Agents and Chemotherapy*, 42(12), p. 3053.

Turner, A. K., De la Cruz, F. and Grinsted, J. (1990) "Temperature sensitivity of transposition of class II transposons," *Journal of General Microbiology*, 136(1), pp. 65–67.

Ubben, D. and Schmitt, R. (1987) "A transposable promoter and transposable promoter probes derived from Tn1721," *Gene*, 53(1), pp. 127–134.

US EPA, O. (2020) *What are Antimicrobial Pesticides?*

Vaser, R., Sovic, I., Nagarajan, N. and Sikic, M. (2017) "Fast and accurate de novo genome assembly from long uncorrected reads," *Genome Research*, 27(5), p. gr.214270.116.

Venkatesan, A. K. and Halden, R. U. (2015) "Wastewater Treatment Plants as Chemical Observatories to Forecast Ecological and Human Health Risks of Manmade Chemicals," *Scientific Reports*, 4(1), p. 3731.

Vogwill, T. and MacLean, R. C. (2015) "The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach.," *Evolutionary applications*, 8(3), pp. 284–95.

Vos, M., Hesselman, M. C., Beek, T. A. te, Passel, M. W. J. van and Eyre-Walker, A. (2015) "Rates of Lateral Gene Transfer in Prokaryotes: High but Why?," *Trends in Microbiology*, 23(10), pp. 598–605.

Waglechner, N., McArthur, A. G. and Wright, G. D. (2019) "Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance," *Nature Microbiology*, 4(11), pp. 1862–1871.

Wagner, A. (2006) "Cooperation is Fleeting in the World of Transposable Elements," *PLoS Comput Biol*, 2(12), pp. 1522–1529.

Wang, C., Cai, P., Guo, Y. and Mi, Z. (2007) "Distribution of the antiseptic-resistance genes qacEΔ1 in 331 clinical isolates of Pseudomonas aeruginosa in China," *Journal of Hospital Infection*, 66(1), pp. 93–95.

Wang, M., Tran, J., Jacoby, G., Zhang, Y., Wang, F. and Hooper, D. (2003) "Plasmid-mediated quinolone resistance in clinical isolates of Escherichia coli from Shanghai, China," *Antimicrobial agents and chemotherapy*, 47(7), pp. 2242–2248.

Ward, E. and Grinsted, J. (1987) "The nucleotide sequence of the tnpA gene of Tn21," *Nucleic Acids Research*, 15(4), pp. 1799–1806.

Water Resources England and Wales (2017) *The Water Environment (Water Framework Directive) (England and Wales) Regulations 2017*.

Weill, F.-X., Demartin, M., Tandé, D., Espié, E., Rakotoarivony, I. and Grimont, P. A. D. (2004) "SHV-12-Like Extended-Spectrum-β-Lactamase-Producing Strains of Salmonella enterica Serotypes Babelsberg and Enteritidis Isolated in France among Infants Adopted from Mali," *Journal of Clinical Microbiology*, 42(6), p. 2432.

WHO (2017) *WHO publishes list of bacteria for which new antibiotics are urgently needed*. Available at: https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed (Accessed: September 23, 2021).

WHO (2018) *Antimicrobial Resistance Fact Sheet*.

Wick, R. R., Judd, L. M., Gorrie, C. L. and Holt, K. E. (2017) "Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads," *PLOS Computational Biology*, 13(6), p. e1005595.

Xu, Z., Li, N., Shirtliff, M. E., Alam, M. J., Yamasaki, S. and Shi, L. (2009) "Occurrence and characteristics of class 1 and 2 integrons in pseudomonas aeruginosa isolates from patients in southern china," *Journal of Clinical Microbiology*, 47(1), pp. 230–234.

Yamamoto, T., Koyama, Y., Matsumoto, M., Sonoda, E., Nakayama, S., Uchimura, M., Paveenkittiporn, W., Tamura, K., Yokota, T. and Echeverria, P. (1992) "Localized, Aggregative, and Diffuse Adherence to HeLa Cells, Plastic, and Human Small Intestines by Escherichi coli Isolated from Patients with Diarrhea," *Journal of Infectious Diseases*, 166(6), pp. 1295–1310.

Yamanaka, H., Kobayashi, H., Takahashi, E. and Okamoto, K. (2008) "MacAB Is Involved in the Secretion of Escherichia coli Heat-Stable Enterotoxin II," *Journal of Bacteriology*, 190(23), p. 7693.

Yi, H., Cho, Y.-J., Yong, D. and Chun, J. (2012) "Genome sequence of Escherichia coli J53, a reference strain for genetic studies.," *Journal of bacteriology*, 194(14), pp. 3742–3.

Yoon, M. Y. and Yoon, S. S. (2018) "Disruption of the Gut Ecosystem by Antibiotics," *Yonsei Medical Journal*, 59(1), pp. 4–12.

Yu, H., Chu, L. and Misra, T. K. (1996) "Intracellular inducer Hg2+ concentration is rate determining for the expression of the mercury-resistance operon in cells.," *Journal of Bacteriology*, 178(9), p. 2712.

Yuan, L., Li, Z. H., Zhang, M. Q., Shao, W., Fan, Y. Y. and Sheng, G. P. (2019) "Mercury/silver resistance genes and their association with antibiotic resistance genes and microbial community in a municipal wastewater treatment plant," *Science of the Total Environment*, 657, pp. 1014–1022.

Zaniani, F. R., Meshkat, Z., Nasab, M. N., Khaje-Karamadini, M., Ghazvini, K., Rezaee, A., Esmaily, H., Nabavinia, M. S. and Hoseini, M. D. (2012) "The Prevalence of TEM and SHV Genes among Extended-Spectrum Beta-Lactamases Producing Escherichia Coli and Klebsiella Pneumoniae," *Iranian Journal of Basic Medical Sciences*, 15(1), p. 654.

Zankari, E., Allesøe, R., Joensen, K., Cavaco, L., Lund, O. and Aarestrup, F. M. (2017) "PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens," *The Journal of antimicrobial chemotherapy*, 72(10), pp. 2764–2768.

Zhang, B., Yu, Q., Yan, G., Zhu, H., Xu, X. Y. and Zhu, L. (2018) "Seasonal bacterial community succession in four typical wastewater treatment plants: correlations between core microbes and process performance," *Scientific Reports 2018 8:1*, 8(1), pp. 1–11.

Zhang, W., Zhu, Y., Wang, C., Liu, W., Li, R., Chen, F., Luan, T., Zhang, Y., Schwarz, S. and Liu, S. (2019) "Characterization of a Multidrug-Resistant Porcine Klebsiella pneumoniae Sequence Type 11 Strain Coharboring blaKPC-2 and fosA3 on Two Novel Hybrid Plasmids," *mSphere*, 4(5).

Zhang, X. F., Liu, Z. G., Shen, W. and Gurunathan, S. (2016) "Silver Nanoparticles: Synthesis, Characterization, Properties, Applications, and Therapeutic Approaches," *International Journal of Molecular Sciences*, 17(9).

Zhao, Q.-Y., Zhu, J.-H., Cai, R.-M., Zheng, X.-R., Zhang, L.-J., Chang, M.-X., Lu, Y.-W., Fang, L.-X., Sun, J. and Jiang, H.-X. (2021) " IS 26 Is Responsible for the Evolution and Transmission of bla NDM -Harboring Plasmids in Escherichia coli of Poultry Origin in China ," *mSystems*.

Zhou, H., Liu, W., Qin, T., Liu, C. and Ren, H. (2017) "Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of Klebsiella pneumoniae," *Frontiers in Microbiology*, 8(MAR), p. 371.

Zhou, K., Zhou, L., Lim, Q., Zou, R., Stephanopoulos, G. and Too, H. P. (2011) "Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR," *BMC Molecular Biology*, 12(18).

Zhou, Z., Alikhan, N.-F., Mohamed, K., Fan, Y., Group, A. S. and Achtman, M. (2020) "The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny and Escherichia core genomic diversity," *Genome Research*, 30(1), p. gr.251678.119.

Zhu, Y. G., Gillings, M. R., Simonet, P., Stekel, D., Banwart, S. and Penuelas, J. (2017) "Microbial mass movements," *Science*, 357(6356), pp. 1099–1100.

Zhu, Y.-G., Johnson, T. A., Su, J.-Q., Qiao, M., Guo, G.-X., Stedtfeld, R. D., Hashsham, S. A. and Tiedje, J. M. (2013) "Diverse and abundant antibiotic resistance genes in Chinese swine farms," *Proceedings of the National Academy of Sciences of the United States of America*, 110(9), p. 3435.

Zinchenko, A. A., Sergeyev, V. G., Yamabe, K., Murata, S. and Yoshikawa, K. (2004) "DNA Compaction by Divalent Cations: Structural Specificity Revealed by the Potentiality of Designed Quaternary Diammonium Salts," *ChemBioChem*, 5(3), pp. 360–368.

Zühlsdorf, M. T. and Wiedemann, B. (1992) "Tn21-specific structures in gram-negative bacteria from clinical isolates.," *Antimicrobial Agents and Chemotherapy*, 36(9), p. 1915.

Publications Arising from this Work

Laboratory Stock Variants of the Archetype Silver Resistance Plasmid pMG101 Demonstrate Plasmid Fusion, Loss of Transmissibility, and Transposition of Tn7/*pco*/*sil* Into the Host Chromosome

Steven P.T. Hooton[1,2], Alexander C.W. Pritchard[1], Karishma Asiani[1†], Charlotte J. Gray-Hammerton[1], Dov J. Stekel[1], Lisa C. Crossman[3,4], Andrew D. Millard[2] and Jon L. Hobman[1*]

[1]School of Biosciences, University of Nottingham, Sutton Bonington Campus, Sutton Bonington, United Kingdom, [2]Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom, [3]Sequenceanalysis.Co.uk, Innovation Centre, Norwich, United Kingdom, [4]School of Biological Sciences, University of East Anglia, Norwich, United Kingdom

Salmonella Typhimurium carrying the multidrug resistance (MDR) plasmid pMG101 was isolated from three burns patients in Boston United States in 1973. pMG101 was transferrable into other *Salmonella* spp. and *Escherichia coli* hosts and carried what was a novel and unusual combination of AMR genes and silver resistance. Previously published short-read DNA sequence of pMG101 showed that it was a 183.5 Kb IncHI plasmid, where a Tn7-mediated transposition of *pco*/*sil* resistance genes into the chromosome of the *E. coli* K-12 J53 host strain had occurred. We noticed differences in streptomycin resistance and plasmid size between two stocks of *E. coli* K-12 J53 pMG101 we possessed, which had been obtained from two different laboratories (pMG101-A and pMG101-B). Long-read sequencing (PacBio) of the two strains unexpectedly revealed plasmid and chromosomal rearrangements in both. pMG101-A is a non-transmissible 383 Kb closed-circular plasmid consisting of an IncHI2 plasmid sequence fused to an IncFI/FIIA plasmid. pMG101-B is a mobile closed-circular 154 Kb IncFI/FIIA plasmid. Sequence identity of pMG101-B with the fused IncFI/IncFIIA region of pMG101-A was >99%. Assembled host sequence reads of pMG101-B showed Tn7-mediated transposition of *pco*/*sil* into the *E. coli* J53 chromosome between *yhiM* and *yhiN*. Long read sequence data in combination with laboratory experiments have demonstrated large scale changes in pMG101. Loss of conjugation function and movement of resistance genes into the chromosome suggest that even under long-term laboratory storage, mobile genetic elements such as transposons and insertion sequences can drive the evolution of plasmids and host. This study emphasises the importance of utilising long read sequencing technologies of plasmids and host strains at the earliest opportunity.

Keywords: plasmid, silver resistance, pMG101, PacBio, recombination

360

INTRODUCTION

Clinical resistance to silver nitrate ($AgNO_3$) in *Salmonella* Typhimurium was first reported by McHugh and colleagues in 1975 (Mchugh et al., 1975). The *S.* Typhimurium strain had been responsible for three deaths within the burns ward of Massachusetts General Hospital in 1973. Each of the three patients who died was receiving topical applications of 0.5% $AgNO_3$ solution during treatment of severe burns. Blood- and wound-cultures all consistently produced $AgNO_3$-resistant *S.* Typhimurium, which was also resistant to ampicillin, tetracycline, chloramphenicol, sulphonamides, streptomycin, and mercuric chloride ($HgCl_2$). Conjugation studies demonstrated transmission of metal ($AgNO_3$/$HgCl_2$) and antibiotic resistances to other *Salmonella* spp. and to *Escherichia coli* K-12 J53. The plasmid responsible was designated pMG101 and is historically important as Mchugh et al. (1975) gave the first description of transmissible $AgNO_3$-resistance in the clinic. Although the original *S.* Typhimurium isolate has not been the focal point of further published research (and may no longer exist in available culture collections), the *E. coli* J53/pMG101 transconjugant strain has been studied further. Pulse-field gel electrophoresis (PFGE) of pMG101 DNA indicated an approximate size of 180 Kb, and a novel silver resistance operon was identified (Gupta et al., 1999). Assignment of pMG101 to the IncHI group and comparison of $AgNO_3$-resistance encoding genes of the *sil* operon with other plasmids has given insights into the distribution of similar elements in IncH plasmids (Gupta et al., 2001). A whole genome sequence of a stock of *E. coli* J53/pMG101 (NCTC 50110) obtained from NCTC (National Collection of Type Cultures, United Kingdom) using Illumina MiSeq technology (Accession No. ASRI00000000) was previously published (Randall et al., 2015). The calculated size of pMG101 from the published short read sequence assembly was estimated as 183.5 Kb, which agreed with the S1-endonuclease PFGE size estimate for pMG101 of 180 Kb (Gupta et al., 1999). The sequence data showed evidence of Tn7-mediated transposition of the silver (*sil*) and copper (*pco*) resistance operons as a single unit (Tn7/*sil*/*pco*) into the *E. coli* J53 chromosome (Randall et al., 2015). The antimicrobial resistance gene (ARG) and metal resistance gene (MRG) content of pMG101 was however, congruent with the observed phenotypic resistances originally reported by Mchugh et al. (1975). A class D β-lactamase (bla-$_{OXA-1}$), tetracycline resistance (*tetA*), chloramphenicol acetyltransferase (*catA*), sulphonamide resistance (*folP*), and the mercury resistance (*mer*) operon were all reported as being present in pMG101 (Randall et al., 2015).

IncHI plasmids constitute a diverse collection of large (75–400 Kb), self-transmissible, low-copy number mobile genetic elements commonly found in Gram negative *Enterobacteriaceae* (Rozwandowicz et al., 2018). Molecular typing of IncHI plasmids provides discrimination between subgroups IncHI1, IncHI2, and IncHI3. IncHI2 plasmids are prevalent in *Enterobacteriaceae* associated with causing disease in humans and animals. As such, plasmids from this incompatibility group constitute a threat to public health and animal welfare. Of significance, *Salmonella* spp., *E. coli*, *Klebsiella pneumoniae*, and *Enterobacter*

cloacae have all been identified as carriers of IncHI2 plasmids (Wong et al., 2016; Matamoros et al., 2017; Zhao et al., 2018; Chaudhry et al., 2020). Gene operons encoding metal resistances, such as tellurium (*terZABCDE/Y3Y2XY1W*), mercury (*merRTPCADE*), copper (*pcoABCDRSE*), and silver (*silESRCBAP*) are also common-place in this plasmid family (Fang et al., 2016). Co-occurrence of MRGs with multiple ARGs is widely reported (Pal et al., 2015; Li et al., 2017), and a recent study has highlighted the diverse array of ARGs encoded in IncHI2 plasmids isolated from poultry, ducks, and pigs in China (Fang et al., 2016). Extended spectrum β-lactamases (ESBL), plasmid-mediated quinolone resistance determinants (PMQRs), and Tn7/*pco*/*sil* transposition units are frequently seen in IncHI plasmids (Fang et al., 2016; Wang et al., 2017; Billman-Jacobe et al., 2018; Branger et al., 2018). Further expansion of the genetic content of IncHI plasmids can be driven *via* the acquisition of IncFI, IncFII, and IncN plasmids co-residing in the same bacterial cell. Fusion between two separate incompatibility groups (e.g., IncHI2/IncFII) may provide unique opportunities for plasmid host-range expansion allowing dissemination into novel hosts.

Given the historical interest of pMG101 as an archetype of $AgNO_3$-resistance within the clinic and its relevance to current trends in AMR research, this mobile genetic element represents an important aspect of IncHI plasmid biology. Technological developments in long-read DNA sequencing provide a useful tool for resolving features of bacterial chromosomes and plasmids. Resolution of genomic repeat regions, multi-copy insertion sequences, and transposable elements is quite often beyond the reach of short-read sequencing platforms (Rhoads and Au, 2015), but long-read sequencing can overcome such issues, and in order to build upon the previously published sequence data of pMG101 (Randall et al., 2015) we decided to reanalyze this important mobile genetic element. Due to the advances in long-read and hybrid assembly technologies, it is possible to resolve some of these repeat regions. The technology may also be able to allow identification of genomic rearrangements and changes in the strain genome after longer term storage. The changes after long term storage would be expected to be small but support the theory that strains will have the capacity to undergo recombination after years in culture. Genome sequencing of microorganisms offers a snapshot to the genome and plasmidome as part of a continuous landscape (Hobman et al., 2007).

Using two separate stocks of *E. coli* J53/pMG101 obtained from different laboratories, PacBio sequencing of both strains and plasmids was used to provide complete closed-circular chromosomal and plasmid DNA sequences. Data obtained provided confirmation of previously published work regarding pMG101 such as ARG content and Tn7/*pco*/*sil* transposition into the *E. coli* chromosome. However, previously undescribed features of pMG101 including two distinct genotypes were identified from the sequence data, and further confirmed *via* S1-endonuclease PFGE. Here, we present analysis of each isoform – pMG101-A is a 383 Kb IncHI2/IncFI/IncFII plasmid, whereas pMG101-B is a 154 Kb IncFI/IncFII plasmid. Evidence obtained in the resequencing of these two plasmids, has allowed

us to suggest a model for the transposition of the Tn7/*pco*/*sil* module into the chromosome of *E. coli* K-12 J53.

Given the medical relevance of silver as an antimicrobial able to enhance the effects of antibiotics, surveillance of resistance operons such as *sil* in a Tn7 transposable element across conjugative plasmids and novel hosts in lab and environmental hosts is of significance (Zhou et al., 2015).

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

Two separate stocks of *E. coli* J53 harbouring pMG101 were obtained from Anne Summers, University of Georgia, United States (pMG101-A) and Simon Silver, University of Chicago, United States (pMG101-B). Frozen (−80°C) glycerol stocks of both strains were used to inoculate LB agar plates prior to incubation overnight at 37°C. *Escherichia coli* J53 carrying a chromosomal insertion of *mrfp1* and *kan*ʳ was used as a recipient in conjugation experiments (J53 RFP). For liquid growth, a single well-isolated colony was used to inoculate LB broth prior to overnight incubation at 37°C with shaking, unless stated otherwise.

Pulse Field Gel Electrophoresis

Pulse-field gel electrophoresis was used to determine the sizes of the two plasmids pMG101-A and pMG101-B (Hooton et al., 2011). A single colony of *E. coli* J53 pMG101-A and *E. coli* J53 pMG101-B were inoculated in to separate 10 ml Miller's lysogeny broth -LB (Merck, United Kingdom). Broth cultures were incubated at 25°C for 48 h and shaking at 180 rpm. Following incubation cultures were centrifuged at 4,400×*g* for 5 min, the supernatant was decanted, pellets were re-suspended in 1 ml LB Broth, and 55 μl of cell suspension was transferred to a 1.5 ml micro-centrifuge tube. For PFGE plug preparation, 5 μl of 20 mg/ml proteinase K (Sigma Aldrich, United Kingdom) was added followed by mixing 50 μl of molten 1.2% PFGE grade agarose (Biorad, United States) in TE [10 mM Tris-HCl (pH 7.5), 1 mM EDTA] buffer. The suspension was mixed thoroughly and then set in PFGE plug moulds (Biorad, United States). The plugs were then incubated overnight at 55°C shaking at 300 rpm in 1 ml lysis buffer [50 mM Tris-HCl (pH 8), 50 mM EDTA, 1% *N*-lauroyl sarcosine, and 100 μg/ml proteinase K]. Plugs were then washed three times for 1 h in 1 ml wash buffer [20 mM Tris-HCl (pH 8), 50 mM EDTA] at 55°C shaking at 300 rpm. A 3 mm slice of each plug was then loaded to a 1% PFGE grade agarose gel (100 ml 1 X TAE buffer) along with PFGE lambda ladder (New England Biolabs, United States). The lanes were sealed with 1% PFGE grade agarose. A CHEF-DR II system was used for the run with a 10–30 s switch time over 18 h at 6 V/cm, and 1 X TAE running buffer was circulated at 14°C. The gel was then stained using ethidium bromide (1 μg/ml) and imaged using a Biorad ChemiDoc MP System (BioRad, United States).

Conjugation of pMG101-A and pMG101-B

A single colony of *E. coli* J53 pMG101-A, *E. coli* J53 pMG101-B and *E. coli* J53 RFP were inoculated in to separate 20 ml Miller's

LB and incubated at 25°C with shaking until an OD$_{600nm}$ 1.0 was reached. About 1 ml of each culture was then pelleted at 16,000 *g* for 1 min, the supernatant was removed before being resuspended and washed in 1 ml maximum recovery diluent – MRD (Oxoid, United Kingdom). The wash step was repeated once more. Around 50 μl of each culture containing pMG101 was added to 50 μl J53 RFP, mixed thoroughly by pipetting, and deposited on to non-selective LB agar plates and incubated at 25°C for 48 h. The lawn of growth was then scraped off each plate and resuspended in 1 ml MRD and vortexed thoroughly to end mating. The cells were then pelleted at 16,000 *g* for 1 min and supernatant removed and resuspended in 1 ml MRD. The wash step was repeated once. The OD$_{600nm}$ was then measured and adjusted to 1.0 with MRD. About 10-fold dilutions of the cell suspension were then made and spread plated in triplicate onto LB agar supplemented with 25 μg/ml HgCl$_2$ and incubated at 25°C for 48 h. Successful transconjugants, that produced red pigment, and were able to grow in the presence of HgCl$_2$, were counted. Transconjugants were then selected at random, and patch plated to LB agar supplemented with 25 μg/ml HgCl$_2$ and non-selective LB agar to confirm conjugation.

DNA Extraction

Genomic DNA was extracted from overnight *E. coli* J53 LB broth cultures using a GenElute™ Bacterial Genomic DNA Kit (Sigma Aldrich, United Kingdom) as per manufacturer's instructions. DNA integrity was checked by electrophoresing 5 μl of each preparation in a 1% agarose gel with Quick-Load® 1 Kb DNA ladder (New England Biolabs, United Kingdom) as a molecular weight marker. Genomic DNA extracts were quantified using an Invitrogen™ Qubit™ 4 Fluorometer (Fisher Scientific, United Kingdom) to determine overall yield.

Pacific Biosciences Library Preparation and Sequencing

Genomic DNA samples were purified with 1x cleaned AMPure beads (Agencourt, United Kingdom), and the quantity and quality was assessed using a Nanodrop spectrophotometer and Qubit assay. Additionally, the Fragment Analyser (Agilent, United Kingdom; using a high sensitivity genomic kit) was used to determine the average size of the DNA and the extent of degradation. This procedure was also used at the steps indicated below to determine average fragment size of the DNA. One microgram of DNA for each sample was sheared to approximately 10 Kb using a Covaris G tube (Beckman Coulter, United Kingdom), as per manufacturer's instruction. The average size was checked using a Fragment Analyser (Agilent, United Kingdom). Samples were bead purified and subjected to a DNA damage repair reaction at 37°C for 1 h followed by an end repair reaction at 25°C for 5 min. Bead cleaning was performed as described above. Samples were individually ligated to barcoded blunt-end adapters overnight at 25°C. After a heat inactivation step at 65°C for 10 min, the samples were pooled and treated with exonucleases 3 and 7 to remove any non-circular DNA. The library was size selected on Sage BluePippin (Beverly MA, United States) using a 0.75%

cassette between 5 and 50 Kb to remove any shorter material. The final library concentration was measured by Qubit assay, and DNA sizes were determined by Fragment analyser. Two SMRT cells were run for this DNA using diffusion loading and sequenced on the PacBio RS system. Libraries were sequenced using 600-min movies at the Centre for Genomic Research, University of Liverpool, Crown Street, Liverpool, L69 3BX, United Kingdom.

Sequence Assembly and Bioinformatics

The resulting reads had adapters removed prior to the trimmed reads being assembled using the "canu" assembler tool, and subsequently polished using the "arrow" algorithm (Chin et al., 2013; Koren et al., 2017). Polished assemblies were quality checked for contiguity using "QUAST" with "busco" being employed to determine completeness of assemblies using Benchmarking Universal Single-Copy Orthologs (Gurevich et al., 2013; Simão et al., 2015). In this case, BUSCO V2 was used with the BUSCO proteobacteria database used as the reference. The assembled genomes for both E. coli K-12 J53 pMG101 variants were subsequently annotated using Prokka (Seemann, 2014) and screened for ARGs and MRGs using ResFinder (Bortolaia et al., 2020) and BacMet 2.0 database (Pal et al., 2014). Plasmid incompatibility groups were identified using PlasmidFinder (Carattoli et al., 2014).

Chromosome and plasmid sequences have been deposited at NCBI under Bioprojects PRJNA 701544 and PRJNA701546. Individual Genbank accession numbers host strain and plasmid sequences are E. coli J53/pMG101-A (JAFFIC010000001.1), pMG101-A (JAFFIC010000003.1), E. coli J53/pMG101-B/Tn7 insertion (CP070962), and pMG101-B (CP070963).

Antimicrobial Susceptibility Testing

The two E. coli J53 pMG101 variants were tested for antibiotic susceptibility according to the Clinical and Laboratory Standards Institute standard protocol (CLSI, 2019). The following antibiotics were tested: ampicillin 10 μg (AMP), amoxicillin-clavulanic acid 20 and 10 μg (AMC), cefoxitin 30 μg (FOX), ceftazidime 30 μg (CAZ), cefpodoxime 10 μg (CPD), aztreonam 30 μg (ATM), imipenem 10 μg (IPM), streptomycin 10 μg (S10), tetracycline 30 μg (T), ciprofloxacin 5 μg (CIP), nalidixic Acid 30 μg (NA), trimethoprim-sulfamethoxazole 1.25 μg and 23.75 μg (SXT), chloramphenicol 30 μg (C), nitrofurantoin 300 μg (F), and azithromycin 15 μg (AZM). All discs were obtained from Pro-Lab Diagnostics (Birkenhead, United Kingdom). Antimicrobial metal resistance testing for $HgCl_2$ (184 μM) and $AgNO_3$ (400 μM) was carried out according to standard techniques (Yang et al., 2020).

RESULTS

Pulsed Field Gel Electrophoresis

Imaging of the plasmids extracted from the two J53 strains by PFGE showed a significant difference in size between the two large, low copy number plasmids. pMG101-A was identified as approximately 380 Kb, more than double the size of pMG101-B at approximately 170 Kb. Due to the size and the copy number of the plasmids, the band intensity was significantly reduced although was still clearly visible (**Supplementary Figure S1**, Supplementary data).

Sequencing of IncHI2/IncFI/II Plasmid pMG101-A and IncFI/FII Plasmid pMG101-B

Previously, we had sequenced the ~380 Kb pMG101-A plasmid using Illumina MiSeq. Assembly of short reads produced an approximately 300 Kb plasmid sequence, albeit fragmented in 27 contigs ranging from 120 to 34 and 220 bp in length (Unpublished data). To improve assembly of pMG101-A, we utilised long-read PacBio sequencing. Assembly of 10 Kb PacBio reads yielded a single closed-circular 383,246 bp dsDNA molecule representative of pMG101-A. The size is in agreement with S1-endonuclease PFGE analysis of cultures of the E. coli J53 strain harbouring pMG101-A (**Figure 1**). Analysis of the nucleotide composition revealed a GC content of 47.8%. Identification of plasmid replicons in pMG101-A was achieved using the online tool PlasmidFinder (Carattoli et al., 2014). Five separate replicons were identified in pMG101-A representing incompatibility groups IncH and IncF (**Tables 1, 2**). Several features haracteristic of IncH plasmids are apparent in pMG101-A (**Figure 1A**) including the tellurium resistance operon (terZABCDE/terY3Y2XY1W), smr loci 0018/0199, IncH conjugal transfer (tra1/tra2) regions, and plasmid replicons of the IncHI2/IncHI2A sequence types (Carattoli et al., 2014). A single 4,634,108 bp contig for the host strain E. coli K-12 J53 harbouring pMG101-A was obtained.

As was found with pMG101-A, Illumina sequencing of plasmid pMG101-B yielded a fragmented assembly (Unpublished data) containing gaps, unresolvable small contigs etc. The pco/sil genes were noticeably absent from the plasmid assembly. We used PacBio sequencing of the J53 pMG101-B strain to understand the biology of the plasmid and to confirm chromosomal integration of the pco/sil operon. A single closed-circular 154,497 bp dsDNA molecule representative of pMG101-B was assembled from PacBio reads as described above (**Figure 1B**). A single 4,683,867 bp contig representative of the host E. coli K12 J53/Tn7 insertion harbouring pMG101-B was also assembled. Analysis of the nucleotide content of pMG101-B indicates a higher GC content (53.28%) compared to pMG101-A (47.8% GC). As can be seen from **Table 1**, pMG101-B is an IncF plasmid containing replicon-types IncFIA, IncFIB, and IncFII.

pMG101-A and pMG101-B ARG and MRG Content

Resistance to toxic metals and antibiotics is encoded by MRGs and ARGs present in both pMG101-A and pMG101-B. **Table 3** shows resistances, nucleotide positions, and associated mobile genetic elements (e.g., transposons/insertion sequences). Antibiotic sensitivity tests showed that resistance phenotypes of E. coli J53 carrying pMG101-A and pMG101-B were congruent with ARGs identified in the PacBio sequence assembly. Both J53

FIGURE 1 | (A) Circular genetic map of IncH/IncF plasmid pMG101-A – the IncHI2/HI2A replicon (yellow), IncFIA/FIB/FII component (blue), Tn7/*pco*/*sil* mobile unit (grey), *ter* and *mer* metal resistance operons (red), Tn3/*strAB* antimicrobial resistance gene (ARGs; black), *tetBCD*/*catA1* (light blue), class I integron (light green), and two copies (dark green) of ISVSa5 (IS4-like transposase) are all highlighted. **(B)** Circular genetic map of IncF plasmid pMG101-B – the IncFIA/FIB/FII backbone of pMG101-B (blue), class I integron (light green), *tetBCD*/*catA1* (light blue), *mer* operon (red), and a single copy (dark green) of ISVSa5 (IS4-like transposase).

pMG101 strains were resistant to ampicillin, aminoglycosides, tetracycline, chloramphenicol, nalidixic acid, and sulphonamides, as well as AgNO₃ and HgCl₂. The discernible difference between the two variant pMG101 plasmids was *strAB*-encoded streptomycin resistance in pMG101-A, which was absent in pMG101-B.

364

TABLE 1 | Plasmid replicon typing of pMG101-A and pMG101-B.

Locus	pMG101-A		pMG101-B	
	Start	End	Start	End
IncFIB	182,235	182,916	117,399	118,080
IncFIA	215,984	216,371	83,944	84,331
IncFII	276,913	277,173	23,142	23,402
IncHI2	335,674	336,000	–	–
IncHI2A	350,868	351,497	–	–

TABLE 2 | Plasmid double locus sequence typing of pMG101-A.

Locus	Allele	Length	Start	End
FIA	1	384	215,964	216,347
FIA	6	329	216,019	216,347
FIB	1	373	182,494	182,866
FII	1	157	276,999	277,155
smr0018	2	330	369,393	369,722
smr0199	2	460	133,673	134,132

TABLE 3 | Locations of ARGs and metal resistance genes (MRGs) within pMG101-A and pMG101-B.

Gene(s)	Resistance	pMG101-A	pMG101-B
sul1*	Sulphonamide	294,833–295,672	13,551–14,390
qacEΔ1*	Quaternary ammonium compounds	295,666–296,013	14,384–14,731
aadA1*	Aminoglycoside	296,177–296,956	14,895–15,674
bla$_{OXA-1}$*	β-lactam	297,081–297,956	15,799–16,674
catA1	Chloramphenicol	303,987–304,709	153,227–154,886
tetRBCD	Tetracycline	306,933–309,549	147,819–150,940
strAB	Aminoglycoside	71,740–73,489	–
terZABCDE	Tellurium	60,310–66,529	–
merRTPCADE	Mercury(II)	282,742–286,695	1,451–5,413
silESRCFBA	Silver(I)	95,329–104,678	4,560,169–4,572,614^{chr}
pcoEABCDRS	Copper(II)	109,736–116,591	4,574,582–4,581,431^{chr}

*resistance gene belongs to gene cassette within Tn21-like integron.

^{chr}location in the bacterial chromosome of E. coli K-12 J53.

The shared ARG content of pMG101-A and pMG101-B consists of a classical Tn21-family mercury resistance operon and associated class I integron harbouring antibiotic resistance gene cassettes (**Figure 2**). The mer operon (merRTPCADE) and class I integron structure is 20,676bp in length and forms a composite transposable element of the Tn3 family (Tn21 subgroup). Both sides of the transposon are flanked by 38bp imperfect inverted repeats (IRs) with the left-IR positioned immediately upstream of merR and the right-IR immediately downstream of the Tn21 transposon. Moving inwards from the mer operon, a further set of embedded transposable elements is seen. A Mu-like transposon (tniA/tniB) and an IS26-like element (istA/istB) constitute the region between the mer operon and the class I integron. The 3'-conserved segment of the class I integron contains a typical arrangement of sul1/qacE that encodes resistance to sulphonamides and quaternary ammonium

compounds, respectively. Resistance to aminoglycoside antibiotics is encoded by the aadA1 gene cassette and is located immediately downstream of sul1. A class D β-lactamase gene, bla$_{OXA-1}$, a progenitor of the well-established and studied oxacillinases is the next antibiotic resistance gene cassette in the array of the class I integron. This suggests bla$_{OXA-1}$ was the most recently captured cassette in the integron cassette array. Finally, the diagnostic feature of the class I integron, the intI1 integrase, is located at the 5'-region. Immediately downstream of intI1 is the previously described Tn21 family transposon unit. Three genes encoding the modulation function (tnpM), a resolvase (tnpR), and the transposase (tnpA) form the Tn21 transposable element. This region is followed by the right-IR, which defines the end of the composite transposon. Immediately, downstream of the Tn21 unit is a chloramphenicol acetyltransferase (catA1) gene that confers resistance to phenicol antibiotics. IS1-like transposon genes (insA/insB) and a tetracycline resistance transposable unit (tetRBCD) are the final ARGs in this multidrug and metal resistance region.

IS-Mediated Fusion of IncHI/IncF pMG101-A

As discussed previously, the 383,246bp pMG101-A is composed of IncH1/H12A/FIA/FII/FIB plasmid sequences. Based on sequence analysis and mapping of insertion sequences and transposable elements, it is possible to hypothesise a chain of events leading to fusion of the various Inc. plasmid regions. Downstream of conjugal transfer region 2 of the IncH1/H12A element is an ISVSa5 (IS4-like transposon) that defines the beginning of the IncF region. Approximately, 157 Kb further downstream is a further ISVSa5 transposon that delineates the end of the IncF plasmid sequence. This ISVSa5 element is located immediately downstream of tetD. Due to the presence of these elements, at the junctions of the IncH components of the plasmid, it is possible that the integration event has been driven by ISVSa5-mediated recombination. An interesting feature of the IncF region is its similarity to a chromosomal genomic island (GI-DT12) previously observed in S. Typhimurium T000240. Chromosomal insertion of this large genomic island is reported as being mediated by IS1-like transposable elements (Izumiya et al., 2011). The major differences between GI-DT12 and the homologous regions found in both pMG101 variants are the presence of IncF replicons in the plasmid nucleotide sequences. The two copies of ISVSa5 that flank the IncF region of pMG101-A may have arisen following a duplication event that occurred during recombination. Analysis of pMG101-B shows that there is only a single copy of ISVSa5 present in this plasmid, which is positioned downstream of tetD.

Factors Influencing Conjugal Transfer of pMG101-A

Multiple attempts to transfer pMG101-A via conjugation into different E. coli recipient strains at 25 and 37°C were undetected, whereas pMG101-B transferred into E. coli recipients at 25°C. The pMG101-B plasmid was transferred into the recipient strain

FIGURE 2 | Genetic map of Tn*21* present in both pMG101-A and pMG101-B – the backbone consisting of the *mer* operon (red) and transposase genes (green). The insertion of the Tn*402*-like element containing Class I integron with cassette array encoding the following resistance genes: *bla*CXA-7, *aadA1*, *qacE*Δ*1*, and *sul1*. The Tn*21* mobile genetic element is flanked by two 38 bp imperfect inverted repeats.

E. coli K-12 J53 RFP kan^r at a frequency of 1.25×10^{-3} conjugants/ml. Analysis of the conjugal transfer regions of pMG101-A highlighted the presence of an IS*Ec27* element that had integrated into *orf009* of the transfer region 2 (*tra2*). It has recently been shown that *orf009* (herein designated *rsp*) encodes an essential component of the conjugation apparatus – the RSP/R27-secreted protein (Huttener et al., 2019), suggestive that IS*Ec27* has inactivated the transfer ability of pMG101-A. IS*Ec27* is found as multiple copies in both *E. coli* K12 J53 chromosomes with a total of nine in the strain harbouring pMG101-A (plus one insertion in pMG101-A *rsp*), and seven copies in *E. coli* J53 pMG101-B. No insertions of IS*Ec27* were identified in pMG101-B. Expansion of mobile genetic elements such as IS*Ec27* in long-term stocks derived from strains such as *E. coli* K12 J53 should be monitored for such events.

Tn7-Mediated *sil/pco* Chromosomal Integration

As previously reported the NCTC stock of *E. coli* J53 pMG101 strain carries a Tn*7/sil/pco* module chromosomally, conferring resistance to copper- and silver-containing compounds (Randall et al., 2015). In most instances, mobility of Tn*7*-elements is driven *via* Tn*7*-mediated transposition into the canonical *att*Tn*7* site located at the 3'-end of *glmS* in the *E. coli* chromosome. Analysis of the Tn*7/sil/pco* region in pMG101-A shows that the complete module (*tnsABCDE/silESRCFBA/pcoEABCDRS*) is located in the plasmid DNA assembly – specifically within the IncHI region of the plasmid. IncHI plasmids and their relatives are recognised as providing an environmental reservoir of MRGs, including *pco/sil* within the family *Enterobacteriaceae* (Gupta et al., 2001; Fang et al., 2016). For pMG101-B, the Tn*7/sil/pco* module has transposed into the chromosome of the *E. coli* J53 host. The transposition event has mediated the integration of Tn*7/sil/pco* in between *yhiN* (NAD[P]/FAD-dependent oxidoreductase) and *yhiM* (inner membrane protein; **Figure 3**). The overall effect of this is that *yhiN/yhiM* is now split by the Tn*7/sil/pco* module such that the two genes are separated by a genomic island of approximately 33.7 Kb in length. The absence of Tn*7/sil/pco* in the plasmid sequence of pMG101-B is also coupled to the complete absence of all the recognisable features of the IncH component of pMG101-A. Analysis of *glmS* in *E. coli* J53/pMG101-B shows that the site is unoccupied with respect to the presence of Tn*7*-like elements. The use of alternate insertion sites in the *E. coli* chromosome for the Tn*7/sil/pco* that are not the canonical *att*Tn*7* has also been observed in pig *E. coli* isolates (Chalmers et al., 2018).

DISCUSSION

Observable differences in antibiotic resistances and PFGE size profiles of plasmid pMG101 between what was expected and previously reported data (Randall et al., 2015) drove us to investigate the underlying mechanisms driving these changes. Previously, published Illumina MiSeq data for pMG101 hinted at some unusual biology but the fragmented assembly (Randall et al., 2015) and difficulties faced when resolving repeat sequences (Cahill et al., 2010) suggested a long-read sequencing approach may improve our current understanding of pMG101. Therefore, we applied PacBio long-read sequencing in order to resolve the important features of pMG101, gain complete closure of the circular genetic element, and determine any alterations to the host *E. coli* K12 J53 chromosome. Working from two separately obtained stocks of pMG101, it was possible to show evidence of plasmid recombination between IncH and IncF plasmids (pMG101-A and pMG101-B), Tn*7*-mediated transposition of MRGs into a non-canonical *att*Tn*7* site in the chromosome of *E. coli* J53 (pMG101-B), and loss of mobilisation (pMG101-A). The transposition event of pMG101-B Tn*7/sil/pco* is consistent with other reports that the Tn*7* element can insert into the chromosome of other *E. coli* strains at non-canonical *att*Tn*7* sites (Chalmers et al., 2018). Transposition of the Tn*7* element resulted in the host strain gaining chromosomal resistance to silver and copper in the form of *silESRCFBAGP* and *pcoEABCDRS*, which form with the Tn*7* transposition machinery – a recognised, likely ancient, and widely distributed, mobile genetic element (Hao et al., 2015; Hobman and Crossman, 2015; Staehlin et al., 2016). The use of this *att*Tn*7* suggests that the chromosomal insertion occurred *via* the TnsABC+D rather than the TnsABC+E system. Whilst both of the mechanisms are able to induce transposition to the chromosome, TnsABC+D promotes insertion by "cut and paste" mechanism to the specific attTn*7* sites (5'-CCCGC-3'), which is duplicated and normally 23 bases from the C-terminal end of the *glmS* gene (Waddell and Craig, 1989; Bainton et al., 1993; Mitra et al., 2010). In the case of pMG101-B, the same *att*Tn*7* site is observed between the *yhiN* and *M* genes in the *E. coli* J53 chromosome. Tn*7* and Tn*7*-like elements undergo chromosomal insertion by inserting downstream (3') adjacent to *glmS* within the host chromosome, as it is considered a safe site that does not interfere with host cell function and is therefore more likely to propagate to the daughter cells (Peters and Craig, 2001). The mechanism for insertion has been exploited and used to produce high

FIGURE 3 | Tn7-mediated transposition of metal resistance operons *pco* and *sil* into *Escherichia coli* J53 chromosome – the transposition event has resulted in insertion of the mobile unit between *yhiN* and *yhiM*.
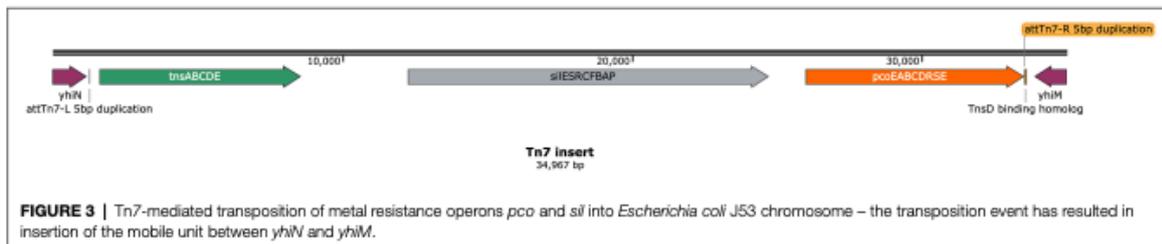
copy number expression vectors such as pRM2, a pUC18 backbone with a Tn7 insertion site also presenting resistance to ampicillin (Mckown et al., 1988).

From the assembled data, it is seen here that from the original location in pMG101-B, Tn7/*pco*/*sil* has inserted using this mechanism between *yhiN* and *yhiM*. Chromosomal insertion of Tn7 is documented to be less frequent than transposition to conjugal plasmids. In this case, pMG101-A has lost mobilisation ability due to *orf009*, a part of the *tra* genes needed for conjugative plasmid transfer, having an integrative and conjugative element (ICE) inserting in the middle of it and separating it into two parts. Over time, it is possible that the loss of conjugative ability of the plasmid may have increased the likelihood that the TnsABC+D mechanism for chromosomal insertion of the Tn7 element as a means to make sure it is passed onto daughter cells (Craig, 1996). This hypothesis is in agreement with the reported function of TnsC, which may evaluate insertion target sites for *attTn7* sites acting as a regulator for the TnsD and E proteins. TnsD then binds to a consensus sequence allowing for insertion of the Tn7 (Hauer and Shapiro, 1984). While a similar chain of events has allowed pMG101-B/Tn7 to integrate at an alternative chromosomal locus, limited sequence homology between *attTn7*/*glmS* and the insertion site of *yhiNM* is observed.

Insertion of mobile genetic elements to host DNA may be advantageous, but equally may have less favourable implications. Transmissibility of IncHI plasmids from host to recipient cells is determined by the presence of specific transfer regions – Tra1 and Tra2. In this instance, the ISEc27 element that has undergone expansion in numbers in the *E. coli* K12 J53 (pMG101-A) host strain is inserted on the opposite strand of Tra2 *rsp*, is transcribed in the opposite direction to all other genes encoded in the gene operon and may well also be exerting polar effects on genes downstream that are recognised as being essential for conjugal transfer (*trhFWUN*). For IncHI plasmids, R27 isolated from *S.* Typhi and prototypical founder for the IncHI1 plasmid group remains one of the best studied plasmids of this type (Sherburne et al., 2000; Lawley et al., 2003; Huttener et al., 2019; Luque et al., 2019). In *S.* Typhimurium SL1344, the 155.4 kDa RSP was initially identified in cell-free supernatants and found to be expressed in a temperature-dependent manner (optimal expression 25°C). RSP is exported to the extracellular surface *via* the plasmid-encoded type IV secretion system in a *trhC*-dependent manner. Externally RSP has been identified as primarily interacting with flagella, which consequently reduces cell-motility, an initial step towards bacterial conjugation. However, parallel studies with non-flagellated cells also indicate other external cellular components may interact with RSP to reduce motility (Huttener et al., 2019). Disruption of essential transfer genes such as *rsp* would therefore likely have a negative effect on conjugation ability as observed for *E. coli* K12 J53 (pMG101-A).

From the two *E. coli* J53 pMG101 cell stocks we analysed and the similarity of one stock (of pMG101-B) to the NCTC50110 stock sequenced by Randall et al. (2015), it is quite clear that there are some significant variations between the two plasmids and host strains. The fusion of a whole plasmid replicon larger than the first reported analysis of pMG101 in one stock (Gupta et al., 2001) and loss of transmissibility resulting in chromosomal insertion of the Tn7/*pco*/*sil* mobile element in the other suggest that bacterial evolution during long term storage and use in laboratories has occurred. This has broader, but unsurprising, implications for our understanding of isolates of plasmid containing strains, suggesting that in the real world these are likely to be transient forms, which can also evolve in the laboratory. Whilst sequence analysis demonstrates what the plasmid structure is, what the sequence of events that led up to these significant differences in pMG101 was, is less easy to reconstruct in a timeline. Over 20 years passed between the original reports of transfer of pMG101 into *E. coli* K-12 and reports of the size of the plasmid by PFGE. It is clear that at some point in the history of pMG101, two plasmids fused, one of which carried a Tn7/*sil*/*pco* mobile genetic element and the other of which carried SGI11. It is also clear that the Tn7/*sil*/*pco* element transposed into a non-standard *attTn7* site in the *E. coli* chromosome. As there is now cheap and readily accessible long-read sequencing available for bacterial strains, we argue that researchers should take advantage of this technology for early sequencing of minimally manipulated isolates and use sequencing as a tool to regularly check strain integrity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary material**.

AUTHOR CONTRIBUTIONS

SH, AM, DS, and JH conceived and designed the experiments. SH, AP, and C-GH performed experimental work. SH, AP, LC, C-GH, AM, and JH: bioinformatic analysis. SH, AP, KA, C-GH, LC, DS, AM, and JH: wrote the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.723322/full#supplementary-material

Supplementary Figure S1 | PFGE analysis of S1-endonuclease digestion of pMG101-A (left image) and pMG101-B (right image).

REFERENCES

Bainton, R. J., Kubo, K. M., Feng, J. N., and Craig, N. L. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. *Cell* 72, 931–943. doi: 10.1016/0092-8674(93)90581-A

Billman-Jacobe, H., Liu, Y., Haites, R., Weaver, T., Robinson, L., Marenda, M., et al. (2018). pSTM6-275, a conjugative IncHI2 plasmid of salmonella enterica That confers antibiotic and heavy-metal resistance under changing physiological conditions. *Antimicrob. Agents Chemother.* 62, e02357–e02317. doi: 10.1128/AAC.02357-17

Bortolaia, V., Kaas, R. F., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., et al. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* 75, 3491–3500. doi: 10.1093/jac/dkaa345

Branger, C., Ledda, A., Billard-Pomares, T., Doublet, B., Fouteau, S., Barbe, V., et al. (2018). Extended-spectrum beta-lactamase-encoding genes are spreading on a wide range of *Escherichia coli* plasmids existing prior to the use of third-generation cephalosporins. *Microb. Genom.* 4:e000203. doi: 10.1099/mgen.0.000203

Cahill, M. J., Koser, C. U., Ross, N. E., and Archer, J. A. (2010). Read length and repeat resolution: exploring prokaryote genomes using next-generation sequencing technologies. *PLoS One* 5:e11518. doi: 10.1371/journal.pone.0011518

Carattoli, A., Zankari, E., Garcia-Fernandez, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Chalmers, G., Rozas, K. M., Amachawadi, R. G., Scott, H. M., Norman, K. N., Nagaraja, T. G., et al. (2018). Distribution of the pco gene cluster and associated genetic determinants among swine *Escherichia coli* from a controlled feeding trial. *Gene* 9:504. doi: 10.3390/genes9100504

Chaudhry, T. H., Aslam, B., Arshad, M. I., Alvi, R. F., Muzammil, S., Yasmeen, N., et al. (2020). Emergence of Bla NDM-1 harboring *Klebsiella pneumoniae* ST29 and ST11 in veterinary settings and waste of Pakistan. *Infect. Drug Resist.* 13, 3033–3043. doi: 10.2147/IDR.S248091

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

CLSI (2019). Clinical and Laboratory Standards Institute CLSI performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals; second informational supplement; VET01-S2; Clinical and Laboratory Standards Institute. Available at: https://clsi.org/standards/products/veterinary-medicine/documents/vet01/ (Accessed July 02, 2019).

Craig, N. L. (1996). Transposon Tn7. *Curr. Top. Microbiol. Immunol.* 204, 27–48. doi: 10.1007/978-3-642-79795-8_2

Fang, L., Li, X., Li, L., Li, S., Liao, X., Sun, J., et al. (2016). Co-spread of metal and antibiotic resistance within ST3-IncHI2 plasmids from *E. coli* isolates of food-producing animals. *Sci. Rep.* 6:25312. doi: 10.1038/srep25312

Gupta, A., Matsui, K., Lo, J. F., and Silver, S. (1999). Molecular basis for resistance to silver cations in salmonella. *Nat. Med.* 5, 183–188. doi: 10.1038/5545

Gupta, A., Phung, L. T., Taylor, D. E., and Silver, S. (2001). Diversity of silver resistance genes in IncH incompatibility group plasmids. *Microbiology* 147, 3393–3402. doi: 10.1099/00221287-147-12-3393

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Hao, X., Luthje, F. L., Qin, Y., Mcdevitt, S. F., Lutay, N., Hobman, J. L., et al. (2015). Survival in amoeba--a major selection pressure on the presence of bacterial copper and zinc resistance determinants? Identification of a "copper pathogenicity island." *Appl. Microbiol. Biotechnol.* 99, 5817–5824. doi: 10.1007/s00253-015-6749-0

Hauer, B., and Shapiro, J. A. (1984). Control of Tn7 transposition. *Mol. Gen. Genet.* 194, 149–158. doi: 10.1007/BF00383510

Hobman, J. L., and Crossman, L. C. (2015). Bacterial antimicrobial metal ion resistance. *J. Med. Microbiol.* 64, 471–497. doi: 10.1099/jmm.0.023036-0

Hobman, J. L., Penn, C. W., and Pallen, M. J. (2007). Laboratory strains of *Escherichia coli*: model citizens or deceitful delinquents growing old disgracefully? *Mol. Microbiol.* 64, 881–885. doi: 10.1111/j.1365-2958.2007.05710.x

Hooton, S. P., Atterbury, R. J., and Connerton, I. F. (2011). Application of a bacteriophage cocktail to reduce salmonella Typhimurium U288 contamination on pig skin. *Int. J. Food Microbiol.* 151, 157–163. doi: 10.1016/j.ijfoodmicro.2011.08.015

Huttener, M., Prieto, A., Aznar, S., Bernabeu, M., Glaria, E., Valledor, A. F., et al. (2019). Expression of a novel class of bacterial Ig-like proteins is required for IncHI plasmid conjugation. *PLoS Genet.* 15:e1008399. doi: 10.1371/journal.pgen.1008399

Izumiya, H., Sekizuka, T., Nakaya, H., Taguchi, M., Oguchi, A., Ichikawa, N., et al. (2011). Whole-genome analysis of salmonella enterica serovar Typhimurium T000240 reveals the acquisition of a genomic island involved in multidrug resistance via IS1 derivatives on the chromosome. *Antimicrob. Agents Chemother.* 55, 623–630. doi: 10.1128/AAC.01215-10

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Lawley, T. D., Klimke, W. A., Gubbins, M. J., and Frost, L. S. (2003). F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* 224, 1–15. doi: 10.1016/S0378-1097(03)00430-0

Li, L. G., Xia, Y., and Zhang, T. (2017). Co-occurrence of antibiotic and metal resistance genes revealed in complete genome collection. *ISME J.* 11, 651–662. doi: 10.1038/ismej.2016.155

Luque, A., Paytubi, S., Sanchez-Montejo, J., Gibert, M., Balsalobre, C., and Madrid, C. (2019). Crosstalk between bacterial conjugation and motility is

368

mediated by plasmid-borne regulators. *Environ. Microbiol. Rep.* 11, 708–717. doi: 10.1111/1758-2229.12784

Matamoros, S., Van Hattem, J. M., Arcilla, M. S., Willemse, N., Melles, D. C., Penders, J., et al. (2017). Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction. *Sci. Rep.* 7:15364. doi: 10.1038/s41598-017-15539-7

Mchugh, G. L., Moellering, R. C., Hopkins, C. C., and Swartz, M. N. (1975). Salmonella typhimurium resistant to silver nitrate, chloramphenicol, and ampicillin. *Lancet* 1, 235–240. doi: 10.1016/s0140-6736(75)91138-1

Mckown, R. L., Orle, K. A., Chen, T., and Craig, N. L. (1988). Sequence requirements of Escherichia coli attTn7, a specific site of transposon Tn7 insertion. *J. Bacteriol.* 170, 352–358. doi: 10.1128/jb.170.1.352-358.1988

Mitra, R., Mckenzie, G. J., Yi, L., Lee, C. A., and Craig, N. L. (2010). Characterization of the TnsD-attTn7 complex that promotes site-specific insertion of Tn7. *Mob. DNA* 1:18. doi: 10.1186/1759-8753-1-18

Pal, C., Bengtsson-Palme, J., Kristiansson, E., and Larsson, D. G. (2015). Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics* 16:964. doi: 10.1186/s12864-015-2153-5

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., and Larsson, D. G. (2014). BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.* 42, D737–D743. doi: 10.1093/nar/gkt1252

Peters, J. E., and Craig, N. L. (2001). Tn7: smarter than we thought. *Nat. Rev. Mol. Cell Biol.* 2, 806–814. doi: 10.1038/35099006

Randall, C. P., Gupta, A., Jackson, N., Busse, D., and O'neill, A. J. (2015). Silver resistance in Gram-negative bacteria: a dissection of endogenous and exogenous mechanisms. *J. Antimicrob. Chemother.* 70, 1037–1046. doi: 10.1093/jac/dku523

Rhoads, A., and Au, K. F. (2015). PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002

Rozwandowicz, M., Brouwer, M. S. M., Fischer, J., Wagenaar, J. A., Gonzalez-Zorn, B., Guerra, B., et al. (2018). Plasmids carrying antimicrobial resistance genes in *Enterobacteriaceae*. *J. Antimicrob. Chemother.* 73, 1121–1137. doi: 10.1093/jac/dkx488

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sherburne, C. K., Lawley, T. D., Gilmour, M. W., Blattner, F. R., Burland, V., Grotbeck, E., et al. (2000). The complete DNA sequence and analysis of R27, a large IncHI plasmid from salmonella typhi that is temperature sensitive for transfer. *Nucleic Acids Res.* 28, 2177–2186. doi: 10.1093/nar/28.10.2177

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Staehlin, B. M., Gibbons, J. G., Rokas, A., O'halloran, T. V., and Slot, J. C. (2016). Evolution of a heavy metal homeostasis/resistance island reflects increasing copper stress in Enterobacteria. *Genome Biol. Evol.* 8, 811–826. doi: 10.1093/gbe/evw031

Waddell, C. S., and Craig, N. L. (1989). Tn7 transposition: recognition of the attTn7 target sequence. *Proc. Natl. Acad. Sci. U. S. A.* 86, 3958–3962.

Wang, J., Li, Y., Xu, X., Liang, B., Wu, F., Yang, X., et al. (2017). Antimicrobial resistance of *Salmonella* enterica Serovar Typhimurium in Shanghai, China. *Front. Microbiol.* 8:510. doi: 10.3389/fmicb.2017.00510

Wong, M. H., Chan, E. W., Xie, L., Li, R., and Chen, S. (2016). IncHI2 plasmids are the key vectors responsible for oqxAB transmission among *Salmonella* species. *Antimicrob. Agents Chemother.* 60, 6911–6915. doi: 10.1128/AAC.01555-16

Yang, H., Wei, S. H., Hobman, J. L., and Dodd, C. E. R. (2020). Antibiotic and metal resistance in *Escherichia coli* isolated from pig slaughterhouses in the United Kingdom. *Antibiotics* 9:746. doi: 10.3390/antibiotics9110746

Zhao, H., Chen, W., Xu, X., Zhou, X., and Shi, C. (2018). Transmissible ST3-IncHI2 plasmids are predominant carriers of diverse complex IS26-class 1 integron arrangements in multidrug-resistant *Salmonella*. *Front. Microbiol.* 9:2492. doi: 10.3389/fmicb.2018.02492

Zhou, Y., Xu, Y. B., Xu, J. X., Zhang, X. H., Xu, S. H., and Du, Q. P. (2015). Combined toxic effects of heavy metals and antibiotics on a *Pseudomonas fluorescens* strain ZY2 isolated from swine wastewater. *Int. J. Mol. Sci.* 16, 2839–2850. doi: 10.3390/ijms16022839

Conflict of Interest: LC is the director of SequenceAnalysis.co.uk and was paid to perform this work.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

369

19 **Abstract**

20

21 Integrons are bacterial genetic elements that can integrate mobile gene cassettes. They are

22 mostly known for spreading antibiotic resistance cassettes among human pathogens.

23 However, beyond clinical settings, gene cassettes encode an extraordinarily diverse range of

24 functions important for bacterial adaptation. The recovery and sequencing of cassettes has

25 promising applications, including: surveillance of clinically important genes, particularly

26 antibiotic resistance determinants; investigating the functional diversity of integron-carrying

27 bacteria; and novel enzyme discovery. Although gene cassettes can be directly recovered

28 using PCR, there are no standardised methods for their amplification and, importantly, for

29 validating sequences as genuine integron gene cassettes. Here, we present reproducible

30 methods for the PCR amplification, sequence processing, and validation of gene cassette

31 amplicons from complex communities. We describe two different PCR assays that either

32 amplify cassettes together with integron integrases, or gene cassettes together within cassette

33 arrays. We compare the use of Nanopore and Illumina sequencing, and present bioinformatic

34 pipelines that filter sequences to ensure that they represent amplicons from genuine integrons.

35 Using a diverse set of environmental DNAs, we show that our approach can consistently

36 recover thousands of unique cassettes per sample and up to hundreds of different integron

37 integrases. Recovered cassettes confer a wide range of functions, including antibiotic

38 resistance, with as many as 300 resistance cassettes found in a single sample. In particular,

39 we show that class 1 integrons appear to be collecting and concentrating antibiotic resistance

40 genes out of the broader diversity of cassette functions. The methods described here can be

41 applied to any environmental or clinical microbiome sample.

42

43

44 **Introduction**

45

46 Integrons are bacterial genetic elements that can capture, mobilise and rearrange gene

47 cassettes [1, 2]. They are mostly known for spreading a diverse repertoire of gene cassettes

48 that collectively confer resistance to almost all classes of antibiotics [3]. Beyond clinical

49 settings, however, integrons play a crucial role in bacterial evolution by rapidly generating

50 genomic diversity [4, 5]. Functional integrons are characterised by their flagship gene, the

51 integron integrase (*intI*), which encodes a site-specific tyrosine recombinase (IntI). IntI

52 mediates the insertion of gene cassettes at the integron recombination site (*attI*), which acts as

53 the insertion site of captured gene cassettes [6]. Gene cassettes, prior to their insertion, are

54 circular molecules, which possess a cassette recombination site (*attC*). Their insertion

55 involves IntI-mediated recombination between the *attI* site of the integron and the *attC* site of

56 the cassette [7-11]. Multiple cassettes can be inserted to form a linear cassette array, which

57 can vary in size from zero or one cassette to hundreds [12, 13]. IntI activity is induced by

58 DNA damage, often triggered by environmental stress [14, 15]. Integrons can therefore

59 provide genomic diversity at precisely the moment when it is needed the most, thus

60 facilitating 'adaptation on demand' [16].

61   Recovery and sequence analysis of integron gene cassettes serve several purposes.

62 First, screening gene cassettes can provide a direct method for surveillance of resistance

63 genes that are prevalent in an environment of interest. It has also been proposed that

64 surveying gene cassettes can help detect novel functions that might be harmful to human

65 health, such as increased pathogenicity or resistance to novel antibiotics [17]. In particular,

66 class 1 integrons, due to their mobility, abundance and distribution [18, 19], are primed to

67 play a crucial role in dissemination of these genes. Finally, exploring gene cassettes provides

68 a window into the functional diversity of the bacterial pangenome. Gene cassettes have been

371

69 found to be extraordinarily abundant and diverse in every environment surveyed [20-26].

70 Further, many cassettes with known functions act as single-gene/single-trait entities. As such,

71 they need minimal integration into metabolic networks and can likely function in a relatively

72 wide range of genomic contexts. These traits make them highly valuable commodities for

73 synthetic biology and biotechnological applications, particularly for the discovery of diverse

74 enzymatic activities [17].

75 Currently, gene cassettes can be recovered from genome sequencing of cultured

76 isolates, whole metagenomic sequencing, or amplicon sequencing of *attC*-associated genes.

77 Since most bacteria are yet to be cultured, cassettes identified from isolate genomes

78 inevitably reflect only a small proportion of all gene cassettes, exacerbated by the fact that

79 different strains of the same species can vary widely in cassette content [4]. Whole

80 metagenomic sequencing, although potentially a less biased approach, can be challenging, as

81 gene cassettes often exist at very low abundances and can contain repeat sequences. A

82 targeted amplicon sequencing approach, however, can overcome these issues and could

83 provide the most efficient method for recovering diverse gene cassettes from complex

84 microbial communities [27].

85 Cassette-targeted amplicon sequencing has been used previously, with varying returns

86 in gene cassette recovery [20-27]. As sequencing technologies have improved, the ability to

87 capture a greater diversity of gene cassettes has also increased [20, 26]. However, such

88 studies lack standardised methods for amplifying and, importantly, validating amplicon

89 sequences as part of genuine cassettes arrays. Here, we present standardised and reproducible

90 methods for amplifying, sequencing, and bioinformatic filtering of genuine gene cassettes

91 from mixed microbial communities.

92 We applied two different PCR assays using DNA isolated from diverse environmental

93 samples with the aim of recovering integron integrases and gene cassettes. PCR products

94   were sequenced with both long-read Nanopore (ONT) and short-read Illumina MiSeq

95   sequencing technologies. Importantly, we present bioinformatic pipelines that filter

96   sequences for complete *attC* sites or *intI* genes. We show that after filtering, we can

97   consistently recover thousands of gene cassettes from a single sample. We find that recovered

98   genes display a diverse suite of functional traits, including antibiotic resistance.

99

100   **Methods**

101

102   *Sample collection and DNA extraction*

103   Duplicate samples were collected from six different sites (3 x terrestrial and 3 x aquatic

104   environments). Terrestrial sites consisted of urban parkland soil from Macquarie University

105   (Sydney, New South Wales, Australia) [20], hot desert soil from Sturt National Park (North-

106   western New South Wales) [28, 29], and Antarctic soil from Herring Island [20]. The aquatic

107   sites consisted of river sediment (Lane Cove River, New South Wales), freshwater biofilms

108   (Mars Creek, New South Wales) [30], and estuarine sediment (Paramatta River Estuary, New

109   South Wales). From each of the 12 samples, DNA was extracted from 0.3 g of material using

110   a standard bead-beating protocol [31]. Each resulting DNA sample was used as the template

111   for two different PCR assays, described below, and all were subsequently sequenced using

112   long-read Nanopore (ONT) and short-read MiSeq technologies (Fig. 1).

113

114   *PCR amplification and DNA sequencing*

115   For each sample, we conducted two different PCR assays (Fig. 1A). The first used the

116   primers HS287 and HS286 [27], which target *attC* recombination sites in opposing directions

117   to amplify intervening gene cassettes. The second primer set, intI-R / HS286, amplifies

118   approximately 800bp of the integron integrase gene as well as adjacent gene cassettes. The

119 primer intI-R (5'- GCG AAC GAR TGB CGV AGV GTG TG -3') was designed to target

120 diverse integron integrases and was based on an alignment of 174 complete *intI* sequences

121 containing a functional catalytic site, as compiled by Cambray et al. [14]. Importantly, the

122 last 6 bp of the 3' end of intI-R exactly matches 75% of aligned *intI* sequences. For

123 amplification, we used Phusion Hot Start II DNA Polymerase (ThermoFisher Scientific,

124 Waltham, MA, USA), which is a long-range DNA polymerase, chosen to facilitate the

125 amplification of large segments of integron cassette arrays, known to reach more than 100

126 kilobases in length [12]. The PCRs were carried out in 50 μL volumes containing a final

127 concentration of 1 x GC Phusion Buffer, 0.2 mM dNTPs, 0.5 μM of each primer, 3% DMSO

128 and 2 U of Phusion DNA polymerase. All PCRs were performed using GeneReleaser®

129 (Bioventures, Murfreesboro, TN, USA) as previously described [32]. Triplicate PCRs were

130 performed for each sample to increase the chances of capturing rare gene cassettes that might

131 otherwise escape amplification due to the stochastic nature of PCR.

132   For the HS287 / HS286 primer set, the following thermal cycling program was used:

133 98 °C for 3 min for 1 cycle; 98 °C for 10 s, 60 °C for 30 s, 72 °C for 3 min 30 s for 35 cycles;

134 and a final extension step at 72 °C for 10 min. For the intI-R / HS286 primer set, the

135 following thermal cycling program was used: 98 °C for 3 min for 1 cycle; 98 °C for 10 s, 65 °C

136 for 30 s, 72 °C for 3 min 30 s for 35 cycles; and a final extension step at 72 °C for 10 min.

137 PCR efficiency was assessed using 2% agarose gel electrophoresis. Triplicate PCRs were

138 pooled and then purified with AMPure XP beads (Beckman Coulter, Danvers, MA, USA)

139 according to the manufacturer's protocol.

140   For long-read sequencing, the 24 PCR products (representing the 12 samples

141 amplified with each primer set) were multiplexed in a single DNA library using the ONT

142 Ligation Sequencing Kit (SQK-LSK109) and the ONT Native Barcoding Expansion Kits

143 (EXP-NBD104 and EXP-NBD114) according to the manufacturer's protocol. The DNA

374

144　library was sequenced using a MinION MK 1B sequencing device on an R10.3 flow cell.

145　Sequencing was allowed to run for 24 hours. Basecalling was carried out with Guppy v.4.3.4

146　with default parameters using the high accuracy (HAC) basecalling model.

147　　　For short-read sequencing, the 24 PCR products underwent an Illumina DNA shotgun

148　library preparation using the Nextera XT protocol and then sequenced with MiSeq 300 bp

149　paired-end sequencing on a single lane. Illumina sequencing and library preparation were

150　carried out at the Australian Genome Research Facility (Melbourne, Australia).

151

152　*Sequence processing and attC filtering: HS287 / HS286 PCRs*

153　To compare short- and long-read sequencing technologies, we sequenced HS287 / HS286

154　PCRs on both Nanopore (ONT) and MiSeq platforms. The respective workflows and

155　software used for sequence processing and filtering are summarised in Figure 1B.

156　　　For ONT sequences of the HS287 / HS286 PCRs, we first filtered reads based on

157　average quality (q) scores. We removed any reads with an average q score below 10 using

158　NanoFilt v2.8 [33] [parameters: -q 10]. Although each read spans the length of an entire

159　amplicon, many amplicons represent overlapping subsections of larger potential templates.

160　Thus, an assembly of these initial reads into larger cassette arrays was carried out using Canu

161　v2.0 [34] [parameters: genomeSize=5m minReadLength=250 minOverlapLength=200

162　corMinCoverage=0 corOutCoverage=20000 corMhapSensitivity=high

163　maxInputCoverage=20000 batMemory=125 redMemory=32 oeaMemory=32 batThreads=24

164　purgeOverlaps=aggressive]. Assembled contigs and unassembled reads were then extracted

165　together using the tgStoreDump script within Canu [parameters: -consensus -fasta]. Any

166　redundancies were removed using dedupe.sh, available from the BBTools package v35 [35]

167　with default parameters. Consensus sequences were then corrected with 4 rounds of polishing

168　using Racon v1.4.20 [36]. Each round of Racon polishing involved read mapping with

169 minimap2 v2.20-r1061 [37] [parameters: -x map-ont -t 24] and error correction with Racon

170 [parameters: -m 8 -x 6 -g -8 -w 500 -t 24].

171   For MiSeq sequence data of the HS287 / HS286 PCRs, paired-end reads first

172 underwent quality trimming and adapter clipping using Trimmomatic v0.38 [38] [parameters:

173 -phred33 ILLUMINACLIP:adapters.fa:2:30:10 LEADING:3 TRAILING:3

174 SLIDINGWINDOW:4:15 MINLEN:30] where 'adapters.fa' is a fasta-formatted file

175 containing all commonly used Illumina adapter sequences. If only one end of paired reads

176 had acceptable quality, it was used as a single read during assembly. Surviving paired-end

177 reads and single reads were assembled together using SPAdes v3.14.1 [39-41] [parameters -k

178 21,33,55,77,99,127 --only-assembler --careful].

179   The resulting ONT and MiSeq sequences were both filtered based on the presence of

180 internal cassette recombination sites (attCs). Given the degenerate nature of the primers, off-

181 target amplicons might constitute a significant portion of the reads. Filtering for sequences

182 that have internal attC sites is thus an essential step when analysing cassette amplicon data. It

183 should be noted that filtering for attCs in this way may discard some genuine amplicons

184 which consist of single gene cassettes, since they do not possess a complete attC site (Fig.

185 1A). Nevertheless, we consider that for obtaining meaningful ecological data, the removal of

186 potential false positives is more important than the loss of some true positives. To filter for

187 attC sites, we used an in-house script, attC-screening.sh (available:

188 https://github.com/timghaly/integron-filtering), with default parameters. The script uses the

189 HattCI [42] + Infernal [43] pipeline that has been previously described [44, 45]. In short,

190 attC-screening.sh searches for the sequence and secondary structures conserved among attCs

191 and retains any input sequence that has at least one attC site. The script can be used on data

192 generated from any sequencing technology.

193

376

194 *intI-R / HS286 PCRs: sequence processing and IntI filtering*

195 All intI-R / HS286 PCRs were also sequenced on both ONT and MiSeq platforms (Fig. 1B).

196 For Nanopore sequencing, basecalled reads were first quality filtered using NanoFilt v2.8

197 [33] [parameters: -q 10]. Reads representing concatemers and chimeras were removed using

198 yacrd v0.6.2 [46] with default parameters for ONT data. Since all amplicons should be

199 anchored on one end to the *intI* gene, an assembly would not be suitable. Instead, we

200 clustered reads into amplicon 'types' using isONclust v0.0.6.1 [47]. Each cluster was then

201 individually corrected using isONcorrect v0.0.8 [48] with default parameters. Unlike error-

202 correction of genomic data, isONcorrect takes into account uneven coverage within the same

203 read as well as structural variation among similar reads from different clusters (e.g., reads

204 that represent true biological rearrangements of the same gene cassettes). From each

205 corrected cluster, a consensus sequence was then generated using spoa v4.0.7 [36]

206 [parameter: -r 0]. Any redundancies were removed using the BBTools v35 [35] script

207 dedupe.sh with default parameters.

208 MiSeq sequences were processed in the same manner as described above for the

209 HS287 / HS286 data. This involved quality trimming and adapter clipping using

210 Trimmomatic v0.38 [38], followed by an assembly of the reads using SPAdes v3.14.1 [39-

211 41].

212 The resulting ONT and MiSeq sequences were both filtered based on the presence of

213 IntI protein sequences. To detect sequences that encoded IntI, we used an in-house script,

214 intI-screening.sh (available: https://github.com/timghaly/integron-filtering), with default

215 parameters. The script uses a profile hidden Markov model (HMM) provided by Cury et al.

216 [13] to detect the additional domain that is unique to integron integrases, separating them

217 from other tyrosine recombinases [49]. The intI-screening.sh pipeline first uses Prodigal [50]

218 to predict all encoded protein sequences, and then screens them for the IntI-specific domain

377

219 using hmmsearch from the HMMER v3 software package [51]. Any sequences that do not

220 contain a recognisable integron integrase are discarded. Similarly, intI-screening.sh can be

221 used on data generated from any sequencing technology.

222

223 *Protein prediction and functional classification of gene cassettes*

224 Cassette open reading frames (ORFs) and their translated protein sequences were predicted

225 using Prodigal v2.6.3 [50] in metagenomic mode [parameters: -p meta].

226 To assess the broad-scale functional diversity of gene cassettes, we used the Clusters

227 of Orthologs Groups (COGs) database [52]. COG functions were assigned to cassette-

228 encoded protein sequences using eggNOG-mapper v2.0.1b [53, 54] executed in DIAMOND

229 [55] mode with default parameters. To detect cassette-encoded antimicrobial resistance genes

230 (ARGs), we used ABRicate v0.8 [56] to search against the Comprehensive Antibiotic

231 Resistance Database (CARD) [57] [Downloaded: 2021-Apr-21].

232

233 *Taxonomic classification of attC sites*

234 The gene cassettes of sedentary chromosomal integrons (SCIs) generally possess highly

235 similar *attC* sites, and this conservation spans the SCIs of different bacteria within the same

236 taxon [4, 58, 59]. We have recently modelled the conserved sequence and structure of *attC*

237 sites from the chromosomal integrons of 11 bacterial taxa [44]. These included six

238 Gammaproteobacterial orders (Alteromonadales, Methylococcales, Oceanospirillales,

239 Pseudomonadales, Vibrionales, Xanthomonadales) and an additional five phyla

240 (Acidobacteria, Cyanobacteria, Deltaproteobacteria, Planctomycetes, Spirochaetes). A

241 covariance model (CM) was generated separately for each taxon, and this can be used to

242 correctly identify the source taxon of *attC* sites with high specificity (98-100%) [44].

378

243   Here, we used an in-house script, attC-taxa.sh (available:

244   https://github.com/timghaly/attC-taxa), with default parameters to detect any *attC* sites that

245   have originated in the SCIs of one of the 11 taxa. The attC-taxa.sh pipeline incorporates all

246   11 CMs and uses cmsearch [parameters: --notrunc --max] from the Infernal software package

247   [43] to classify *attC*s. It is important to note that each taxon-specific model exhibits different

248   sensitivities in detecting true positives and thus the relative proportion of different taxa

249   cannot be compared within the same sample. However, the relative proportion of the same

250   taxon can be compared between different samples.

251

252   *ONT – MiSeq comparisons*

253   For comparisons of the cassette and integrase diversity recovered between ONT and MiSeq

254   technologies, we first considered differences in sequencing depth. To do this, we randomly

255   subsampled 50 Mb of raw reads from each sample using rasusa v0.3.0 [60] [parameters: --

256   coverage 50 --genome-size 1Mb]. After subsampling, all sequence processing and filtering

257   steps were repeated as described above.

258   All formal comparisons were made using two-sample T-tests (or Wilcoxon rank sum

259   tests if the data were not normally distributed) using the rstatix v0.7.0 R package [61]. To

260   determine if the data were normally distributed, Shapiro-Wilk tests were carried out using

261   rstatix v0.7.0 [61], as well as visually inspected the distributions against their theoretical

262   normal distributions using Q-Q plots generated with the R package ggpubr v0.4.0 [62].

263   To assess the overlap in recovered ORFs between ONT and MiSeq, we mapped the

264   cassette ORFs from one technology to the reads of the other using minimap2 v2.20-r1061

265   [37]. We considered the ORF to be present if it had a mean coverage depth of at least 1x that

266   spanned at least 98% of the ORF. For ONT and MiSeq read mapping, we used the minimap2

267   presets [-ax map-ont -t 8] and [-ax sr -t 8], respectively. Coverage statistics were extracted

268 from the mapping alignments using the 'sort' and 'coverage' programs within the SAMtools

269 software package [63, 64].

270

271 **Results and Discussion**

272

273 Here, we present a stringent pipeline for PCR amplifying, sequencing and analysing integron

274 integrases and gene cassettes from diverse microbial communities (Fig. 1). For this, we used

275 two different PCR primer sets, HS287 / HS286 and intI-R / HS286 (Fig. 1A). The sample

276 types consisted of a wide variety of soils (from an urban parkland, an Australian desert, and

277 an Antarctic island), as well as river and estuarine sediments, and freshwater biofilms.

278 To assess the suitability of long- and short-read sequencing technologies, we

279 sequenced amplicons from both PCR assays using Nanopore (ONT) and Illumina MiSeq

280 platforms, respectively. Average ONT yield was 181 Mb (100 – 358 Mb per sample) for the

281 HS287 / HS286 primer set and 216 Mb (62 – 502 Mb per sample) for the intI-R / HS286

282 primer set. The average MiSeq yield was 418 Mb (228 – 720 Mb per sample) and 663 Mb

283 (275 – 1,247 Mb per sample), respectively for these primer sets.

284 To ensure amplicons were part of genuine integrons, we filtered the HS287 / HS286

285 data for *attC* sites, and the intI-R / HS286 data for IntI protein sequences (Fig. 1B). For the

286 HS287 / HS286 data, an average of 23.8% and 19.0% of amplicon sequences were retained

287 after filtering for ONT and MiSeq, respectively (Supplementary Fig. S1A). While, for the

288 intI-R / HS286 data, an average of 1.2% and 1.5% of sequences remained after filtering for

289 ONT and MiSeq, respectively (Supplementary Fig. S1B). The difference in proportions of

290 sequences retained after filtering between ONT and MiSeq were not statistically significant

291 for either primer set. The low proportion of surviving sequences for the intI-R / HS286 data is

292 likely a result of the intI-R primer binding to other tyrosine recombinases. While many

380

293   sequences were filtered out, the data retained from this primer set, as described below,

294   include a large, diverse set of both known and entirely novel integron integrases and gene

295   cassettes.

296   The lengths of the recovered sequences for both primer sets were significantly larger

297   for ONT sequencing compared to MiSeq (Supplementary Fig. S2). For the HS287 / HS286

298   set, sequence lengths ranged from 500 bp to more than 25,304 bp for ONT, and 500 bp to

299   19,244 bp for MiSeq. For the intI-R / HS286 data, sequence lengths ranged from 803bp to

300   16,179bp for ONT and 800bp to 7,432bp for MiSeq.

301

302   *Recovered diversity of gene cassette ORFs*

303   We assessed the efficiency of both primer sets in recovering gene cassette open reading

304   frames (ORFs). Among all 12 samples, the HS287 / HS286 primers amplified 33,854 and

305   62,118 non-redundant cassette-encoded proteins when sequenced with ONT and MiSeq,

306   respectively (Fig. 2A). After adjusting for sequencing depth, there was no significant

307   difference in cassette recovery between the two sequencing technologies (Fig. 2B). On

308   average, we observed that ~50% of cassette ORFs sequenced with one technology were also

309   recovered by the other (Supplementary Fig. S3A).

310   The HS287 / HS286 primer set is preferred in order to recover the greatest diversity of

311   gene cassettes. Indeed, the recovery rate of gene cassettes using the methods described here

312   surpasses any previously described approach. Notably, Pereira et al. [45] conducted an

313   impressive survey of gene cassettes from 10 terabases of metagenomic data obtained from 14

314   public databases. Across all datasets, they identified an average of 0.03 unique cassette ORFs

315   per 500 kilobases of assembled data. Here, we recover 218 and 265 ORFs per 500 kilobases

316   of assembled data when sequenced with ONT and MiSeq, respectively. Although screening

317   metagenomes may provide a relatively unbiased approach in analysing gene cassettes, it

318   clearly requires much deeper sequencing to recover sufficient cassette data for in-depth

319   ecological or evolutionary analyses. Studies of integrons and their associated genetic cargo

320   will therefore continue to benefit from the use of amplicon sequencing approaches, such as

321   described in the present study.

322   For the intI-R / HS286 primer set, we recovered a total of 9,641 and 3,742 non-

323   redundant cassette ORFs when sequenced with ONT and MiSeq, respectively (Fig. 2C). ONT

324   sequencing recovered significantly more (P<0.0001) of the cassette ORF diversity of each

325   sample than MiSeq (Fig. 2D). This was despite all ONT cassette ORFs being covered by the

326   MiSeq reads (Supplementary Fig. S3B). This shows that although the MiSeq reads cover all

327   the cassettes being amplified by the primers, recovery of the cassettes is sub-optimal, most

328   likely due to difficulties in their assembly. In particular, different cassette arrays associated

329   with the same or similar integron integrase are likely to be problematic for a short-read

330   assembly approach. While the intI-R / HS286 primer pair does not recover as much diversity

331   as the HS287 / HS286 set, it does provide additional key information on IntI diversity

332   (discussed further below) and indicates which gene cassettes are associated with which *intI*

333   genes.

334   The intI-R / HS286 primer set can also reveal which gene cassettes are located

335   towards the start of a cassette array (Fig. 1A). This is of biological and ecological

336   significance, since the first cassettes in arrays are the most recently inserted cassettes and are

337   likely to be strongly expressed [65]. During environmental perturbations, integron integrase

338   activity leads to the acquisition of novel and rearrangement of existing cassettes, inserting

339   them at the start of the array where strong expression is guaranteed [17, 66, 67]. Selection

340   fixes lineages with first-position cassettes that confer significant advantages. Thus, gene

341   cassettes recovered from the intI-R / HS286 primer set might provide important ecological

342   insights at the time of sampling.

343

344   *Recovered diversity of integron integrases*

345   Using the intI-R / HS286 primers, we recovered a total of 1,413 and 1,867 different integron

346   integrase genes among the 12 samples when sequenced with ONT and MiSeq, respectively

347   (Fig. 3A). There was no significant difference in integron-integrase recovery between the two

348   sequencing technologies, with or without adjusting for sequencing depth (Figs. 3A-B). Both

349   sequencing technologies could recover an impressive diversity of IntIs from the 12 samples.

350   In comparison, a comprehensive screening of 2,484 bacterial genomes recovered only 215

351   different IntIs [13]. This shows that despite the low rate of intI-R / HS286 sequences that are

352   retained after filtering, a significant number of novel integron sequences are recovered.

353   To determine how many classes of integrons these IntIs represented, we sought to

354   define the amino acid clustering threshold for an integron class. To do this, we used the most

355   abundant and widely distributed IntI, the class 1 integron integrase (IntI1) [68]. Here, we

356   iteratively set decreasing amino acid clustering thresholds for our library of IntIs using CD-

357   HIT v4.6 [69, 70] [parameters: -n 5 -d 0 -g 1 -t 0]. We continued until all IntI1s in our dataset

358   were grouped into a single cluster while ensuring all non-IntI1s were excluded

359   (Supplementary Fig. S4A). This resulted in a 94% amino acid identity being selected as the

360   most appropriate clustering threshold for IntI1s. Although this might not reflect the ideal

361   threshold for all classes, it nevertheless provides a semi-quantitative approach to defining an

362   integron class based on amino acid homology.

363   Using a 94% clustering threshold, we recovered a total of 984 and 1,646 integron

364   classes among our dataset when sequenced with ONT and MiSeq, respectively

365   (Supplementary Fig. S4B). There was no significant difference in integron class recovery

366   between the two sequencing technologies, with or without adjusting for sequencing depth

367   (Figs. 3C-D). In addition, we examined the most prevalent integron classes; defined here as

383

368 those being IntIs that were present in at least one-third of all samples (Table 1). This

369 identified ten prevalent IntI classes, found to be 60-70% similar to endogenous IntIs from

370 diverse bacterial phyla (Table 1). Not surprisingly, class 1 integrons were the only class to be

371 found in every sample, including those from Antarctica and outback Australia.

372

373 *Functional diversity of gene cassettes*

374 Here we show that gene cassette ORFs largely encode proteins of unknown functions (Fig.

375 4A). This is in agreement with previous functional analyses of gene cassettes [5, 20, 24, 25].

376 On average, only ~20% of cassette-encoded proteins amplified with HS287 / HS286 could be

377 assigned a COG functional category, approximately half of which could be assigned a

378 category of known function (Fig. 4A). The dominant COG categories were 'Transcription',

379 'Replication, recombination and repair', and 'Amino acid transport and metabolism' (Fig.

380 4B). We show that our methods are capable of recovering gene cassettes that confer a wide

381 range of traits spanning many functional classes.

382

383 *Cassette-encoded antibiotic resistance*

384 For a more specific functional characterisation, we focused on antimicrobial resistance, since

385 these phenotypes are often conferred by integron gene cassettes in clinical settings [3, 71,

386 72]. Interestingly, we found that for either primer set, ONT sequencing could recover many

387 more ARGs than MiSeq, the latter recovering no ARGs for most samples (Fig. 5A). In

388 contrast, ONT sequencing recovered as many as 300 ARG cassettes within a single sample.

389 We suspect that this discrepancy is an artefact caused by the high similarity between different

390 ARG types, and multiple arrangements of the same ARGs in class 1 cassette arrays that

391 makes their assembly difficult from short-read data. In total, we recovered 106 different

392 ARGs from both primer sets when sequenced with ONT (Fig. 5B). Almost all ARG cassettes

393 encoded proteins known to confer resistance to β-lactam and aminoglycoside antibiotics,

394 these being the most commonly observed integron-mediated resistance types [3].

395   Upon examining all cassette ORFs associated with class 1 integron integrases

396 recovered using intI-R / HS286 primers (Table 1), we found that 162 of 462 (34.6%) were

397 known ARGs. In comparison, only 586 of the 10,385 (5.6%) total cassette ORFs amplified

398 with this primer set were known ARGs. These findings show that class 1 integrons are

399 collecting and concentrating ARG cassettes out of the broader diversity of cassette functions.

400 This enrichment strongly supports the idea that class 1 integrons are key vectors for

401 acquisition and dissemination of antibiotic resistance [3, 73, 74].

402

403 *Taxonomic classification of attC sites*

404 We could identify the likely taxonomic sources of 5,998 *attC*s (18.8%) and 10,257 *attC*s

405 (20%) sequenced with ONT and MiSeq, respectively. For taxonomic classification, we used

406 models that capture the sequence and structural homology of chromosomal *attC*s from 11

407 different taxa. These included six Gammaproteobacterial orders (Alteromonadales,

408 Methylococcales, Oceanospirillales, Pseudomonadales, Vibrionales, Xanthomonadales) and

409 an additional five phyla (Acidobacteria, Cyanobacteria, Deltaproteobacteria, Planctomycetes,

410 Spirochaetes). It should be noted that although the specificity (ability to reject false positives)

411 of each model is very high (98-100%), they exhibit a wide range of sensitivities (proportion

412 of true positive detected) [44]. Therefore, relative abundance of each taxon cannot be

413 compared within the same sample, however, the same taxon can be compared between

414 different samples. It also indicates that the relative abundances of each taxon are likely to be

415 lower-bound estimates.

416   Here, we show that the relative abundance of each taxon varied across the different

417 sampled environments (Fig. 6). For example, both the Cyanobacteria- and Methylococcales-

385

418   type *attC*s were most abundant in urban parkland soil (Fig. 6C-D), while Vibrionales-type

419   *attC*s were more abundant in estuarine sediments and freshwater biofilm samples (Fig. 6F

420   and Supplementary Figure S5 for a comparison of all 11 taxa). Such data can provide useful

421   information on the taxonomic contribution to gene cassette pools among different samples.

422

423   **Conclusions**

424

425   Here, we present experimental and bioinformatic methods for the PCR amplification, DNA

426   sequencing and analysis of integrons from microbial communities. We describe approaches

427   using two different PCR assays and compare the outputs from ONT and MiSeq sequencing.

428   We find that, relative to sequencing depth, ONT generally outperforms or performs the same

429   as MiSeq regarding the recovery of gene cassettes and integron integrases. Most notably,

430   ONT outperforms MiSeq in the recovery of complete ARG gene cassette sequences. We also

431   find that the primer set HS287 / HS286 is efficient at amplifying a wide range of gene

432   cassettes, encompassing extensive *attC* and functional diversity. However, the intI-R / HS286

433   primer set can provide additional useful information in linking gene cassettes with an

434   integron class. For example, we show that class 1 integrons are collecting and concentrating

435   ARGs relative to the broader cassette pool.

436   Our described methods can recover key information on the diverse pool of gene

437   cassettes that are helping drive adaptation and niche specialisation in bacteria [4, 16, 67].

438   Such an approach allows us to investigate the potential traits that are available to integron-

439   carrying bacteria, and to understand the role that gene cassettes play in mediating

440   evolutionary responses under environmental or clinical selection pressures. In addition, the

441   large proportion of cassettes with unknown functions provides an important resource for the

442   discovery of novel enzymatic activities [17].

443

Data availability

445 All sequence data generated in this study are available from the NCBI SRA database under

446 the BioSample accessions SAMN21354384 to SAMN21354431. All BioSamples are linked

447 to the NCBI BioProject PRJNA761546.

448

Code availability

450 The code used for filtering sequences to ensure that they represent amplicons from genuine

451 integrons are available at https://github.com/timghaly/integron-filtering. The code used to

452 predict the taxonomic sources of gene cassette recombination sites (*attC*s) is available at

453 https://github.com/timghaly/attC-taxa.

454

Acknowledgements

458

References

460 1. Gillings MR: **Integrons: past, present, and future.** *Microbiology and Molecular*
461 *Biology Reviews* 2014, **78**:257-277.
462 2. Mazel D: **Integrons: agents of bacterial evolution.** *Nature Reviews Microbiology*
463 2006, **4**:608-620.
464 3. Partridge SR, Tsafnat G, Coiera E, Iredell JR: **Gene cassettes and cassette arrays in**
465 **mobile resistance integrons.** *FEMS Microbiology Reviews* 2009, **33**:757-784.
466 4. Gillings MR, Holley MP, Stokes H, Holmes AJ: **Integrons in *Xanthomonas*: a**
467 **source of species genome diversity.** *Proceedings of the National Academy of*
468 *Sciences* 2005, **102**:4419-4424.
469 5. Boucher Y, Labbate M, Koenig JE, Stokes H: **Integrons: mobilizable platforms**
470 **that promote genetic diversity in bacteria.** *Trends in Microbiology* 2007, **15**:301-
471 309.
472 6. Stokes H, Hall RM: **A novel family of potentially mobile DNA elements encoding**
473 **site-specific gene-integration functions: integrons.** *Molecular Microbiology* 1989,
474 **3**:1669-1683.

475   7.   Vit C, Richard E, Fournes F, Whiteway C, Eyer X, Lapaillerie D, Parissi V, Mazel D,
476   Loot C: **Cassette recruitment in the chromosomal Integron of *Vibrio cholerae*.**
477   *Nucleic Acids Research* 2021, **49**:5654-5670.
478   8.   Bouvier M, Demarre G, Mazel D: **Integron cassette insertion: a recombination**
479   **process involving a folded single strand substrate.** *The EMBO Journal* 2005,
480   **24**:4356-4367.
481   9.   Loot C, Ducos-Galand M, Escudero JA, Bouvier M, Mazel D: **Replicative resolution**
482   **of integron cassette insertion.** *Nucleic Acids Research* 2012, **40**:8361-8370.
483   10.   Mukhortava A, Pöge M, Grieb MS, Nivina A, Loot C, Mazel D, Schlierf M:
484   **Structural heterogeneity of *attC* integron recombination sites revealed by optical**
485   **tweezers.** *Nucleic Acids Research* 2018, **47**:1861-1870.
486   11.   Nivina A, Escudero JA, Vit C, Mazel D, Loot C: **Efficiency of integron cassette**
487   **insertion in correct orientation is ensured by the interplay of the three unpaired**
488   **features of *attC* recombination sites.** *Nucleic Acids Research* 2016, **44**:7792-7803.
489   12.   Chen C-Y, Wu K-M, Chang Y-C, Chang C-H, Tsai H-C, Liao T-L, Liu Y-M, Chen
490   H-J, Shen AB-T, Li J-C: **Comparative genome analysis of *Vibrio vulnificus*, a**
491   **marine pathogen.** *Genome research* 2003, **13**:2577-2587.
492   13.   Cury J, Jové T, Touchon M, Néron B, Rocha EP: **Identification and analysis of**
493   **integrons and cassette arrays in bacterial genomes.** *Nucleic Acids Research* 2016,
494   **44**:4539-4550.
495   14.   Cambray G, Sanchez-Alberola N, Campoy S, Guerin É, Da Re S, González-Zorn B,
496   Ploy M-C, Barbé J, Mazel D, Erill I: **Prevalence of SOS-mediated control of**
497   **integron integrase expression as an adaptive trait of chromosomal and mobile**
498   **integrons.** *Mobile DNA* 2011, **2**:6.
499   15.   Guerin É, Cambray G, Sanchez-Alberola N, Campoy S, Erill I, Da Re S, Gonzalez-
500   Zorn B, Barbé J, Ploy M-C, Mazel D: **The SOS response controls integron**
501   **recombination.** *Science* 2009, **324**:1034-1034.
502   16.   Escudero JA, Loot C, Nivina A, Mazel D: **The integron: adaptation on demand.**
503   *Microbiology Spectrum* 2015, **3**:MDNA3-0019-2014.
504   17.   Ghaly TM, Geoghegan JL, Tetu SG, Gillings MR: **The peril and promise of**
505   **integrons: beyond antibiotic resistance.** *Trends in Microbiology* 2020, **28**:455-464.
506   18.   Ghaly TM, Chow L, Asher AJ, Waldron LS, Gillings MR: **Evolution of class 1**
507   **integrons: Mobilization and dispersal via food-borne bacteria.** *PLoS One* 2017,
508   **12**:e0179169.
509   19.   Gillings MR: **Class 1 integrons as invasive species.** *Current Opinion in*
510   *Microbiology* 2017, **38**:10-15.
511   20.   Ghaly TM, Geoghegan JL, Alroy J, Gillings MR: **High diversity and rapid spatial**
512   **turnover of integron gene cassettes in soil.** *Environmental Microbiology* 2019,
513   **21**:1567-1574.
514   21.   Elsaied H, Stokes H, Nakamura T, Kitamura K, Fuse H, Maruyama A: **Novel and**
515   **diverse integron integrase genes and integron-like gene cassettes are prevalent in**
516   **deep-sea hydrothermal vents.** *Environmental Microbiology* 2007, **9**:2298-2312.
517   22.   Elsaied H, Stokes HW, Kitamura K, Kurusu Y, Kamagata Y, Maruyama A: **Marine**
518   **integrons containing novel integrase genes, attachment sites, *attI*, and associated**
519   **gene cassettes in polluted sediments from Suez and Tokyo Bays.** *The ISME*
520   *Journal* 2011, **5**:1162-1177.
521   23.   Elsaied H, Stokes HW, Kitamura K, Kurusu Y, Kamagata Y, Maruyama A: **Marine**
522   **integrons containing novel integrase genes, attachment sites, *attI*, and associated**
523   **gene cassettes in polluted sediments from Suez and Tokyo Bays.** *The ISME ournal*
524   2011, **5**:1162-1177.

525 24. Koenig JE, Boucher Y, Charlebois RL, Nesbø C, Zhaxybayeva O, Bapteste E,
526   Spencer M, Joss MJ, Stokes HW, Doolittle WF: **Integron-associated gene cassettes**
527   **in Halifax Harbour: assessment of a mobile gene pool in marine sediments.**
528   *Environmental Microbiology* 2008, **10**:1024-1038.
529 25. Koenig JE, Sharp C, Dlutek M, Curtis B, Joss M, Boucher Y, Doolittle WF: **Integron**
530   **gene cassettes and degradation of compounds associated with industrial waste:**
531   **the case of the Sydney tar ponds.** *PLoS One* 2009, **4**:e5276.
532 26. Dias MF, de Castro GM, de Paiva MC, de Paula Reis M, Facchin S, do Carmo AO,
533   Alves MS, Suhadolnik ML, de Moraes Motta A, Henriques I, et al: **Exploring**
534   **antibiotic resistance in environmental integron-cassettes through *intI-attC***
535   **amplicons deep sequencing.** *Brazilian Journal of Microbiology* 2021, **52**:363-372.
536 27. Stokes HW, Holmes AJ, Nield BS, Holley MP, Nevalainen KH, Mabbutt BC,
537   Gillings MR: **Gene cassette PCR: sequence-independent recovery of entire genes**
538   **from environmental DNA.** *Applied and Environmental Microbiology* 2001,
539   **67**:5240-5246.
540 28. Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, Gillings M,
541   Beattie AJ: **Spatial scaling of microbial eukaryote diversity.** *Nature* 2004, **432**:747-
542   750.
543 29. Oliver I, Holmes A, Dangerfield JM, Gillings M, Pik AJ, Britton DR, Holley M,
544   Montgomery ME, Raison M, Logan V: **Land systems as surrogates for biodiversity**
545   **in conservation planning.** *Ecological Applications* 2004, **14**:485-503.
546 30. Gillings MR, Krishnan S, Worden PJ, Hardwick SA: **Recovery of diverse genes for**
547   **class 1 integron-integrases from environmental DNA samples.** *FEMS*
548   *Microbiology Letters* 2008, **287**:56-62.
549 31. Yeates C, Gillings M, Davison A, Altavilla N, Veal D: **Methods for microbial DNA**
550   **extraction from soil for PCR amplification.** *Biological Procedures Online* 1998,
551   **1**:40-47.
552 32. Chow L, Waldron L, Gillings M: **Potential impacts of aquatic pollutants: sub-**
553   **clinical antibiotic concentrations induce genome changes and promote antibiotic**
554   **resistance.** *Frontiers in Microbiology* 2015, **6**.
555 33. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C: **NanoPack:**
556   **visualizing and processing long-read sequencing data.** *Bioinformatics* 2018,
557   **34**:2666-2669.
558 34. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM: **Canu:**
559   **scalable and accurate long-read assembly via adaptive k-mer weighting and**
560   **repeat separation.** *Genome Research* 2017, **27**:722-736.
561 35. Bushnell B: **BBTools software package.** vol. 578. pp. 579.
562   http://sourceforge.net/projects/bbmap; 2014:579.
563 36. Vaser R, Sović I, Nagarajan N, Šikić M: **Fast and accurate *de novo* genome**
564   **assembly from long uncorrected reads.** *Genome Research* 2017, **27**:737-746.
565 37. Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics*
566   2018, **34**:3094-3100.
567 38. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina**
568   **sequence data.** *Bioinformatics* 2014, **30**:2114-2120.
569 39. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
570   Nikolenko SI, Pham S, Prjibelski AD: **SPAdes: a new genome assembly algorithm**
571   **and its applications to single-cell sequencing.** *Journal of Computational Biology*
572   2012, **19**:455-477.
573 40. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A,
574   Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y, et al: **Assembling Genomes and**

575   **Mini-metagenomes from Highly Chimeric Reads.** In*; Berlin, Heidelberg*. Springer
576   Berlin Heidelberg; 2013: 158-170.

577 41. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A: **Using SPAdes**
578   ***de novo* assembler.** *Current Protocols in Bioinformatics* 2020, **70**:e102.

579 42. Pereira MB, Wallroth M, Kristiansson E, Axelson-Fisk M: **HattCI: Fast and**
580   **accurate *attC* site identification using hidden Markov models.** *Journal of*
581   *Computational Biology* 2016, **23**:891-902.

582 43. Nawrocki EP, Eddy SR: **Infernal 1.1: 100-fold faster RNA homology searches.**
583   *Bioinformatics* 2013, **29**:2933-2935.

584 44. Ghaly TM, Tetu SG, Gillings MR: **Predicting the taxonomic and environmental**
585   **sources of integron gene cassettes using structural and sequence homology of**
586   ***attC* sites.** *Communications Biology* 2021, **4**:946.

587 45. Pereira MB, Österlund T, Eriksson KM, Backhaus T, Axelson-Fisk M, Kristiansson
588   E: **A comprehensive survey of integron-associated genes present in metagenomes.**
589   *BMC Genomics* 2020, **21**:1-14.

590 46. Marijon P, Chikhi R, Varré J-S: **yacrd and fpa: upstream tools for long-read**
591   **genome assembly.** *Bioinformatics* 2020, **36**:3894-3896.

592 47. Sahlin K, Medvedev P: ***De novo* clustering of long-read transcriptome data using**
593   **a greedy, quality value-based algorithm.** *Journal of Computational Biology* 2020,
594   **27**:472-484.

595 48. Sahlin K, Medvedev P: **Error correction enables use of Oxford Nanopore**
596   **technology for reference-free transcriptome analysis.** *Nature Communications*
597   2021, **12**:2.

598 49. Messier N, Roy PH: **Integron integrases possess a unique additional domain**
599   **necessary for activity.** *Journal of Bacteriology* 2001, **183**:6699-6706.

600 50. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal:**
601   **prokaryotic gene recognition and translation initiation site identification.** *BMC*
602   *Bioinformatics* 2010, **11**:119.

603 51. Eddy SR: **HMMER 3.2: Biosequence analysis using profile hidden Markov**
604   **models.** *URL: http://hmmerorg/* 2018.

605 52. Tatusov RL, Galperin MY, Natale DA, Koonin EV: **The COG database: a tool for**
606   **genome-scale analysis of protein functions and evolution.** *Nucleic Acids Research*
607   2000, **28**:33-36.

608 53. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C,
609   Bork P: **Fast genome-wide functional annotation through orthology assignment**
610   **by eggNOG-mapper.** *Molecular Biology and Evolution* 2017, **34**:2115-2122.

611 54. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H,
612   Mende DR, Letunic I, Rattei T, Jensen LJ: **eggNOG 5.0: a hierarchical, functionally**
613   **and phylogenetically annotated orthology resource based on 5090 organisms and**
614   **2502 viruses.** *Nucleic Acids Research* 2019, **47**:D309-D314.

615 55. Buchfink B, Xie C, Huson DH: **Fast and sensitive protein alignment using**
616   **DIAMOND.** *Nature Methods* 2015, **12**:59-60.

617 56. Seemann T: **ABRicate: mass screening of contigs for antimicrobial and virulence**
618   **genes.** *Department of Microbiology and Immunology, The University of Melbourne,*
619   *Melbourne, Australia Available online: https://githubcom/tseemann/abricate* 2018.

620 57. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave
621   BM, Pereira S, Sharma AN: **CARD 2017: expansion and model-centric curation of**
622   **the comprehensive antibiotic resistance database.** *Nucleic Acids Research*
623   2016:gkw1004.

624 58. Rowe-Magnus DA, Guerout A-M, Biskri L, Bouige P, Mazel D: **Comparative**
625   **analysis of superintegrons: engineering extensive genetic diversity in the**
626   **Vibrionaceae.** *Genome Research* 2003, **13**:428-442.
627 59. Vaisvila R, Morgan RD, Posfai J, Raleigh EA: **Discovery and distribution of super-**
628   **integrons among Pseudomonads.** *Molecular Microbiology* 2001, **42**:587-601.
629 60. Hall M: **Rasusa: Randomly subsample sequencing reads to a specified coverage.**
630   2019.
631 61. Kassambara A: **rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R**
632   **package version 0.7.0.** 2021.
633 62. Kassambara A: **ggpubr: 'ggplot2' Based Publication Ready Plots. R package**
634   **version 0.4.0.** 2020.
635 63. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
636   Durbin R: **The sequence alignment/map format and SAMtools.** *Bioinformatics*
637   2009, **25**:2078-2079.
638 64. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A,
639   Keane T, McCarthy SA, Davies RM: **Twelve years of SAMtools and BCFtools.**
640   *Gigascience* 2021, **10**:giab008.
641 65. Collis CM, Hall RM: **Expression of antibiotic resistance genes in the integrated**
642   **cassettes of integrons.** *Antimicrobial Agents and Chemotherapy* 1995, **39**:155-162.
643 66. Souque C, Escudero JA, MacLean RC: **Integron activity accelerates the evolution**
644   **of antibiotic resistance.** *Elife* 2021, **10**:e62474.
645 67. Escudero JA, Loot C, Mazel D: **Integrons as adaptive devices.** In *Molecular*
646   *Mechanisms of Microbial Evolution.* Edited by Rampelotto PH: Springer International
647   Publishing; 2018: 199-239
648 68. Zhu Y-G, Gillings M, Simonet P, Stekel D, Banwart S, Penuelas J: **Microbial mass**
649   **movements.** *Science* 2017, **357**:1099-1100.
650 69. Li W, Jaroszewski L, Godzik A: **Clustering of highly homologous sequences to**
651   **reduce the size of large protein databases.** *Bioinformatics* 2001, **17**:282-283.
652 70. Li W, Jaroszewski L, Godzik A: **Tolerating some redundancy significantly speeds**
653   **up clustering of large protein databases.** *Bioinformatics* 2002, **18**:77-82.
654 71. Zhu Y-G, Zhao Y, Li B, Huang C-L, Zhang S-Y, Yu S, Chen Y-S, Zhang T, Gillings
655   MR, Su J-Q: **Continental-scale pollution of estuaries with antibiotic resistance**
656   **genes.** *Nature Microbiology* 2017, **2**:16270.
657 72. Ghaly TM, Paulsen IT, Sajjad A, Tetu SG, Gillings MR: **A novel family of**
658   *Acinetobacter* **mega-plasmids are disseminating multi-drug resistance across the**
659   **globe while acquiring location-specific accessory genes.** *Frontiers in Microbiology*
660   2020, **11**:605952.
661 73. Rowe-Magnus DA, Mazel D: **The role of integrons in antibiotic resistance gene**
662   **capture.** *International Journal of Medical Microbiology* 2002, **292**:115-125.
663 74. Stalder T, Barraud O, Casellas M, Dagot C, Ploy M-C: **Integron involvement in**
664   **environmental spread of antibiotic resistance.** *Frontiers in Microbiology* 2012,
665   **3**:119.
666

667   **Figures and Table**
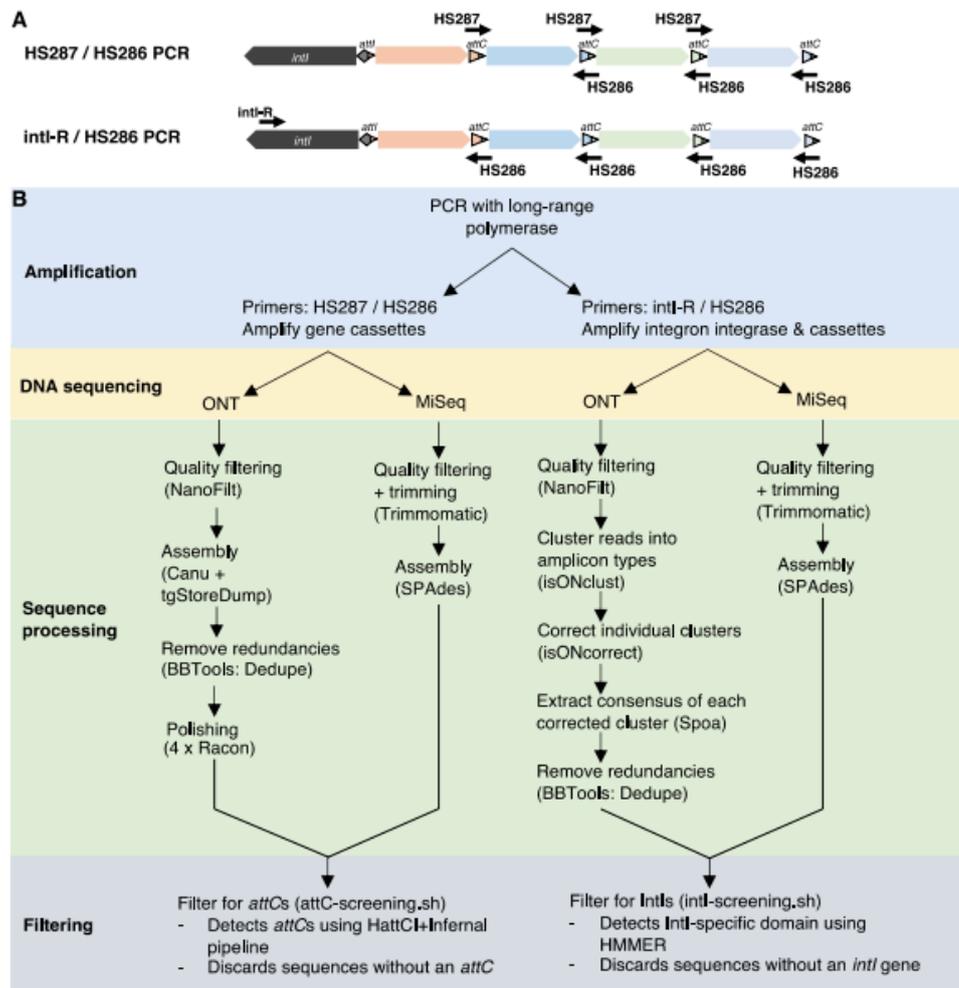
668   **Figure 1. Experimental and bioinformatic workflow for gene cassette amplicon**
669   **sequencing**. (**A**) Components of integrons amplified by the two PCR assays. The primer set
670   HS287 / HS286 targets cassettes that lie between two *attC* sites. Potentially any gene
671   cassette(s) can be amplified by this primer set. The primer set intI-R / HS286 targets diverse
672   integron integrases (*intI*) and cassette recombination sites (*attC*). The resulting amplicons
673   include ~800 bp of *intI* and at least the first cassette(s) of an array. (**B**) The bioinformatic
674   steps and software (in parentheses) used to process and filter amplicon data. Methods are
675   shown for both primer sets sequenced with either Nanopore (ONT) or Illumina MiSeq.
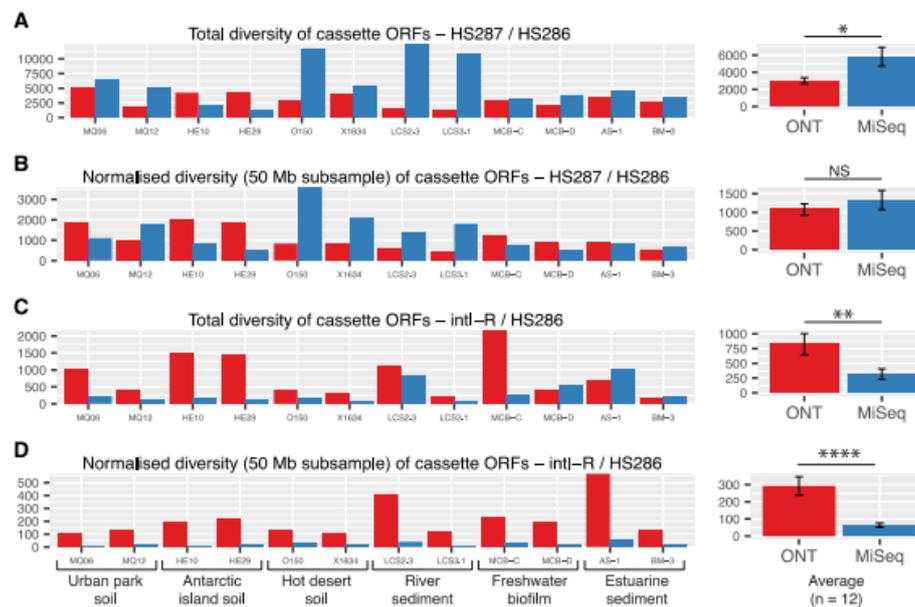
676 **Figure 2. Diversity of recovered gene cassette ORFs**. Redundancy was removed using a
677 100% amino acid identity of translated protein sequences. (**A**) Total non-redundant cassette
678 ORFs amplified using the primers HS287 / HS286. (**B**) Cassette ORF diversity was
679 normalised for HS287 / HS286 sequencing depth (based on a 50 Mb subsample of sequence
680 reads). Total (**C**) and normalised (**D**) cassette ORF diversity are shown for the intI-R / HS286
681 primer set. Average (± 1 S.E) diversity for each analysis are shown on the right-hand side of
682 each panel. The degree of statistical significance is shown by asterisks as determined by two-
683 sample T-tests or Wilcoxon rank sum tests (depending on the normality of the data). NS:
684 $P > 0.05$, *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$, ****: $P < 0.0001$.

393
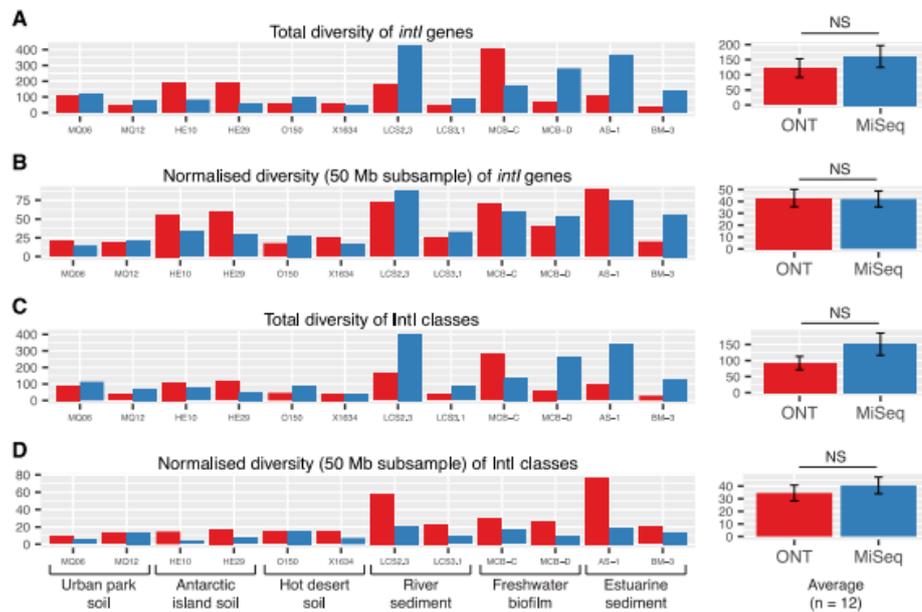
685 **Figure 3. Diversity of integron integrases recovered by the intI-R / HS286 primer set.**
686 (**A**) Total non-redundant (100% amino acid identity) integron integrases (IntIs) recovered.
687 (**B**) IntI diversity was normalised for sequencing depth (based on a 50 Mb subsample of
688 sequence reads). Total (**C**) and normalised (**D**) diversity of IntI classes (using a 94% amino
689 acid clustering threshold) are shown. Average (± 1 S.E) diversity for each analysis are shown
690 on the right-hand side of each panel. Differences between Nanopore (ONT) and Illumina
691 MiSeq technologies were not significant (NS) as determined by Wilcoxon rank sum tests.
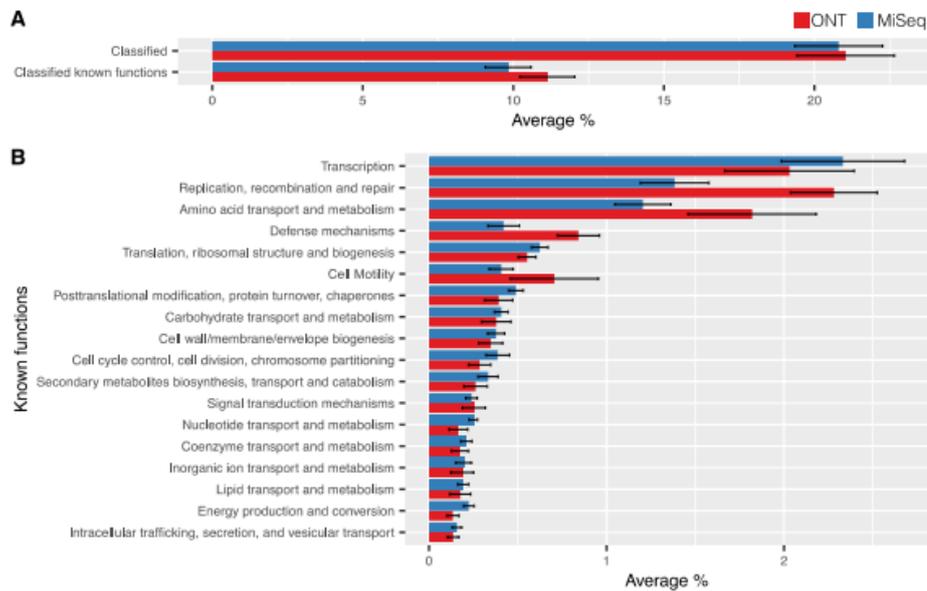
692 **Figure 4. COG functional analysis of cassette-encoded proteins recovered with the**
693 **HS287 / HS286 primer set.** (**A**) Average (± 1 S.E) percentage of proteins per sample (n=12)
694 that can be classified into functional categories. On average ~20% of cassette-encoded
695 proteins can be classified by a COG category, half of which fall into categories of known
696 function. (**B**) The average (± 1 S.E) proportion of proteins within a sample assigned to each
697 of the known functional categories.

Figure 5. Abundance and diversity of antibiotic resistance gene (ARG) cassettes. (A) Abundance of ARGs recovered from either primer set. **(B)** The average (± 1 S.E) abundance of each ARG type recovered from Nanopore (ONT) sequencing per sample (n=12).
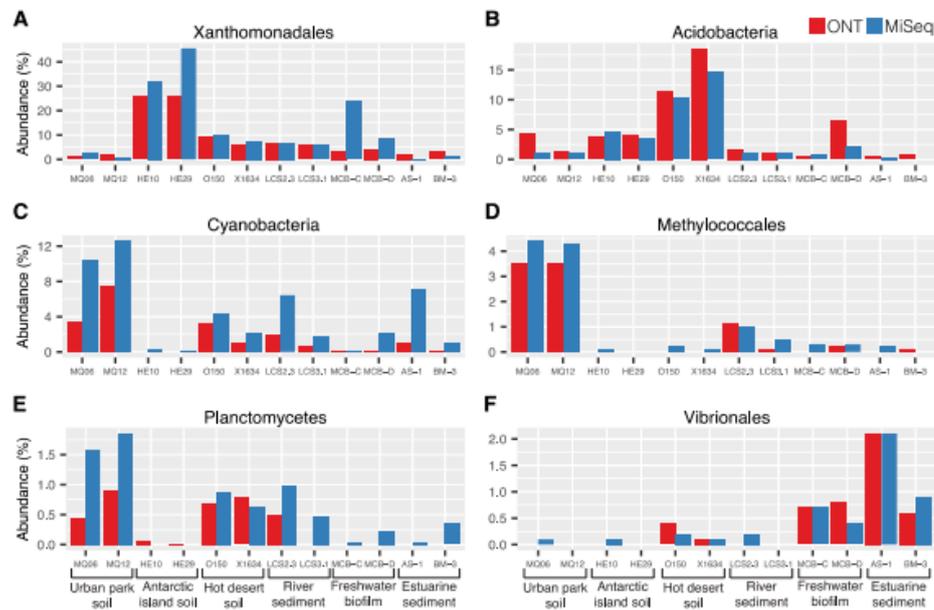
701 **Figure 6. Taxonomic classification of gene cassette recombination sites (*attC*s).**
702 Taxonomic predictions are based on a selection of six (**A-F**) of the eleven available
703 taxonomic models of chromosomal *attC*s. Each figure panel shows the proportion of *attC*s
704 across each sample that exhibit sequence and structure conserved among that taxon. For a
705 comparison of all eleven taxa, see Supplementary Figure S5.

706   **Table 1. Most prevalent integron integrase (IntI) classes**

Number of IntIs in cluster/class	Percentage of samples present in	BLASTP taxa	BLASTP amino acid %
94	100	Class 1 integron - Multispecies	99.3
71	91.7	Xanthomonadales (Rhodanobacteraceae & Xanthomonadaceae)	~70
19	91.7	Multiple phyla (Deltaproteobacteria, Nitrospinae, Chloroflexi)	~60
26	66.7	Xanthomonadaceae (*Lysobacter, Vulcaniibacterium, Thermomonas, Luteimonas, Pseudoxanthomonas*)	~70
10	58.3	Betaproteobacteria	~80
13	41.7	'IntI1-like' - Multispecies	~91
13	41.7	Rhodanobacteraceae	~75
12	41.7	Xanthomonadales (Rhodanobacteraceae & Xanthomonadaceae)	~74
8	41.7	Xanthomonadales (Rhodanobacteraceae & Xanthomonadaceae)	~72
7	41.7	Planctomycetes	~67

398

707 **Supplementary Figures**
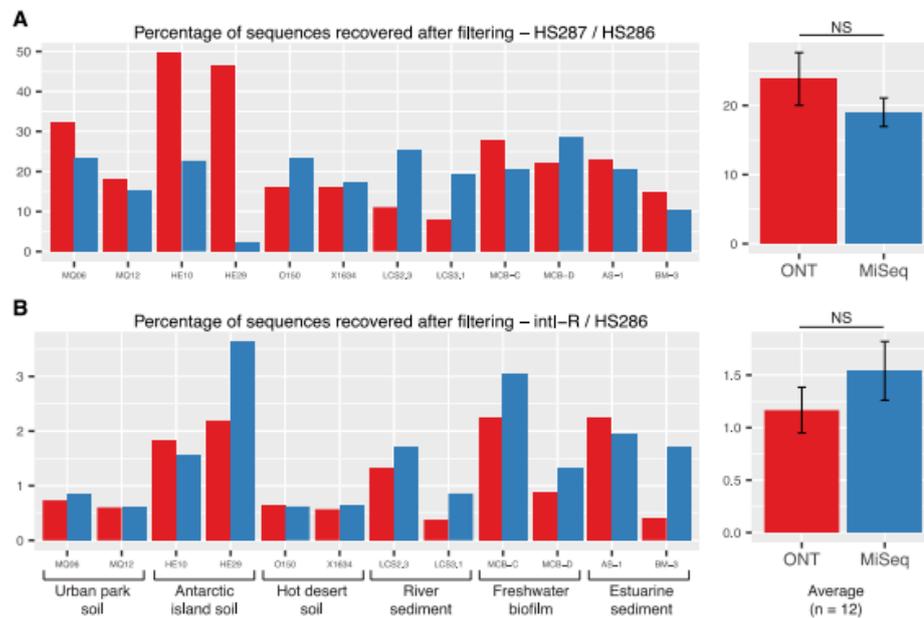
708 **Figure S1. Percentages of sequences recovered after bioinformatic filtering.** Filtering
709 involved screening (**A**) HS287 / HS286 sequences for cassette recombination sites (*attC*s)
710 and (**B**) intI-R / HS286 sequences for integron integrase (IntI) encoding genes to ensure that
711 they represented amplicons of genuine integrons. Average (± 1 S.E) diversity for each
712 analysis are shown on the right-hand side of each panel. Differences between Nanopore
713 (ONT) and Illumina MiSeq technologies were not significant (NS) as determined by two-
714 sample T-tests.
715
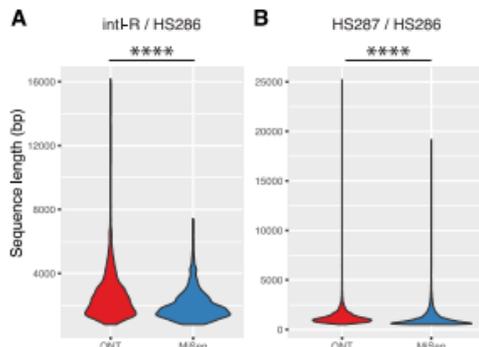
Figure S2. Sequence lengths recovered by each sequencing technology. The violin plots show the range of sequence lengths (bp) for (**A**) intI-R / HS286 and (**B**) HS287 / HS286 primer sets. The width of each curve represents the relative density of datum points across the ranges. For both primer sets, the average length of recovered amplicons (n=12) is significantly larger (Wilcoxon rank sum test, P<0.0001) when sequenced with Nanopore (ONT) compared MiSeq sequencing.
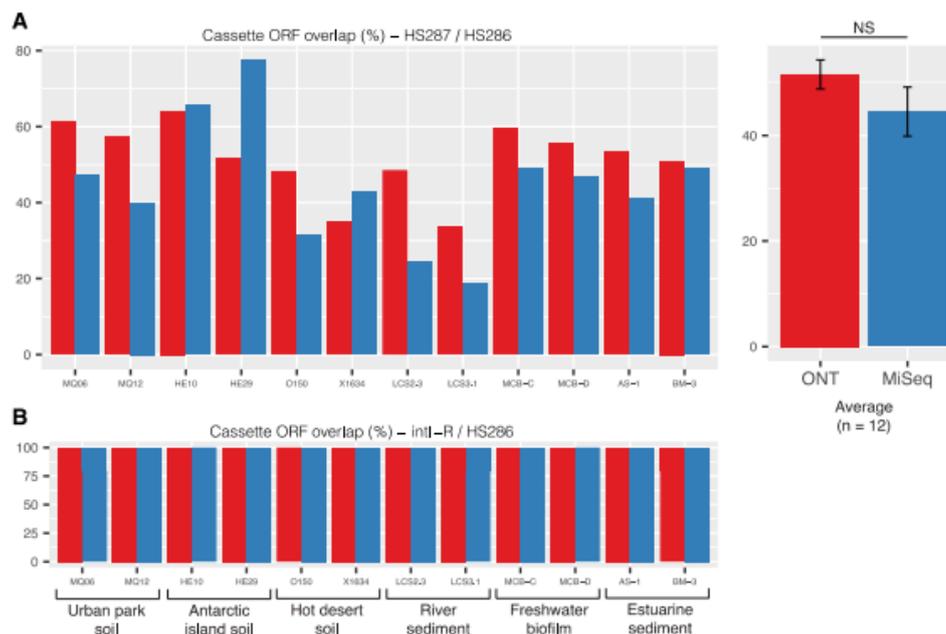
Figure S3. Overlap in recovered ORFs between Nanopore (ONT) and MiSeq sequencing technologies. The percentage of ORFs recovered from one sequencing technology that were covered by reads from the other technology are shown for (**A**) HS287 / HS286 and (**B**) intI-R / HS286 primer sets. ORFs considered to present in the opposite sequencing technology had to have a mean coverage depth of at least 1x that spanned at least 98% of the ORF. The average (± 1 S.E) percentage overlap for HS287 / HS286 data is shown on the right-hand side of panel (**A**). There was no significant (NS) difference between ONT and MiSeq (Two-sample T-test, P=0.209).

400
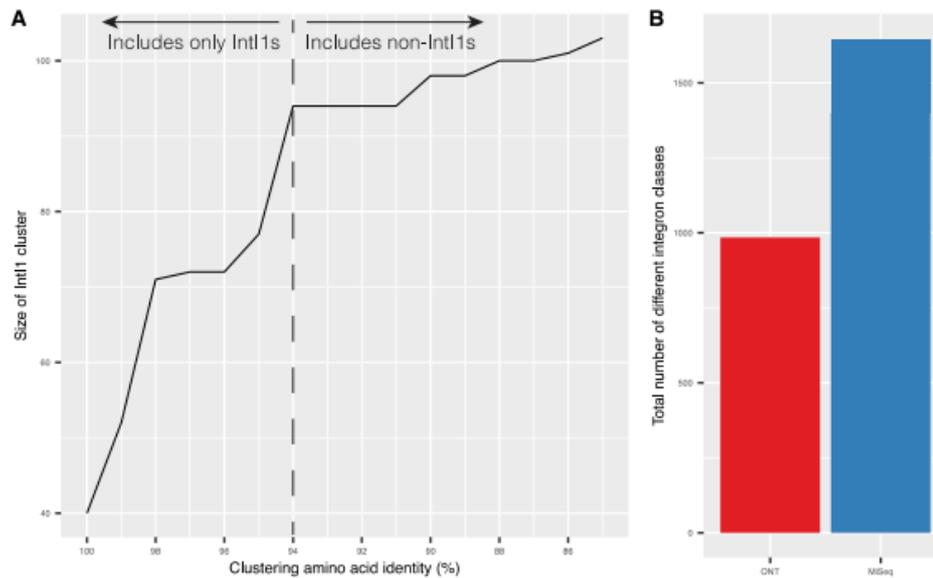
730   **Figure S4. Diversity of integron classes.** (**A**) The amino acid clustering threshold for
731   integron classes was determined using class 1 integron integrases (IntI1) present in our
732   dataset. Decreasing amino acid clustering thresholds were iteratively set until all IntI1s were
733   grouped in the same cluster and all non-IntI1s were excluded. A protein sequence was
734   considered to be IntI1 if it aligned with any previously characterised class 1 integron in
735   GenBank using BLASTP (>98% amino acid identity and >70% subject cover). An amino
736   acid clustering threshold of 94% was found to include all IntI1s (n=94) and exclude all non-
737   IntI1s. (**B**) The total number of integron classes (based on a 94% amino acid clustering
738   threshold) recovered for all samples (n=12).
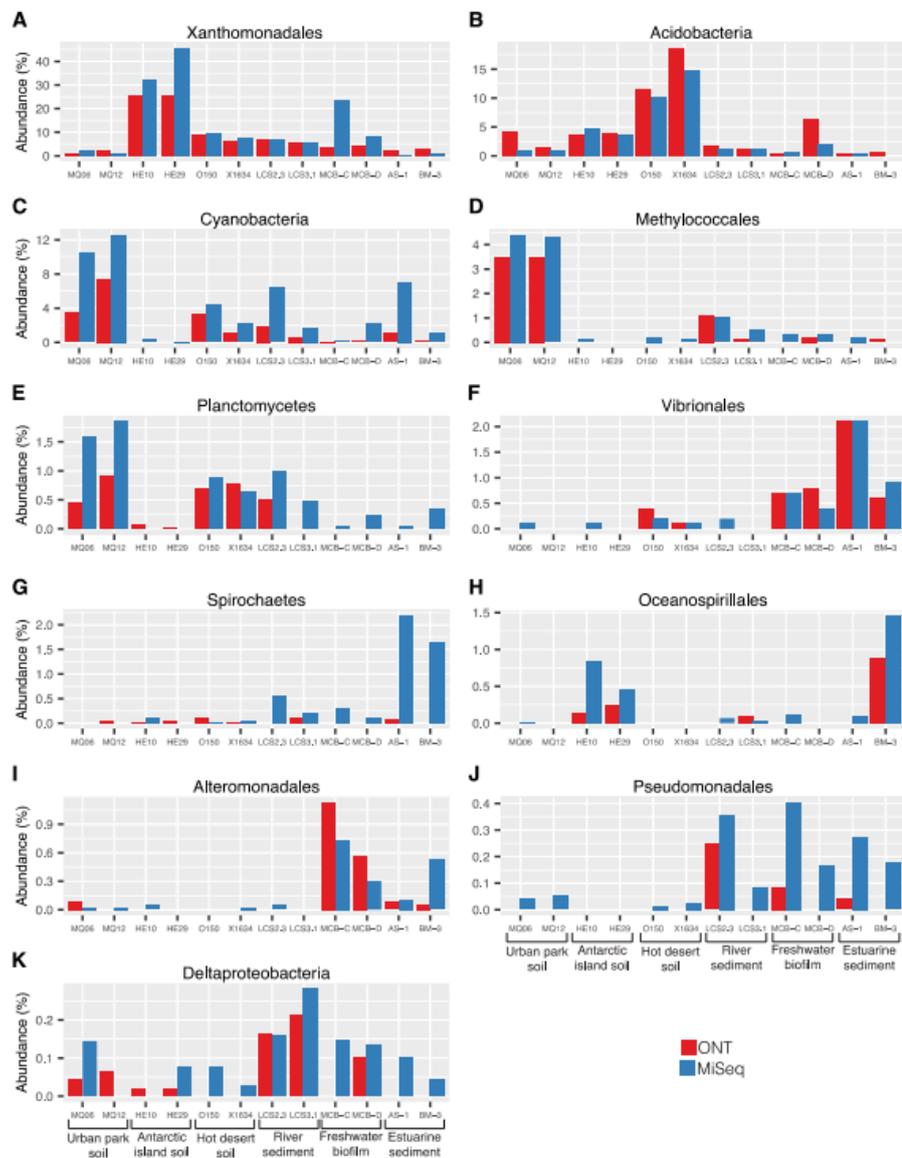
Figure S5. Taxonomic classification of gene cassette recombination sites (*attC*s).
Taxonomic predictions are based on all eleven (**A-K**) available taxonomic models of chromosomal *attC*s. Each figure panel shows the proportion of *attC*s across each sample that exhibit sequence and structure conserved among that taxon.

402

Chapter 8: Appendix

Breakdown of Every Mercury Resistant Characterised Bacterial Isolate Collected from Wastewater

Figure 8.1 Mercury resistant bacteria from campus wastewater influent and their results from antibiotic sensitivity testing using CLSI disc diffusion assays.

Bacteria were isolated from campus wastewater influent by spread plating ten-fold dilutions onto TBX agar, the resultant presumptive *Enterobacteriaceae* were then further screened for resistance to mercury (II) chloride by growing on LB agar supplemented with 25 μg mL^{-1} HgCl$_2$. Resistant isolates were then screened for sensitivity to 16 different antibiotics using CLSI disc diffusion assays. Red indicates resistance, yellow indicates intermediate susceptibility and green susceptible.

Antibiotic Class	Penicillin	B-Lactam and Inhibitor	Second Gen Ceph	Third Gen Ceph	Third Gen Ceph	Third Gen Ceph	Monobactam	Carbapenem	Amino-glycoside	Tetracycline	Fluoro-quinolone	Quinolone	Folate Pathway Inhibitor	Phenicol	Nitrofuran	Macrolide
Antibiotic	AMP	AMC	FOX	CAZ	CTX	CPD	ATM	IPM	S10	OT	CIP	NA	SXT	C	F	AZM
Antibiotic Disc Concentration / µg	10	20 & 10	30	30	30	10	30	10	10	30	5	30	1.25 & 23.75	30	300	15

Sample	Taxon
V30	*Escherichia coli*
VH32	*Citrobacter feundii*
VH31	*Klebsiella pnuemoniae*
VH28	*Enterobacter sp.*
VH11	*Enterobacter cloacae*
VH24	*Escherichia coli*
11V8	Presumptive *Enterobacteriaceae*
11V25	Presumptive *Enterobacteriaceae*
11V39	Presumptive *Enterobacteriaceae*
11V41	Presumptive *Enterobacteriaceae*
11V42	Presumptive *Enterobacteriaceae*
11V48	Presumptive *Enterobacteriaceae*
0801V1	Presumptive *Enterobacteriaceae*
0801V5	Presumptive *Enterobacteriaceae*
0801V7	Presumptive *Enterobacteriaceae*
0801V9	Presumptive *Enterobacteriaceae*
0801V11	Presumptive *Enterobacteriaceae*
0801V15	Presumptive *Enterobacteriaceae*
0801V20	Presumptive *Enterobacteriaceae*
0801V21	Presumptive *Enterobacteriaceae*
2001V3	Presumptive *Enterobacteriaceae*
2001V7	Presumptive *Enterobacteriaceae*
2001V10	Presumptive *Enterobacteriaceae*
2001V15	Presumptive *Enterobacteriaceae*
2001V21	Presumptive *Enterobacteriaceae*
2001V22	Presumptive *Enterobacteriaceae*
2001V24	Presumptive *Enterobacteriaceae*
2001V27	Presumptive *Enterobacteriaceae*
2001V39	Presumptive *Enterobacteriaceae*
2001V41	Presumptive *Enterobacteriaceae*
2001V42	Presumptive *Enterobacteriaceae*
2001V47	Presumptive *Enterobacteriaceae*

Figure 8.2 Mercury resistant bacteria from village wastewater influent and their results from antibiotic sensitivity testing using CLSI disc diffusion assays.

Bacteria were isolated from campus wastewater influent by spread plating ten-fold dilutions onto TBX agar, the resultant presumptive *Enterobacteriaceae* were then further screened for resistance to mercury (II) chloride by growing on LB agar supplemented with 25 µg mL^{-1} HgCl$_2$. Resistant isolates were then

screened for sensitivity to 16 different antibiotics using CLSI disc diffusion assays. Red indicates resistance, yellow indicates intermediate susceptibility and green susceptible.

Antibiotic Class		Penicillin	B-Lactam and Inhibitor	Second Gen Ceph	Third Gen Ceph	Third Gen Ceph	Third Gen Ceph	Monobactam	Carbapenem	Amino-glycoside	Tetracycline	Fluoro-quinolone	Quinolone	Folate Pathway Inhibitor	Phenicol	Nitrofuran	Macrolide
Antibiotic		AMP	AMC	FOX	CAZ	CTX	CPD	ATM	IPM	S10	OT	CIP	NA	SXT	C	F	AZM
Antibiotic Disc Concentration / µg		10	20 & 10	30	30	30	10	30	10	10	30	5	30	1.25 & 23.75	30	300	15
Sample	Taxon																
BPF1-1	Escherichia coli																
BPW2-3	Escherichia coli																
BPW2-4	Escherichia coli																
NT1	Escherichia coli																
NT7	Escherichia coli																
NT13	Escherichia coli																
NT31	Escherichia coli																
NT38	Escherichia coli																
NT49	Escherichia coli																
NT50	Escherichia coli																
NT55	Escherichia coli																
NT65	Escherichia coli																
NT67	Citrobacter feundii																
TT1	Escherichia coli																
TT2	Escherichia coli																
2001E1	Presumptive Enterobacteriaceae																
2001E12	Presumptive Enterobacteriaceae																
2001E15	Presumptive Enterobacteriaceae																
2001E17	Presumptive Enterobacteriaceae																
2001E20	Presumptive Enterobacteriaceae																
2001E21	Presumptive Enterobacteriaceae																
2001E22	Presumptive Enterobacteriaceae																
2001E23	Presumptive Enterobacteriaceae																
2001E37	Presumptive Enterobacteriaceae																
2001E39	Presumptive Enterobacteriaceae																
2001E43	Presumptive Enterobacteriaceae																
2001E45	Presumptive Enterobacteriaceae																

Figure 8.3 Mercury resistant bacteria from wastewater effluent and downstream of the effluent pipe and their results from antibiotic sensitivity testing using CLSI disc diffusion assays.

Bacteria were isolated from campus wastewater influent by spread plating ten-fold dilutions onto TBX agar, the resultant presumptive *Enterobacteriaceae* were then further screened for resistance to mercury (II) chloride by growing on LB agar supplemented with 25 µg mL^{-1} HgCl$_2$. Resistant isolates were then screened for sensitivity to 16 different antibiotics using CLSI disc diffusion assays. Red indicates resistance, yellow indicates intermediate susceptibility and green susceptible.

Breakdown of Every ESBL Characterised Bacterial Isolate Collected from Wastewater

Figure 8.4 ESBL Producing isolates from campus and village wastewater influent.

Wastewater was spread plate in ten-fold dilutions onto CHROMagar ESBL agar plates. Presumptive *E. coli* were then picked at random and screened for resistance to 16 antibiotics using disc diffusion techniques as per the CLSI standards. Red indicates resistance, yellow indicates intermediate susceptibility and green susceptible. The same isolates were then screened for determinants of ESBL resistance, bla_{SHV} (not shown as no isolates possessed the gene), bla_{CTX-M}, bla_{TEM}, bla_{OXA-2}.

All Isolates in NCBI Containing a Tn*21* or Tn*21*-like Transposable Element

Accession Number	Organism	Location	Isolation Date	Transposase Genes	Mercury Resistance Genes	In2	Gene Cassettes	Reference or NCBI BioProject
KU315015	*Serratia marcescens*	USA	1980	*tnpA, tnpR, tnpM*	*merRTPCADE*	IS*6100*, *orf5, orfQ, orfP, intI1*, IR$_i$	*sul1, qacEΔ1, aadA1, aacC1*	Blackwell *et al.*, 2016
JN247441	*Acinetobacter baumannii*	Netherlands	1982	*tnpA, tnpR, tnpM*	*urf2, merTPCADE*	IR*i, intI1, orfP, orfQ, orf5*, IS*6100*, IR*t*	*aacC1/aacA1, aadA1, qacEΔ1, sul1*	Nigro and Hall, 2012
Y18360	*Pseudomonas sp.*	Russia	1988	*tnpA*	*urf2, merRTPACDE*	*N/A*	N/A	Mindlin *et al.*, 2001
DQ364638	*Escherichia coli*	N/A	1988	*tnpM, tnpR, tnpA*	*urf2, merRTPCADE*	*intI1, orf5, istB, istA, orfAB, tniBΔ1, tniA*	*sul1, qacEΔ1, aadA1*	Williams *et al.*, 2006
U12338	*Pseudomonas aeruginosa*	USA	1990	*res* site, *tnpR, tnpA,*	*urf2Y, merRTPCADE*	IR*t, orf5, orfE, intI1*, IR$_i$	*sul1, qacEΔ1, cmlA1, aadA2, aacC1*	Bissonnette *et al.*, 1991
MG764534	*Klebsiella aerogenes*	France	1995	1st: *tnpA, tnpR*	1st: *urf2Y, merRTPCADE*	1st: IS*6100, orf6, orf5, intI1*	1st sequence: *sul1, qacEΔ1, aacA4;*	Siebor *et al.*, 2018
				2nd: *tnpA, tnpR, tnpM*	2nd: *urf2, IR, merRTPCADE*	2nd: IR$_i$, *intI1, orf5, istB, istA, tniBΔ1, tniA*	2nd sequence: *aadA1, qacEΔ1, sul1*	Post and Hall, 2009

FJ172370	*Acinetobacter baumannii*	Australia	1997	*tnpA*	*urf2Y, merRTPCADE*	*IRi, intI1, orfP, orfQ, orf5, resX, trbL, lspA, ISL3, orf5, IRt,* IS*6100*	*aacA1, aadA1, qacEΔ1, sul1*	Post and Hall, 2009
FTZQ01000073	*Shigella sonnei*	Peru	1999	*tnpR, tnpA, urf2*	*merTPADE*	IS*21 orf2,* IS*21 orfA,* IS*150 orf2, tniB, tnsB,* cAMP phosphodiesterase	*emrE, sul1*	Novais *et al.*, 2010
GQ293498	*Escherichia coli*	Spain	2002	*tnpA, tnpR, tnpM*	*urf2, merRPCADE*	*tniA, tniBΔ1, istA, istB, orf5, intI1*	*aadA1, qacEΔ1, sul1*	Call *et al.*, 2010
FJ621588	*Escherichia coli*	USA	2002	*tnpA, tnpR, tnpM*	*urf2M, merRTPABDE*	*gorS, groL, insE, insF, istB, istA, insG, tniB, tniA, insB3, insB4*	*aadA, aacC qacEΔ1, sul1*	Novais *et al.*, 2010
GQ293499	*Escherichia coli*	Spain	2002	*tnpA, tnpR*	*urf2Y, merRTPCADE*	*intI1, orf5,* IS*6100*	*aacA4, qacEΔ1, su1*	Holt *et al.*, 2007
AM412236	*Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU_12601*	Pakistan	2002	*tnpA, tnpR, tnpM*	*merRTPCADE*	*repA, repC, tniAΔ1*	*sul1, bla*TEM-1*, sul2, strA, strB*	PRJNA471337
CP029645	*Salmonella enterica subsp. enterica serovar Typhi*	N/A	2002	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*tniA, repC,* IS*6, intI1*	*sul2, strA, strB, bla*TEM-1*, sul1, ebr, dfrA*	Han *et al.*, 2012

411

JN983043	*Salmonella enterica subsp. enterica serovar Heidelberg*	USA	2002	*tnpR, tnpA*	*urf2, merRTPABDE*	*intI1*	*dfrA16*	PRJEB5065
FLDB01000015	*Klebsiella pneumoniae*	UK	2002	*tnpA, tnpR*	*merRTPCDE*	*intI1, tniB*	*emrE, sul1*	PRJEB5065
FLID01000025	*Klebsiella pneumoniae*	UK	2002	*tnpA, tnpR*	*merR, merTPCDE*	*intI1, tniB*	*ereB, emrE, sul1*	Wibberg *et al.,* 2013
JX127248	Uncultured bacterium	Canada	2003	*tnpR, tnpA*	*urf2, merRTPCADE*	*tniA*, IS26, *repC, repA*	*aph, strB, strA, sul2, bla*	Tamamura *et al.,* 2013
AB605179	*Salmonella enterica subsp. enterica serovar Typhimurium*	Japan	2005	*tnpA, tnpR, tnpM*	*urf2, merADE*	*intI1, istB, istA, merR, tniA*	*aadA, qacEΔ1* fusion protein, *sul, tetR, tetA*	PRJEB5065
PRJEB5065	*Klebsiella pneumoniae*	UK	2005	*tnpA, tnpR*	*merRTPCADE*	*N/A*	N/A	De Curraize *et al.,* 2020
MN148427	*Proteus vulgaris*	France	2005	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*intI1, orf5, istB, istA, tniBΔ1, tniA*	*aadA1, qacEΔ1, sul1*	PRJEB5065
FLBJ01000030	*Klebsiella pneumoniae*	UK	2006	*tnpA, tnpR*	*merRTPCADE*	*N/A*	N/A	PRJEB5065
FLHL01000023	*Klebsiella pneumoniae*	UK	2006	*tnpA, tnpR*	*merRTPCADE*	*N/A*	N/A	Chiou *et al.,* 2015

KM023773	*Salmonella enterica subsp. enterica serovar Typhi*	Bangladesh	2007	*tnpA, tnpR, tnpM*	*merRTPCADE*	*intI1*, IS26, *repC, repA, tniA*	*sul1, bla$_{TEM-1}$, strB, strA, sul2*	Cuzon *et al.,* 2016
KX230795	*Enterobacter cloacae subsp. cloacae*	Belgium	2007	*tnpA*	N/A	*intI1*, GNCS-acetyltransferase	*bla$_{GES-7}$, sul1, aadA1, qacEΔ1*	Moran *et al.,* 2006
KR827684	*Escherichia coli*	Australia	2010	*tnpA, tnpR, tnpM*	*urf2M, merRTPCADE*	*IR$_i$, intI1, orfF*, IS440, *orfA, orfB*, IS26, *IR$_t$*	*aadA2, cmlA1, aadA1, qacH, sul3, mefB*	Ho *et al.,* 2014
KF250428	*Klebsiella pneumoniae*	Philippines	2010	*tnpA*	*urf2, merRTPCADE*	*tniA, tniB*	*ampR, bla$_{DHA-1}$*	Martinez *et al.,* 2012
JN559393	*Pseudomonas aeruginosa*	Australia	2010	*tnpA*	*urf2, merRTPCADE*	*tniA, tniBΔ1, orf5, intI1Δ*	*sul1, qacEΔ1, aacA4, bla$_{OXA-129}$, catB3, aadB*	Martinez *et al.,* 2012
JF826500	*Pseudomonas aeruginosa*	Australia	2010	*tnpA*	*urf2, merRTPCADE*	*tniA, tniBΔ1, orf5, intI1Δ*	*sul1, qacEΔ1, aacA4, bla$_{OXA-129}$, catB3, aadB*	Adamczuk *et al.,* 2015
AB715422	*Klebsiella oxytoca*	Japan	2011	*tnpA, tnpR*	*urf2, merRTPCADE*	*intI1, orf5, tniB, tniA*	*Ebt(qacF), aacA4, bla$_{IMP-34}$, ebr(qacE2), qacEΔ1, sul1*	Chavda *et al.,* 2015

413

KP345882	*Escherichia coli*	USA	2011	*tnpR, tnpA*	*urf2, merRTPCADE*	*IntIPac, tniB, tniA,*	PAC, sul1, emrE, cat, ant(9)-Ia, bla_{PSE-1}	Villa *et al.,* 2013
JX442976	*Klebsiella pneumoniae*	France	2011	*tnpA, tnpR, tnpM*	*urf2, merRTCA*	*tniA, tniB,* IS*10*R, ISCR1, *intI1,* IS*26*	qnrA6, sul1, qacEΔ1, aadA1, bla_{OXA-10,} cmlA7, arr	Du *et al.,* 2016
KU302801	*Enterobacter cloacae*	China	2012	*tnpR, tnpA*	*urf2, merRTCADE*	*N/A*	N/A	SAMN14228030
MH909327	*Klebsiella pneumoniae*	China	2012	*tnpA*	*urf2, merRTPCADE*	*tniA, intI1,*	dhfr, qacEΔ1, sul1, PAC	Welch *et al.,* 2007
CP000603	*Yersinia pestis biovar Orientalis str. IP275*	Madagascar	2012	*tnpA, tnpM*	*urf2, merRTPCADE*	*tniA, tniB, insC, insB7, insB8, insB9, insB, istB, intI1*	bla_{SHV-1}, tetA, tetR, sul1, qacEΔ1, aadA	Irrgang *et al.,* 2019
HG530658	*Escherichia coli R178*	Germany	2013	*tnpM, tnpR, tnpA*	*urf2, merRTPCADE*	*tniA, tniB, orfAB, istA, istB, orf5, intI1*	sul1, aadA1, addA4, bla_{VIM-1}	Wang *et al.,* 2014
KJ484626	*Escherichia coli*	Switzerland	2013	N/A	*urf2, merRTCADE*	*tniA, tniB,* alpha/beta family (hydolase), *GGDEF, tniB, istB, istA, intI1*	tetR, bla, sul1, emrE, aadA1b	Rahube *et al.,* 2014
JX486126	Uncultured bacterium	Canada	2013	*tnpA, tnpR, tnpM*	*N/A*	*intI1, orf5*	aadA2, qacEΔ1, sul1	Wyrsch *et al.,* 2019

MN334220	*Salmonella enterica subsp. enterica serovar Typhimurium*	Australia	2013	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*intI1, tniA*, IS*26*	N/A	Feng *et al.*, 2017
KY270850	*Klebsiella pneumoniae*	China	2013	*tnpA*	*urf2, merRTPCADE*	*tniA*, IS*26, intI1, orf6*	*dfrA25, sul1*	Cheng *et al.*, 2019
MH917285	*Klebsiella pneumoniae*	China	2013	*tnpA*	*urf2, merRTPCADE*	*tniA, intI1, BsuBI-PstI*	*DHFR, qacEΔ1, sul1, PAC*	Kizny Gordon *et al.*, 2020
FJWW01000015	*Enterobacter hormaechei*	UK	2014	*tnpR*	*urf2, merRTPCADE*	Acetyltransferase, *intI1*	*PAC, sul1, emrE, aacD-aphD*	Guo *et al.*, 2016
LN555650	*Salmonella enterica subsp. enterica serovar Infantis*	Germany	2015	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*tniA, tniBΔ1, orfAB, istA, istB, orf5, intI1*	*sul1, qacEΔ1, aadA1, bla*_{VIM-1}	PRJEB8776
UDIU01000003	*Escherichia coli*	N/A	2015	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*intI1, istB, istA, tniB, tniA,*	*ebr, sul1, bla, tetR, mphB*	PRJEB8774
UDDS01000003								Dang, Mao and Luo 2016
UCZN01000013	*Escherichia coli*	N/A	2015	*tnpR, tnpM*	*merRT*	*intI1*	*aadA, qacE, sul1*	Oliva *et al.*, 2018

KU238092	Uncultured bacterium	China	2015	*tnpA, tnpR*	*N/A*	Relaxase, *intI1*	*tetR, strB, strA, sul1, qacEΔ1, bla*GES-5	PRJEB11403
MH257753	*Salmonella enterica subsp. enterica serovar Typhimurium*	Italy	2015	*tnpA, tnpR*	*urf2, merRTPCADE*	IS*26*, IS*440*, *intIPac*	*emrE, aadA, cat*	Papagiannitsis *et al.*, 2017
FLWR01000011	*Klebsiella pneumoniae*	Thailand	2015	*tnpA, tnpR*	*merRTPADE*	*N/A*	N/A	Papagiannitsis *et al.*, 2017
KY860570	*Pseudomonas aeruginosa*	Czech Republic	2015	*tnpR, tnpA*	*urf2, merRTADE*	IR$_i$, *intI1, orf5*, IR$_t$, IS*1600, resX*	*bla*VIM-1, *qacEΔ1, sul1*	Yoshii *et al.*, 2015
KY860569	*Pseudomonas aeruginosa*	Czech Republic	2015	*tnpR, res* site	*urf2, merRTADE*	*intI1*, IS*6100*, IR$_t$	*aadB, bla*IMP-1, *aadA1, qacEΔ1, sul1*	Paskova *et al.*, 2018
AB852526	*Acidovorax avenae subsp. avenae*	Japan	2016	tnpR, tnpA	*merRTPADE*	*intI1*	*strS, strB*	Zeng *et al.*, 2018
MG252895	*Escherichia coli*	Czech	2016	*tnpA*	*urf2, merRTPADE*	IR$_i$, *intI1, orf513*, IS*Aba125, orf5, tniA, tniB*	*arr-2, cmlA1, bla*OXA-10, *qacEΔ1, sul1, aphA6, bla*NDM-7	Ghaly *et al.*, 2017
MH457126	*Vibrio alginolyticus*	China	2016	*tnpR, tnpA*	*urf2, merRTPABDE*	*intI1*, ISCR1, IS*Aba24*, ISCR27, *groL, groS, cutA, dsbC, trpF*, IS*5075*, IS*110, chrA, padR*	*dfrA15, qacEΔ1, sul1,*	Karimi *et al.*, 2020

416

KY126370	*Enterobacter cloacae subsp. cloacae*	Australia	2016	tnpR, tnpA	*urf2, merRTPABDE*	*intI1, tniR, tniQ,* uracil-DNA glycosylase, *Lys, tniA*	*ahpD, tetR, qacE2*	PRJEB22665
CABWMY010000 0061	*Sphingomonas sp. T1*	Australia	2017	*tnpA*	*merRTPA*	*N/A*	N/A	PRJNA38919
CABVLK01000001 2	*Hoeflea sp. EC-HK425*	N/A	2017	*tnpM, tnpR, tnpA*	*urf2, merRTPCADE*	*tniA, tniB, intI1*	*sul1, qacEΔ1*	Li *et al.,* 2020
ADIY01000014	*Escherichia coli*	N/A	2017	*tnpM*	*urf2, merRTPCAD*	GNAT family, integrase/recombinase	*sul1, qacE, aad, aacC1*	PRJEB24625
MN583270	*Pseudomonas aeruginosa*	China	2018	N/A	*urf2, merRTPADE*	*tniA, tniB, intI1*	*sul1, qacEΔ1, cmlA, aadB, PAC*	PRJEB24625
LT985225	*Escherichia coli*	France	2018	*tnpR, tnpA*	*urf2M, merRTPCADE*	*tniA, tniBΔ1, istA, itsB, intI1*	*sul1, qacEΔ1, aacA4*	PRJEB24625
LT985227	*Escherichia coli*	France	2018	*tnpR, tnpA*	*urf2, merAD*	*tniA*	*aacA4, bla*OXA-9	PRJEB24625
LT985222	*Escherichia coli*	France	2018	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*intI1, istB, istA, tniBΔ1, tniA*	*aadA1, qacEΔ1, sul1*	Clerissi *et al.,* 2018
LT985220	*Escherichia coli*	France	2018	*tnpR, tnpA*	*urf2M, merRTPCADE*	*intI1, istB, istA, tniA, tniBΔ1*	*aacA4, qacEΔ1, sul1*	Bonnin *et al.,* 2020
OGUS01000066	*Cupriavidus oxalaticus*	India	2018	tnpR, tnpA	*merRTPDE*	*N/A*	N/A	PRJNA224116
OGUS01000115								Li *et al.,* 2020

MN699848	*Citrobacter pasteurii*	France	2019	Res site, *urf2, tnpA*	*urf2, merRTPCADE*	*intI1*, GNAT acetyltransferase, *tniB, tniA*	*aacA7, aacC14, bla*OXA-198*, cmlA1, sul1*	PRJEB40413
KT033470	*Aeromonas salmonicida subsp. salmonicida*	N/A	2020	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*tniA, tniB, ISCR, intI1*	*cat, aadA, qacEΔ1, sul1*	Moradigaravand *et al.,* 2016
MT320534	*Proteus mirabilis*	China	2020	*tnpA, tnpR*	*urf2, merRTPBDE*	IS*1326*, IS*1353*, IS*26*	*aac(6')-Ib-cr, arr-3, qacEΔ1, sul1*	Kizny Gordon *et al.,* 2020
LR883013	*Escherichia coli*	UK	2020	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*intI1*	*aadA1, qacEΔ1, sul1, tetA*	Yang *et al.,* 2020
FKYU01000056	*Klebsiella oxytoca*	UK	2003	*tnpA, tnpR*	*merRTPACDE*	*intI1, orf5, tniA, tniB,*	*aadA, emrE, sul1*	Gao *et al.,* 2020; Wang *et al.,* 2017
FKZH01000040			2008					Vincent *et al.,* 2014
MH909330	*Leclercia adecarboxylata*	China	2012-2015	*tnpA,* mobile element	*urf2, merRTPCADE*	*intI1*, tryptophan synthease	*aac(6')-Ib/aac(6')-II, bla*OXA-1*, catB, arr, bla*TEM*, bla*CTX-M*, sul1, PAC*	Siebor and Neuwirth 2014
MF168406	*Klebsiella pneumoniae*	China	2012-2016	*tnpR, tnpA*	*urf2, merRTPCADE*	IS*26*, ISKpn27, KorC, klcA, orf279, orf396	*bla*KPC-2	Herrero, A., *et al* 2008
KY913898	*Klebsiella oxytoca*	China	2012-2017	resA, *tnpA*	*urf2, merTCARE*	Tn*3*, IS*3*, IS*110*, IS*5*/IS*1182*	*bla*TEM-1*, aadA1,*	Stoesser *et al.,* 2016

					bla~OXA~, emrE, sul1			
KJ909290	*Aeromonas salmonicida*	Canada	N/A	*tnpA, tnpR, tnpM*	*urf2, merRTPABDE*	*intI1, orf166a, orf95c,* IS*6100*	*aadA, qacEΔ1, sul1*	PRJEB33308
KJ439039	*Proteus mirabilis*	France	N/A	*tnpA, tnpR*	*urf2, merRTPFADE*	*intI1, IR~i~*	*qacEΔ1, sul1, aadA1*	PRJEB33308
AM991977	*Salmonella enterica subsp. enterica serovar Typhimurium*	Spain	N/A	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*tniA, tniBΔ1, orfAB, istA, istB, orf5, intI1*	*sul1, qacEΔ1, aadA1, bla~OXA-1~*	PRJEB33308
LN864819	*Klebsiella pneumoniae*	Lebanon	N/A	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*intI1, orf5, istB, istA, orfAB, tniB, tniA*	*bla~OXA-1~, aadA1, qacEΔ1, sul1*	PRJEB33308
CABFYB010000005 CABFYK010000004 CABFWU0100000003 CABFYT010000008 CABFWY0100000003	*Klebsiella pneumoniae*	Europe	N/A	*tnpA, tnpR*	*merRTPCADE*	*N/A*	N/A	Molina *et al.*, 2014
CABFXS0100000003	*Klebsiella pneumoniae*	Europe	N/A	*tnpA, tnpR*	*merEDCT*	IS*26, tnsB*	*Bla*	Driver *et al.*, 1983

CABFXN01000000 2	*Klebsiella pneumoniae*	Europe	N/A	*tnpA, tnpR*	*merRTPCAD E*	*intI1, tnsB, cph2*	*tetR, mrx, mph, bla, sul1, str, strB*	PRJEB33308
CABFWZ0100000003	*Klebsiella pneumoniae*	Europe	N/A	*tnpA, tnpR*	*merRTPCAD E*	Integron found but not in usual location: *intI1, tniB*	*aacA-aphD, bla$_{NDM-1}$, emrE, sul1, strB, str*	PRJEB24625
CABFWR0100000003								Fricke *et al.*, 2011
CABFXC01000000 3	*Klebsiella pneumoniae*	Europe	N/A	*tnpA, tnpR*	*merRTPCAD E*	*intI1, tniB, tnsB, cph2*	*bla$_{NDM-1}$, emrE, sul2*	PRJNA61147
CP003739	*Pseudomonas putida*	France	N/A	tnpR, tnpA	*urf2, merRTPADE*	COG1734 DnaK suppressor, *intI1*, IS*116*/IS*110*/IS*902*	*sul1, aad, qacE*	PRJEB6403
JQ899055	*Salmonella enterica subsp. enterica serovar Infantis*	Italy	N/A	*tnpA, tnpR, tnpM*	*urf2, merRTPCDE*	*tniA, tniBΔ1, istAΔ*	*bla$_{TEM-1}$, qacEΔ1, aadA4, catB3, Sul1*	PRJEB10018
UKAT01000018	*Klebsiella variicola*	N/A	N/A	*tnpR, tnpA*	*merRTPACD E*	2-phosphoglycerate kinase, *tniB, tnsB*	*emrE, sul1*	PRJEB10018
AP000342.1	*Shigella flexneri 2b*	N/A	N/A	*tnpM, tnpR, tnpA*	*urf2, merRTPCAD*	*tniA, tniΔ1, orfAB, istA, istB, intR*	*sul1, qacEΔ1, aadA1*	PRJEB10018
AF071413	*Escherichia coli*	N/A	N/A	*tnpM, tnpR, tnpA*	*urf2, merRTPCDE*	*tniAΔ*, IS*26*, IS*6100*, *intI1*	*aph, strB, strA, sul2, bla$_{TEM-1b}$, mph(A)*	PRJEB10018
MN254970	*Escherichia coli*	N/A	N/A	*tnpA, tnpR*	*urf2, merRTPABD E*	IS*26, orf3, orf2, tniA, tniB,* IS*1353,* IS*1326, intI1*	*aacII, qacEΔ1, sul1, PAC,*	PRJEB10018

							mphR(A), mrx	
LT985215	*Escherichia coli*	N/A	N/A	N/A	*urf2, merRTPCADE*	*tniAΔ*, IS26, *insB, insA*	N/A	PRJEB10018
CP000604	*Salmonella enterica subsp. enterica serovar Newport str. SL254*	N/A	N/A	*tnpA, tnpR*	*urf2, meRTPABDE*	*intI1, groS, groL, insE, insF, orf5, istB, istA, insG, tniB, tniA, insB3*	*aadA, aacC, qacEΔ1, sul1, ble, bla_{NDM-1,} sul1*	PRJEB10018
AYDN01000064	*Salmonella enterica subsp. enterica serovar Agona str. ATCC BAA-707*	N/A	N/A	*tnpA, tnpR, tnpM*	*urf2, merRTPCADE*	*tniA, tniB, intI1*	*sul1, qacEΔ1*	PRJEB6403
CABDWH010000001	*Klebsiella pneumoniae*	N/A	N/A	*tnpA, tnpR*	*merRTPCADE*	*N/A*	N/A	Blackwell *et al.,* 2016
UIWA01000031	*Klebsiella pneumoniae*	N/A	N/A	*tnpA, tnpR*	*merRTPCADE*	*N/A*	N/A	Kenyon and Hall 2013
ULBB01000031								Nigro and Hall 2012
ULCI01000033								Mindlin *et al.,* 2001
UJXK01000043								Williams *et al.,* 2006

421

UJXA01000042								Bissonnette *et al.,* 1991
ULBV01000034								Siebor *et al.,* 2018
UJXI01000040								Kenyon and Hall 2013
UJXG01000042								Hamidian *et al.,* 2013
UJXM01000040								Post and Hall, 2009
UJDH01000034								PRJEB10018
UJOS01000024								Novais *et al.,* 2010
ULBE01000028								Call *et al.,* 2010
ULBJ01000029								Novais *et al.,* 2010
UJKO01000036								Holt *et al.,* 2007
UJHZ01000036								Han *et al.,* 2012
UJIF01000036								Wibberg *et al.,* 2013
UJML01000038								Tamamura *et al.,* 2013
UJBW01000027								De Curraize *et al.,* 2020
ULBI01000032								PRJEB5065
UJKB01000014								PRJEB5065
ULCP01000034								Chiou *et al.,* 2015
UIZL01000030								PRJEB10018

UKZM01000032								Cuzon *et al.,* 2016
UKNQ01000024								Moran *et al.,* 2006
UIXX01000034.1								Ho *et al.,* 2014
UJVH01000029	*Klebsiella pneumoniae*	N/A	N/A	*tnpA, tnpR*	*merRTPCADE*	IS*1, orf1, insB, tniQ, tniB, tnsB*	*cmlA2*	Martinez *et al.,* 2012
UJCL01000025								Martinez *et al.,* 2012
UJKQ01000014								Adamczuk *et al.,* 2015
UIRM01000032								Shimada *et al.,* 2016
UIRS01000033								Chavda *et al.,* 2015
UJPG01000024								Villa *et al.,* 2013
UKCJ01000034	*Klebsiella pneumoniae*	N/A	N/A	*tnpA, tnpR*	*merRTPCADE*	*intI1, sul1, str, strB*	*sul1, str, strB*	Du *et al.,* 2016
UKHC01000033								Welch *et al.,* 2007
UISD01000038								Irrgang *et al.,* 2019
UIVC01000027								Wang *et al.,* 2014
UIXW01000036								Rahube *et al.,* 2014
UIRN01000034								Wyrsch *et al.,* 2019
UKKT01000023	*Klebsiella pneumoniae*	N/A	N/A	*tnpA, tnpR*	*urf2, merRTPCADE*	*tnsB, tniB, tniQ, insB*	*cmlA2*	Feng m*et al.,* 2017

UKTX01000036	*Klebsiella pneumoniae*	N/A	N/A	*tnpA, tnpR*	*merRTPCADE*	*intI1, tniB, tnsB*	*ant1, emrE, sul1*	Cheng *et al.,* 2019
UJSA01000023	*Klebsiella pneumoniae*	N/A	N/A	*tnpA, tnpR*	*merRTPCDE*	*intI1, tniB*	*ereb, emrE, sul2*	Kizny Gordon *et al.,* 2020
UJFZ01000018	*Klebsiella pneumoniae*	N/A	N/A	*tnpA, tnpR*	*merRTPADE*	*intI1, tniB, tnsB, cph2*	*bla*$_{NDM-1}$, *aacA-aphD, emrE, sul1*	Guo *et al.,* 2016
UGMG01000002	*Klebsiella pneumoniae*	N/A	N/A	*tnpA, tnpR*	*merRTPCADE*	*intI1, IS, tniB, tnsB, cph2*	*aadA1, emrE, sul1*	PRJEB8776

Table 8.1 A comprehensive list of previously published Tn*21* and Tn*21*-like transposable elements from the NCBI database used for phylogenetic analysis of isolates.

Each sequence had their gene components listed broken down into transposase genes, mercury resistance genes insertion sequences identified

in the In2 and gene cassettes identified in the class I integron.