The interplay between plasmids in *Shigella sonnei*

Qian Lin Lee

Sir William Dunn School of Pathology

and

Lincoln College

A thesis submitted for the degree of

Doctor of Philosophy

at the University of Oxford, Trinity Term 2024

Abstract

Shigella sonnei causes bacillary dysentery in higher-income countries, with emerging antimicrobial resistance spread by plasmids. My thesis aimed to investigate the interplay between *S. sonnei* pINV, which is required for virulence, and resistance plasmids. pRES, a naturally occurring 114 kbp conjugative resistance plasmid, was used as an exemplar to investigate the transfer the 228 kbp virulence plasmid pINV, which was previously considered non-mobilisable.

pRES belongs to incompatibility group IncB/O/K/Z and is stably maintained during laboratory culture. The co-existence of pINV and pRES does not affect their maintenance within a bacterial population. pRES is a conjugative plasmid which can mobilises a 5.1 kbp colicin-E1 encoding plasmid as well as pINV. During pINV transfer, pINV and pRES form a single 334 kbp fusion plasmid, and/or hybrid plasmids, *via* RecA-dependent and -independent recombination, and insertion sequence transposition. Copies of IS*21* on pINV and pRES, and a conserved 199 bp region (199R) downstream of the plasmid replicons were detected at the pINV/pRES junctions in fusion and hybrid molecules.

The 199R is conserved in plasmids from IncFII and IncI-complex groups in *S. sonnei* which replicate unidirectionally. The sequence includes the single-stranded initiation site for synthesis of the leading strand during plasmid replication. Removal of the 199R increases pINV and pRES loss. The first 99 bp of the 199R (99O) overlaps with a previously described 218 bp (218B) deletion on the R1 plasmid, which involves the *terR* site for replication termination. Deletion of 218B or 99O also increases plasmid loss. RNAfold prediction suggest that the 99O can form secondary structures when single-stranded. SNPs were introduced into the 99O, disrupting these structures and increasing plasmid loss, corresponding to the degree of disruption. Secondary structures in the 199R might influence plasmid replication and inter-plasmid recombination, contributing to the mosaic architecture of many plasmids.

Acknowledgement

I am deeply grateful for the opportunity to pursue a DPhil in the Tang lab and have appreciated every moment of this journey. The journey has been challenging, especially starting during the COVID-19 pandemic, but it has also been incredibly rewarding, and provided me with a unique experience that I will never forget. Throughout this period, I received immense support from many people, to whom I want to express my heartfelt thanks. Without their help, I would not have been able to complete my DPhil.

First and foremost, I want to express my enormous gratitude to my main supervisor, Professor Christoph Tang. He provided me with the chance to start my research journey, guiding me with warmth, patience, and unwavering support. I have learned so much from him and believe that I will continue to do so in the future.

I also want to thank my second supervisor, Dr. Rachel Exley. Her patience and determination to achieve excellence in every aspect of research have been invaluable. Her logical thinking and cautious approach to challenges have greatly aided me during my research, and I deeply appreciate her guidance.

My sincere thanks also go to my day-to-day supervisor, Dr. Giulia Pilla. Her brightness and expertise were crucial during the initial period of my DPhil, especially as I began with limited knowledge of molecular biology and bacteriology.

Lastly, I want to thank my family for their constant mental support throughout this journey.

While I cannot mention everyone by name, I also wish to thank all current and previous lab members, friends, and the staff in the department and college for their support.

Together, these individuals have shaped my unique DPhil experience and prepared me for the future. I will cherish this experience throughout my life.

Contents

Abbreviations

bp/kbp	Base pair/kilobase pairs
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
$tRNA^{fMet}$	Transfer ribonucleic acid (formylated methionine)
RNase	Ribonuclease
gDNA	Genomic DNA
OD_{600}	Optical density measured at the absorbance of 600 nm
CFU	Colony forming unit
CRISPR	Clustered regularly interspaced short palindromic repeats

Unless stated otherwise, all figures and tables were created by Qian Lin Lee. If a figure or table has been taken and adapted, the source is clearly indicated.

1 Introduction

1.1 Introduction of *Shigella*

During an outbreak of dysentery in Japan in 1897, Kiyoshi Shiga discovered a bacterium, originally termed Shiga bacillus, now classified as *Shigella dysenteriae* 1. This bacterium is now taxonomically categorised as a member of the *Shigella* genus (1-3). *Shigella* spp. are rod shaped, non-spore-forming, Gram-negative bacteria from the family Enterobacteriaceae, and are non-motile, non- or late-lactose-fermenting (4-6). The genus *Shigella* consists of four species: *Shigella sonnei*, *Shigella flexneri*, *Shigella dysenteriae* and *Shigella boydii*, each of which has distinct lineages. Initially, sequencing of eight housekeeping chromosomal genes classified *Shigella* into three major clades (C1, C2 and C3). However, *S. sonnei* does not belong to any of these clades, and represents a separate species (7). The four species of *Shigella* are further classified into different serotypes according to the chemical composition of their O-antigen (O-Ag) which is the external component of lipopolysaccharide (LPS) (8). *S. dysenteriae* consists of 15 serotypes, *S. flexneri* of 19, and *S. boydii* of 20, while *S. sonnei* only expresses a single serotype (5, 9-11).

Shigella causes shigellosis, a form of dysentery, which is responsible for a high medical burden globally (12-17). Shigellosis typically presents with diarrhoea, fever, vomiting, and abdominal pain. In severe cases, patients exhibit dysentery, characterised by the presence of blood and/or mucus in the stool. (3). Together the four species cause over 200,000 deaths around the globe annually, with over 30% of cases in children younger than five years old (16, 18). Among the four species, *S. sonnei* and *S. flexneri* are responsible for the majority of cases. *S. flexneri* is the leading cause of bacillary dysentery in low- and middle-income countries (LMICs) due to poor hygiene and sanitation, causing > 60% of shigellosis in some parts of Asia and sub-Saharan Africa (19). *S. sonnei* is responsible for approximately 80% of cases in higher-income countries (HICs), causing only about 25% of cases in LMICs (19-23). *S. dysenteriae* and *S. boydii* together

contribute to less than 5% of shigellosis globally. *S. dysenteriae* secretes Shiga toxin and has caused epidemics and pandemics with high mortality in the past century (10, 19, 24, 25). *S. boydii*, while relatively uncommon, is typically confined to South Asia, with a few exceptions linked to travel (14).

Antibiotic treatments are used to reduce the length of infection and the risk of serious complications casued by shigellosis (3). The recommended first-line treatment by the World Health Organisation (WHO) in 2005 is ciprofloxacin, and the second line treatments, pivmecillinam, ceftriaxone and azithromycin, are used when the strain is resistant to ciprofloxacin (26). These treatments are effective in treating shigellosis in LMICs, however, there are reports of increasing resistance against these antibiotics and these resistance can be spread to other bacteria (19, 27-32). In addition, reduced efficacy of other antibiotic treatments such as fluoroquinolone due to the presence of resistance genes in *Shigella* strains was also observed in countries including Vietnam and Kolkata in recent years (33, 34). Therefore, the usage of antibiotics needs to be carefully regulated due to the rapid emergence of antimicrobial resistance (AMR) Shigella after the treatments (26).

1.1.1 Evolution of *Shigella*

Shigella emerged from *Escherichia coli* between 35,000-270,000 years ago (20). It has been suggested that *Shigella* should be re-classified as a pathovar of *E. coli*. However, due to the distinct host-pathogen interactions and biochemical characteristics, *Shigella* has remained a distinct genus (35). Several methods have been employed to define the evolutionary relationship between *Shigella* and *E. coli*, including whole genome sequencing (WGS) with single nucleotide polymorphism (SNP) analysis or examination of k-mers for alignment-free sequence analyses (36).

Examination of genome sequences reveals differences between *Shigella* and *E. coli*, contributing to a better understanding of the evolution of *Shigella* (10, 37). *Shigella* has become highly adapted to their human host, and employs characteristic mechanisms of pathogenesis following multiple genetic acquisition, loss, and rearrangement events, leading to changes in genome structure and loss of certain biochemical functions (37).

Insertion sequences (ISs) have shaped the evolution of *Shigella* by causing insertions, deletions, inversions, translocations, and other genetic rearrangements, resulting in the plasticity of *Shigella* genomes (38). Between 5-8% of the chromosomes of *Shigella* consists of ISs, which is significantly higher proportion than *E. coli* (37, 39). For example, *Shigella* chromosomal replication origins and termini are inversed compared to *E. coli* MG1655, and these sites are suggested to be hot spots of recombination (40). This may have involved multiple IS-mediated events in *Shigella* (37). There are five IS-families (IS*1*, IS*2*, IS*4*, IS*600* and IS*911*) that have undergone expansion in all *Shigella* species. This expansion has led to strain-to-strain variation and has contributed to convergent patterns of gene loss within and between *Shigella* species (39).

S. dysenteriae has undergone the most extensive IS expansion and genome degradation, while *S. sonnei* has the least (37). Gene loss contributes to the convergent loss of metabolic capabilities in *Shigella* (37). There are also different IS insertion sites in the four species of *Shigella*, reflecting divergent evolution and population diversity generated by ISs (39). This demonstrates a closer evolutionary relationship between *S. sonnei* and *S. flexneri*, which followed a separate evolutionary trajectory from *S. dysenteriae* (7, 39). IS*1* predominates in all *Shigella* chromosomes except *S. dysenteriae* which instead harbours a specific IS*1* variant, IS*1N* (37, 39), consistent with the unique evolutionary path of this species. Intact IS*21* and IS*630* are only found in *S. sonnei*, while IS*Sbo6* is mainly found in *S. boydii* (37).

Shigella emerged from *E. coli* after acquisition of a ~220 kbp, low copy number virulence plasmid pINV (**p**lasmid for **inv**asion). The plasmid has a mosaic architecture evidenced by the presence of genes derived from multiple plasmid origins (7, 10, 41). Plasmids are extrachromosomal DNA molecules which encode important features for their bacterial hosts. Although plasmids are beneficial for bacterial hosts under certain circumstances, they are not essential for bacterial survival, and are thus considered as parasitic genetic elements (42). Plasmids are classified into different incompatibility (Inc) groups based on their replication and maintenance mechanisms, plasmids of the same Inc group cannot co-exist in the same host cell (43). All *Shigella* harbour pINV with an IncFII group plasmid replicon (R1-like), which the replicon is defined as the smallest unit containing essential elements for regulation of plasmid replication. All *Shigella* pINV also carries a VapBC toxin-antitoxin (TA) system, and a ParAB partitioning system (37). A CcdAB TA system is found in pINV from all *Shigella* except *S. sonnei*, demonstrating that pINVs differ between *Shigella* species (37). Importantly, pINV carries a conserved 30 kbp pathogenicity island (PAI) containing genes encoding the Mxi-Spa Type 3 Secretion System (T3SS) and Ipa proteins which are required for *Shigella* virulence (37). It was hypothesised that pINV in all *Shigella* species were acquired from a single common ancestor, or the T3SS was acquired by pINV from related sources as the locus is flanked by IS*100* and IS*600* (37). This again highlights the contribution of ISs to the evolution of *Shigella*.

The IS content of pINV of *S. sonnei* and *S. flexneri* is different. *S. sonnei* pINV contains 26 ISs from 13 different families, with IS*600* being the most abundant (44). In contrast, *S. flexneri* pINV contains 29 ISs from 12 families, with IS*629* being the most prevalent (44). The presence of multiple copies of homologous ISs facilitates deletion of T3SS PAI, leading to the emergence of avirulent *Shigella* (44). Due to the different IS content on *Shigella* pINV, loss of the T3SS PAI is caused by different ISs in different *Shigella* (44). IS-mediated mechanisms therefore contribute to the acquisition and loss of virulence, playing a crucial role in shaping the emergence of *Shigella* as pathogen. As *Shigella* chromosomes and pINVs share IS elements, inter- and intra-

replicon translocation and replication are responsible for the large numbers of IS elements observed in the genomes of *Shigella* (37, 39, 44). This observation supports the hypothesis that pINV introduced IS variants into the chromosome, leading to the parallel expansion of IS, genetic plasticity in *Shigella*, and the acquisition of new genetic traits essential for virulence and host adaptation (37, 39, 44).

In addition to differences caused by ISs, other genetic changes in *Shigella* species differentiate them from *E. coli*. Changes in chromosomal genes contribute to the different biochemical properties of the two genera (37). For example, *Shigella* is non- or late-lactose-fermenting (6). Interestingly, this property is caused by different events in different species of *Shigella*. *S. flexneri* and *S. boydii* do not harbour *lacY*, *lacA* and *lacZ* required for lactose fermentation, while *S. dysenteriae* 1 contains *lacY* and *lacA* but no *lacZ* (45). Although *S. sonnei* harbours all three *lac* genes, it is late-lactose-fermenting due to a lack of lactose permease activity (37, 45). Further loss of metabolic pathways in *Shigella* (46), including *nadA/nadB* for nicotinic acid biosynthesis (47), *cadA* encoding lysine decarboxylase (48), and *speG* for spermidine deactivation (49), make the genus biochemically distinct from *E. coli*. In addition, *Shigella* is non-motile due to absence of flagella caused by a deletion in the *fliF* operon or an IS*1* insertion disrupting the *flhD* operon (37). These genetic changes reflect the adaptation and virulence of *Shigella* during infection of their human hosts (37).

Together, the differences and similarities between the genomes of *Shigella* and *E. coli* demonstrate convergent and divergent evolution in *Shigella*, and shed light on the effect of ISs during evolution.

1.1.2 Epidemiology and transmission of *Shigella*

Humans are the only natural host for *Shigella* with no known animal reservoir (3). The pathogen is mainly spread from person-to-person *via* the faeco-oral route, predominantly among

individuals residing in densely populated areas with poor hygiene and limited access to clean water (50, 51). Although less common, *Shigella* can also be spread *via* contaminated water or food (52). In addition, houseflies can contribute to the transmission of *Shigella* in settings where hygiene is poor (53, 54). Annually, approximately 10-40% travellers to LMICs develop diarrhoea, with the proportion of cases caused by *Shigella* varying on the geographical location and other factors (55). For example, *Shigella* accounted for approximately 13% of traveller-associated diarrhoea in the United States between 2004 to 2009 (56). A recent review also reports that travellers' diarrhoea is a major medical concern for military personnel (57). Furthermore, international travel is a key factor contributing to the global dissemination of antimicrobial resistant (AMR) *Shigella* (58, 59). However, *Shigella* can be underdiagnosed due to limited access to diagnostic tools, lack of patients' stool samples, and underreporting of mild cases. Therefore, there is an urgent need for effective preventive measures and surveillance strategies.

The distribution of *Shigella* worldwide differs geographically. In the Global Burden of Disease Study between 1990 and 2016, the highest burden of shigellosis was observed in LMICs(16). *Shigella* was identified as the second leading cause of diarrhoeal mortality among all ages, accounting for 212,438 deaths (95% uncertainty interval, 136,979-326,913) and about 13.2% (9.2-17.4) of all diarrhoea-related deaths. *Shigella* was responsible for 63,713 deaths (41,191-93,611) among children under 5 years of age. There was no significant difference in the mortality rate of males and females. The highest mortality rate (~10 per 100,000 people per year) was observed in sub-Saharan Africa (18). In contrast, the lowest mortality in children was seen in western Europe. Infection and mortality caused by *Shigella* has declined in recent years, which is likely due to improving sanitation and the use of antibiotics (19, 60).

As described earlier, the dominance of *Shigella* in different countries is linked to their economic status (3, 12, 18, 61). There is evidence that *S. sonnei* is increasingly replacing *S. flexneri* in economically transitioning LMICs (62, 63). This phenomenon has been observed in Vietnam (64),

Thailand (65), Bangladesh (66), and China (67). One explanation is that people living in LMICs are naturally immune to *S. sonnei* due to exposure to *Pleisiomonas shigelloides* O17 in contaminated water. *P. shigelloides* O17 expresses an O-Ag which is virtually identical to that of *S. sonnei*. As water quality improves, exposure to *P. shigelloides* O17 decreases, potentially reducing cross immunity and allowing *S. sonnei* to become more prevalent (68, 69). Alternatively, *S. sonnei* can be engulfed by the ubiquitous, free-living amoeba *Acanthamoebae castellanii*, which protects the bacterium from the effects of water chlorination (70). In contrast, *S. flexneri* is not protected as it is lethal to *A. castellanii* (71). Consequently, improvement in the water quality, coupled with economic development, exerts a selective pressure against *S. flexneri* and favours the proliferation of *S. sonnei*. Another explanation is that *S. sonnei* has acquired more mobile genetic elements (MGEs) conferring AMR than *S. flexneri*, which provides a survival advantage under antibiotic selection (62).

Shigella transmission can occur among men who have sex with men (MSM) through contact with contaminated body fluids during sexual activity (72). Most of the sexually transmitted cases are caused by *S. sonnei*, with some caused by *S. flexneri* serotype 2a or 3a (22, 73). There are increasing reports of *S. sonnei* outbreaks among MSM in HICs, including the United States (74), United Kingdom (22, 75), France (76), Belgium (77), Netherlands (78), Spain (79), and Australia (80). This could result from the misuse of antibiotics and the development of AMR in *S. sonnei*, making infection harder to treat and control. The prevalence of recurrent and persistent infections among MSM highlights the urgent need for innovative strategies to address this health concern (81).

1.1.3 Pathogenesis of *Shigella*

The common symptoms of shigellosis include blood and mucus in faeces, which indicates extensive damage of intestinal epithelium and a high degree of inflammation. The pathogenesis of *Shigella* has been studied extensively (81-84).

All *Shigella* are intracellular pathogens that cause mucosal inflammation in the gastrointestinal tract, particularly affecting the large intestine. The ability of the bacterium to invade epithelial cells depends on the T3SS encoded on the ~30 kbp PAI on pINV. Multiple pathogens express a T3SS to assist their pathogenesis, such as *Salmonella Typhimurium*, *Yersinia enterocolitica*, and *Yersinia pestis* (85). The T3SS consists of several key components: (1) a needle-like apparatus known as the injectisome that extends from the bacterial surface and serves as a conduit for delivering proteins into host cells; (2) the basal body structure spanning the bacterial inner and outer membranes, consisting a series of proteins that anchor the system to the bacterial cell; (3) the cytoplasmic export apparatus for selecting and exporting proteins that will be injected into host cells; (4) and regulatory proteins controlling the assembly and function of the T3SS, ensuring that it is activated only when required (86-88). The T3SS delivers effector proteins from the bacterial cytoplasm into the local environment and target cells. These effector proteins facilitate the invasion and spread of *Shigella* within host cells (81).

The expression of the T3SS is carefully controlled by three regulators: the histone-like nucleoid structure protein H-NS, VirB, and VirF (86, 89). H-NS is a global transcriptional regulator found in many Gram-negative bacteria. H-NS preferentially binds to AT-rich regions of DNA and can repress the expression of genes in these regions. The high AT content of the genes encoding the T3SS on pINV allows H-NS to repress their expression at temperatures below 32°C (86, 89). This leads to the repression of the pINV encoded transcriptional activators VirB and VirF as well as genes encoding the T3SS apparatus, and conserves energy and resources (90-94). At temperatures over 32°C (such as in the human body), the binding of H-NS to DNA is reduced,

weakening its repressive effect, and allowing production of the transcriptional activator VirF which facilitates the expression of virulence genes during infection (86, 93, 94). Higher temperatures also promote the expression of *virB* and lead to activation of genes encoding the T3SS and its substrates (86). The temperature-dependent regulation of gene expression is a strategy employed by *Shigella* to adapt to different environments.

In general, *Shigella*'s pathogenesis is a well-established multi-step process (Figure 1.1.3A): (1) bacteria enter the gastrointestinal tract and reach the intestinal epithelium, attach and enter the basolateral side of the epithelium after transcytosis through microfold cells; (2) bacteria induce their phagocytosis by resident macrophages, and cause pyroptosis of macrophages, a highly inflammatory form of programmed cell death triggered by intracellular pathogens (95); (3) bacteria enter intestinal epithelial cells, where they actively replicate in the cytosol and spread to neighbouring cells by actin polymerisation. This process has been extensively studied and reviewed (81, 82).

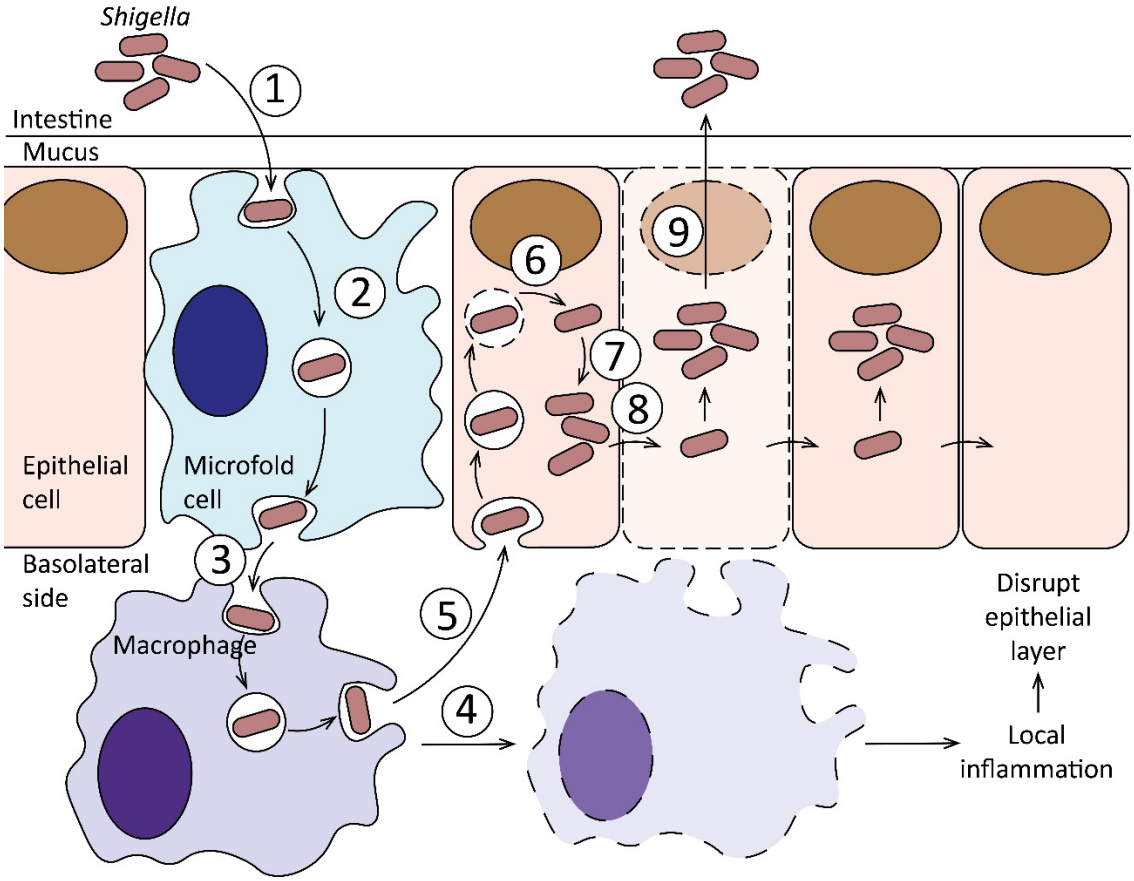


Figure 1.1.3A: General multi-step infection cycle of *Shigella*.
Figure not to scale. (1) *Shigella* employs multiple strategies to bypass the obstacles such as dramatic changes in pH and bile salts to reach the epithelial surface of intestinal mucosa. It degrades the mucus layer to gain access into the apical side of the intestinal epithelium. (2) *Shigella* is then transcytosed by microfold cells to enter the sub-epithelial compartment. (3) *Shigella* is endocytosed by macrophages, and utilises effectors encoded by the T3SS to evade the immune responses by macrophages. (4) *Shigella* causes pyroptosis of macrophages. This leads to local inflammation that recruits immune cells, with the migration of immune cells disrupting the epithelial barrier, favouring further entry of *Shigella* into the basolateral compartment. (5) *Shigella* gains entry into epithelial cells from the basolateral side, and is contained in the vacuole. (6) In the cells, *Shigella* escapes from the vacuole and enters the cytosol. (7) *Shigella* replicates in the cytosol. (8) *Shigella* utilises effectors secreted from the T3SS to cause cellular actin rearrangement and move from one cell to another. (9) During the process, the epithelial cells may die and release more *Shigella* bacteria, with the cycle repeating until *Shigella* is cleared by the immune system.

The human gastrointestinal tract poses multiple obstacles for *Shigella*. *Shigella* needs to tolerate pH changes along the gastrointestinal tract (96, 97). *Shigella* has an acid tolerance response that maintains internal pH homeostasis and repairs acid damaged proteins, enabling the bacterium to survive in the acidic conditions in the stomach (98). The acidic environment also triggers expression of T3SS and biofilm formation (99-101). In addition, *Shigella* utilises LPS and the AcrAB efflux pump to resist the antibacterial effect of bile salts. *S. flexneri* and *S. sonnei* respond

to bile salts differently. The primary and secondary bile salts, cholate and deoxycholate respectively, promote biofilm formation in the presence of glucose. This biofilm protects *S. flexneri* until it reaches the small intestine (102). Bile salts also induce assembly of T3SS in *S. flexneri* and IcsA-mediated cell adhesion which is important for pathogenesis and biofilm formation (103-106). However, in *S. sonnei*, IcsA is downregulated by deoxycholate which reduces cell invasion (107). Adhesins, which facilitate the attachment of bacteria to host cells, also play a role in the virulence of *S. flexneri* and *S. sonnei* (108, 109). However, the specific adhesins and their roles can vary between these two species (103, 107). More studies are required to understand differences in pathogenesis between different *Shigella* species.

Shigella employs several strategies to penetrate the mucus layer protecting the intestinal epithelium. *Shigella* triggers mucus production by goblet cells and induces local inflammation as part of the host immune responses to infection (110). The excessive mucus restricts oxygen diffusion and alters the oxygen gradient at the site of invasion, which in turn induces T3SS expression for invasion (111). To cross the mucus layer, *Shigella* secretes mucinases to degrade the mucus layer to reach the epithelial surface, which in turn promotes further mucus secretion (112, 113).

After crossing the mucus layer, *Shigella* reaches the apical side of the intestinal epithelium. *Shigella* is then transcytosed by microfold cells, specialised epithelial cells responsible for the sampling of intestinal contents. Upon encountering pathogens, microfold cells present the antigens of the invading bacterium, thereby initiating immune responses (114, 115). The process of transcytosis by *Shigella* is not harmful to microfold cells, and facilitates *Shigella* access to the basolateral side of the epithelium (116). Upon entry into the sub-epithelial compartment, *Shigella* is endocytosed by macrophages, and enters vacuoles (117). The T3SS induces vacuole lysis and facilitates escape of *Shigella* from the vacuoles, and entry into the cytoplasm. Subsequently, *Shigella* induces pyroptosis of macrophages. This protects *Shigella* from the

antimicrobial activities of macrophages, such as phagosome-lysosome fusion, and production of reactive oxygen species (ROS) and nitric oxide (118-121). There is evidence that macrophages are less efficient at internalising *S. sonnei*, resulting in reduced pyroptosis compared to *S. flexneri*, causing reduced inflammation. It was hypothesised that the O-Ag and the group 4 capsule (G4C, a high molecular weight immunogenic structure on the surface of the bacteria) shield the *S. sonnei* T3SS, thus reducing host immune responses (122-125).

Pyroptosis of macrophages triggers local inflammation due to the release of pro-inflammatory cytokines, leading to the recruitment of immune cells, including neutrophils (81, 126). Neutrophils are more active at killing *Shigella* than macrophages, and employ multiple antimicrobial strategies, including the formation of neutrophil extracellular traps, and the production of ROS and antimicrobial peptides (127). However, the migration of neutrophils to the inflamed site can disrupt the epithelial layer by increasing vascular permeability and physical breaching of the epithelial monolayer during their transmigration (128, 129). This allows further invasion of *Shigella* at the basolateral side of the epithelial layer (127, 130-134). Neutrophils can also be killed by *S. sonnei* and *S. flexneri* by inducing necrosis of neutrophils, while the *S. sonnei* O-Ag protects it from neutrophil killing (135, 136).

Following pyroptosis of macrophages, *Shigella* reaches the basolateral side of the intestinal epithelium, and delivers the T3SS effectors into the host cells using the T3SS apparatus (137, 138). These effectors cause rearrangements of actin and change the host cytoskeletal architecture, leading to entry of *Shigella* into the cells by endocytosis (139, 140).

Upon entering intestinal epithelial cells, *Shigella* rapidly escapes from the endocytic vacuoles and enters the cytosol (141-143). Once in the cytosol, *Shigella* induces changes in the cytosolic environment and manipulates changes in the actin polymerisation to facilitate its spread to neighbouring cells (144-146). This manipulation is primarily through the activity of *Shigella*'s IscA protein, which recruits and activates host N-WASP, a protein that is crucial in actin

cytoskeleton dynamics. The activation of IscA leads to actin polymerisation at one pole of the bacterium and propels the bacterium through the cytoplasm and into neighbouring cells (145, 147, 148). It is important to note that *S. sonnei* exhibits reduced invasion into epithelial cells, and the underlying reasons for this remain to be determined (81, 123, 125).

To ensure its replicative niche, *Shigella* manipulates various cell death pathways, including apoptosis and pyroptosis, to prevent epithelial cell death (149). This manipulation involves the secretion of T3SS effectors, such as OspC1 and OspD3 (150). The contents released by dying infected cells further contribute to local inflammation, leading to damage of surrounding cells, and the symptoms such as diarrhoea containing blood and mucus (83).

1.2 *S. sonnei* virulence and antimicrobial resistance

S. sonnei was discovered in 1915 by Carl Sonne (151). Analyses of WGS of *S. sonnei* suggest that *S. sonnei* emerged from a common ancestor in Europe less than 500 years ago (20). Analyses of SNPs or core genome sequences have classified *S. sonnei* into five lineages and inferred their evolutionary history and spread (20, 152). This typing system reveals the low levels of genomic diversity of *S. sonnei*, and demonstrates clonal expansion of the species (20, 153, 154). Among the five lineages, Lineage III is currently the most widespread lineage (20, 153). Lineage III *S. sonnei* is characterised by the presence of a distinct class II integron (In*2*) conferring resistance to trimethoprim, streptothricin, and streptomycin (155). Higher resolution sequence analysis further classifies these lineages into different clades and subclades (153). This hierarchical classification of lineages can reflect localised transmission dynamics and adaptation. In addition, sequence analyses allow surveillance of the dissemination of AMR genes in *S. sonnei* globally, and can inform research priorities and policies to counter the spread of the pathogen (153).

1.2.1 *S. sonnei* virulence plasmid pINV

S. sonnei harbours a 220 kbp pINV which encodes a T3SS. The T3SS means that plasmid exerts a high fitness cost to its bacterial host, potentially due to the energetic requirements for T3SS regulation and assembly (156, 157). In addition, *S. sonnei* pINV harbours a 13 kbp O-Ag gene cluster which contributes to its virulence and serotyping (69, 81, 84, 123, 136). Horizontal acquisition of the O-Ag gene cluster from *P. shigelloides* contributed to the emergence of *S. sonnei* (69). The O-Ag gene cluster is flanked by IS*1* upstream of *wzz* and an IS*629* ORF downstream of *aqpZ*, and consists of 18 ORFs including an IS*630* found between *wzy* and *wbgV* (8, 84, 158). The genes are required for O-Ag synthesis and translocation across the inner membrane. The O-Ag is the outermost component of the LPS molecule and consists of repeating

oligosaccharide units (8, 81, 159). The O-Ag saccharides are also incorporated in the G4C of *S. sonnei* (125).

As the plasmid is vital for *S. sonnei* virulence, studies have investigated the maintenance of *S. sonnei* pINV. As mentioned in Chapter 1.1.1, all *Shigella* pINV carries an IncFII replicon carrying genes required for plasmid replication (84). In addition, plasmids encode the maintenance systems, such as TA and partitioning systems, to ensure their vertical transmission in a bacterial population (160-162). TA systems induce post-segregational killing (PSK) of plasmid-free bacteria by utilising a stable toxin and a less stable antitoxin (163-165). If the plasmid is improperly segregated, bacterial cells without the plasmid will be killed by the toxin due to the absence of the neutralising antitoxin, ensuring the maintenance of plasmid in the bacterial population (163-165). Partitioning systems act to ensure segregation of the plasmids prior to cell division (166-168).

All *Shigella* pINV carries genes encoding a VapBC (previously known as MvpAT) as the major TA system contributing to plasmid maintenance. The VapC toxin cleaves the initiator tRNA, $tRNA^{fMet}$, inhibiting translation initiation; *vapB* encodes the antitoxin VapB which binds to VapC to form a hetero-octameric complex. VapB is readily degraded by the protease Lon. Therefore, in the absence of *de novo* production of VapB due to the loss of pINV, VapC will cleave $tRNA^{fMet}$ and lead to cell death/growth arrest (44). However, *vapBC* from *S. sonnei* pINV differs to that from *S. flexneri* pINV by SNPs in both *vapC* and *vapB*, and the single amino acid polymorphism in VapC (K32R) affects the function of the TA system (44). The absence of genes encoding the *ccdAB* and *gmvAT* TA systems means that *S. sonnei* pINV is frequently lost compared to *S. flexneri* pINV (156, 169). In addition, *S. sonnei* pINV harbours a functional *relBE* TA system which is not found on *S. flexneri* pINV, but does not contribute to *S. sonnei* pINV maintenance during laboratory culture (44). Furthermore, *S. sonnei* pINV harbours genes encoding a ParAB partitioning system, as *S. flexneri* pINV, to promote plasmid segregation during cell division (37).

In addition to the maintenance systems, *S. sonnei* pINV also harbours genes derived from the conjugation system found in *E. coli* F plasmid. These genes encode for various proteins including TraI, a relaxase, TraD, a Type 4 coupling protein (T4CP), TraX, a pilin acetylation protein, and FinO, a conjugation repressor protein. (84, 170, 171). The TraI is involved in the initiation of conjugation, though TraI on *S. sonnei* pINV is hypothesised to be non-functional due to insertion of IS*600* and absence of certain amino acids (84, 172). Due to the absence of the remainder of the conjugation system, the plasmid is thought to be non-conjugative (20, 84, 169, 173). Interestingly, the majority of ISs on *S. sonnei* pINV are expected to be inactive for IS transposition due to mutations (84). However, ISs can still contribute genetic changes as the sites of homologous recombination (HR) and loss of T3SS PAI (163). Due to these genetic changes, further understanding of horizontal gene transfer (HGT) related to pINV is required.

1.2.2 Antimicrobial resistance (AMR) in *S. sonnei*

The World Health Organisation has declared *S. sonnei* as a priority pathogen due to the emergence of AMR (174). *S. sonnei* has acquired multiple AMR genes *via* MGEs such as plasmids, transposons, ISs, and genomic islands (175). There is evidence demonstrating that commensal *E. coli* can serve as a reservoir of AMR genes which are then transferred to *S. sonnei* making it a multi-drug resistant (MDR) pathogen that resists multiple antibiotic categories (176). The frequent spread of these AMR determinants *via* plasmid conjugation emphasises the importance of studying HGT in *S. sonnei*.

S. sonnei is often resistant to antibiotics such as ampicillin, trimethoprim-sulfamethoxazole, chloramphenicol, streptomycin and tetracycline (177). Ciprofloxacin-resistant *S. sonnei* arose in Asia in the 1990s after the use of ciprofloxacin and other fluoroquinolones (178). Furthermore, a study from Chile reported significant variation in antibiotic susceptibility patterns among *S.*

sonnei strains over several periods: 1995-1997, 2002-2004, 2008-2009, and 2010-2013 (179), with *S. sonnei* developing resistance against antibiotics soon after their introduction (63, 177).

In recent years, there have been increasing reports of MDR *S. sonnei* with different resistance profiles around the world (177). Outbreaks of MDR *S. sonnei* have occurred in the UK, specifically of *S. sonnei* clade 5 (180). In Belgium, phylogenetic analysis revealed four independent clusters of persistently circulating MDR strains internationally, which were associated with the MSM population (77). In France, there has been a dramatic increase in *S. sonnei* resistant to ciprofloxacin, third-generation cephalosporins, and azithromycin since 2015 (21). MDR Shigella sonnei with reduced susceptibility to fluoroquinolones has been circulating in China since 2010, further limiting treatment options (181). Colistin resistance was detected retrospectively, with the first report of the mobile colistin resistance gene (*mcr-1*) in 2015 (182). It was hypothesised that misuse or overuse of antibiotics acts as a selective pressure, and contribute to development of extensive drug resistance in *S. sonnei* (180, 183).

S. sonnei employs several strategies to resist killing by antibiotic, including expression of drug efflux pumps, reduced cellular permeability to antibiotics, enzymatic inactivation of antibiotics, and mutation of antibiotic targets (63). Efflux pumps can expel a broad range of antibiotics, reducing the concentration of antibiotics within the cell (184). Reduction in cellular permeability to antibiotics, by biofilm formation or reducing number of porins, increases the minimum inhibitory concentration (MIC) of antibiotics (185). These mechanisms can be conferred by chromosomal mutations or acquisition of genes by HGT through MGEs such as plasmids, transposons, and ISs (Table 7.1) (186). For example, *S. sonnei* develops resistance against fluoroquinolones after accumulation of point mutations in genes encoding DNA gyrase and topoisomerase IV, which are the targets of the antibiotic (187). Expression of efflux pumps and alteration of membrane permeability also increase resistance against fluoroquinolones (188). In addition, acquisition of plasmid-mediated quinolone resistance genes, encoding proteins that

protect DNA gyrase from cleavage by fluoroquinolones can also lead to the resistance (187). Studies have shown that *S. sonnei* harbours plasmids carrying resistance determinants, with some plasmids sharing backbones with those found in *E. coli*, differentiated by the insertion of resistance gene cassettes associated with MGEs such as transposons Tn*3*, Tn*7*, and Tn*10* (79, 179, 189).

These resistance genes can be arranged in specific gene clusters and associated with integrons, particularly class 1 (In*1*) and class 2 (In*2*), capturing and expressing gene cassettes that confer AMR (10, 20, 189-196). Integrons have been utilised in the investigation of the epidemiology of *S. sonnei* around the world (10, 20, 190-196). In*2*, often associated with Tn*7*, are frequently found in *S. sonnei* and typically contain gene cassettes like *dfrA1*, *sat1* (streptothricin resistance), and *aadA1* (20, 197). Other transposons, including Tn*3*, Tn*10* and Tn*21*, are also involved in spreading associated AMR genes, such as Tn*3*-associated bla_{TEM-1A}, through insertion into plasmids and chromosomal DNA (197). ISs further contribute to the mobility of resistance genes, enabling their integration into various genomic contexts within *S. sonnei* (39).

The prevalence of loci containing resistance genes varied across different time periods and geographic regions (10, 179). For instance, a study investigated AMR dynamics in Chilean *S. sonnei* strains between 1995 and 2013 at different locations showed that *S. sonnei* can harbour a chromosomal *Shigella* resistance locus pathogenicity island (SRL PAI) containing genes encoding resistance to antibiotics including ampicillin (bla_{OXA-1}), streptomycin (*aadA1*), trimethoprim (*dfrA14*), and chloramphenicol (*cat*) (179). This locus can carry an In*1* containing the bla_{OXA-1}-*aadA1*-IS*1* gene cluster. Alternatively, *S. sonnei* strains without SRL PAI can harbour the *dfrA1-aadA1-qacEΔ1-sul1* gene cluster (179).

The interplay between these MGEs and chromosomal elements underscores the complex mechanisms driving the emergence and spread of multidrug-resistant *S. sonnei* strains. The rapid emergence of AMR in *S. sonnei* poses a serious threat to public health by reducing

treatment options. Epidemiological studies contribute to understanding the current AMR trends in *S. sonnei*. Therefore, genomic surveillance is important for tracking the spread of AMR and evolution of *S. sonnei*.

1.2.3 Importance of plasmids in the dissemination of AMR in *S. sonnei*

Multiple mechanisms of HGT contribute to the spread of AMR in *S. sonnei*, with plasmids being a major vehicle for the dissemination of AMR genes among *S. sonnei* isolates (176, 198). For example, by harbouring a gene *psiB* encoding the SOS response inhibitor on the low-fitness cost plasmids, *S. sonnei* exhibits enhanced resistance against ciprofloxacin by inhibiting its SOS response to low concentrations of ciprofloxacin (23).

Plasmids are generally classified into three categories based on their transferability: non-mobilisable, mobilisable, and conjugative. Non-mobilisable plasmids lack the capacity for horizontal transfer. Mobilisable plasmids can be transferred by utilising the conjugation system provided by a co-resident conjugative plasmid. Conjugative plasmids possess a comprehensive set of genes that encode for the conjugation machinery, enabling them to self-transfer between bacteria in a contact-dependent manner (199-202). The HGT of plasmids serves as a mechanism for the spread of AMR genes that are encoded on either small mobilisable plasmids or large conjugative plasmids (203, 204). Larger plasmids (> 50 kbp) are likely to be conjugative or encode more AMR genes (205). Conjugative plasmids can harbour multiple AMR genes which are of clinical concern, including genes encoding beta-lactamases (206). These plasmids are increasingly identified in *S. sonnei* from various countries (21, 22, 79, 176, 180, 181, 207, 208). Conjugation is considered as the main mechanism for the spread of AMR in *S. sonnei* (176). The identification of low-fitness cost resistance plasmids in *S. sonnei* is a further concern as bacteria can carry these plasmids without a significant reduction in their growth or survival (23).

In addition, DNA crossover, where segments of DNA are exchanged between homologous sequences, can occur during HR. The process of DNA crossover contributes to the evolution of resistance plasmids and the acquisition of AMR genes (198, 209-211). HR is a common mechanism utilised by bacteria to maintain genomic integrity by repairing DNA damage, such as double-strand breaks (DSBs) and collapsed replication forks (212). The HR machinery utilises single-stranded DNA (ssDNA) at the site of lesion to search for homologous double-stranded DNA (dsDNA) sequences as the template for DNA repair. The mechanisms of RecA-dependent and -independent HR contributing to DNA repair have been extensively studied and reviewed (211-213). The emergence of novel resistance plasmids with virulence genes arising through HR has been detected in other clinically important pathogens such as *Klebsiella pneumoniae* and *Salmonella enterica*, contribute to the spread of these genetic traits (214-216).

1.2.4 Horizontal transfer of plasmids in *S. sonnei*

Plasmids can undergo HGT through transduction, transformation, and conjugation (217-219). Transduction is mainly utilised by bacteriophages to transfer genetic material (220). Transformation involves uptake of DNA from the surrounding environment (219). Conjugation requires direct bacteria-bacteria contact, and the transfer relies on a conjugation system often encoded by the plasmid (Figure 1.2.4A) (206, 217). In addition to the plasmid transferring itself, conjugative plasmids can also mediate transfer of other plasmids provided those plasmids carry an *oriT* which can be recognised and cleaved by a mobilisable plasmid-encoded relaxase, and delivers the plasmid ssDNA to the conjugation system assembly (199). Plasmids without an *oriT* and a relaxase are normally considered to be non-mobilisable.

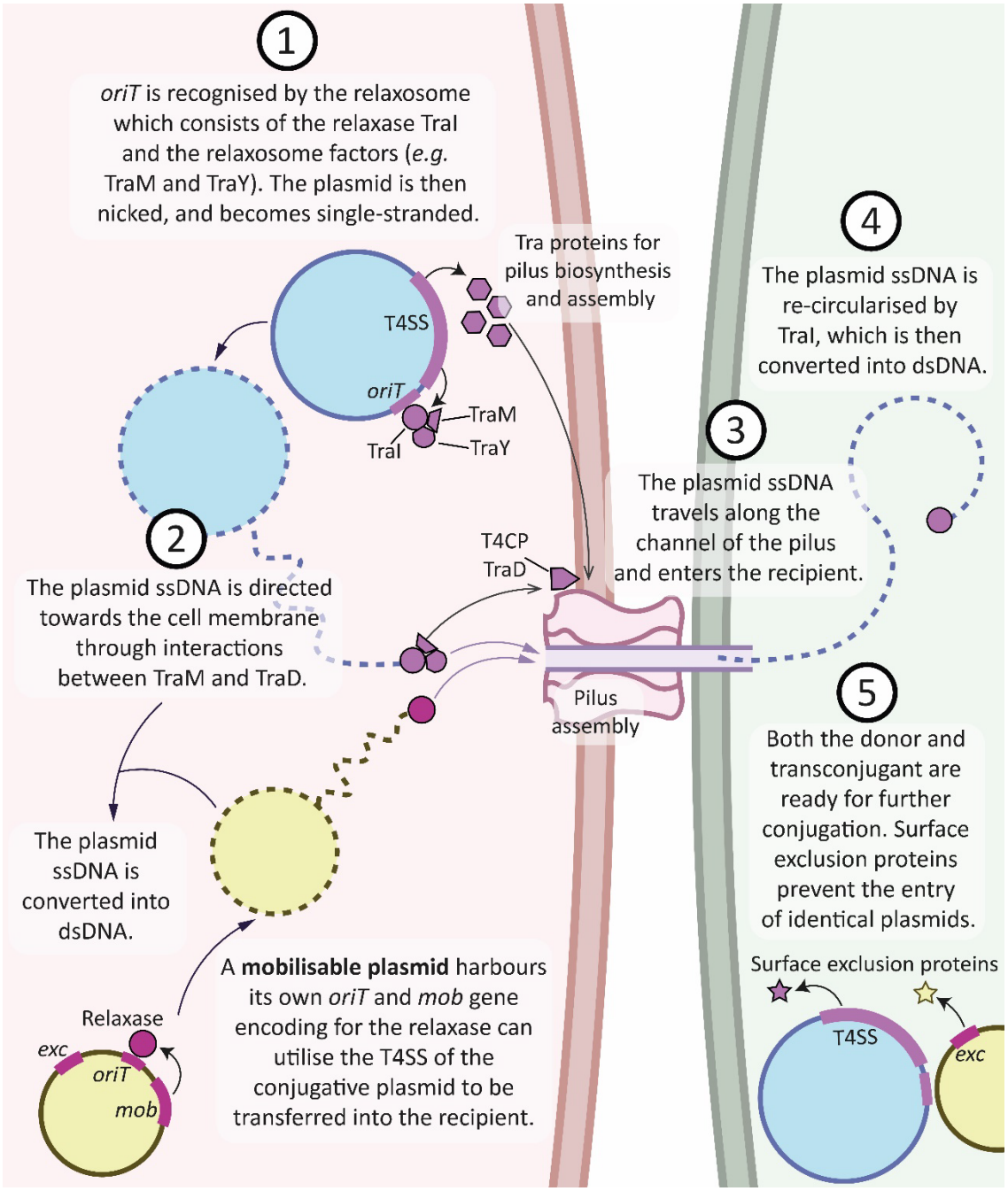


Figure 1.2.4A: General process of plasmid conjugation and mobilisation.
Figure not to scale. Plasmids are transferred from the donor bacterium (pink) to the recipient bacterium (green), with T4SS factors shown in purple. (1) In the donor, the relaxosome (formed by TraI, TraM, and TraY) recognises and binds to the *oriT* on the conjugative plasmid (blue). The relaxase nicks the plasmid dsDNA at *oriT*, converting plasmid dsDNA to ssDNA. (2) TraM directs the plasmid ssDNA towards the cell membrane *via* interactions with the T4CP, TraD (221). A mobilisable plasmid (yellow) with *oriT* can also be transferred using the conjugative plasmid's T4SS. The plasmid ssDNA is then converted into dsDNA. (3) The linear plasmid ssDNA travels through the pilus and enters the recipient. (4) In the recipient, the plasmid ssDNA is recircularised by the relaxase and converted into dsDNA. (5) At the end of the process, both donor and recipient (now transconjugant) harbour the same plasmids. Plasmid-encoded surface exclusion proteins prevent the entry of identical plasmids.

As many AMR plasmids are conjugative, I will describe the mechanism of conjugation in more detail, using the well-studied F plasmid as an example. Conjugation involves a donor bacterium transferring a plasmid into a recipient bacterium. An intact conjugation system typically includes an origin of transfer (*oriT*), a relaxase gene (*e.g.* TraI from the F plasmid) that recognises *oriT*, and genes encoding a Type 4 Secretion System (T4SS), which forms a pilus for plasmid DNA transfer. To initiate conjugation, in the donor the relaxase TraI with other factors such as TraM and TraY forms a relaxosome, which recognises *oriT* and nicks one strand of the duplex DNA. The DNA is then unwound by helicase activity, which is an intrinsic activity of some relaxases, to generate single stranded plasmid DNA (222). The plasmid ssDNA is covalently bound to the relaxosome, and directed towards bacterial inner membrane through interactions between relaxosome and the T4CP, such as TraM-TraD interactions during transfer of F plasmid (221). Upon reaching the T4SS, the relaxase:plasmid ssDNA travels through the channel connecting the donor with the recipient, and enters the recipient. Plasmid ssDNA then re-circularises in both the donor and recipient, and is converted into dsDNA (217, 223, 224). The conjugation system can be utilised by a mobilisable plasmid to mediate its horizontal transfer. At the end of conjugation, as both donor and transconjugant (a recipient that received the plasmid) harbour the same plasmids, and can act as donors in further conjugations. Plasmid-encoded exclusion protein(s), such as TraS and TraT from the F plasmid and Exc from mobilisable colicin-E1 plasmids, prevents the acquisition of related plasmids (225, 226). Conjugation is regulated by chromosomal factors like H-NS and the plasmid-encoded fertility inhibition system, which includes an antisense RNA, *finP*, that represses expression of the transcriptional activator TraJ, and an RNA chaperone, FinO, that protects *finP* from RNase E degradation and stabilises the *finP-traJ* mRNA duplex (171, 227-232).

T4SS and conjugation systems can differ between plasmids and bacterial species and have been characterised into different families according to the mobilisation elements such as the relaxase (Mob-typing) (205, 233, 234). These typing schemes contribute to the classification of

conjugative plasmids. This typing scheme has been used alongside replicon typing (Inc typing), with some association between Mob-type and Inc-type, and further assisted by the plasmid multi-locus sequence typing (pMLST) scheme for complicated plasmids by categorising the allelic profiles of multiple housekeeping genes (205, 235, 236).

1.2.5 Other mobile genetic elements in *S. sonnei*

In addition to HGT of the entire plasmid, HGT can also occur of smaller genetic elements. For example, IS can mediate genetic change through transposition and/or HR (38). *S. sonnei* pINV has been considered as a mosaic plasmid because sequences on the plasmid are from different origins, which were acquired *via* HGT and/or HR involving ISs (69). ISs have contributed to extensive genomic rearrangements in plasmids, are highly plastic and can mobilise large sections of DNA. ISs account for approximately one-third of the coding sequence on pINV from a Lineage III Chinese clinical isolate *S. sonnei* Ss046, where over two-third of the ISs are partial copies due to truncated ends and interactions with other ISs leading to deletion and/or insertion of the elements and adjacent DNA (84). There are 30 intact ISs existing in different abundance on Ss046 pINV, including IS*600*, IS*630*, IS*1294*, IS*629*, IS*1*, IS*21*, IS*2*, IS*Sfl2*, IS*3*, IS*4*, and IS*91*. Similarly, a common laboratory strain Lineage II *S. sonnei* 53G, initially isolated from Japan, harbours 26 intact ISs from 13 different types on pINV, including IS*600*, IS*21*, IS*629*, IS*91*, IS*100,* IS*630*, IS*1294*, IS*4*, IS*2*, IS*1*, IS*3*, IS*110*, and IS*Sso4* (20, 44, 237).

Together these ISs could contribute to shape the evolution of *S. sonnei* pINV. An example would be the 13 kbp O-Ag gene cluster which is divided into two parts by an intact IS*630*, and flanked by an intact IS*1294*, a truncated IS*630* and a partial IS*629* (84). Due to the sequence identity, this cluster was thought to be horizontally acquired from *P. shigelloides* O17 due to IS-mediated events (84). In addition, ISs could also contribute to the loss of important genetic traits, such as the T3SS PAI and/or the gene *virF* responsible for regulating the expression of T3SS. Loss of *S.*

sonnei virulence could be mediated by recombination between homologous copies of IS*21* (leading to the deletion of the T3SS PAI and *virF*), IS*1* (causing deletion of the T3SS PAI and *virB*), and IS*1294* (single-ended rolling transposition followed by HR and the subsequent deletion of *virF*) (44). These ISs are well-documented and will be discussed in more details (238, 239).

Different ISs may exhibit different features and therefore different mechanisms for transposition (Figure 1.2.5A). IS*630* harbours short terminal inverted repeats (IRs) and genes encoding for a transposase, allowing the excision of itself and preferentially insert into the sequence 5´-NTAN-3´, leading to duplication of a TA dinucleotide at the site of insertion (240). IS*630* is able to mediate transposition of antibiotic resistance cassette flanked by the IRs of IS*630*, but there was no evidence that IS*630* contributes to the formation of cointegrates between plasmids (241). IS*21* harbours *istA* and *istB* encoding for a transposase and an ATPase, and uses a different transposition mechanism compared to IS*630* (242-244). IS*21* preferentially targets 5´-TCGG-3´ and requires a circular intermediate or tandem duplication of itself to promote active transposition. This typically leads to the duplication of up to four base pairs at the site of insertion (242, 245, 246). Unlike IS*630*, IS*21* has been involved in the formation of cointegrates between plasmids (242). Similar to IS*21*, IS*1*, harbouring IRs and InsAB transposase and preferentially targets AT-rich region of the DNA, leads to simple insertion and formation of cointegrates when two copies of IS*1* flanked a non-transposable region (247-249). These ISs could contribute to the insertion of the entire plasmid into another with targeting sites at optimal conditions (246, 247). IS*1294*, an IS*91*-like element, harbours an origin (*oriIS*) and a termination site (*terIS*) flanking genes encoding the transposase (250). IS*91*-like elements specifically target 5´-GTTC-3´ or 5´-CTTG-3´ without causing duplication of sequences at the site of transposition (251, 252). IS*91*-family elements transpose *via* single ended rolling-circle mechanism, similar to that for plasmid replication, and thus, they are capable of mobilising adjacent gene clusters (250, 253, 254). As IS*91* elements are commonly found to be adjacent to pathogenicity- and virulence-related genes/islands, it is thought to be centrally involved in the

dissemination of these important genetic traits and generate novel virulence plasmid variants (255). As copies of intact IS*1294* were found to flank the O-Ag gene cluster, and adjacent to T3SS PAI, IS*1294* could contribute to the acquisition of the genetic traits and shape the plasmid backbone of *S. sonnei* pINV by either RecA-dependent HR, or transposition/insertion/deletion of IS*1294*, or both (44, 84, 253).

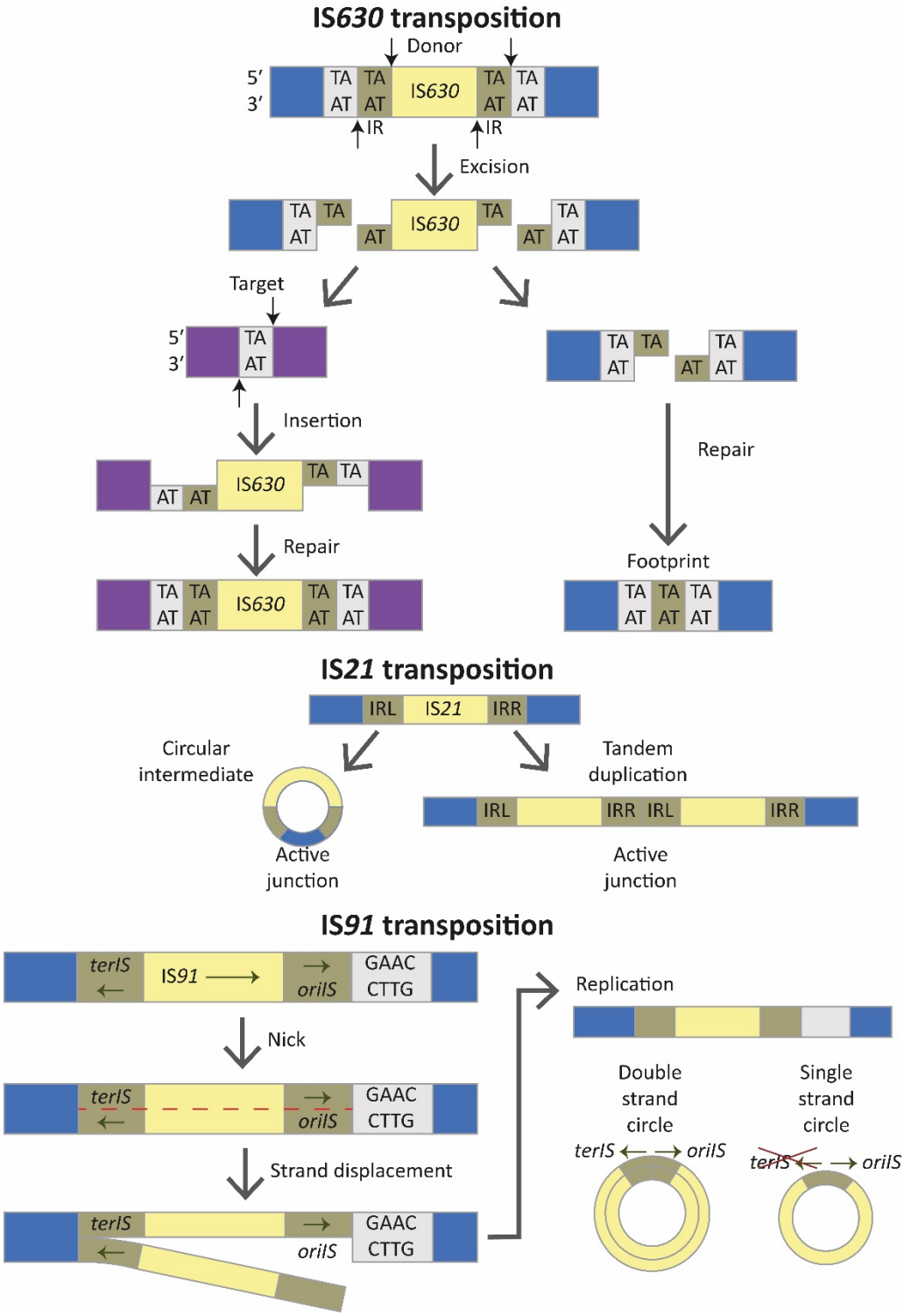


Figure 1.2.5A: General mechanisms of IS transposition.
The figures of models of IS transposition were taken from TnCentral: Figure IS630.4, Figure IS21.1 and Figure IS91.10 (238). IS*630* transposes by excising and inserting itself at the sequence 5′-NTAN-3′, resulting in duplication of "TA" at the site of insertion, while leaving a "TATATA" footprint at the donor backbone. IS*21* requires an active junction formed by generation of a circular intermediate or tandem replication of itself for transposition. Transposase of IS*91* nicks at the *oriIS* and the strands peel off. Replication occurs from the 3′-OH created by the cleavage of transposase, and the transposition terminates at *terIS*, which both ends are then covalently joined to generate a single-strand circle. The single-strand circle is then converted into double-strand by rolling-circle replication. Cointegration caused by IS*91* could be due to one-ended transposition where termination occurs at the conserved 5′-GTTC-3′ or 5′-CTTG-3′ in the donor plasmid, and could lead to the transposition of the adjacent sequences.

In addition to ISs, transposons, which are are generally larger than IS, often carry accessory genes, such as those encoding AMR, and contribute to the spread of these genes (256). Transposons can be classified into two classes according to their characteristics. Class I transposons typically harbours a 5´-untranslated region (5´-UTR) with internal promoter for the following open reading frames encoding a nucleic acid binding protein and a protein with endonuclease and reverse transcriptase activities, and a poly(A) tail. Class II transposons contains a gene encoding for transposase which is flanked by terminal inverted repeats (TIRs). Similar to ISs, transposons can move within the genome by excising themselves from one location and integrating into another, or insert a copy of itself at a new location with the original copy unaltered (238, 257). The widespread Lineage III *S. sonnei* is often associated with the carriage of transposon Tn*7* and class II integrons (In*2*), which can also carry a small plasmid harbouring genes conferring resistance against tetracycline (*tetAR*), streptomycin (*strAB*) and sulphonamides (*sul2*) (20).

Integrons are genetic units containing an *intI* gene encoding an integron integrase, and an *attI* integron-associated recombination site (193). Integrons are characterised by their ability to capture and incorporate gene cassettes, such as AMR genes (195, 258). Class I integrons (In1) is well described in both Gram-positive and -negative bacteria, and often associates with *aadA1* (conferring streptomycin and spectinomycin resistance) and *dfrA1* (conferring trimethoprim resistance) (196). In*2*, which is frequently detected in Lineage III *S. sonnei* and other Gram-negative bacteria, is commonly associated with Tn*7* and its derivatives (20, 196). In2 shares a common pool of AMR genes with In*1*, including *dfrA1*, *aadA1* and *sat1* (conferring streptothricin resistance) (259). These cassettes can be expressed from a strong promoter within the integron, leading to the expression of the resistance genes (190, 192). These features make integron a significant concern for the evolution and dissemination of AMR in Gram-negative bacteria (191). Analysis of integrons also allows the classification of *S. sonnei* strains (20). The diversity and

distribution of MGEs can be determined using databases such as TnCentral and ISfinder, which are helpful resources for research (238, 239).

In summary, understanding of plasmid biology and HGT is informative about the evolution of *S. sonnei*, and might be beneficial in tackling the dissemination of AMR genes (260).

1.2.6 Inter-bacterial competition by *S. sonnei*

S. sonnei can harbour a Type 6 Secretion System (T6SS) which enhances its competition against other bacteria occupying the same niche (261). T6SS is a contact-dependent syringe-like molecular machinery for inter-bacterial competition found in several Gram-negative bacteria (262, 263). *S. sonnei* T6SS can target *S. flexneri* and *E. coli*, but has no obvious effect on its human host (136). The presence of a functional T6SS in *S. sonnei* could contribute to the observed shift in dominance of *S. sonnei* over *S. flexneri* in LMICs (261). The T6SS is beneficial for *S. sonnei* during gut colonisation (81, 264). However, there is some functional degradation in the T6SS of *S. sonnei* due to SNPs, indels and ISs, so the T6SS is not functional across *S. sonnei* Lineage I, II (*e.g. S. sonnei* 53G), III, and V (265, 266).

In addition to contact-dependent mechanisms, *S. sonnei* can synthesise colicins, the antimicrobial proteins that kill or inhibit the growth of closely related bacterial strains by binding to specific receptors and disrupting cellular processes in a contact-independent manner (267, 268). Colicins contribute to the anti-bacterial activity of *S. sonnei* and can be beneficial for its colonisation (265). Multiple types of colicins and their mechanisms of action have been extensively studied (269). Colicin operons consist of three genes; a colicin-encoding gene, a gene encoding an immunity protein against the colicin, and a gene encoding a lysis protein necessary for secretion of colicin into the extracellular environment (269-271). After being released into the surrounding environment, colicins enter a target cell by diffusion or active transport, and induce lysis of the bacteria (272). Colicins are not toxic to the producing bacteria as they are

protected by the immunity protein (273). Colicin operons are normally carried on plasmids, either on small, high-copy number plasmids with mobilisation elements, or on large, low copy number conjugative plasmids (274). Understanding of these colicinogenic plasmids has contributed to different applications. For example, pColE1 has become a common cloning vector (275). In addition, pColIb-P9 (Accession no. AB021078), first isolated from *S. sonnei*, is commonly used to study the IncI1 group plasmid replicon (276-278). Due to the high prevalence of colicinogenic plasmids in *S. sonnei*, colicin typing has been used in the past to track epidemic strains of *S. sonnei* (279, 280).

Overall, plasmids are important for virulence, competition and AMR in *S. sonnei*. Therefore, *S. sonnei* is a good model to understand the interplay between plasmids and its effect on plasmid and bacterial evolution. Better understanding of plasmid biology could also lead to the development of novel applications to prevent and treat infection.

1.3 Understanding plasmid biology is important for *S. sonnei*

Understand plasmid biology is crucial because plasmids are important for virulence and AMR in *S. sonnei* and other pathogens, including *E. coli*, *K. pneumoniae*, and *S. enterica* (215, 216, 218, 281-284). There is evidence suggesting the association of AMR genes to specific Inc groups of plasmids, with the knowledge used to tackle the spread of plasmids and AMR genes (205).

1.3.1 Regulation of plasmid replication and termination

Plasmids replicate autonomously of the chromosome, and the elements required for replication are arranged in plasmid replicons (285). The IncFII plasmid replicon on *S. sonnei* pINV is very similar to the well-studied IncFII R1 replicon (84, 286). The IncFII replicon typically encodes a copy number regulatory system CopAB, a leader peptide Tap, a replication initiation protein RepA, and contains an origin of replication (*ori*), allowing unidirectional replication according to the Class A Theta mode which is characterised according to the formation of theta-shaped DNA intermediates during replication, the reliance of the host's replication machinery, and the tightly regulated initiation of replication (166, 287).

Initiation of plasmid replication involves the expression of RepA (Figure 1.3.1A). The mRNA of RepA is transcribed along with the transcript for the leader peptide *tap*, which aids in the translation of RepA by disrupting a hairpin structure that conceals the Shine-Dalgarno (SD) sequence, thereby enabling ribosome binding to *repA*-mRNA (288). A high number of RepA molecules (about 20) is required to bind the *ori*, causing melting of dsDNA and relaxing the AT-rich region of the *ori* (289). This leads to the recruitment of proteins for replication, such as DnaB helicase, DnaG primase and DNA polymerase III holoenzyme (PolIII-HE). DnaB unwinds the plasmid dsDNA. DnaG synthesises a short RNA primer to allow DNA chain elongation of the leading strand which is catalysed by PolIII-HE. Replication of the lagging strand occurs through synthesis of Okazaki fragments (290-292). It is hypothesised that once bound, RepA remains at

the *ori* and blocks the progression of the lagging strand synthesis which is in the opposite direction of the replication fork, ensuring unidirectional replication; *in vitro* evidence indicates that replication of the lagging strand terminates at the RepA binding site (293). Leading strand synthesis starts at the single-stranded initiation (*ssi*) site which contains a site for synthesis of an RNA primer (290). Once DnaG synthesises a RNA primer at the *ssi*, PolIII-HE and other proteins involved in replication are recruited for leading strand synthesis (290, 294). There is a *terR* sequence located between plasmid replicon and the *ssi*, which can be recognised by the chromosomally encoded protein, Tus (295, 296). Tus binds to *terR* and forms a physical barrier to the progression of DNA replication fork, thereby slowing down and stopping DNA replication (297). This mechanism of termination of replication is shared among plasmids replicate unidirectionally *via* similar mechanisms (295, 296).

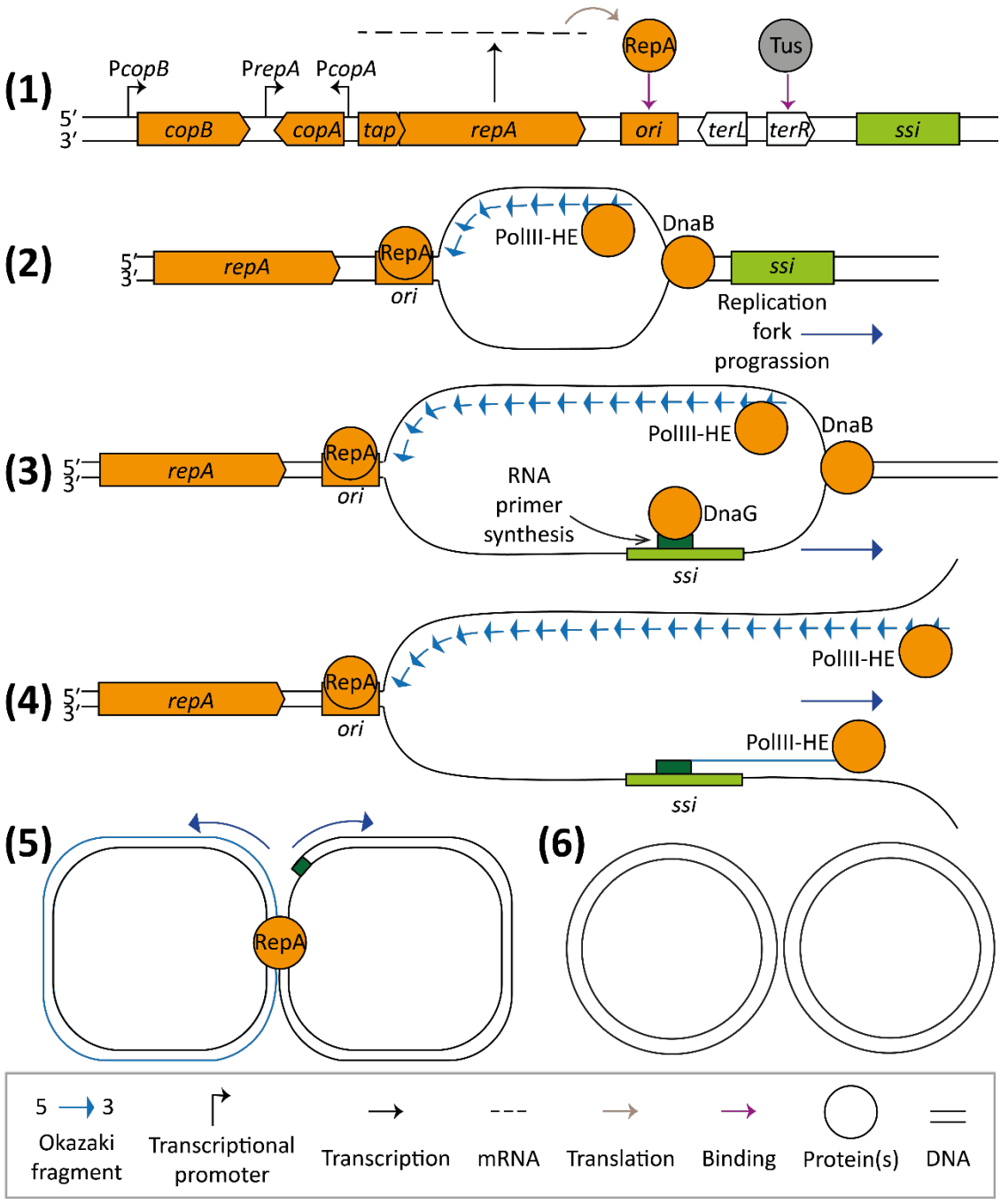


Figure 1.3.1A: Model of IncFII plasmid replication.
The model is adapted from (289, 294), not to scale. The elements in the IncFII plasmid replicon are shown in (1). IncFII plasmid replication involves multiple steps: (1) RepA is expressed and binds to *ori*. (2) Binding of RepA to *ori* recruits DnaB to unwind plasmid dsDNA. During the process, lagging strand replication by Okazaki fragments and PolIII-HE initiates at the exposed ssDNA. RepA stays bound to *ori* and forms a physical obstacle for the progression of the replication fork towards *repA* to ensure unidirectional replication as labelled. (3) Once the *ssi* is exposed as ssDNA, DnaG binds to the *ssi* and synthesises primers for leading strand replication. (4) PolIII-HE is involved in the replication of the leading strand. (5) A theta-shaped intermediate is formed at later stages of replication. (6) At the end of replication, two identical plasmids are created.

Replication of IncFII plasmids is tightly controlled to maintain the plasmid copy number (Figure 1.3.1B). IncFII plasmids are normally maintained at < 6 copies relative to the chromosome (167). The genes involved in copy number control *copAB* are transcribed in the opposite direction of *tap* and *repA* (287). The gene *copA* generates an antisense RNA with a region complementary to the *repA* mRNA (287). The *copA* antisense RNA binds to the *repA* mRNA, thereby inhibiting translation of RepA and preventing the initiation of replication (298). Another layer of regulation involves the protein CopB binding to a site within a 60 bp region of the *repA* promoter and repressing *repA* transcription (299). The concentration of CopB determines the extent of repression as the complex is only active upon formation of homo-tetramers (300). An increase in the number of plasmid copies leads to a corresponding increase in the number of CopB monomers, and the likelihood of tetramer formation, which then inhibits the transcription of *repA* mRNA, providing a negative feedback loop. These regulatory mechanisms ensure the low copy number of the plasmid with potential benefits including a low fitness burden to the bacterial host (301).

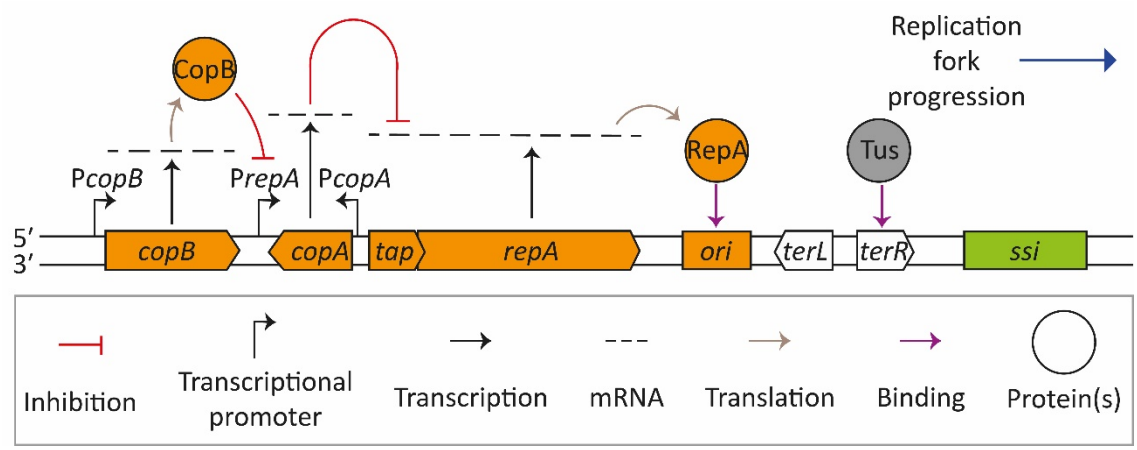


Figure 1.3.1B: Control of IncFII plasmid copy number.
IncFII plasmid copy number is tightly controlled. Production of RepA is required for initiation of plasmid replication. *copA* encodes an antisense RNA binds to the mRNA of *tap-repA* and blocks the translation of RepA, thereby reducing plasmid replication. Alternatively, CopB forms a homo-tetramer that binds to promoter of *repA* (P*repA*), and blocks transcription of the gene. The illustration is not to scale and the model is adapted from (287).

Plasmid replication regulated by antisense RNA is also commonly found in plasmids of other Inc groups, such as I-complex plasmids (including IncI1, IncB/O/K/Z and IncI2), which are widespread in *S. sonnei* (21). The IncI1 replicon consists of an *inc* antisense RNA, a gene encoding a leader peptide RepY, a gene for a replication initiation protein RepZ (or RepA), a CIS region for the termination of *repZ* transcription, and an origin of replication *oriV* (or *ori*) (276).

IncI1 plasmids also replicate unidirectionally through similar mechanisms as that of IncFII plasmids (Figure 1.3.1C). In IncI1 plasmids, RepZ translation is coupled with the leader peptide RepY, and RepZ binds to the *oriV* to unwind DNA duplex and initiate DNA replication upon synthesis of an RNA primer at the *ssi* (302, 303). However, the mechanisms controlling plasmid copy number are different to those in IncFII plasmids (Figure 1.3.1C) (304). There are stem-loop structures embedded in the mRNA of *inc*, *repY* and *repZ*, and pseudoknot structures formed by these stem-loop structures is required for successful translation of RepZ, as the structure exposes the SD sequence and the start codon of *repZ*-mRNA. This process can be inhibited by the antisense RNA *inc*, which binds to the stem-loop structure and prevents the formation of the pseudoknot structure (278). This highlights the importance of nucleic acid secondary structures in the regulation of plasmid replication.

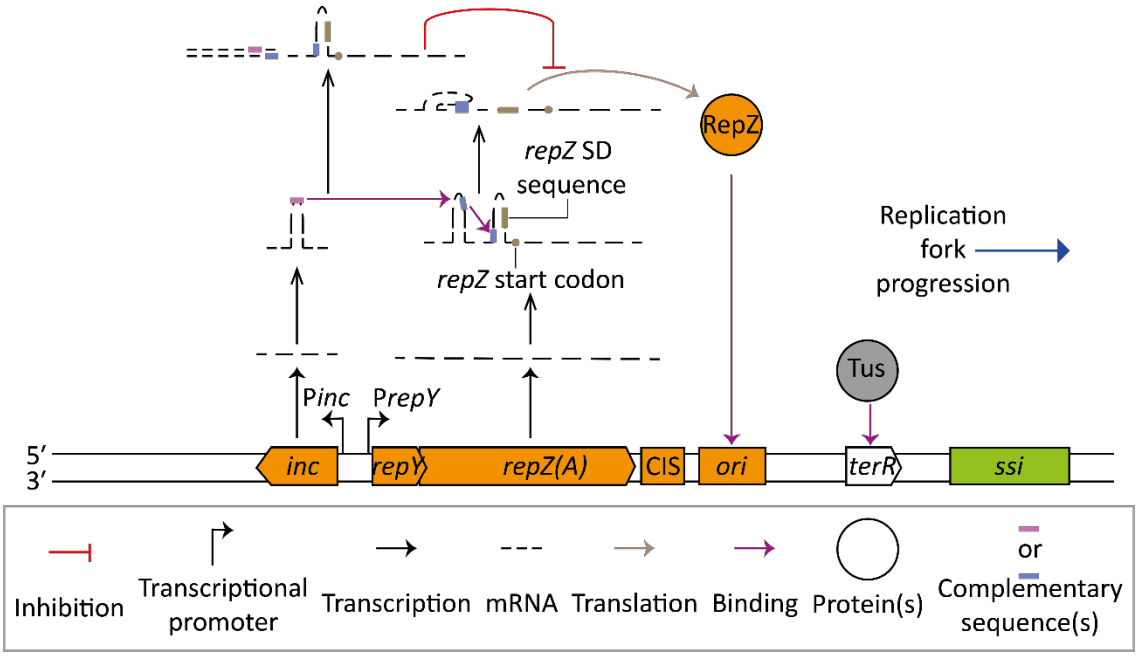


Figure 1.3.1C: Regulation of IncI1 plasmid replication.
The model is adapted from (278, 304). Figure not to scale. *repY-repZ* mRNA forms stem-loop structures which block the *repZ* SD sequence required for RepZ translation. This SD sequence is exposed after formation of a pseudoknot structure, which is formed by the binding of complementary sequences in the stem loops of the *repY-repZ* mRNA, allowing RepZ to be translated and initiate plasmid replication. This process is regulated by the *inc* antisense RNA which also forms a stem loop structure and binds to the complementary sequence on *repY-repZ* mRNA, and inhibits the pseudoknot structure that is essential for RepZ translation.

Understanding of these mechanisms has led to several applications. The plasmid replicon has been used for plasmid typing. Initially, plasmids were classified into Inc groups according to their ability to co-exist in the same bacterium with other plasmids, with another typing scheme using sequences of the basic plasmid replicon was developed (285, 305-307). Another typing system has been developed which relies on the hybridisation of 19 DNA probes which recognise specific replicons (308). Based on this scheme, a further system which is less labour-intensive was developed using multiplex PCR to detect *inc*/*rep* of plasmids, which is able to characterise the majority of the Inc groups of plasmids circulating in members of the family Enterobacteriaceae (309, 310). All isolates of *S. sonnei* carry the IncFII pINV virulence plasmid, with an increasing proportion also harbouring AMR plasmids belong to other Inc groups.

1.4 Aims of the thesis

S. sonnei pINV has been considered as a non-mobilisable plasmid as no *oriT* has been identified. However, in a study that defined the contribution of *S. sonnei* pINV to virulence, pINV was transferred from *S. sonnei* into *E. coli* with a lab-adapted conjugative plasmid R386 (169). This shows that although pINV is non-conjugative, it is mobilisable in the presence of a helper conjugative plasmid.

Therefore, the thesis has several aims:

1. To characterise a naturally-occurring resistance plasmid, pRES, identified from a Brazilian *S. sonnei* clinical isolate CS6, which was used as a model AMR plasmid in this work.
2. To investigate whether pRES can mediate the transfer of pINV in *S. sonnei*, and characterise the mechanism of transfer.
3. To understand a 199 bp region of homology located downstream of plasmid replicons which was involved in the formation of fusion/hybrid plasmids.

2 Materials and Methods

2.1 Bacterial culture

Bacterial strains used in this study are listed in Table 2.1. Bacteria were grown in liquid Lennox broth (LB, 20 g/l, Invitrogen) or tryptic soy broth (TSB, 30 g/l, Sigma-Aldrich). For solid media, 15 g/l agar (Oxoid) was added for LB agar (LBA) or TSB agar (TSA). For Congo Red TSA (CRA), 0.01% Congo Red (CR, w/v) was added (311). Antibiotics were included as needed at the following final concentrations: chloramphenicol (20 µg/ml), carbenicillin (100 µg/ml), kanamycin (100 µg/ml), and hygromycin (100 µg/ml). For *sacB* counter selection, sterile 10% sucrose (w/v) was added to autoclaved salt-free agar (10 g/l tryptone, 5 g/l yeast extract, 15 g/l agar). For *pheS* counter-selection, 4 mM 4-chlorophenylalanine (4CP) was added to the solid media (10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, 15 g/l agar). When culturing *E. coli* MFD *λpirΔhsdRΔdapA*, a final concentration of 300 µM 2,6-diaminopimelic acid (DAP, Sigma-Aldrich) was included with/without antibiotics. Unless otherwise specified, bacteria grown on solid media were incubated at 37°C, while liquid cultures were incubated at 37°C with 180 rpm.

Bacteria	Strain	Genotype	Reference
***S. sonnei* CS6**	QLSS2	pINV pRES:: $bla_{TEM-116}$	This work
***S. sonnei* 53G**	QLSS8	*aroG::aph(3′)-Ia*	This work
	QLSS12	pINV::*cat* pRES::$bla_{TEM-116}$	This work
	QLSS25	pINV::*cat* pRES::$bla_{TEM-116}$::Δ*nikB*	This work
	QLSS29	pRES:: $bla_{TEM-116}$::Δ*nikB*	This work
	QLSS32	pRES::$bla_{TEM-116}$	This work
	GMCT323	pINV::*cat*	This work
	QLSS43	Δ*recA* pINV::*cat* pRES::$bla_{TEM-116}$	This work
	QLSS45	Δ*recA aroG::aph(3′)-Ia*	This work
	QLSS47	Δ*recA* pRES::$bla_{TEM-116}$	This work
	QLSS51	Δ*recA aroG::aph(3′)-Ia* hybrid plasmids type 1	This work
	QLSS55	Δ*recA aroG::aph(3′)-Ia* hybrid plasmids type 3	This work
	QLSS58	Δ*recA aroG::aph(3′)-Ia* hybrid plasmids type 3	This work
	QLSS60	Δ*recA aroG::aph(3′)-Ia* fusion plasmid formed by IS*21* transposition	This work
	QLSS61	pINV::*cat* pRES::$bla_{TEM-116}$::Δ*oriT*	This work
	QLSS63	pINV::*cat* pRES::$bla_{TEM-116}$::*sacB*	This work
	QLSS64	pRES::$bla_{TEM-116}$::Δ*oriT*	This work

	QLSS65	pRES::*bla*$_{TEM-116}$::*sacB*	This work
	QLSS74	*aroG::aph(3′)-Ia* hybrid plasmids type 1 (TC4^{+})	This work
	QLSS77	pINV::*cat*::Δ199R	This work
	QLSS83	pRES::*bla*$_{TEM-116}$::*sacB*::Δ199R	This work
	QLSS86	pINV::*cat*::Δ99O	This work
	QLSS87	pINV::*cat*::99O^{Mut1}	This work
	QLSS89	pINV::*cat*::99O^{Mut3}	This work
	QLSS92	pINV::*cat*::99O^{Mut2}	This work
***S. sonnei* 53G transconjugants**	TC1 (010(4)-3-7)	*aroG::aph(3′)-Ia* pINV::*cat*	This work
	TC1 (010(4)-2-8)	*aroG::aph(3′)-Ia* pINV::*cat* pRES::*bla*$_{TEM-116}$	This work
	TC1 (010(4)-1-6)	*aroG::aph(3′)-Ia* 90 kbp hybrid plasmid type 1	This work
	TC1 (CR^{-} QLSS74)	*aroG::aph(3′)-Ia* 90 kbp hybrid plasmid type 1	This work
	015-1-2	*aroG::aph(3′)-Ia* hybrid plasmids type 2	This work
	015-1-3	*aroG::aph(3′)-Ia* hybrid plasmids type 1	This work
	015-1-4	*aroG::aph(3′)-Ia* hybrid plasmids type 3	This work
	015-2-1	*aroG::aph(3′)-Ia* hybrid plasmids type 1	This work
	015-2-3	*aroG::aph(3′)-Ia* pINV::*cat* pRES::*bla*$_{TEM-116}$	This work
	015-2-4	*aroG::aph(3′)-Ia* hybrid plasmids type 4	This work
	015-2-8	*aroG::aph(3′)-Ia* hybrid plasmids type 4	This work
	015-3-1	*aroG::aph(3′)-Ia* hybrid plasmids type 3	This work
	015-3-2	*aroG::aph(3′)-Ia* pINV::*cat*	This work
	015-3-3	*aroG::aph(3′)-Ia* pINV::*cat*	This work
	010(6)-1-1	*aroG::aph(3′)-Ia* pINV::*cat*	This work
	010(6)-1-6	*aroG::aph(3′)-Ia* pINV::*cat* pRES::*bla*$_{TEM-116}$	This work
	010(6)-1-7	*aroG::aph(3′)-Ia* pINV::*cat* pRES::*bla*$_{TEM-116}$	This work
	010(6)-1-8	*aroG::aph(3′)-Ia* hybrid plasmids type 1	This work
	010(6)-2-2	*aroG::aph(3′)-Ia* hybrid plasmids type 3	This work
	010(6)-2-3	*aroG::aph(3′)-Ia* hybrid plasmids type 3	This work
	010(6)-3-1	*aroG::aph(3′)-Ia* pINV::*cat* (integrated into chromosome)	This work
	010(6)-3-2	*aroG::aph(3′)-Ia* hybrid plasmids type 1	This work
	010(6)-3-5	*aroG::aph(3′)-Ia* pINV::*cat* pRES::*bla*$_{TEM-116}$	This work
	010(6)-3-10	*aroG::aph(3′)-Ia* hybrid plasmids type 1	This work
	010(6)-3-13	*aroG::aph(3′)-Ia* hybrid plasmids type 1	This work
E. coli	SM10	pCONJ4s	(312, 313)
	QLEC4	pCONJ4s with construct for 53G chromosome *aroG::aph(3′)-Ia* insertion	This work
	QLEC6	pCONJ4s with construct for *bla*$_{TEM-116}$ insertion into pRES	This work
	MFD	pCONJ5K	(312)
		pCONJ5H	
		pCONJ5HP	
	QLEC49	pCONJ5K with construct for deletion of *nikB* on pRES	This work

	QLEC53	pCONJ5H with construct for deletion of *recA* on 53G chromosome	This work
	QLEC58	pCONJ5K with construct for deletion of *oriT* on pRES	This work
	QLEC61	pCONJ5HP with construct for insertion of *sacB* on pRES	This work
	QLEC64	pCONJ5K with construct for deletion of 199R on pRES	This work
	QLEC76	pCONJ5K with construct for deletion of 199R on pINV	This work
	QLEC90	pCONJ5K with construct for deletion of 99O on pINV	This work
	QLEC92	pCONJ5K with construct for mutation 99O^{Mut1} on pINV	This work
	QLEC93	pCONJ5K with construct for mutation 99O^{Mut3} on pINV	This work
	QLEC94	pCONJ5K with construct for mutation 99O^{Mut2} on pINV	This work

Table 2.1: Bacterial strains used in this study.

2.2 Bacterial strain construction

Genetic manipulation was performed using pCONJ-based plasmids as previously described for markerless mutations in the strains of interest (312). Vector plasmids (pCONJx) were purified from overnight liquid culture of *E. coli* strains using GenElute™ (Sigma-Aldrich), and the plasmid sequences were amplified with Phanta Max (Vazyme) according to the manufacturer's instructions to create linearised plasmid opening at the *Pme*I restriction site. The remaining circular plasmid templates in the PCR reactions were digested with *Dpn*I (New England BioLabs). The linear templates were purified using the Wizard® SV System (Promega).

Approximately 500 bp upstream and downstream of the target site were amplified with Herculase II Fusion DNA polymerase (Agilent) using primers designed to include 30 bp overhangs that were homologous to the fragments to be joined together. The amplified fragments were assembled with linearised pCONJx vector plasmid using NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs) following the manufacturer's instructions. Assembled plasmids were transformed into *E. coli* MFD *λpir*Δ*hsdR*Δ*dapA* *via* heat shock, and the transformants were selected using antibiotics (100 µg/ml carbenicillin, 100 µg/ml kanamycin, or 100 µg/ml

hygromycin), and the plasmid constructs were verified by PCR and sequencing (Source BioScience).

Conjugation was performed with *E. coli* MFD *λpir*Δ*hsdR*Δ*dapA* carrying plasmid constructs as donors, and the recipients which were the strains of interest. Both donor and recipient strains were cultured overnight in 5 ml LB with corresponding antibiotics, with *E. coli* supplemented with 300 µM DAP. 1 ml of the overnight liquid culture was centrifuged at 16,000 *xg* for one minute and the supernatant was removed. The bacteria pellets were washed and resuspended in 1 ml LB. The resuspension of both donor and recipient was mixed at a 1:1 ratio (in a 40 µl volume), dried on LBA with 300 µM DAP, and incubated at 37°C for four-hour conjugation. Conjugation mixes were harvested into 1 ml LB or TSB, and 100 µl of the suspension was inoculated onto LBA or CRA with antibiotics selecting for transconjugants with the pCONJx constructs integrated into the site of interest. Transconjugants with the insertion of plasmid constructs were verified by PCR, and resuspended in 5 ml LB with antibiotics for overnight growth at 37°C. Cultures were then washed with LB, diluted 1:20, and sub-cultured at 37°C for four hours. The sub-culture was washed with Dulbecco's phosphate buffered saline (PBS, Sigma-Aldrich), and 5 µl was inoculated onto solid media with 10% sucrose (w/v) or 4mM 4CP for counter-selection to remove pCONJx plasmid backbone. Resulting colonies were re-inoculated onto CRA with antibiotics, and mutations at the site of interest were verified by PCR and sequencing. The verified strains were stored in 40% (w/v) glycerol (Fisher Scientific™) at -70°C.

Oligo	Purpose	Sequence	Reference
***vapBC* (forward)**	Detection of the presence of pINV	GGAAACCACCGTATTTCTCA	(163)
***vapBC* (reverse)**		GGGCAAGTTCTGCTCTTATC	
***hns* (forward)**	Detection of the presence of pINV and characterise genetic events leading to emergence of CR$^-$ colonies. *hns* is the	GCTCAACAGTATGCACAGAA	
hns* *(reverse)		TTGCAAAGGCGTTGAATTA	

***virB* (forward)**	chromosomal control, *virB* and *virF* are T3SS-related genes.	ACATCAGAGCTCCACAAGAA	
virB* *(reverse)		AGACGATAGATGGCGAGAAA	
***virF* (forward)**		CTTAGCTTGTTGCACAGAGA	
virF* *(reverse)		AAGATGGGCTTGATATTCCG	
QL1	Detection of the origin and gene *repA* on pRES	TGCTGCGTTTGTGTGGTG	This work
QL2		CTGTAACTCTGAGAAATCGTGTCG	
QL5	Insertion of $bla_{TEM-116}$ into pRES	CGGAGTATAGTAAAACGACGGCCAGTGTT TCTGTCAGGGCAAGCGAAAATTG	This work
QL6		CAAATAGGGGTTCCGCGTGTCTAACAATT CGTTCAAGCCG	
QL7		CGCGGAACCCCTATTTGTTTATTTTTC	This work
QL8		CTCTACGCCGGACGCATC	
QL9		GATGCGTCCGGCGTAGAGGCTCTCTAACG CTTGAGTTAAGC	This work
QL10		CTATACTAGCAGGAAACAGCTATGACGTT TGAAGTCACCATTGTTGTGCACG	
QL11	Checking $bla_{TEM-116}$ insertion on pRES	CTTGATGAAACAACGCGGCG	This work
QL12		CAGCTTGCGACTAGATGTTGAGG	
pGL987	Detection of *aroG::aph(3')-Ia* insertion	AAAGTGGCTATCGATGCC	(312)
pGL988		TATGGCGACGAGCAGGTG	
QL27	Detection of the gene *cea*	CTCCAGTATTCATGCCCGTG	This work
QL28		CTTAACCGCCTCCAGTGC	
QL45	Deletion of *nikB* on pRES	CGGAGTATAGTAAAACGACGGCCAGTGTT TTCACGATCAAACAATACAGGAGC	This work
QL46		GAAGGAGAGTGTGAAGGAAGAGC	
QL47		GACAACCGCTCTTCCTTCACACTCTCCTTCT TATCTCAGCGTCCCATTAAGTCT	
QL48		CTATACTAGCAGGAAACAGCTATGACGTT TAAAAACGGGACAGGATGTTGG	
QL49	Checking *nikB* deletion on pRES	ATCCACTCCTCATCCCTACATTCC	
QL50		CTCTGCCGTCTGCATAAGAC	
QL55	Checking *oriT* deletion on pRES	CGGAATAATGGCATTCATCAGCA	This work
QL56		CTGAGTCTGTCTTTTTAGGTGCTG	
QL73	Detection of pRES *traU*	CCACACAGCCTGAAACACAC	This work
QL74		TGTTCTGTTGTGATGACGCC	
QL75	Detection of pRES DNA primase	ATGAATACCTGACGCGCAAG	This work
QL76		AACTGTGCATTCATGGCATTC	
QL77	Detection of pRES *pilO2*	TTTCTCTCTCGCCATGCTTATCC	This work
QL78		GTCACTGCATCCCCGGATAAAA	
QL79	Detection of pRES *pilM*	GGGTTACGGGTTTCATTATGG	This work
QL80		AGCTCAGCGATCATTTCCC	
QL81	Detection of pRES *pilP*	CTTTCCGTCCTGATCCTGCC	This work
QL82		CCGGGGGAATAATCTGTTCC	
QL83	Detection of pRES *pndAC*	AACACGTGTCAGCCGTCC	This work
QL84		AGCGGAGAAAGGCAGGG	

QL85	Detection of pRES IS*21*	GATGCAGTTCCAGAAAATCCC	This work
QL86		ATGCTGCTGATTTGATGGAGG	
QL87	Detection of 53G pINV IS*21* *(1)*	GCTGTGGGTGGCTGATTTT	This work
QL88	CS6 pRES upstream origin	GACTGTGTTCCCGCCCTTC	This work
QL89	53G pINV downstream origin	GAAGCCGATTTAGTTACAACATGC	This work
QL90	Detection of 53G pINV IS*21* (1)	CCCAGCCTTTCCAGCAATC	This work
QL95	Deletion of *recA*	CGGAGTATAGTAAAACGACGGCCAGTGTT TTTTTTGACTGACTGAGGTTGTGATG	This work
QL96		AACGCTTTCTGTTTGTTTTCGTC	
QL97		GGCTATCGACGAAAACAAACAGAAAGCGT TGATAATGTGTGCACGGACGG	
QL98		CTATACTAGCAGGAAACAGCTATGACGTT TTCTTCCATCAGATAGCCACGATAG	
QL99	Checking deletion of *recA*	ACGCCAACACCATCTTCC	This work
QL100		GGGAAGTAAAATACCGTATGCGT	
QL101	Detection of 53G pINV *relBE*	CTCAGACAGAATGATGTTAGGCAT	This work
QL102	Checking 240 kbp hybrid plasmid join near origin	CGCGGAACAACAACTACACAC	This work
QL103		TATGCACCTCCCACGCTGA	
QL104	Checking 90 kbp hybrid plasmid join near origin	CCCCTCCTCATAACTGAAAAGCG	This work
QL107	Checking hybrid plasmid join near origin	TGTTGTAACTACATCATCATCGCTG	This work
QL127	Detection of 53G pINV IS*21* (1)	GCCGGTCGTCTGTTTTGAC	This work
QL128		CAGGCAAACAGGTCATTCACTAT	
QL129	Detection of 53G pINV IS*21* (3)	CTCCCTTGTTGCGACGCA	This work
QL130		GTAACGCCGGTTCGGAAATT	
QL139	Deletion of *oriT* on pRES	CGGAGTATAGTAAAACGACGGCCAGTGTT TATGGTTTTTTCGATTCTGCTTCCC	This work
QL140		CCATTATAGCCTGTTATTTTTGTGCATC	
QL141		ATGATGCACAAAAATAACAGGCTATAATG GACGCGCACTCAGACATTC	
QL142		CTATACTAGCAGGAAACAGCTATGACGTT TATTAAGTCTCTTCGCAGAGCATTG	
pGL1341	Amplification of pCONJ5 plasmid backbone	AAACGTCATAGCTGTTTCCTGCTAGTATAG	(312)
pGL1342		AAACACTGGCCGTCGTTTTACTATACTCC	
QL68	Insertion *sacB* into pRES	CTATACTAGCAGGAAACAGCTATGACGTT TCTGCATAATTCTCTTACTGTCATGCC	This work, Zhang et al., unpublished
QL121		CGGAGTATAGTAAAACGACGGCCAGTGTT TGCCATTTATGACGACCAATGCC	
QL122		CATAGTACCTAGGACTGAGCTAGCCGTAA ATGTCTAACAATTCGTTCAAGCCG	
QL135		GAACAAGGACAATTAACAGTTAACAAATA ACTTGAACGAATTGTTAGACACGCG	
YZ03		CTAGCTCAGTCCTAGGTACTATGCTAGC	
YZ04		TTATTTGTTAACTGTTAATTGTCCTTGTTCA AGGATGC	

QL114	Deletion of the 199R on pRES	CGGAGTATAGTAAAACGACGGCCAGTGTT TAAAAAACAGCCTCATCATCTGCC	This work
QL115		CTCCGTTGTTGTTACATATGGCACAGAAAC GAGGACTTCCAGCCAGAAGA	
QL116		GTTTCTGTGCCATATGTAACAACAAC	
QL117		CTATACTAGCAGGAAACAGCTATGACGTT TGGCCAGCACCAGAAAATCAT	
QL118		GCTTCCGCACGTCTGAAG	
QL119		AGAATGACCGGAGCCTGT	
QL120		GAAGCTCTGTCATGGCTGAA	
QL126		CGGAGTATAGTAAAACGACGGCCAGTGTT TCTCCCCCTGTTTTTCACGC	
QL136		CTATACTAGCAGGAAACAGCTATGACGTT TTCGATAACCTGTCCGGATGTC	
QL185	Deletion of the 199R on pINV	CGGAGTATAGTAAAACGACGGCCAGTGTT TCCTCACAGACGGGGAATTTTG	This work
QL186		GCACGCTGCGGTTACATGTGATACCGGAG TCGATACGTCCAGCCAGAAGAC	
QL187		ACTCCGGTATCACATGTAACCG	
QL188		CTATACTAGCAGGAAACAGCTATGACGTT TCACAAACTGTTGTTGAATGGTCTTATCC	
QL189		GCATTCGCACGCCTGAAG	
QL190		CAATCACCTGATAAACCAGACGAAATC	
QL211	Deletion of the 99O on pINV	CGGAGTATAGTAAAACGACGGCCAGTGTT TCCTGTTTTTGATACAAAACACGCC	This work
QL212		CGATACGTCCAGCCAGAAG	
QL213		TCGCTGTCAGTCTTCTGGCTGGACGTATCG CTCTGTCATGGCTGAAAACGG	
QL214		CTATACTAGCAGGAAACAGCTATGACGTT TCTGATTGCTGCGTTCAGTAACC	
QL215		CGGTATATGAGGATTTTCGCTGC	
QL227	Mutation of the 99O on pINV	GACTGACAGCGATGATGATGTAG	This work
QL228		GTTGTAACTACATCATCATCGCTGTCAGTC TTCTGGCTGGACGTATCGAGTACA	
QL229		CCTGCTAAGCCATACCCGTTTTCAGCCATG ACAGAGCTTCTGTGCGCGGTC	
QL230		CATGGCTGAAAACGGGTATGG	
QL231	Long oligo for 99O^{Mut1} on pINV	CTGACGGCCCGCTAACGCGGAGATACGCC CCGACTTCAGATAAACCCTCGTCGGGACC ACTCCGACCGCGCACAGAAGCT	
QL232	Long oligo for 99O^{Mut3} on pINV	CTGACGGCCCGCTAACGCGGAGATACGCC CCGACTTCAGATAAACCCTCATTGGGACCA CTCCGACCGCGCACAGAAGCT	
QL233	Mutation of the 99O on pINV	GGCTGGACGTATCGAGTACACGCTCGTAA GCGGCCCTGACGGCCCGCTAACG	
QL234	Long oligo for 99O^{Mut2} on pINV	CTGACGGCCCGCTAACGCGGAGATACGCC CCGACTTCGGGTAAACCCTCATTGGGACC ACTCCGACCGCGCACAGAAGCT	

ori (forward)	Detection of pINV *ori*	GTGACCTCCTCAGAATAATCC	(163)
ori (reverse)		AAAAGATACATTGCACCCTGT	

Table 2.2: Primers used in this study.

2.3 Conjugations

Before conjugation, antibiotic resistance cassettes were inserted into plasmids in donors and the chromosome of recipients without plasmids as mentioned in Chapter 2.2.

Donor and recipient strains were cultured on CRA with antibiotics (20 µg/ml chloramphenicol and/or 100 µg/ml carbenicillin for selection of pINV and/or pRES in the donor, or 100 µg/ml kanamycin selecting for the recipient) from glycerol stocks. After overnight incubation at 37°C, single colonies were resuspended in 5 ml TSB with antibiotics and grown overnight. The OD_{600} of the overnight culture was adjusted to 0.05 in 30 ml fresh TSB with antibiotics and grown to exponential phase (OD_{600}~0.5) in approximately two hours. Subcultures were centrifuged at 4°C, 4000 *xg* for 10 minutes, the supernatant was removed and the bacterial pellets were washed with LB, and resuspended in 1520 µl LB to achieve the concentration at 5 x 10^9 CFU/ml. The donor and the recipient were mixed at desired donor:recipient (D:R) ratio (e.g 1:1, 0.33:1 or 3:1). Conjugation mixes (100 µl) were dried on plain LBA, incubated at 37°C for four hours to allow conjugation but prevent expansion of transconjugants which could affect the results. After incubation, the conjugation mixes were then resuspended in 100 µl TSB. The resuspension was serially diluted in PBS, and 5 µl of each dilution was spotted on CRA with antibiotics to select for recipients (100 µg/ml kanamycin) and transconjugants (100 µg/ml kanamycin and carbenicillin for selection of pRES transfer, or 100 µg/ml kanamycin and 20 µg/ml chloramphenicol for selection of pINV transfer). The transconjugants received pINV were re-inoculated on CRA with

100 µg/ml kanamycin and 20 µg/ml chloramphenicol and verified by PCR detecting *vapBC* from pINV and the *aroG::aph(3')-Ia* marker on the chromosome of the recipient.

2.4 Plasmid loss assay (PLA)

sacB gene was inserted into pRES::$bla_{TEM\text{-}116}$ to allow counter-selection of pRES-less colonies on salt-free agar with 10% sucrose (w/v) (314). The presence of *sacB* confers sensitivity to sucrose due to levansucrase (encoded by *sacB*) generating toxic products from sucrose (315). Bacteria harbouring pRES would be carbenicillin resistant but sucrose sensitive.

PLA was performed in three replicates per *S. sonnei* strain. Strains were initially cultured from frozen stocks on plain CRA at 37°C for ~30 generations of growth over 16 hours to allow pRES loss. For each replicate, three colonies from each strain were resuspended in 100 µl PBS separately, serially diluted, and inoculated on plain LBA or solid media with sucrose for quantification to the entire population and the sucrose resistant population. Colonies emerged from each initial colony on different media were counted, and 30 sucrose-resistant colonies per replicate were re-inoculated onto LBA with 100 µg/ml carbenicillin to check if the colonies still harboured pRES, as sucrose resistance could be caused by spontaneous mutations in *sacB* (316), rather than pRES loss. Colonies surviving on LBA with carbenicillin were re-inoculated back to media with sucrose to confirm the sucrose resistant phenotype. A total of 90 colonies were checked over three biological replicates for each strain. Plasmid carriage of both sucrose and carbenicillin resistant colonies was checked by PCR detecting pRES replicon. The plasmid loss frequency was calculated as the ratio of plasmid-less bacteria (confirmed by sucrose resistance, carbenicillin sensitivity and the absence of pRES replicon according to PCR detection) to the total population.

2.5 Congo Red loss assay (CRLA)

The protocol was adapted from (163) for *S. sonnei*. CRLA was used to predict the presence of *S. sonnei* pINV, by observing the CR binding of the colonies in the presence of the T3SS on pINV, allowing differentiation between virulent and avirulent bacteria (311). CR^+ colonies indicated the presence of the T3SS, which indirectly suggested the presence of pINV in the colonies; CR^- colonies indicated the absence of the T3SS, which could either be caused by the deletion of the T3SS or the loss of the entire pINV (163). Although variability in the red coloration of the colony can be observed visually, it cannot be reliably quantified. Therefore, colonies exhibiting a red region (normally in the center of the colonies) are categorised as CR^+ with the assumption that they originated from bacteria that initially harboured pINV, which may have been lost during early growth of the colony and expanded in the later stage of the growth, resulting in a "white ring" surrounding a "red center".

For CRLA, *S. sonnei* was grown on CRA without antibiotics from glycerol stocks at 37°C for 16 hours to allow pINV loss events. For each replicate, three Congo Red positive (CR^+) colonies from each strain were resuspended in 100 µl PBS separately, serially diluted, and 80 µl of the 10^{-5} dilution was inoculated onto plain CRA and incubated at 37°C overnight. CR^+ and CR^- colonies emerged from each initial colony were counted, and the presence of pINV in 10 CR^- colonies was checked by PCR for *vapBC* and/or by multiplex PCR detecting *hns* (chromosomal control), *virB* and *virF* (regulators of the T3SS). A total of 30 CR^- colonies per replicate were checked. The frequency of pINV loss was calculated as the proportion of the plasmid-less bacteria within the entire population.

2.6 Multiplex PCR

The protocol was adapted from (163) to check the presence of pINV and understand the genetic events leading to the emergence of avirulent bacteria (44). Bacterial colonies were inoculated

into 100 µl nuclease-free water, boiled at 95°C for five minutes. Primer mix (100 µl) contained 50 µM each *hns* primer, 75 µM each *virB* primers, and 25 µM each *virF* primers. Q5® High-Fidelity DNA Polymerase (New England Biolabs) was used with 1 µl primer mix and 1 µl bacterial boilate in a reaction volume of 50 µl, with a primer annealing temperature of 51.2°C and extension time of one minute fifty seconds. 1 kb Hyper ladder (Bioline) was used for gel electrophoresis with 1 in 10,000 SYBR Safe DNA gel stain (w/v, Invitrogen), at 110V for 30 minutes.

2.7 Extraction of large plasmids

The protocol was adapted from (317). Bacterial strains were grown on CRA with antibiotics from glycerol stocks at 37°C overnight. Colonies were collected from the solid media in a 2.0 ml tube, resuspended in 200 µl nuclease-free water, centrifuged at 16,000 *xg* for two minutes, and the CR layer on the top of the bacterial pellet was removed. Bacterial pellets were resuspended in 100 µl resuspension buffer (50 mM glucose, 10 mM EDTA, 10 mM Tris-Cl, pH 8). Lysis solution (200 µl, 0.2 M NaOH, 1% SDS) was added, mixed by inversion, and incubated at room temperature for five minutes. Ice-cold ammonium acetate (150 µl, 7.5 M) and chloroform (150 µl) were added simultaneously, mixed by inversion of the tubes, and incubated on ice for ten minutes. The lysate was centrifuged, cleared lysate transferred to 200 µl precipitation solution (30% polyethylene glycol 8000, 1.5 M NaCl), mixed by inversion, and chilled on ice for 15 minutes. The mixture was centrifuged, supernatant removed, and the DNA pellet washed with 600 µl 70% ethanol. Plasmid DNA was eluted in 50 µl EB buffer (Qiagen), incubated at 65°C for ten minutes, and treated with 1 µl RNase A (Promega) and 1 µl proteinase K (Promega) at 37°C for an hour.

Plasmid DNA samples (15 µl) and 6 µl loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol, w/v) were mixed and loaded on 0.5% agarose gels (Invitrogen) with 1:10,000 SYBR Safe (Invitrogen). Electrophoresis was performed at 4°C, 20V for 16 hours. The

gel was stained again with SYBR Safe and imaged using the Gel Doc XR+ Gel Documentation System (Bio-Rad).

2.8 gDNA extraction

Bacterial strains were grown on CRA/LBA or in TSB/LB with or without antibiotics overnight. gDNA was extracted using the Wizard® HMW DNA Extraction Kit (Promega) or QIAamp DNA Mini Kit (Qiagen) on a QIAcube (Qiagen) per manufacturers' instructions. gDNA samples were eluted in molecular-grade water or EB buffer (Qiagen). The quantity and quality of gDNA were assessed by Qubit (ThermoFisher) and Nanodrop (ThermoFisher).

2.9 Oxford Nanopore Technology (ONT) sequencing

gDNA samples were barcoded using the 96-Rapid Barcoding Kit (SQK-RBK110.96, ONT) or the Rapid Barcoding Kit (SQK-RBK004, ONT). The sequencing libraries were concentrated using SPRI beads (Mag-Bind TotalPure, Omega). The flow cells R9.4 and the sequencing libraries were prepared using the Flow Cell Priming Kit (EXP-FLP002, ONT). The sequencing was performed with MinION for three or 72 hours, using MinKnow with live basecalling and demultiplexing enabled, generating fast5 and fastq files as output for further analyses (318).

2.10 Bioinformatics analysis

Paired-end reads (fastq) from demultiplexed ONT sequencing data were assembled *de novo* using Flye (v2.9.2) with a minimum overlap of 5 kb. For combined ONT and Illumina sequencing data (> 30x coverage, MicrobesNG), Unicycler was used. BAM files from demultiplexed Pacific Biosciences (single molecule real time (SMRT) long read sequencing, Centre for Genomic

Research, University of Liverpool) data were converted to fastq using bam2fastq in SMRTTOOLS and assembled with Flye. Prokka (v1.14.5) was used for gene annotation.

Plasmids were characterised using staramr (v0.9.1) and oriTfinder (319). Mobilisation genes were typed using mob-typer (320) at usegalaxy.eu, including replicon typing. Gene analysis was performed using BLAST, and plasmid sequences were visualised with SnapGene (v7.0.2) and aligned using ClustalOmega (321). Nucleotide alignment was also performed using minimap2 and visualised by IGV viewer (322, 323).

For phylogenetic analysis, sequences were extracted from fasta files containing complete plasmid sequences using BLAST. Sequences were aligned with Prank (324) and analysed by Raxml-ng (325). Trees were generated in RStudio (2024.04.2+764) with R 4.4.1, using packages readxl, tidyverse, ape, ggplot2, ggtree, cowplot, RColorBrewer, phytools, treeio, readr, and phylobase, to show plasmid classification.

2.11 Statistical analysis

Data were analysed using GraphPad Prism 10. For each experiment, Shapiro-Wilk tests were used to assess the normality of the data. If the data were normally distributed, parametric tests were performed further to assess the differences between the data; if the data were not normally distributed, non-parametric tests were performed instead. Differences between two variables were determined using unpaired, two-tailed t-test or Mann-Whitney test. One-way ANOVA or Kruskal-Wallis tests with multiple comparisons were used to compare three or more conditions. A p-value < 0.05 was considered statistically significant.

3 Characterisation of plasmids in *S. sonnei* CS6

S. sonnei pINV is non-conjugative due to the lack of a complete conjugation system and an *oriT* on the plasmid (169). It should also be non-mobilisable, however, was shown to be mobilised by conjugative plasmids in the laboratory (169). As resistance plasmids were detected at a high prevalence in *S. sonnei*, one aim of work in this thesis was to investigate the ability of naturally-occurring resistance plasmids to mediate HGT of pINV (10, 62).

A Brazilian clinical isolate *S. sonnei* CS6 was identified to carries multiple plasmids (20). I utilised bioinformatic tools to analyse short and long read nucleotide sequences for CS6, and revealed that it harbours multiple plasmids including pINV, a plasmid that confers AMR, and a colicin-encoding plasmid. I compared the plasmids in CS6 with well-characterised Lineage III isolate *S. sonnei* Ss046, initially isolated from China, and Lineage II laboratory strain *S. sonnei* 53G, initially isolated from Japan, to gain better understanding of plasmids in the strain (84, 237). I also characterised the plasmids using different assays to investigate their maintenance, including optimised CRLA (for pINV) and PLA (for pRES), and their horizontal transfer using conjugation assays. The resistance plasmid pRES from CS6 was used as a model in *S. sonnei* 53G to investigate its ability to mediate horizontal transfer of pINV and characterise the mechanism(s).

3.1 Bioinformatic analysis of *S. sonnei* CS6 pINV, pRES, and pColE1

As determined by core genome sequence analysis, *S. sonnei* CS6 belongs to Lineage III which is associated with AMR around the world (20). The strain was subjected to long read SMRT sequencing using Pacific Biosciences Sequel® II, and the sequence assembled using Flye (326). Further assembly was performed using Unicycler with short read DNA sequences available for this isolate (accession no. ERP000182) (20, 327). Putative open reading frames (ORFs) were predicted by Prokka (328), then I manually annotated each ORF using BLASTx.

The assembly of the sequence revealed that *S. sonnei* CS6 harbours at least three plasmids: pINV (227,733 bp), a 114,182 bp resistance plasmid which was designated pRES, and a 5,153 bp plasmid encoding colicin E1, named pColE1.

3.1.1 *S. sonnei* CS6 pINV backbone is highly conserved and the plasmid is predicted to be non-mobilisable

S. sonnei pINV belongs to the IncFII replicon group as determined by PlasmidFinder (235). Comparing CS6 pINV to corresponding plasmids in *S. sonnei* 53G (Lineage II, accession no. NC_016833) and *S. sonnei* Ss046 (Lineage III, accession no. CP000039.1) (84), pINV is highly conserved with > 99.9% nucleotide identity over the T3SS PAI and the rest of the plasmid (Figure 3.1.1A). All three pINVs carry identical maintenance systems, including VapBC and RelBE TA systems, and a ParAB partitioning system located at the same position in all plasmids (44) (Figure 7.1). It has been previously shown that ISs account for ~ 23% of all ORFs on *S. sonnei* pINV (44). ISs on CS6 pINV were analysed by performing BLASTn manually against the ISfinder database using a maximum E-value of 1e-100 (239). CS6 pINV harbours ISs from nine families (*i.e.* IS*21*, IS*91*, IS*110*, IS*4*, IS*3*, IS*66*, IS*630*, IS*5* and IS*1*); these elements are highly conserved across pINV from the three strains (Figure 3.1.1A). However, some differences were observed. The 2,638 bp IS*21*-family IS*Sso4* on pINV from *S. sonnei* CS6 (155,779 bp – 158,416 bp) differs from that on *S. sonnei* Ss046 and 53G pINV by a few SNPs, as determined by ClustalOmega (Figure 7.2). The lack of a copy of IS*4* (195,304bp – 196,632bp) on CS6 pINV which is present in Ss046 and 53G pINV also demonstrates the diversity of ISs on pINV.

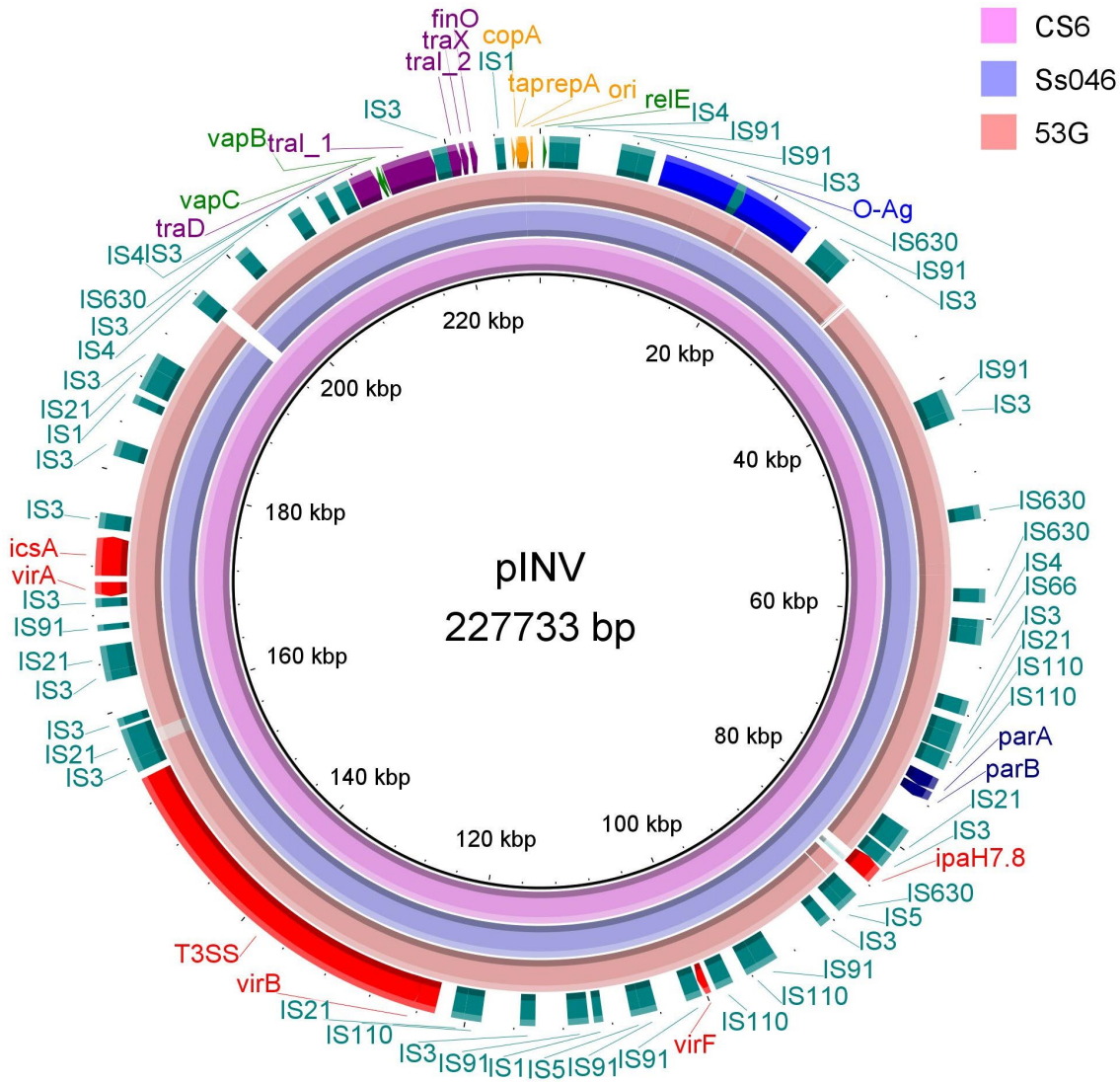


Figure 3.1.1A: Alignment of pINV from *S. sonnei* CS6, Ss046 and 53G.
Figure created using BRIG (329) with BLAST nucleotide identity > 90% aligning plasmids against the reference plasmid CS6 pINV, indicated by the concentric rings. Gaps in the alignment indicate nucleotide identity below 90% when compared to the reference. Innermost ring (pink), CS6 pINV; middle ring (blue), Ss046 pINV; outermost ring (coral), 53G pINV. Annotation of important elements of pINV are colour-coded: the replicon, orange; VapBC and RelBE, green; ParAB, navy; the T3SS and virulence genes, red; ISs, teal; conjugation-related genes, purple. ISs were annotated manually according to ISfinder.

Relevant to HGT, three genes from a MOB_F-type conjugation system are located next to *vapBC* on pINV from these strains, which encode a potential TraD coupling protein, a TraI relaxase, and a TraX pilin acetylase, which are essential for conjugation (Figure 3.1.1A) (170, 172, 221). Comparative analysis using ClustalOmega reveals that the *traD-vapBC-traI-traX* region from CS6, Ss046 and 53G pINV share > 95.9% nucleotide identity with the region from *E. coli* F plasmid which is conjugative (accession no. AP001918, Figure 3.1.1B). TraD exhibits high conservation

of amino acid sequences (> 99% identity) across the three pINVs, and shares 98.6% amino acid identity with F plasmid TraD (Figure 7.3A). The amino acid differences are positioned in their C-termini, consistent with this protein being potentially non-functional as the C-terminal region is involved in protein-protein interactions with the relaxosome factor TraM during conjugation (221).

F plasmid TraI contains three functional domains essential for conjugation, including a relaxase domain in the first ~310 amino acids for cleavage of plasmid DNA at *oriT*, a helicase domain between amino acids 990 and 1450, and a functional domain from amino acid 1450 to the C-terminal which is hypothesised to be involved in interactions with the T4SS apparatus during conjugation (330, 331). Notably, multiple nucleotide mutations in *traI* on CS6 and Ss046 pINV introduce a stop codon at 1428 instead of Glycine, and *traI* is interrupted at nucleotide 4,340 bp by IS*600*. This led to duplication of a 110 bp nucleotide sequence (4230 bp – 4340 bp), corresponding to amino acids 1411 to 1447 of the ORF when compared with F plasmid TraI, potentially affecting the helicase domain. While 53G pINV *traI* is not interrupted by an IS, it also harbours point mutations leading to amino acid changes compared to F plasmid TraI, which potentially impact its function (Figure 7.3B) (330).

The amino acid sequences of TraX from the three pINVs are identical. Mutations and deletions in *traX* when compared to that on F plasmid result in a premature stop codon at amino acid 194, splitting the ORF into two ORFs, and likely rendering it non-functional (Figure 7.3C).

Consequently, TraD, TraI and TraX in these *S. sonnei* strains are probably inactive due to multiple mutations and/or disruption when compared to homologues on the F plasmid, consistent with pINV being non-conjugative (84, 169). Additionally, the absence of an *oriT* on these pINV determined by oriTfinder (319) indicates that the pINV is unlikely to be mobilisable by a co-resident conjugative plasmid.

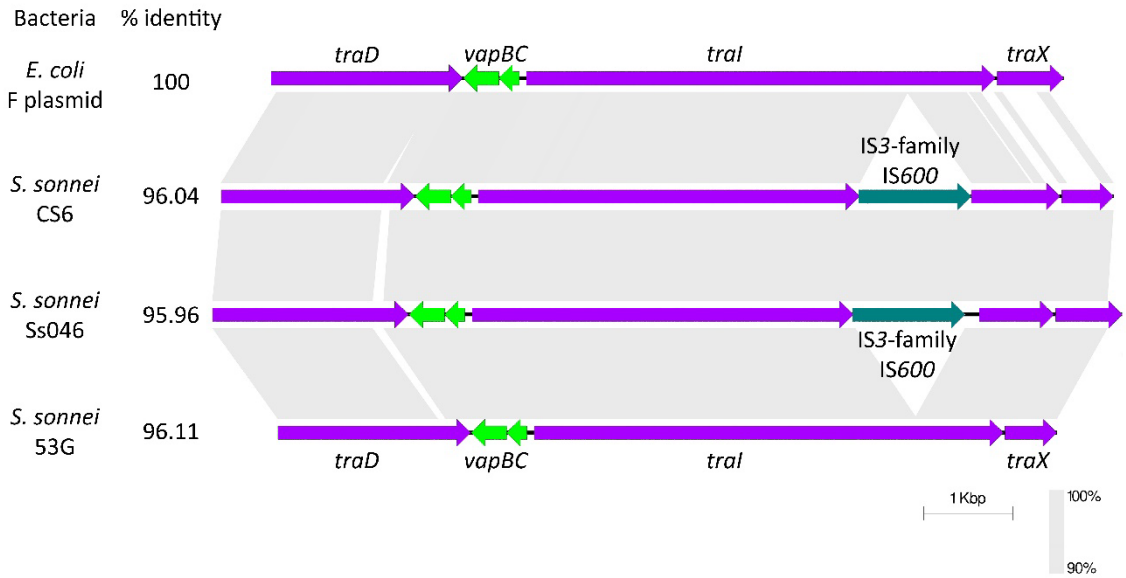


Figure 3.1.1B: Nucleotide alignment of *traI-vapBC-traD-traX* regions from *E. coli* F plasmid, *S. sonnei* CS6 pINV, Ss046 pINV and 53G pINV.

Nucleotide sequence alignment was generated by Easyfig (332), with nucleotide identity of > 90% shown in grey shading. Gaps indicate nucleotide identity of < 90%.

3.1.2 Homologues of partial conjugation genes on *S. sonnei* pINV and chromosome

In addition to the remnants of a conjugation system on pINV, a partial homologue of *traI* (177,974 bp – 178,968 bp) from 53G pINV (accession no. NC_016833) was found on the *S. sonnei* 53G chromosome (accession no. HE616528) with two SNPs identified (C4305T and G4861A); the SNPs result in a synonymous mutation and an amino acid change (G1621V), respectively. The chromosome harbours a region containing genes from a MOB_F-type conjugation system (2,964,538 bp – 2,969,667 bp, Figure 3.1.2A). The region is flanked by IS*3*-family IS*600* at the 5'-end (2,962,609 bp – 2,963,872 bp) and IS*629* at the 3' end (2,969,740 bp – 2,971,045 bp), containing *finO*, *traX*, *traI* (3'-end, 4,276 bp – 5,721 bp), *traW* (5'-end only, 1 bp – 310 bp), *trbI*, *traC* (3'-end only, 1,356 bp – 2,628 bp), *traL* and *traA* (Figure 3.1.2A). The rest of the conjugation system was absent and this entire region was also present on the *S. sonnei* CS6 chromosome (950,313 bp – 955,442 bp).

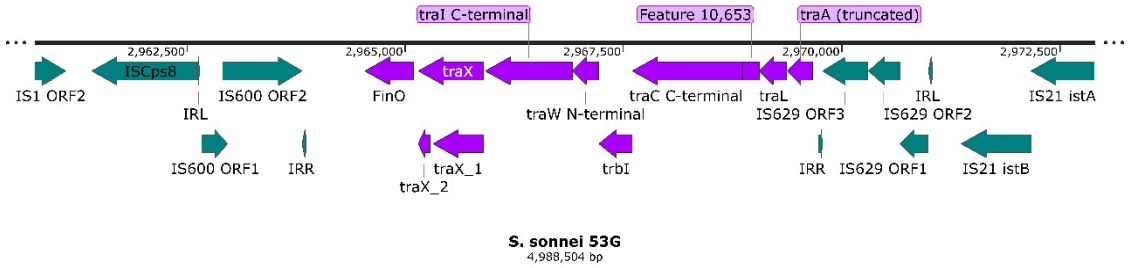


Figure 3.1.2A: The conjugation system-related genes on the *S. sonnei* 53G chromosome.
The figure was exported from Snapgene. Each feature represents the genes as verified by BLASTx. The conjugation system-related genes are in purple and the ISs are in blue. The region is also found on the *S. sonnei* CS6 chromosome.

The arrangement of the conjugation genes on the 53G chromosome is different from F plasmid (66,118 bp - 99,159 bp) (Figure 3.1.2B). The 'Feature 10,653' (184 bp) annotated on the 53G chromosome (Figure 3.1.2A) contains nucleotides from the stop codon of *traL* to 1 - 163 bp of *traE*, which is immediately followed by the 3'-end of *traC*. The 5'-end of *traW* is immediately followed by the 3'-end of *traI*. This suggests that an intact conjugation system was integrated

into the chromosome and lost over time due to genetic rearrangements including deletions. Interestingly, the *traD-vapBC-traI-traX-finO* region on 53G pINV also has IS*600* (169,397 bp - 170,660 bp) upstream of *traD*. Compared to the region on the 53G chromosome, the arrangement of IS*600* is different in the region on pINV (Figure 3.1.2C). Overall, these findings indicate that there was an intact conjugation system on *S. sonnei* pINV and the chromosome, following genetic exchange potentially mediated by ISs. The genes are now predicted to be inactive through degradation and/or loss.

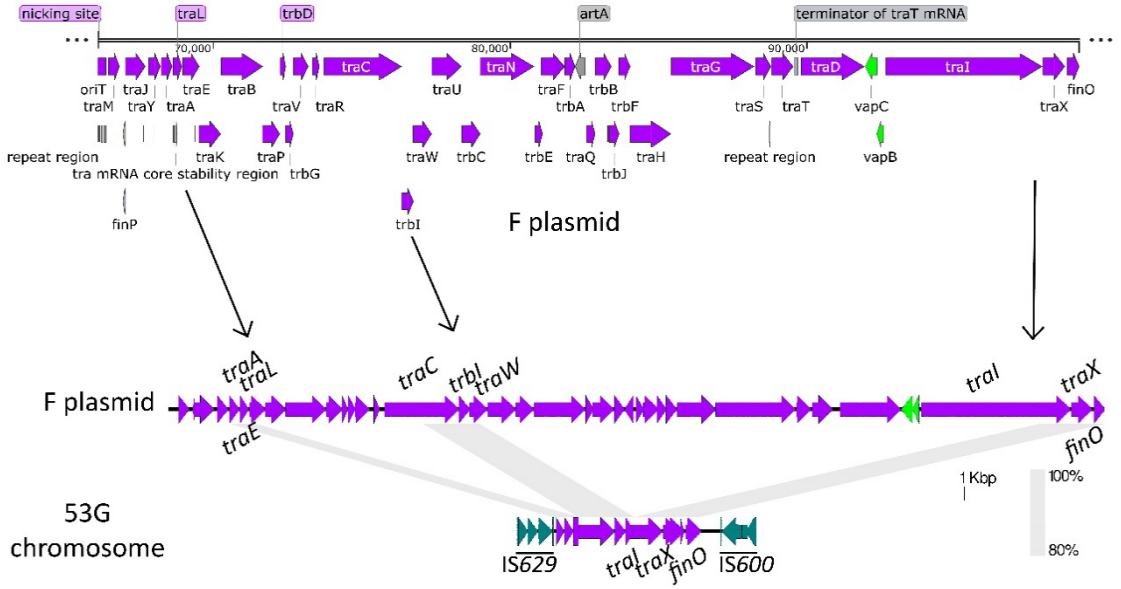


Figure 3.1.2B: Comparison of conjugation system-related genes on the *S. sonnei* 53G chromosome and *E. coli* F plasmid.
The figure was created using Easyfig, and regions with a nucleotide identity of > 80% were aligned (shown in grey).

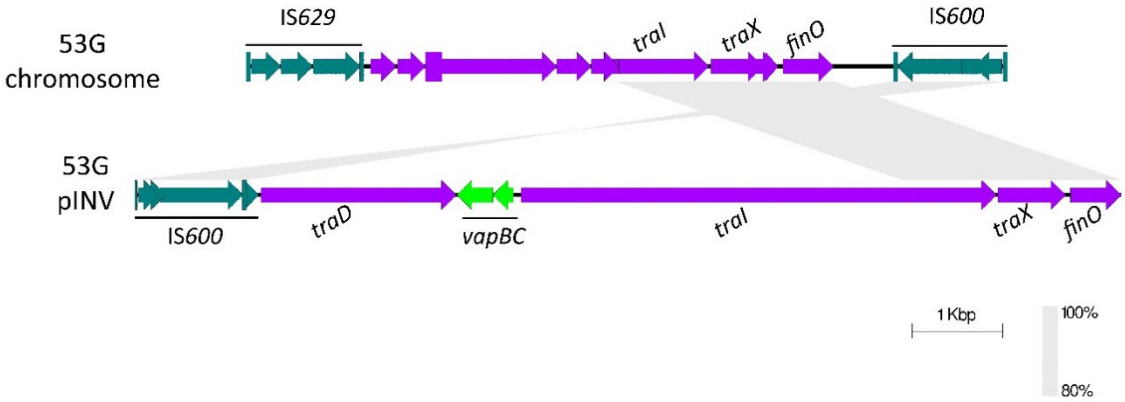


Figure 3.1.2C: Comparison of conjugation system-related genes on *S. sonnei* 53G pINV and the chromosome.
The figure was created using Easyfig, and regions with a nucleotide identity of > 80% were aligned (shown in grey).

3.1.3 *S. sonnei* CS6 pRES harbours multiple AMR genes and an intact conjugation system

In addition to pINV, CS6 also harbours a 114 kbp plasmid pRES (Figure 3.1.3A), which is not present in *S. sonnei* 53G or Ss046. pRES belongs to the IncB/O/K/Z I-complex Inc group (accession no. CU928147) as determined by PlasmidFinder and mobtyper (235, 236, 320). BLAST and ResFinder (333) indicate that pRES harbours multiple genes encoding resistance to antimicrobials (39,687 bp – 62,407 bp). A defective In*2* was identified, which is embedded in a disrupted Tn*7* transposon (accession no. DQ176450), a novel In*2* arrangement detected in *Acinetobacter baumannii* (334); the presence of In*2* is a typical feature of Lineage III *S. sonnei* (20). The In*2* contains *dfrA* (conferring trimethoprim resistance), *sat2* (streptothricin resistance), and *aadA1* (streptomycin resistance). These genes are flanked by *intI2* encoding a predicted integrase (40,623 bp – 41,600 bp) and Tn*7*-*ybfB* (44,908 bp – 45,231 bp) encoding for a hypothetical protein. Tn*7* is disrupted by insertion of Tn*10*-associated *tetB* (tetracycline resistance) and *tetR* (repressor of tetracycline resistance); this insertion could be caused by IS*4*-family IS*Vsa5* mediated transposition (45,241 bp – 52,890 bp). Downstream of this region, there is *sul2* (sulphonamide resistance, 59,529 bp – 60,344 bp) flanked by IS*Vsa3* (58,001 bp – 58,977 bp) and IS*3*-family IS*2* (61,077 bp – 62,407 bp).

pRES carries four potential plasmid maintenance systems. These include two TA systems, RelBE (2,488 bp – 3,041 bp, downstream the plasmid replicon) and PndAC (28,463 bp – 29,109 bp), as well as a putative type-II ParMRC partitioning system (13,017 bp – 14,406 bp), and a putative ParB-family partitioning protein (23,225 bp – 25,183 bp). pRES RelBE system shows 46.1% nucleotide identity with the related system on *S. sonnei* pINV, which has been shown previously to be functional (Figure 7.4A) (163). In pINV and pRES, the *relBE* systems are located close to the origin of replication. A comparative analysis was performed to examine the amino acid sequences of the two RelBE systems, with low similarity observed in both the toxins and antitoxins, which share 16.7% and 13.6% amino acid identity, respectively (ClustalOmega, Figure7.4B).

pRES carries a MOB_P type conjugation system predicted by mobtyper, and this was verified by manual annotation using BLASTp. oriTfinder identified an *oriT* on pRES which is present upstream of *nikA* and *nikB* which encode a potential relaxase consistent with pRES being a conjugative plasmid (202) (Figure 3.1.3A).

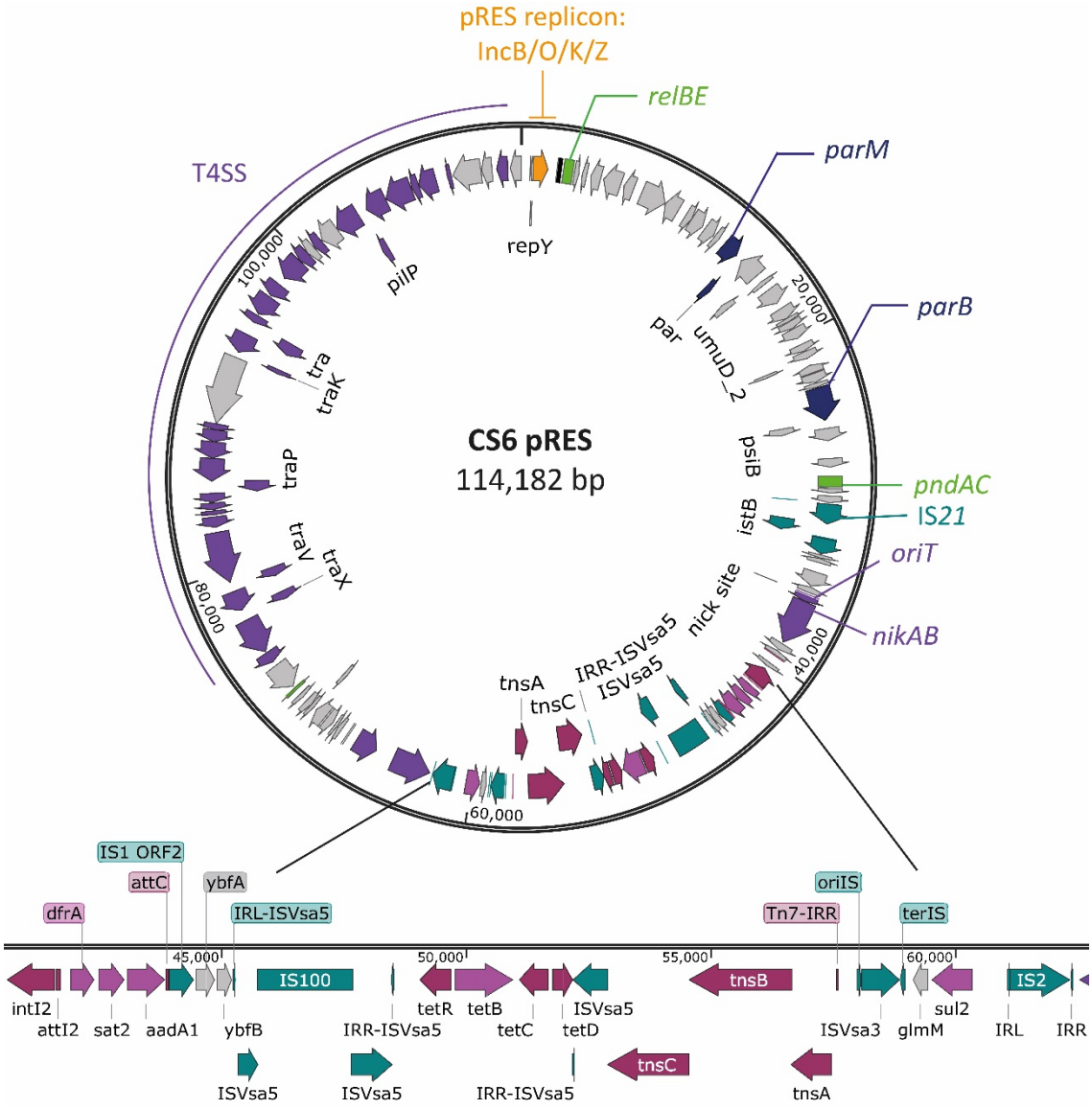


Figure 3.1.3A: Map of pRES.
Elements of pRES are colour-coded: the replicon, orange; TA systems, green; partitioning system, navy; T4SS and associated genes, purple; antimicrobial resistance genes (ARGs), fuchsia; ISs, teal; transposons and integron, plum. The expanded region shows the In*2* which is first detected in *A. baumannii* (334).

To investigate the prevalence of pRES-type plasmids in *S. sonnei*, a collection of 252 complete circularised plasmid sequences was provided by Dr. Gabriele Arcari (La Sapienza, Roma) from

the plasmid database PLSDB, which includes circularised plasmid sequences from the NCBI database with additional meta information (335, 336). I manually excluded sequences of pINV, phage plasmid sequences, and sequences of plasmids of < 40 kbp size as larger plasmids are more likely to carry conjugation systems (205). A total of 134 plasmid sequences were analysed by Mobtyper and PlasmidFinder to predict their mobilisation ability and Inc group, respectively. Mobtyper predicted that 131 out of 134 plasmids (97.8%) are conjugative, one is mobilisable, and two are non-mobilisable. The Inc groups of these plasmids were identified using PlasmidFinder, with a particular focus on IncB/O/K/Z-related plasmids, *i.e.* those with a replicon related to the pRES replicon. Results indicate that I-complex plasmids (IncI1, IncB/O/K/Z, IncI2) are the most commonly detected plasmids, accounting for 81 of 134 plasmids (60.5%) in the collection (Figure 3.1.3B). Among the I-complex plasmids, 30 were of the IncB/O/K/Z type, ranging from 89 kbp to 111 kbp in size. Within this group, 14 plasmids belonged to the same sub-group as pRES, with 13 predicted to be conjugative. These plasmids were isolated from different continents, including Europe (n=1), Asia (n=7), and the country of origin not known for the others (n=6) (Figure 3.1.3B). IncFII-like plasmids are also commonly found in *S. sonnei*, and account for 37.3% of plasmids in the collection.

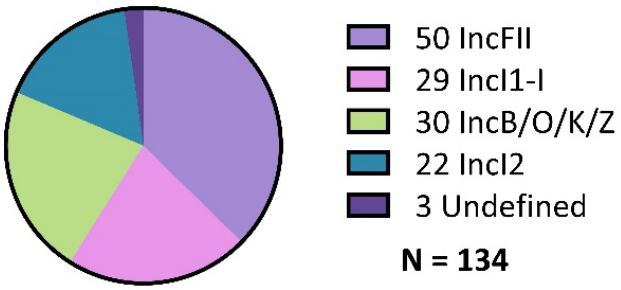


Figure 3.1.3B: Frequency of plasmids of belonging to different Inc groups in a collection of 134 *S. sonnei* plasmids.
Legend shows the number of plasmids belonging to each Inc group.

3.1.4 *S. sonnei* CS6 harbours pColE1 and other small plasmids

Alignment of nucleotide sequences reveals that CS6 also harbours a pColE1 which shares > 99% nucleotide identity with a 5,153 bp plasmid spB in *S. sonnei* Ss046 (accession no. CP000642) (Figure 3.1.4A). This plasmid contains genes *cea* and *imm*, encoding the bacteriocin ColE1 and a cognate immunity protein, respectively, followed by *lys* encoding a lysis protein for releasing ColE1 (226). The *cea-imm-lys* operons of CS6 pColE1 and Ss046 spB are identical (BLASTn), with several SNPs identified in the backbone of the plasmids. Compared with pKY1, a multi-copy ~4.7 kbp non-mobilisable ColE1-like plasmid from *S. sonnei*, the *cea-imm-kil* operon (accession no. M37218) is highly homologous with the corresponding operon on the plasmid in CS6 and Ss046 (337); the predicted amino acid sequences of *lys* from CS6 and Ss046 pColE1 are identical to *kil* from pKY1 (Figure 7.5A). The ColE1 encoded by *cea* on pKY1 is active and shares 94.1% amino acid identity to *cea* on CS6 pColE1 and Ss046 spB (Figure 7.5B). Imm encoded by pKY1 shares 86.6% amino acid identity compared with the versions expressed by the other two plasmids (Figure 7.5C).

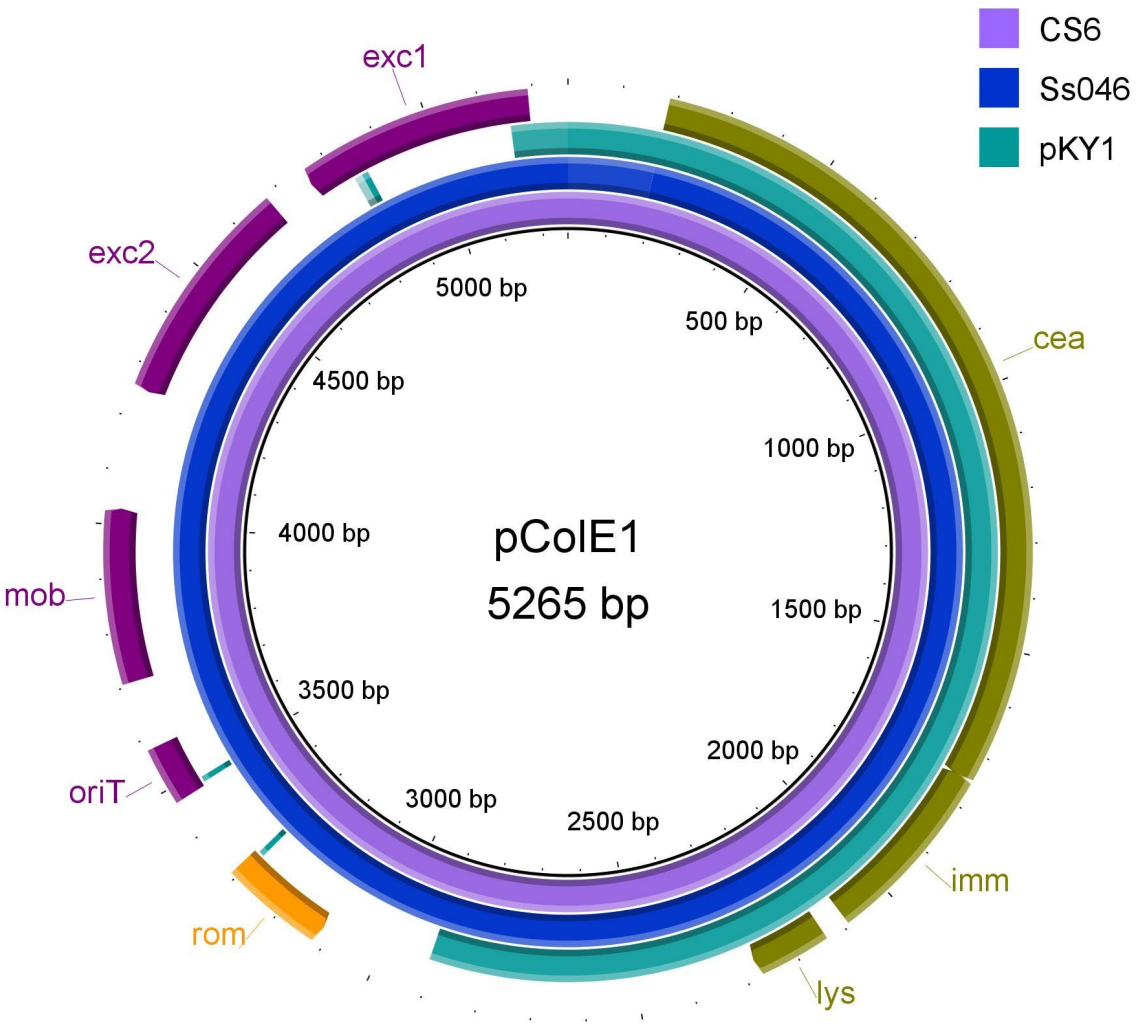


Figure 3.1.4A: Alignment of pColE1 from *S. sonnei* CS6 and Ss046, and the *cae-imm-kil* operon from pKY1.
Figure generated by BRIG using CS6 pColE1 as the reference, with a nucleotide identity of > 90% by BLAST. The rings represent alignment of plasmids. Colour intensities and gaps in the alignment indicate decreasing nucleotide identity from 100% to 90% compared to the reference. Innermost ring (violet), CS6 pColE1; middle ring (navy), Ss046 pColE1 (spB); outermost ring (green), the operon from pKY1. Annotations of pColE1 are colour-coded: the replication modulatory protein Rom (modulator of inhibition of primer formation by RNAI) (338, 339) in orange, conjugation-related genes in purple, and the *cea-imm-lys* operon in olive.

pColE1 also harbours an *oriT* and MOB_F-type mobilisation genes (i.e. *mob*, encoding mobility protein, relaxase; *exc*, encoding plasmid entry exclusion protein) (226) as determined by mobtyper. Together these features suggest that pColE1 is mobilisable and could be transferred in the presence of an active conjugation system.

According to sequencing data from the NCBI database, *S. sonnei* 53G also harbours pColE1 (5,153 bp, accession no. HE616530), with a *cea* gene sharing 99.8% nucleotide identity with CS6 *cea* (Figure 7.6). Using primers (QL27 and QL28) to detect *cea*, a PCR product of 601 bp was amplified from CS6 but not 53G (Figure 3.1.4B), even though the sequencing data indicates that *cea* should be present in *S. sonnei* 53G.

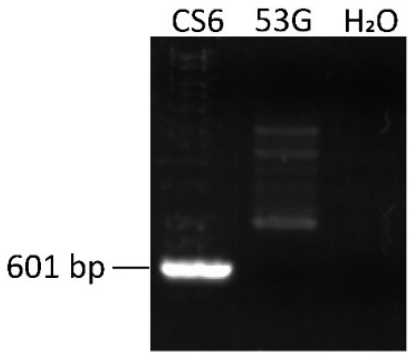


Figure 3.1.4B: PCR detection of *cea* in *S. sonnei* CS6 and 53G.
A 601 bp product was amplified from the boilate of *S. sonnei* CS6 but not 53G. The size of the band is indicated. H_20, water control.

Therefore, plasmid DNA was extracted from these two strains and visualised by gel electrophoresis. Results demonstrate that 53G and CS6 have a different repertoire of small plasmids (Figure 3.1.4C). The ~5 kbp band corresponding to pColE1 was seen in CS6 but surprisingly not 53G, even though 53G is expected to harbour a related plasmid based on sequencing data performed by others (accession no. HE616530). Instead, *S. sonnei* 53G in our laboratory harbours a ~3.5 kbp plasmid which is not predicted by previous sequencing data. One hypothesis is that this plasmid is a remnant of pColE1. Additionally, a smaller plasmid (2,089 bp, accession no. HE616531) was present in *S. sonnei* 53G but not CS6. This plasmid is highly homologous (98.9% nucleotide identity) to *S. sonnei* pKYM (2,083 bp, accession no. NC_001378) which was previously shown to be non-mobilisable (340) (Figure 7.7). There is also a band corresponding to a ~3kb unknown plasmid found in CS6, which was not detected by long read sequencing. These results demonstrate that the version of *S. sonnei* 53G in our laboratory does not harbour pColE1, and there are several cryptic plasmids in both 53G and CS6 which have not been characterised.

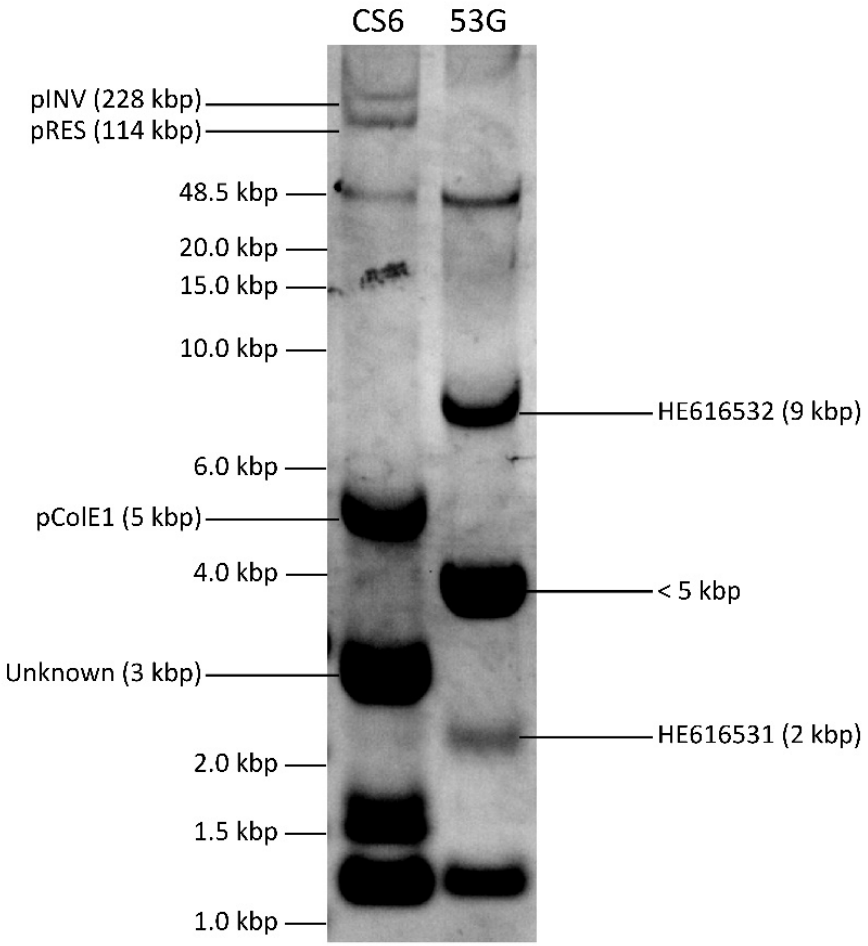


Figure 3.1.4C: Visualisation of plasmids in *S. sonnei* CS6 and 53G by gel electrophoresis.
Plasmid DNA was extracted from bacteria grown on CRA overnight. Sizes of plasmids and their predicted identity are labelled. Plasmids < 2 kbp were not annotated.

3.2 Vertical and horizontal transfer of *S. sonnei* plasmids

Following bioinformatic characterisation of the plasmids in *S. sonnei* CS6, I next sought to understand whether the maintenance of the plasmids is affected by the presence of other plasmids. Therefore, different plasmid loss assays were optimised to assess loss of pINV and pRES.

In addition, the horizontal transfer of plasmids by conjugation was investigated. Based on bioinformatic analysis, CS6 pRES is predicted to be conjugative due to its intact conjugation system with an *oriT* and *nikAB* relaxase (Figure 3.1.3A). A pColE1 harbours an *oriT* and *mob* encoding a potential relaxase which indicate that it could be mobilised by pRES. Therefore, the ability of pRES to mediate its own conjugation and facilitate transfer of other plasmids was investigated. The impact of other plasmids on pRES conjugation frequencies was also explored.

3.2.1 pINV and pRES maintenance in *S. sonnei* is not affected by plasmid co-existence

The ability of *Shigella* spp. to bind the dye CR during growth on solid media has been used to detect the expression of the T3SS and indirectly indicate the presence of *Shigella* pINV (163). CR^+ bacteria express a T3SS, while CR^- bacteria do not. As the T3SS is encoded on pINV, the emergence of CR^- colonies may occur through the loss of pINV, the commonest cause of S. sonnei CR^- colonies (163). To assess pINV loss in *S. sonnei* 53G and CS6, a CRLA was performed following a protocol adapted from Pilla *et al.* (2017) for *S. flexneri* (163). In this method, bacteria are grown from frozen stocks on CRA. During colony growth from a single bacterium, plasmid loss can occur. Since CR^- bacteria appear infrequently in *S. flexneri*, single CR^+ colonies are selected and regrown in liquid media for 16 hours. During this period, further plasmid loss may occur. Bacteria are then inoculated after this sub-culture and the number of CR^+ and CR^- colonies enumerated. CRLA detects plasmid-less populations emerging from the initial growth but does not quantify subsequent plasmid loss during clonal expansion.

S. sonnei 53G and CS6 were cultured on plain CRA from frozen stocks at 37°C for 16 hours to allow plasmid loss. Three single CR^+ colonies, *i.e.* predominantly $pINV^+$, were selected from each strain and cultured overnight in liquid TSB as for plasmid loss assays with *S. flexneri*. The overnight cultures were normalised to OD_{600}=1, then serially diluted in PBS. A 50 µl aliquot of the 10^{-5} dilution was spread on plain CRA and incubated at 37°C for 16 hours to quantify the number of CR^+ and CR^- colonies. To establish whether pINV loss had occurred, eight CR^- colonies per replicate were analysed by PCR for *vapBC* on pINV and by multiplex PCR for chromosomally-encoded *hns* and pINV-encoded *virB* and *virF*, regulators of T3SS expression. Plasmid loss was demonstrated by PCR through the absence of *vapBC, virB*, and *virF*. The frequency of pINV loss (Figure 3.2.1A) was quantified based on PCR analysis of CR^- colonies and shown as the number of bacteria that had lost pINV as a proportion of the total population. Therefore, a pINV loss frequency of 0.1 means that 10% of the population have lost the pINV.

Results indicate that CS6 and 53G pINV are lost at a frequency of 4.0×10^{-2} and 9.0×10^{-3} during laboratory culture, respectively (Figure 3.2.1A). CS6 pINV loss is significantly less than that of 53G ($p = 0.0397$). Given that plasmid loss from different genetic backgrounds was assessed, the loss of pINV was next investigated in an isogenic background using *S. sonnei* 53G in the presence and absence of CS6 pRES.

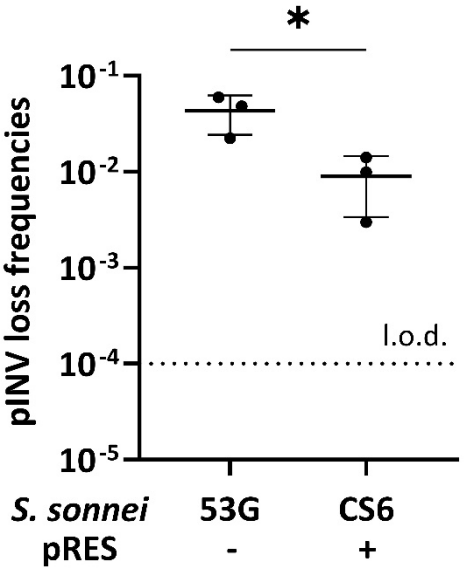


Figure 3.2.1A: pINV is lost more frequently from *S. sonnei* 53G than CS6.
Each dot represents the averaged proportion of CR^- bacteria in a population following sub-culture from three colonies in a replicate. Shapiro-Wilk tests were performed to check the normal distribution of the data. Unpaired t-test was used to compare data from two groups. *, $p = 0.0397$. The limit of detection (l.o.d.) of 10^{-4} was indicated. (N=3)

As *S. sonnei* pINV is frequently lost during prolonged laboratory culture, the CRLA was adapted to investigate the loss of 53G pINV with or without pRES, using an isogenic *S. sonnei* 53G background. $bla_{TEM\text{-}116}$, conferring beta-lactam resistance was inserted downstream of *aadA* on CS6 pRES (Figure 3.1.3A), to ensure the presence of pRES at the start of assays. Subsequently, pRES::$bla_{TEM\text{-}116}$ was transferred from CS6 into pINV$^{+/-}$ 53G by conjugation, and the presence of pRES in 53G confirmed by PCR detecting the pRES replicon (see Table 2.2 for primers).

S. sonnei 53G with pINV ± pRES were grown from frozen stocks on plain CRA and incubated at 37°C for 16 hours. Three CR$^+$ colonies were picked and serially diluted in PBS. An 80 µl aliquot of the 10^{-5} dilution was inoculated on plain CRA and incubated at 37°C for 16 hours, and counted the number of CR$^+$ and CR$^-$ colonies. This assumed that pRES loss frequency is below the detection limit of CRLA and will not affect the results. Eliminating subculture in liquid as in assays with *S. flexneri* enhanced the precision of measuring plasmid loss by focusing on the initial relatively short period of growth, thus minimising the impact of differential growth of pINV$^+$ and pINV$^-$ bacteria and overgrowth of plasmid-less colonies to ensure more consistent and reliable results (341). The results indicate no significant difference in pINV maintenance with or without pRES (p = 0.4172, Figure 3.2.1B). These results indicate that the presence of pRES does not affect pINV maintenance.

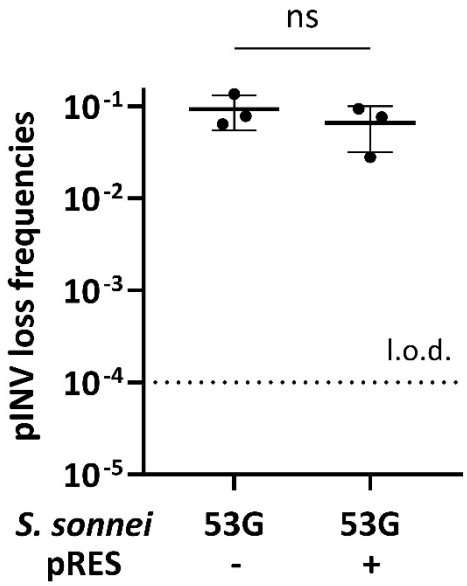


Figure 3.2.1B: pINV loss is not affected by co-existing pRES in isogenic 53G background.
Each dot represents plasmid loss from three single colonies in a replicate. Unpaired t-test was used to assess the differences between two groups. ns, not significant. (N=3)

pRES loss was then assessed during laboratory culture. I-complex plasmids such as pRES are usually stably maintained during bacterial culture due to their maintenance systems such as the TA systems (342). As pRES does not encode a T3SS, CRLA cannot be used to assess pRES loss. The *sacB* gene was introduced into pRES::$bla_{TEM\text{-}116}$ downstream of $bla_{TEM\text{-}116}$; *sacB* allows counter-selection of bacteria that have lost pRES as they can grow on media with 10% (w/v) sucrose (314).

pRES loss was investigated using PLA as indicated in Chapter 2.4 to compare frequencies of pRES loss in the presence or absence of pINV in *S. sonnei* 53G. PCR detecting the pRES replicon in both sucrose and carbenicillin resistant bacteria was performed to confirm the presence of pRES. Of note, the replicon was amplified from in 17 and 13 out of 90 sucrose-resistant colonies of bacteria with and without pINV, respectively. pRES loss occurred in most but not all the sucrose-resistant colonies, and the frequency of plasmid loss, quantified as the proportion of pRES-less bacteria in the population, was adjusted accordingly. Results demonstrate that pRES is stably maintained within the bacterial population with a frequency of loss at 1.5 x 10^{-5} and 9.5 x 10^{-6} in the presence and absence of pINV, respectively (Figure 3.2.1C). Therefore, pRES is infrequently lost during growth of *S. sonnei* 53G in the laboratory. In addition, the presence of pINV does not affect pRES maintenance (p = 0.5377). Taken together, the co-existence of pINV and pRES does not affect maintenance of each other.

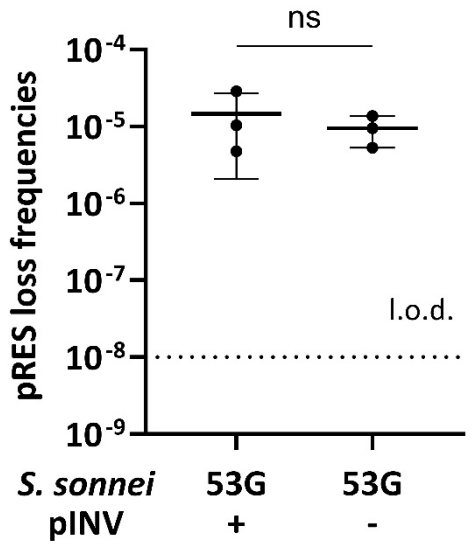


Figure 3.2.1C: pRES loss in 53G.
Each dot represents averaged plasmid loss from three single colonies in a replicate. Unpaired t-test was used to compare data from two groups. ns, not significant. (N=3)

3.2.2 pRES transfer frequency is affected by pColE1 but not pINV

pRES harbours an intact conjugation system that allowed transfer of the plasmid from *S. sonnei* CS6 to 53G; construction of pRES::$bla_{TEM-116}$ in the previous section allowed selection of transconjugants. Next, frequency of pRES transfer by conjugation was investigated in conjugation assays (Chapter 2.3). Initially, conjugation assays were performed using two strains: *S. sonnei* CS6 with pINV and pRES::$bla_{TEM-116}$ as the donor, and *S. sonnei* 53G lacking pINV and pRES but harbouring *aroG::aph(3')-Ia* (conferring kanamycin resistance) as the recipient. The inserted markers allowed identification of 53G transconjugants that had acquired pRES by selection on CRA with 100 µg/ml kanamycin and 100 µg/ml carbenicillin.

For matings, both strains were recovered at 37°C for 16 hours from frozen stocks on CRA with antibiotics to ensure the presence of pRES in the donor at the start of the assays. The donor and recipients were grown separately for 16 hours in 5 ml liquid TSB supplemented with antibiotics. The bacteria were washed with PBS and adjusted to an OD_{600}=0.05 in 30 ml TSB containing antibiotics, then sub-cultured for two hours until the exponential phase (OD_{600}=0.5). Bacteria were then washed with PBS to remove residual antibiotics, and resuspended in PBS to a concentration of approximately 5 x 10^9 CFU/ml (Figure 7.8).

The donor and recipient were mixed at different donor:recipient (D:R) ratios (0.33:1, 1:1 and 3:1), with a fixed volume of 25 µl of the recipient to investigate the impact of the amount of donor on the frequency of pRES transfer. The bacterial mixtures were then spotted onto plain LBA, dried at room temperature, and incubated at 37°C for four hours to allow bacterial conjugation while limiting clonal expansion following early HGT events due to prolonged incubation. After this time, bacteria were harvested and resuspended, serially diluted and inoculated to plain CRA, CRA with 100 µg/ml kanamycin and CRA with 100 µg/ml kanamycin and 100 µg/ml carbenicillin to allow quantification of all bacteria, recipients and transconjugants, respectively. Bacteria were incubated at 37°C for 16 hours, and the number of colonies counted.

The pRES transfer frequency was calculated as the proportion of recipients that were transconjugants at the end of co-incubation with the donor. Thus a frequency of pRES transfer of 0.1 denotes that 10% of recipients received pRES, *i.e.* were transconjugants. Results demonstrate that the pRES transfer frequency increased from 2.4 x 10^{-3} transconjugants per recipient to 5.0 x 10^{-1} as the D:R ratio increased from 0.33:1 to 3:1 ($p = 0.0168$) (Figure 3.2.2A).

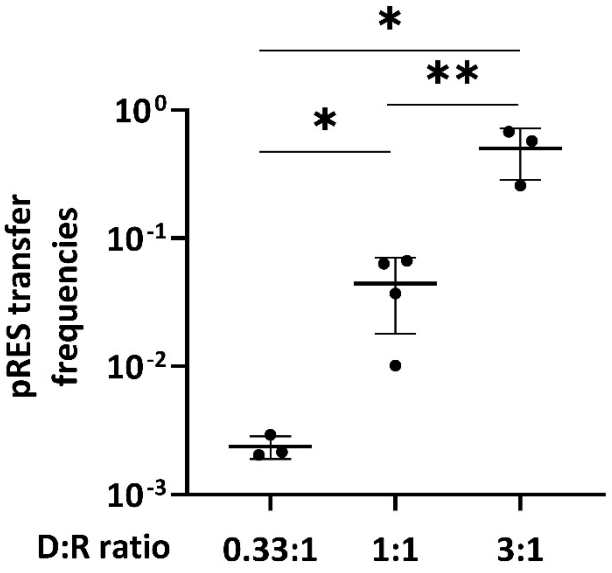


Figure 3.2.2A: pRES transfer frequencies increase with the D:R ratios.
Each dot represents averaged three conjugations in a replicate using pINV$^+$ pRES$^+$ CS6 as the donor and plasmid-less 53G as the recipient. pRES transfer frequencies were shown as the proportion of recipients that became transconjugants at the end of assays. Unpaired t-tests were used to assess the difference in pRES transfer frequencies with different D:R ratios. *, $p = 0.0435$ or 0.0168; **, $p = 0.0079$. (N=3 or 4)

The increasing pRES transfer frequency with increasing D:R ratios was accompanied by a decrease in the number of recipients from 2.3 x 10^9 CFU/ml to 3.3 x 10^4 CFU/ml post-conjugation ($p < 0.0050$, Figure 3.2.2B). This indicates that the recovery of recipients was adversely affected during incubation with the donor *S. sonnei* CS6, suggesting potential inhibitory effects or competitive interactions between the donor and recipient during conjugation, impacting the viability of the recipient.

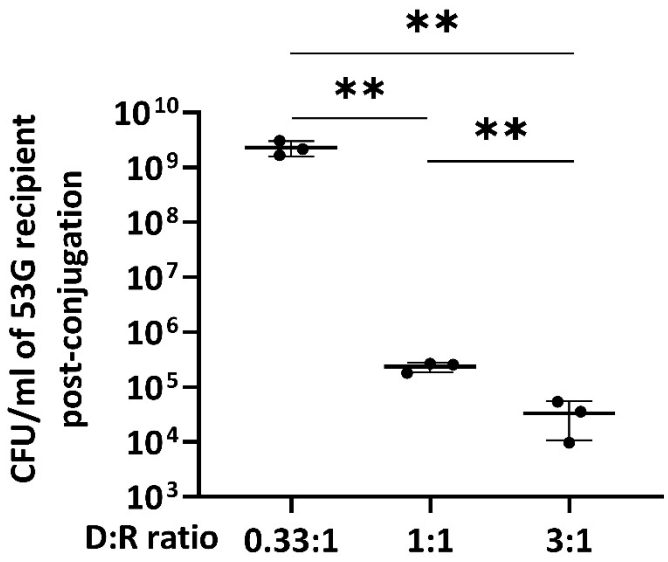


Figure 3.2.2B: Recovery of recipients post-conjugation is reduced with increasing D:R ratios.
Each dot represents three conjugations in a replicate using pINV$^+$ pRES$^+$ CS6 as the donor and plasmid-less 53G as the recipient. Unpaired t-tests were used to analyse the difference in recovery of the recipients post-conjugation under different conjugation conditions. ** p = 0.0050 or 0.0026. (N=3)

A potential explanation for the reduced recovery of recipients was explored. Previous bioinformatic analysis revealed that *S. sonnei* CS6 harbours a potentially mobilisable plasmid, pColE1, encoding ColE1. The recipient *S. sonnei* 53G could have been killed by ColE1 secreted by the donor CS6 containing pColE1. If a recipient acquired pColE1 during conjugation, it would become immune to the effect of ColE1-expressed by the donor. Thus, it was proposed that the recipients recovered post-conjugation might have acquired pColE1 from CS6, allowing their survival in the presence of the high ColE1 concentrations from CS6. The high D:R ratio could exert a selective pressure, favouring recipients with pColE1 and killing those without it.

Therefore, PCRs were conducted to detect the presence of *cea* encoded by pColE1 in nine transconjugants from three independent conjugations to investigate the proportion of pColE1$^+$ transconjugants among the transconjugants. Results demonstrate that only 46.9% of transconjugants contained *cea* when the D:R ratio was 0.33:1. In contrast, *cea* could be amplified from all transconjugants when the D:R ratio was 3:1 (Figure 3.2.2C). This demonstrates that pColE1 is a mobilisable plasmid as predicted by bioinformatic analysis (Figure 3.1.4A), and could affect the frequency of pRES transfer by killing recipients lacking pColE1. The impact of the donor harbouring pColE1 on the survival of the recipient would be expected to be more

pronounced as the D:R ratio increased due to closer bacterial contact and higher secretion of colicin from the donor population. Therefore, the increased pRES transfer frequency in the assay does not reflect the number of HGT events, but instead is an artefact from the death of recipients by the increased number of donors harbouring pColE1.

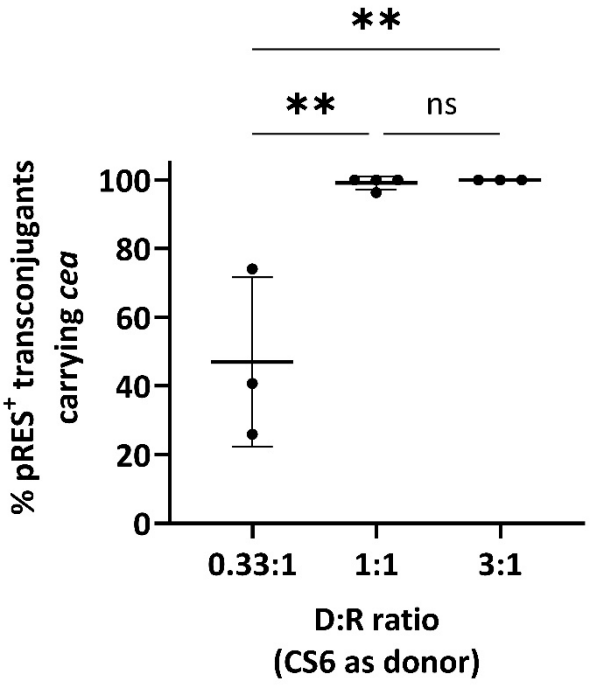


Figure 3.2.2C: Detection of the *cea* gene in the 53G transconjugants.
Each dot represents three conjugations in a replicate. The percentage of 53G with *cea* was determined by sampling from nine transconjugants per conjugation from independent replicates. Ordinary one-way ANOVA was used to examine variations in the proportion of *cea*$^{+}$ transconjugant across different D:R ratios post-conjugation. ** p = 0.0032 or 0.0043; ns, not significant. (N=3)

Plasmids in a donor which mediate killing of the recipient can be a confounding factor in assays to examine the transfer of other genetic elements. Therefore, further conjugation assays were conducted between isogenic *S. sonnei* 53G strains using pINV::*cat* and pRES::$bla_{TEM\text{-}116}$ without pColE1. To generate a donor strain, pRES::$bla_{TEM\text{-}116}$ was transferred from *S. sonnei* CS6 into 53G with pINV::*cat* by conjugation; the presence of pRES::$bla_{TEM\text{-}116}$ and absence of pColE1 in *S. sonnei* 53G was confirmed by PCR. In addition, plasmids were extracted from the newly generated donor strains, then analysed by gel electrophoresis (Figure 3.2.2D). Results confirmed the expected genotypes of the strains.

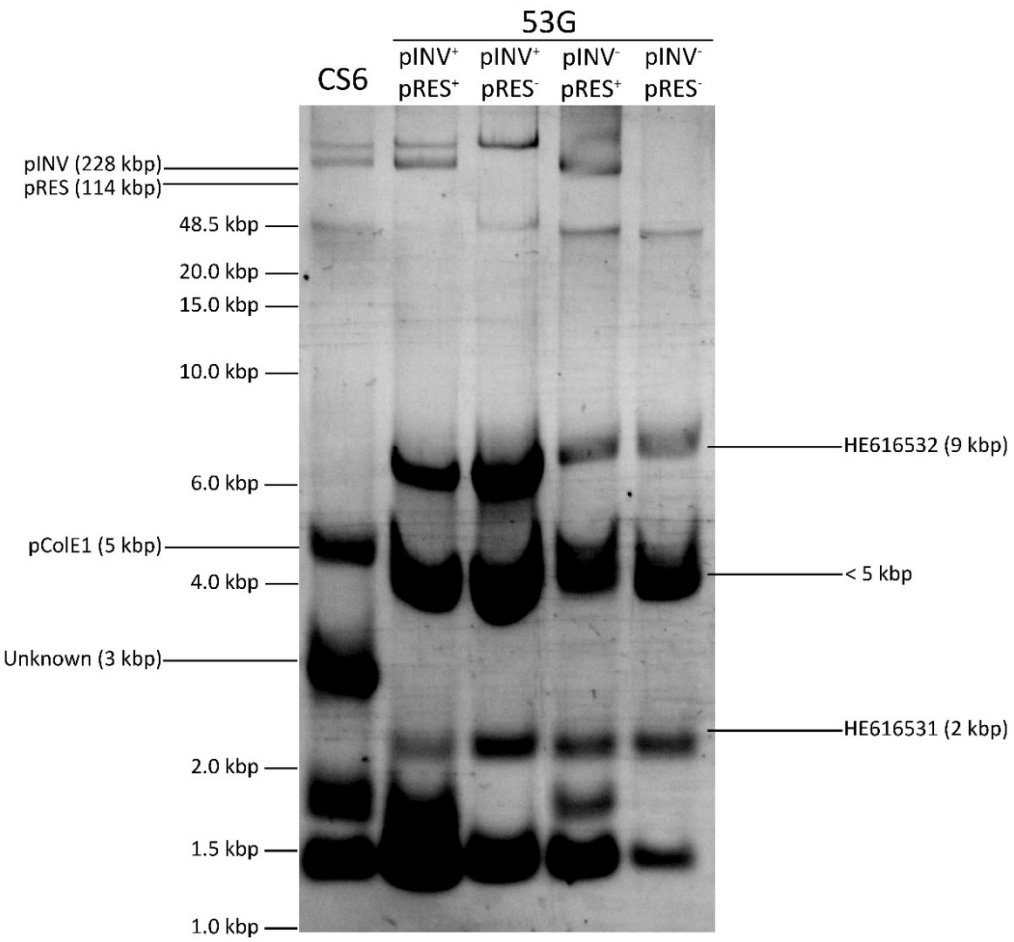


Figure 3.2.2D: Plasmids in wild-type (WT) *S. sonnei* CS6, and 53G with pINV and pRES, with pINV only, with pRES only, and without these plasmids.
The sizes of pINV, pRES, pColE1, HE616532 and HE616531 are labelled.

Conjugation assays were performed using *S. sonnei* 53G pRES::$bla_{TEM-116}$ +/- pINV::*cat* as the donor and *S. sonnei* 53G without these plasmids as the recipient. Conjugation assays were performed at D:R ratio of 1:1. The frequency of pRES transfer was at 8.1 x 10^{-3} and 1.2 x 10^{-2} in the presence and absence of 53G pINV, respectively (p = 0.0891, Figure 3.2.2E). These findings indicate that pRES transfer is unaffected by the presence of pINV. In addition, the presence of the donor did not have a major impact on the recovery of the recipient after matings (Figure 3.2.2F). The recovery of the recipient was not significantly affected when the 53G donor harboured both pINV and pRES (p = 0.0854), while the recovery of recipient was slightly affected when the donor harboured pRES only (p = 0.0171, Figure 3.2.2F). As the variation is minor compared to the previous results affected by the presence of pColE1, it was thought that this

variation could be caused by other reasons, such as slightly lower starting concentration of 53G recipient (Figure 7.8) or competition for nutrients during conjugation.

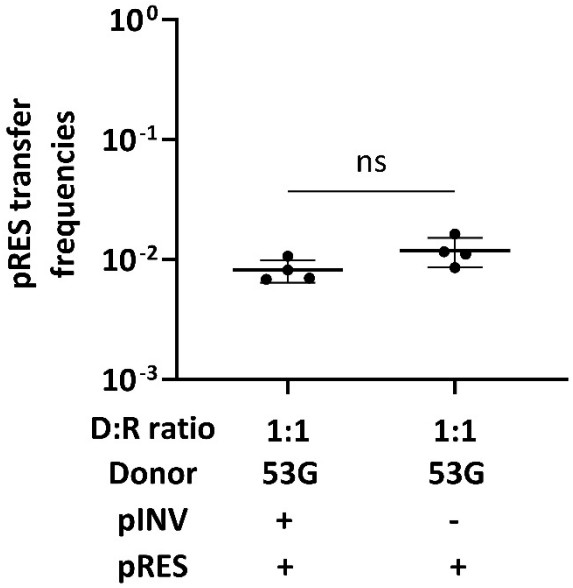


Figure 3.2.2E: The frequency of pRES transfer is not affected by co-existing pINV in isogenic conjugations.
Each dot represents three conjugations in a replicate. Plasmid profiles of the donor are labelled. pRES transfer frequencies were shown as the proportion of $pRES^+$ transconjugants within the recipient population post-conjugation. Unpaired t-test weas used to assess the difference in pRES transfer frequencies under different conditions. ns, not significant. (N=4)

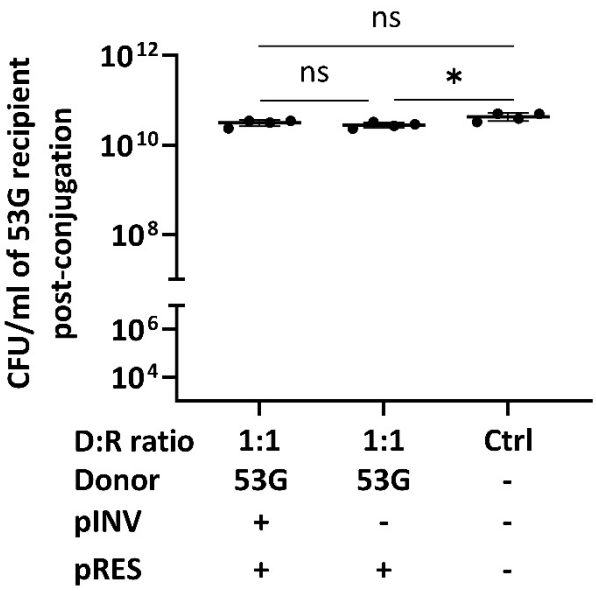


Figure 3.2.2F: Recovery of recipients post-conjugation is unaffected by the presence of pINV in an isogenic 53G background.
Each dot represents three conjugations in a replicate. Plasmid profiles of the donor are labelled. Control (Ctrl) reactions contained only 53G recipients. Unpaired t-tests were used to analyse the difference in the number of recipients post-conjugation under different conjugation conditions. * p = 0.0205; ns, not significant. (N=4)

3.3 Summary

In this chapter, I characterised plasmids in the clinical isolate *S. sonnei* CS6, focusing on their genetic content and organisation, maintenance, and ability to undergo horizontal transfer. *S. sonnei* CS6 harbours multiple plasmids including pINV, pRES, pColE1, and other uncharacterised small genetic elements.

S. sonnei CS6 and Ss046 belong to the Lineage III, which is the most common group of *S. sonnei* isolates circulating currently (20, 84). The conservation of pINV in *S. sonnei* 53G (Lineage II), CS6 and Ss046 was revealed by bioinformatics analysis (Figure 3.1.1A), which is consistent with emergence of *S. sonnei* globally through clonal expansion (20). pINV harbours many ISs, and they are believed to contribute to shaping pINV backbone, such as the acquisition of the O-Ag gene cluster and potential deletion of the T3SS PAI (44, 84). Interestingly, despite the presence of intact ISs on pINV, which could be functional, pINV backbone is strikingly conserved with no further genetic re-assortment observed (20). However, conclusion regarding the conservation of pINV cannot be easily made as the analysis of pINV backbone is limiting due to the lack of detection of pINV in the sequencing data (20). Assessment of the functionality of ISs on pINV could be done to further understand their contribution to shape pINV backbone.

In addition to pINV, *S. sonnei* CS6 (20) also carries an IncB/O/K/Z plasmid, pRES, which harbours multiple AMR-conferring genes. Plasmids of the same Inc group were found in clinical isolates from different parts of the world, and contribute to the abundant resistance plasmids in *S. sonnei*, illustrating the importance of this Inc group of plasmids in the dissemination of AMR (343). IncB/O/K/Z plasmids are also frequently detected in other Enterobacterales (204). The wide distribution and prevalence of plasmids belonging to this Inc group in *S. sonnei* show that this group of plasmids has been successful in this species, and means that my findings can have broad significance for this pathogen. Therefore, pRES from IncB/O/K/Z group is a good model

plasmid for investigation of horizontal transfer of virulence plasmids mediated by resistance plasmids.

S. sonnei can harbour a broad range of resistance plasmids, that are members of other Inc groups, which highlights the ability of this pathogen to acquire AMR genes horizontally (62). Interestingly, plasmids from Inc groups other than IncFII and IncI-complex plasmids are not commonly found in *S. sonnei*. This could be caused by reasons including the potential bias in sequencing strains in the outbreaks and that some infection cases could be underdiagnosed (3). The emergence and spread of resistance plasmids has undermined public health interventions to treat and prevent infections caused by this important human pathogen.

I then examined the interplay between pRES and pINV during plasmid maintenance. pINV is lost more frequently than pRES during laboratory culture (Figure 3.2.1A, Figure 3.2.1B, Figure 3.2.1C). In addition, pINV loss occurs more frequently in *S. sonnei* than *S. flexneri*, so these experiments were performed over a relatively short term. Plasmid loss represents the initial failure of segregation and then the failure of PSK mechanisms. A potential reason why resistance plasmids are being frequently recovered from *S. sonnei* is that the resistance and virulence plasmids might promote the maintenance of each other, through cross-talk between TA and/or partitioning systems. However, my results show that the frequency of pINV or pRES loss is unaffected by the presence of the other plasmid. This suggests that there is no cross talk between the plasmids' TA or partitioning systems, which could affect segregation or PSK. Alternatively, compensatory mutations which offset the fitness costs for one plasmid might also be beneficial for carriage of another unrelated plasmid, which the effect of the plasmids (alone and in combination) on the fitness of their bacterial host could be examined during competition for nutrients between strains with different plasmid content.

While my thesis was predominantly concerned with pINV and pRES, a third plasmid, pColE1, influenced the frequency of conjugation. Results in this chapter demonstrate that plasmid-

encoded ColE1 in *S. sonnei* CS6 had a dramatic effect on the survival of *S. sonnei* 53G. Colicins are widespread in Lineage III *S. sonnei*, and are believed to contribute to the epidemiological success of this lineage (266). This became obvious during matings between the two strains when the D:R ratio was high. In these circustances, the colicin secreted by the *S. sonnei* CS6 donor was highly efficient at killing the 53G recipient, unless it had acquired pColE1 by conjugation and had become resistant. For conjugation, a contact-dependent mechanism of HGT, the influence of antagonistic activities in donors and recipients may form a barrier to genetic transfer, similar to restriction modification systems (RMS). However, in constrast to RMS, resistance to pColE1-mediated killing can be transferred by conjugation (if the plasmid harbouring gene encoding immunity protein against ColE1 is mobilisable). Consequently, the outcome of the interaction between strains will be influenced by the rates and speed of killing and of conjugation. This highlights the impact of small plasmids on inter-bacterial competition and the dynamics of HGT. Given the role of *S. sonnei* colicin plasmids in inter-bacterial competition (268), the mobilisation of pColE1 by pRES could impose a selective advantage for bacteria carrying both pRES and pColE1. The potential impact of these smaller plasmids on the maintenance of other plasmids remains to be elucidated. This could further contribute to the understanding of co-evolutionary dynamics between virulence and resistance determinants on plasmids. To eliminate inter-strain competition as a confounding factor in my experiments, I performed conjugations between isogenic strains lacking pColE1 (*i.e. S. sonnei* 53G).

Although pINV carries several MOB_F-conjugation system genes, it is likely that these genes are non-functional due to mutations compared to the corresponding genes on the F plasmid (Figure 3.1.1B). As pINV lacks an intact conjugation system, pINV is non-conjugative. Alongside the absence of an obvious *oriT* on pINV, this plasmid is also thought to be non-mobilisable in the presence of a conjugative plasmid. However, a previous study demonstrated the transfer of *S. sonnei* pINV into *E. coli* by an IncF conjugative plasmid R386, suggesting that pINV is mobilisable (169). Given the distinct Inc groups between pRES and R386, pRES serves as a good model to

explore mechanisms underlying pINV transfer. As pRES is from Lineage III *S. sonnei* CS6, investigation of potential mobilisation of pINV from Lineage II *S. sonnei* 53G will contribute to understanding of bacterial evolution and potential movement of virulence plasmids across lineages.

4 Characterisation of pINV transfer mediated by pRES

4.1 *S. sonnei* pINV transfer by pRES

In the previous chapter, I characterised plasmids in Lineage III clinical *S. sonnei* isolate CS6. My bioinformatic analyses predicted that pINV is non-mobilisable due to its lack of important mobilisation elements, such as *oriT* (Chapter 3.1.1). However, in a previous work demonstrating that the virulence of *S. sonnei* is mediated by pINV, pINV was mobilised by conjugative plasmids in the laboratory (169). Therefore, given the high prevalence of resistance plasmids in *S. sonnei*, I would like to use a naturally occurring resistance plasmid, pRES from a *S. sonnei* Brazillian clinical isolate CS6, to investigate the potential horizontal transfer of pINV (62).

My previous results demonstrated that the transfer of pRES is unaffected by the presence of pINV in the donor (Figure 3.2.2D). I then optimised the conjugation assays using pRES from CS6 as a model plasmid in an isogenic 53G background, using pINV^{+} pRES$^{+/-}$ 53G as donors and plasmid-less 53G as recipient (Chapter 2.3). I assessed its ability to transfer pINV from *S. sonnei* 53G (Lineage II), and investigated the potential mechanisms of transfer using techniques including biochemical assays (Chapter 2.7), Oxford Nanopore Technology sequencing (Chapter 2.9) and bioinformatic analyses (Chapter 2.10).

4.1.1 Preliminary evidence that *S. sonnei* 53G pINV can be transferred by CS6 pRES

To assess whether pRES can transfer pINV, conjugation assays were performed using a donor *S. sonnei* 53G harbouring pINV with *cat* (allowing for selection with chloramphenicol) and pRES with $bla_{TEM\text{-}116}$ (allowing selection with carbenicillin); the recipient *S. sonnei* 53G with *aroG::aph(3')-Ia* (allowing selection with kanamycin) does not harbour pINV or pRES. Matings were conducted with a D:R ratio of 1:1 (Figure 4.1.1A). The presence of plasmids in the donor was confirmed by PCR; furthermore, the strain does not harbour pColE1 which could affect

quantification of transfer (Figure 3.2.2C). As a control, *S. sonnei* 53G with pINV only was used as a donor in separate conjugations; this strain was also confirmed to lack pColE1.

As in previous conjugation assays, strains were recovered from frozen stocks and grown in liquid TSB with antibiotics overnight. After growth overnight, bacteria were washed and adjusted to an OD_{600}=0.05 then sub-cultured in 30 ml TSB for two hours at 37°C with antibiotics to ensure the maintenance of plasmids prior to conjugation. After reaching the exponential phase (~OD_{600}=0.5), bacteria were washed and resuspended at a concentration of 5 x 10^9 CFU/ml, then mixed at 1:1 ratio in a total volume of 100 µl. The entire volume was spotted onto plain LBA, allowed to dry at room temperature and then incubated at 37°C for four hours. After incubation, bacteria were harvested and resuspended in 100 µl PBS, then spread on CRA with 100 µg/ml kanamycin and 20 µg/ml chloramphenicol to select for transconjugants, *i.e.* recipients containing pINV. The total bacterial population and number of recipients post-conjugation were enumerated by plating onto plain CRA and CRA with kanamycin, respectively. Following overnight incubation, $pINV^+$ transconjugants were re-streaked onto CRA with 100 µg/ml kanamycin and 20 µg/ml chloramphenicol to allow further analyses by PCR. This step was included to prevent carryover of genetic material on initial plates which could lead to false positive PCRs. In parallel, all transconjugants were re-streaked onto CRA with kanamycin, chloramphenicol and carbenicillin to detect the presence of pRES in addition to pINV. These analyses enabled the quantification of the frequency of pINV and pRES transfer, and the partial characterisation of plasmid profiles in transconjugants.

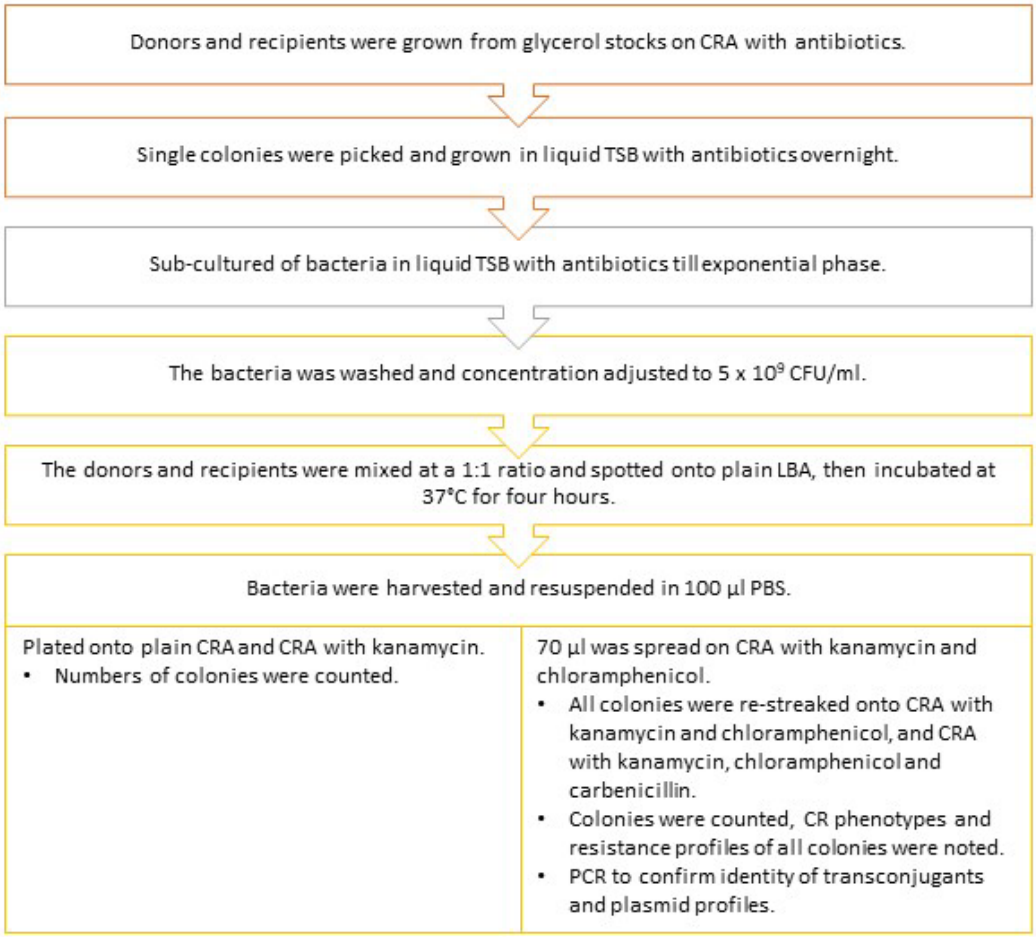


Figure 4.1.1A: Overall strategy to detect pINV transfer (Chapter 2.3).

Consistent with findings by Sansonetti *et al.* (1981), I found that pINV is not conjugative, as no transconjugants were detected when *S. sonnei* 53G with only pINV was used as the donor (Figure 4.1.1B). However, transconjugants were detected at the frequency of 4.5×10^{-9} of the recipient population when the donor also contained pRES. This shows that pINV is mobilisable at a low frequency in the presence of the conjugative plasmid, pRES.

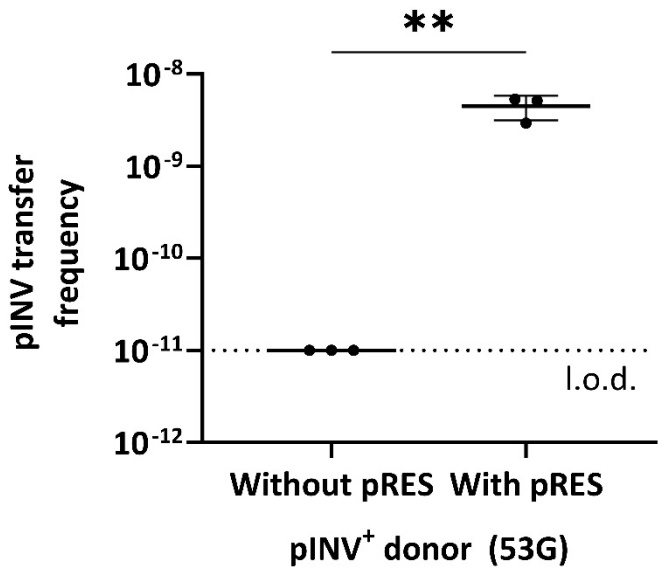


Figure 4.1.1B: pINV can be transferred by pRES.
Each dot represents three conjugations in a replicate. pINV transfer was quantified as the proportion of pINV$^+$ transconjugants within the population of recipients post-conjugation according to PCR results. Unpaired t-test was used to assess the significance. The limit of detection (l.o.d.) of the conjugation assays was 10^{-11}. **, p = 0.0043. (N=3)

I also noted that the colony phenotypes of transconjugants varied when they were re-streaked on CRA with 100 µg/ml kanamycin and 20 µg/ml chloramphenicol (Figure 4.1.1C). Colonies from transconjugants either stayed CR$^+$ with a regular size and shape, or were completely CR$^-$ with a rough surface and irregular shape, or gave rise to a mixture of small CR$^+$ colonies and large CR$^-$ colonies. The growth of transconjugants on CRA with 100 µg/ml kanamycin and 20 µg/ml chloramphenicol indicated the presence of *cat* on pINV; thus the different phenotypes could be the result of the presence and absence of the T3SS PAI and/or O antigen cluster. Results demonstrate that some transconjugants were susceptible to carbenicillin, indicating that they had not acquired pRES during mating. Colonies of different CR phenotypes which emerged from a single transconjugant could also exhibit different resistance profiles, indicating that they had different plasmid profiles.

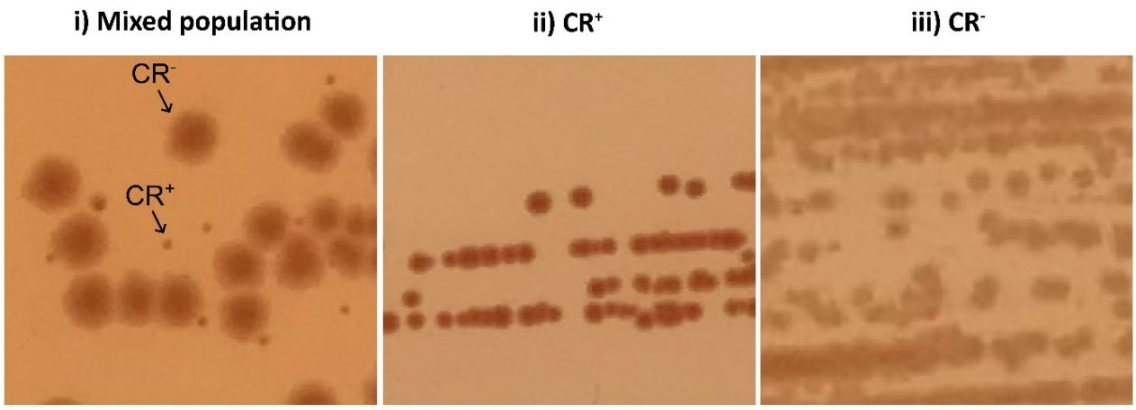


Figure 4.1.1C: Examples of different phenotypes of colonies from pINV$^+$ transconjugants.
Phenotypes of transconjugants observed when re-streaking onto CRA with chloramphenicol and kanamycin. Re-streaked single colonies from three transconjugants are shown to represent the different phenotypes of transconjugants. From left to right, there is i) a mixture of small CR$^+$ colonies and large CR$^-$ colonies emerged after re-streaking a single CR$^+$ transconjugant, ii) transconjugants with consistent CR$^+$ round colonies with regular size, or iii) CR$^-$ colonies with large and irregular shapes after restreaking.

Following examination of the CR phenotypes and resistance profiles of the transconjugants, PCR was conducted to further analyse the plasmid profiles in colonies which emerged from re-streaking transconjugants. PCR was used to detect the CS6 pRES replicon (as evidence for the presence of pRES), the *vapBC* TA system on pINV, and *aroG::aph(3')-Ia* from the chromosome of transconjugants. Four transconjugants (TC1-4) with different resistance profiles and CR phenotypes when re-streaking were selected for assessment. PCR confirmed that TC1-4 were all transconjugants (kanamycin resistant) and harboured pINV (shown by detection of *aroG::aph(3')-Ia* and *vapBC*, respectively). PCR analyses revealed different plasmid profiles (Figure 4.1.1D). TC1 and TC4 colonies that were CR$^+$ on re-streaking were carbenicillin resistant, and the presence of pINV (*vapBC* PCR positive) and pRES (replicon PCR positive) was confirmed by PCR. TC2, TC3 and CR$^-$ colonies derived from TC4 were carbenicillin sensitive, and the absence of the pRES replicon was verified by PCR. Therefore, colonies derived from TC4 with different CR phenotypes harboured different plasmid profiles shown by PCR analyses. Overall, PCR analyses were in agreement with the phenotypic analyses, and suggest that multiple genetic events occurred during pINV transfer.

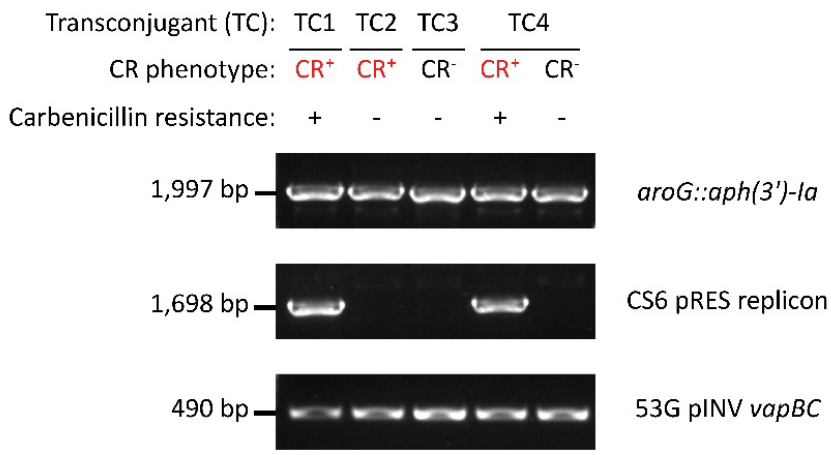


Figure 4.1.1D: Different plasmid profiles in TCs with different phenotypes after re-streaking.
Single colonies from re-streaked TCs with various resistance and CR phenotypes were selected and analysed by PCR. The presence of a 1,997 bp band amplified by primers targeting *aroG::aph(3')-Ia* on the 53G chromosome indicate that all colonies tested were derived from the recipient strain. The chloramphenicol resistance profile and a 490 bp band amplified from adjacent to pINV *vapBC* confirmed the presence of pINV in the TCs. Some TCs were carbenicillin resistant which correlated with the amplification of a 1,698 bp product from the pRES replicon.

Although the presence of pRES in the donor is required for pINV transfer, phenotypic and PCR analyses indicated that pRES is not always found in the transconjugants. To gain insights into the prevalence of pRES in pINV$^+$ transconjugants, PCR was conducted to check for the presence of pRES in transconjugants (amplifying a 1,698 bp product from the pRES replicon). A total of 63 transconjugants of different CR phenotypes and resistance profiles were analysed. Results reveal that 86% of transconjugants also harbour pRES (Figure 4.1.1E). Thus 14% of transconjugants do not contain pRES despite its importance in transferring pINV, and the result was consistent with the resistance profile of the transconjugants (not shown).

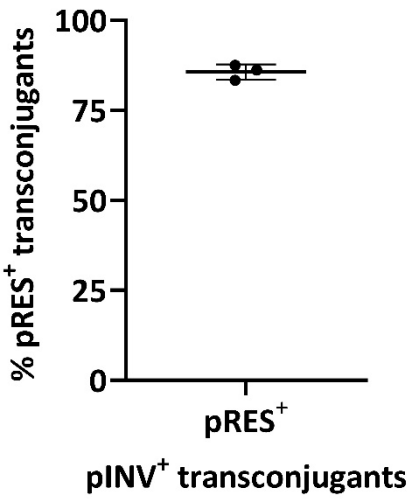


Figure 4.1.1E: The pRES replicon is not always present in transconjugants as detected by PCR.
Each dot represents three conjugations in a replicate. Data was shown as percentage of pRES$^+$ population among all transconjugants as indicated by the detection of pRES replicon. (N=3)

Taken together, these results show that pINV can be transferred in the presence of pRES in the donor. pINV transfer occurs at low frequency, and pRES was not always detected in pINV$^+$ transconjugants. The variety of phenotypes of transconjugants indicates that various genetic events occur during pINV transfer.

4.2 Identification of hybrid pINV/pRES plasmids in transconjugants

After conjugation to allow horizontal transfer of pINV by co-residing pRES in *S. sonnei* 53G, different transconjugants were observed when cultured on CRA with different antibiotics, and PCR analyses revealed the plasmid profiles of transconjugants. Therefore, the plasmid profiles were further investigated by gel electrophoresis and next generation sequencing to characterise genetic events that occurred during pINV transfer.

4.2.1 Hybrid pINV/pRES plasmids were identified in transconjugants with mixed phenotypes

Plasmid DNA was extracted from the transconjugants grown from frozen stocks on CRA with antibiotics and their DNA visualised by gel electrophoresis (Chapter 2.7) (317). Plasmids recovered from *S. sonnei* 53G with pINV and pRES, with pINV only, with pRES only, and without these plasmids were used as size references for pINV and pRES. Transconjugants TC1, TC2, TC3 and TC4 (Figure 4.1.1D) were selected for analysis. As TC4 gave rise to both CR^+ and CR^- colonies, plasmids from colonies of both phenotypes were investigated separately.

PCR analyses suggested that all TCs harbour pINV (*i.e.* PCR positive for *vapBC*), while TC1 and CR^+ TC4 harbour pRES, while TC2, TC3 and CR^- TC4 lack pRES (based on the detection of the pRES replicon by PCR, Figure 4.1.1D). Consistent with the PCR results, gel electrophoresis revealed that TC1 and TC2 harbour a plasmid band of the size corresponding to pINV (Figure 4.2.1A). TC1 harbours an additional band which corresponds to the size of pRES, which is not present in TC2. Interestingly, TC3 and TC4 harbour plasmids which are neither the size of pINV nor pRES. CR^+ and CR^- TC4 both harbour a plasmid of the same size as found in TC3, which is smaller than pRES. However, CR^+ TC4 harbours an additional plasmid which is larger than pINV.

	Controls				Transconjugants (TCs)					Controls			
					TC1	TC2	TC3	TC4					
CR	-	-	+	+	+	+	-	+	-	+	+	-	-
pINV	-	-	+	+	+	+	?	?	?	+	+	-	-
pRES	-	+	-	+	+	-	?	?	?	+	-	+	-

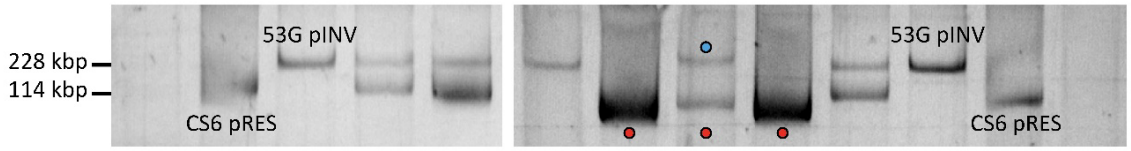


Figure 4.2.1A: Plasmid visualisation by gel electrophoresis.
TC1, TC2, TC3 and TC4 were selected due to their different CR phenotypes, resistance profiles and PCR detection of plasmids as in Figure 4.1.1D. The CR phenotypes, presence of plasmids were annotated for both controls and TCs. The plasmids smaller than pRES (red dot) or larger than pINV (blue dot) are indicated.

To further characterise the plasmids carried by TC4, ONT sequencing was performed using a MinION running for three hours. The demultiplexed Fastq files were subjected to *de novo* assembly by Flye and annotated by Prokka. Annotation of the key elements of the plasmids was manually evaluated by BLASTp. The assembled plasmid files were aligned with wild-type pINV and pRES, and analysed using BLASTn and Snapgene.

The sequencing results revealed that CR^+ TC4 harbours two large hybrid plasmids, each composed of elements from pINV and pRES (Figure 4.2.1B). Bioinformatic analyses demonstrate that the hybrid plasmids were formed by junctures between identical copies of IS*21* on pINV and pRES, and between a 199 bp region which is present in pINV and pRES, located downstream of their replicons. The 199 bp region in pINV and pRES differ by one nucleotide (G161A) (Figure 4.2.1C). The **241** kbp **h**ybrid **p**lasmid (pHP241) harbours the pRES replicon, T4SS, AMR genes, along with pINV *relBE*, the O-Ag gene cluster, *parAB* and the T3SS PAI; the **90** kbp **h**ybrid **p**lasmid (pHP90) harbours the pINV replicon and *vapBC*, as well as pRES *relBE*, *pndAC* and *parB*. CR^- TC4 only harbours pHP90. The result explains the CR phenotypes observed from CR^+ TC4 as the T3SS is encoded on pHP241 which is not present in CR^- TC4. As pHP241 is large and could impose significant fitness costs, the colonies CR^+ TC4 appeared to be smaller than wild-type *S. sonnei*

53G harbouring pINV. Similarly, CR⁻ TC4 only harbours pHP90 which lacks the T3SS PAI and O-Ag cluster, therefore colonies appeared rough and CR⁻. The results also suggest that the lack of pRES replicon detection by PCR may not fully reflect plasmid carriage in transconjugants due to the formation of hybrid plasmids; the replicon could be carried on a plasmid which is not stably maintained by transconjugants.

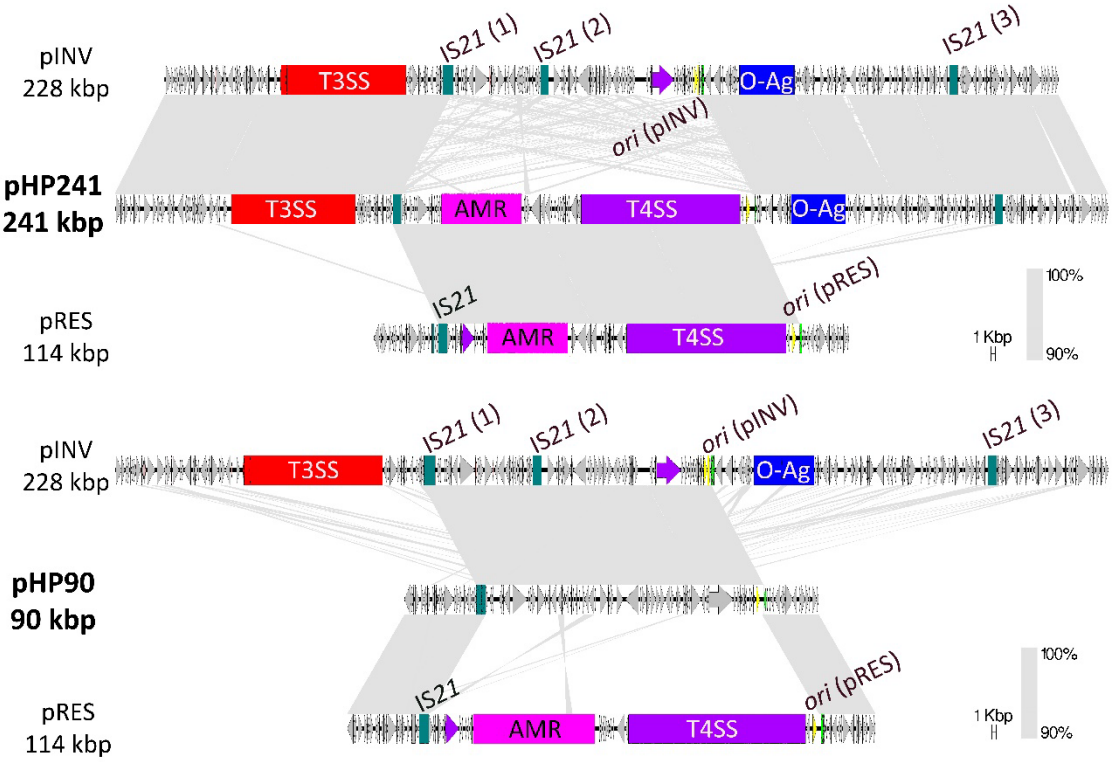


Figure 4.2.1B: Alignment of the hybrid plasmids (pHP241 and pHP90) from TC4 with pINV and pRES.
The nucleotide alignments of hybrid plasmids against pINV and pRES were created by Easyfig, and light grey shows nucleotide identity of > 90%. Important elements are annotated and colour-coded: replicon, yellow; pINV *vapBC* and *relBE*, pRES *relBE* and *pndAC*, green; the T3SS PAI, red; IS*21*s, pale blue; T4SS and conjugation-related genes, purple; O-Ag, dark blue.

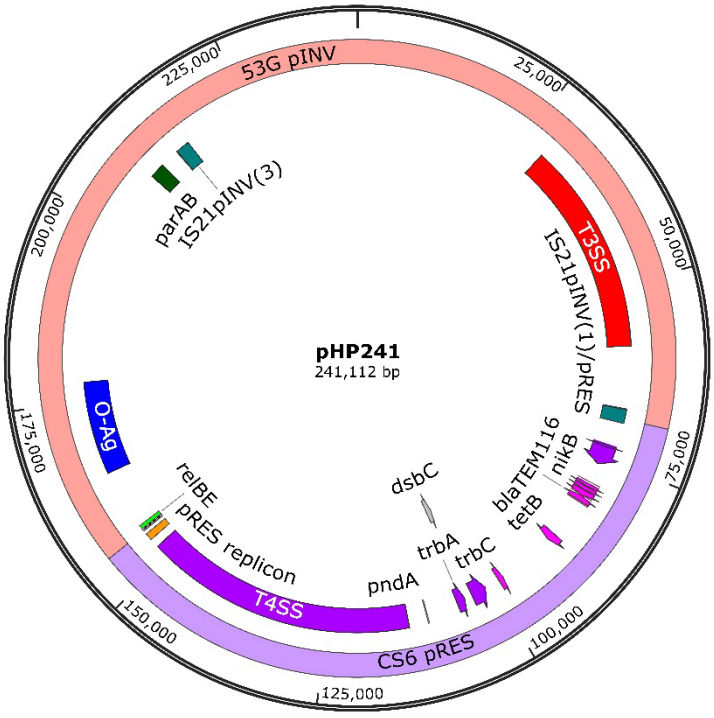
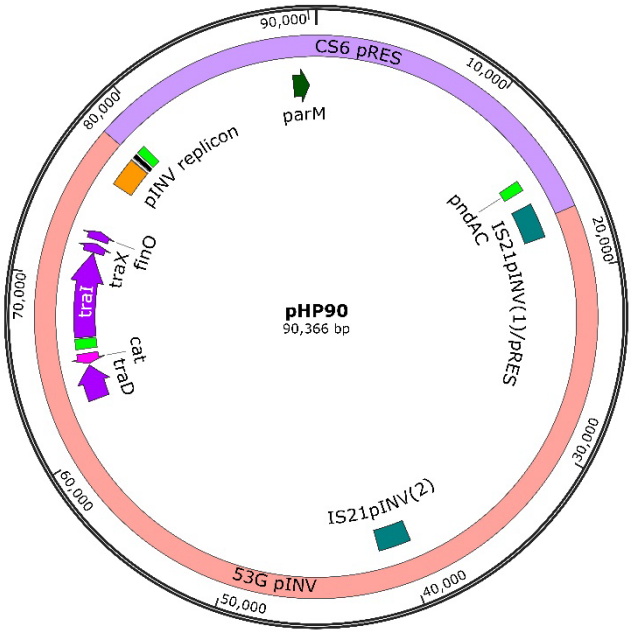
Figure 4.2.1C: The nucleotide sequence of the 199 bp region on 53G pINV.
The SNP compared with pRES is shown.

4.2.2 Identification of hybrid plasmids with different junctions between pINV and pRES

As hybrid plasmids containing elements from pINV and pRES were detected in transconjugants after transfer pINV by pRES, the plasmid profiles of transconjugants from independent conjugations were characterised further. A total of 25 transconjugants with different genotypes, predicted by colony phenotypes and PCR analyses, were selected and sequenced by ONT MinION as well as Illumina (MicrobeNG). Sequencing data were assembled using *de novo* assembler Flye or hybrid assembler Unicycler and annotated by Prokka.

Results reveal that six out of 25 transconjugants harbour regions of both pINV and pRES, while five TCs harbour pINV only (Table 4.1); one harbours only wild-type pINV which is integrated into the chromosome. Based on sequence analysis, this was potentially mediated by HR between copies of IS*1* on the 53G chromosome and pINV.

14 transconjugants were identified harbouring different hybrid plasmids. These hybrid plasmids were classified into groups as Type 1, Type 2, Type 3 and Type 4 according to the pINV/pRES junctions in the hybrid or fusion plasmids. There are seven TCs harbouring Type 1 hybrid plasmids (pHP241 and pHP90), one harbours a Type 2 hybrid/fusion plasmid; four harbour Type 3 hybrid plasmids (pHP146 and pHP186); and two harbour a Type 4 hybrid plasmid (pHP72).

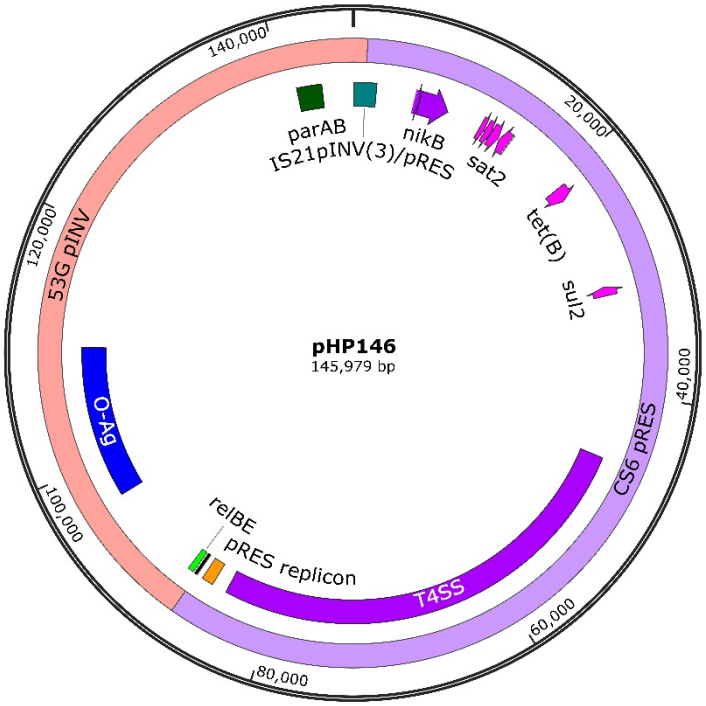
	Plasmid profiles	No. of TCs (n= 25)
WT plasmids	pINV and pRES	6
	pINV only	4
	pINV integrated into the chromosome at IS*1*	1
Hybrid or fusion pINV/pRES plasmids	**Type 1**: Junctions at IS*21*$_{pINV}$(1)/IS*21*$_{pRES}$ and the 199 bp region pHP241: From pRES: replicon, T4SS, AMR genes; from pINV: *parAB*, *relBE*, O-Ag, T3SS PAI  pHP90: From pINV: replicon, *vapBC*; from pRES: *parB*, *pndAC*, *relBE*  	7

Type 2: Junctions at IS*21*$_{pINV}$(2)/IS*21*$_{pRES}$ (fusion or hybrid plasmid detected)	1

Type 3: Junctions at IS*21*$_{pINV}$(3)/IS*21*$_{pRES}$ and the 199 bp region

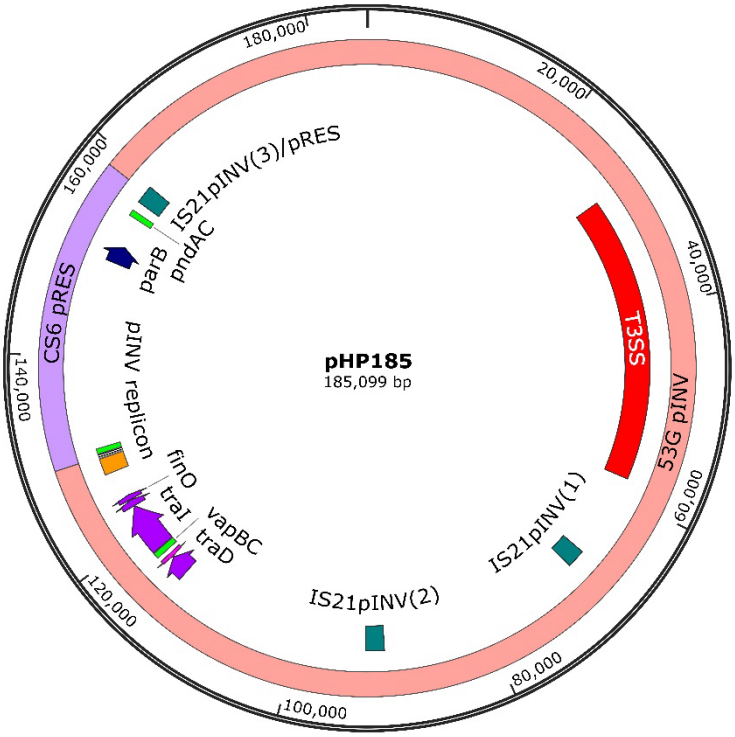
pHP146:

From pRES: replicon, *relBE*, T4SS, AMR genes; from pINV: O-Ag, *parAB*



pHP185: 4

From pINV: replicon, *vapBC*, T3SS PAI; from pRES: *relBE*, *parB*, *pndAC*



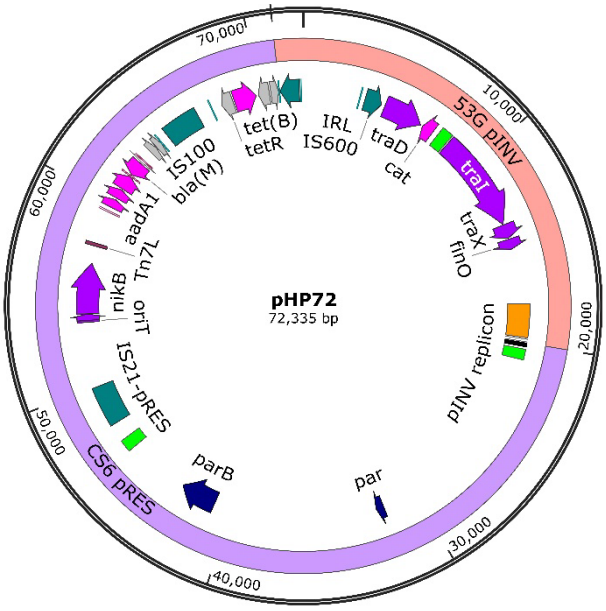
Type 4: Junctions at IS*Vsa5*$_{pRES}$/IS*630*$_{pINV}$ and the 199 bp region pHP72: From pINV: replicon, *vapBC*; from pRES: *relBE*, *parB*, *pndAC*, *oriT* & *nikAB*, AMR genes  	2

Table 4.1: Different transconjugants detected and corresponding joins on hybrid plasmids.
Plasmid profiles and the frequency of detection in selected transconjugants were summarised. IS*21* on pINV were labelled using different numbers according to the order of identification of the copy being involved in forming joins of hybrid plasmids. Elements on different hybrid plasmids were noted.

Further analyses of the plasmids revealed that Type 1, 2 and 3 plasmids contain junctions between the three copies of IS*21* on pINV and the single copy of IS*21* on pRES (Table 4.1). Copies of IS*21* on pINV which were involved in junctions with pRES were numbered for illustrative purposes. Copies of IS*21* accounted for the most of the junctions between pINV and pRES in hybrid plasmids (12 out of 14 hybrid plasmids), potentially due to HR as the ISs are highly homologous. The Type 4 hybrid plasmid was the only plasmid in the transconjugants that did not involve IS*21*. Of note, the 199 bp region was also at pINV/pRES junctions in hybrid plasmids. This site could be involved in HR between the plasmids as it is highly homologous and there is no MGEs encoded close to the sequence.

In the Type 2 plasmid, the *de novo* assembly of long-read sequencing showed discrepancies in results when compared to the hybrid assembler. The hybrid assembler Unicycler failed to circularise the contigs, although there was a contig with a pINV/pRES junction at IS*21*$_{pINV}$(2)-

IS*21*$_{pRES}$. In contrast, *de novo* assembler Flye circularised a single contig into a 333 kbp plasmid containing intact pINV and pRES joined at IS*21*$_{pINV}$(2)-IS*21*$_{pRES}$. Therefore, although the IS*21*$_{pINV}$(2)-IS*21*$_{pRES}$ junction was confirmed by both assembly methods, it was not certain whether the plasmids recombined at another place, resulting in two hybrid plasmids instead of a single fusion plasmid.

The Type 4 hybrid plasmid pHP72 was found in two transconjugants from the same conjugation (Figure 4.2.2A). This plasmid has pINV/pRES junctions at two sites: the 199 bp region and IS*630*$_{pINV}$/IS*4*-family IS*Vsa5*$_{pRES}$. pHP72 harbours pRES *oriT* and *nikAB* encoding the relaxase, with the pINV replicon, *vapBC*, as well as AMR genes from pRES. As it carries an *oriT* and genes encoding the relaxosome, this hybrid plasmid could be mobilisable. Therefore, this hybrid plasmid could be formed within the donor and be transferred by conjugation using the T4SS machinery encoded on another hybrid plasmid. Alternatively, the hybrid plasmid containing the rest of the elements with the pRES replicon could have been lost after transfer into the recipient.

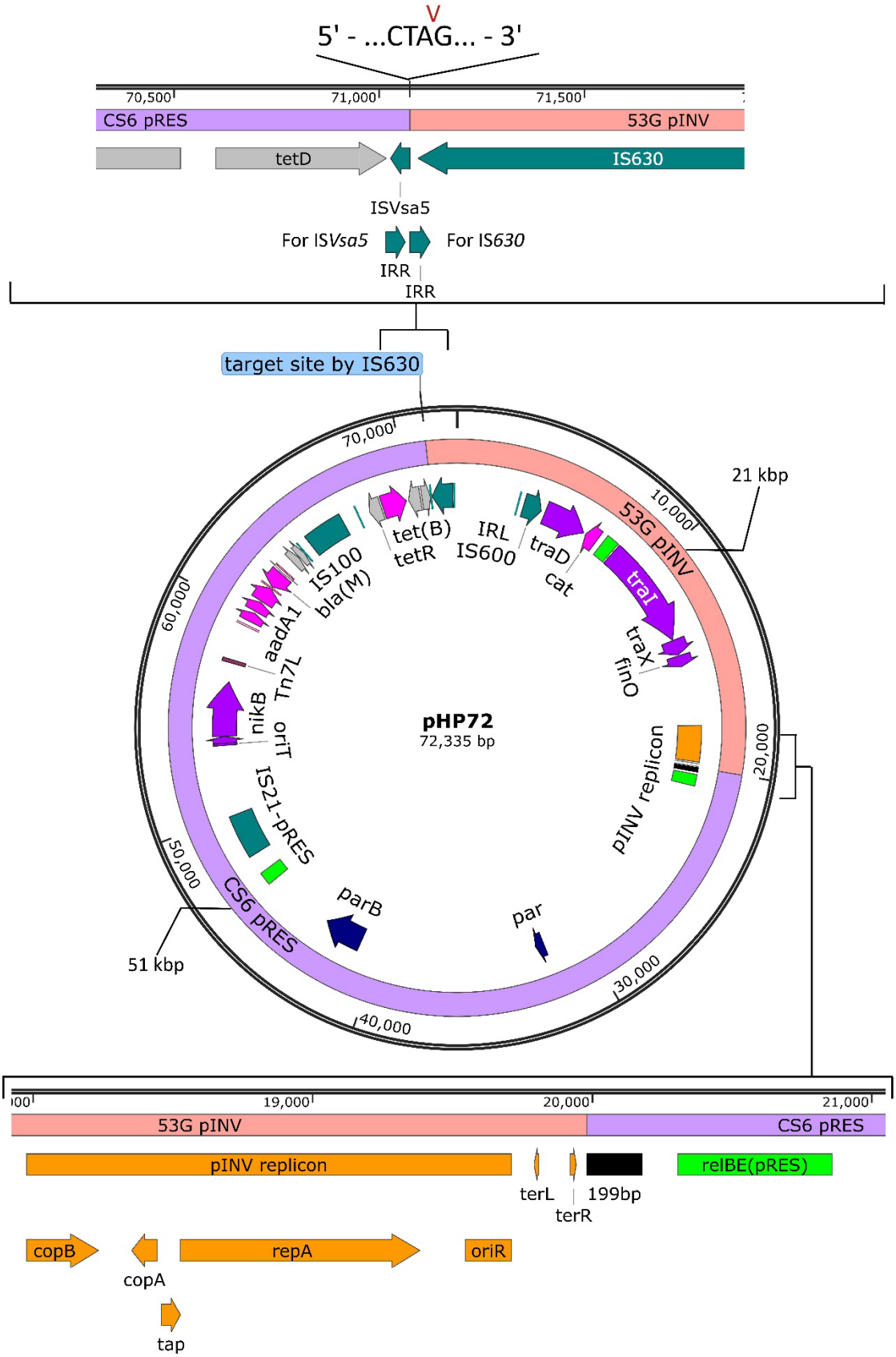


Figure 4.2.2A: Map of pHP72 with pINV/pRES junctions at the 199 bp region downstream pRES replicon and IS*630*$_{pINV}$ transposition.

The plasmid map was exported from Snapgene. Important features including the junctions between the plasmids are annotated. The site of IS*630*$_{pINV}$ transposition was noted as a red arrow.

The junction between IS*630*$_{pINV}$/IS*Vsa5*$_{pRES}$ was only detected in pHP72 (Figure 4.2.2A). These two ISs were analysed further to indicate whether the hybrid plasmid was formed by HR or IS transposition. The two ISs share low level of nucleotide identity (46.34%) as determined by ClustalOmega (Figure 7.9). IS*Vsa5* on pRES (52,198 bp - 52,890 bp) harbours a single copy of inverted repeat(IR), IRR, which is insufficient for its functionality, therefore this IS*Vsa5* is likely to be inactive. In contrast, the IS*630* on 53G pINV (165,114 bp - 166,145 bp, accession no. NC_016833) is predicted to be intact (ISfinder, accession no. X05955). Analysis of the IS*Vsa5* sequence revealed that it harbours a preferential target site 5'-CTAG-3' for IS*630* insertion, which is next to the IRR of IS*630*, suggesting that the join was formed by IS*630* transposition (240). However, PCR and Sanger sequencing (Source BioScience) were employed and there is no duplication of 5'-TA-3', the typical signature of IS*630* transposition, found at the site of insertion of IS*630*$_{pINV}$ in IS*Vsa5*$_{pRES}$ (240). Thus, the formation of pHP72 could be mediated by both HR and/or IS transposition.

In addition to IS-mediated transposition, I hypothesised that HR could form junctions between plasmids prior to horizontal transfer. Bioinformatic analysis was performed to identify sites of homology shared by pINV and pRES with a nucleotide identity of > 85% and a length of > 199 bp using blastn, as RecA-dependent HR is more likely to occur for homolgous sequences of > 200 bp (344). It was found that there are multiple sites of homology between the plasmids other than IS*21*, which be the site of HR between pINV and pRES (Table 4.2). Thus, these further sites of homology could also lead to formation of hybrid plasmids.

No.	Length (bp)	% identity	CS6 pRES			53G pINV (NC_016833.1)		
			Start (bp)	End (bp)	Element	Start (bp)	End (bp)	Element
1	2,130	99.95	30,032	32,161	IS*21*	29,972	32,101	IS*21* (3)
2	2,131	99.91				123,801	125,931	IS*21* (1)
3	2,133	99.86				149,680	147,548	IS*21* (2)
4	1,696	85.32	741	2,418	*repA* and downstream nucleotides (including the 199 bp region)	183,873	185,554	*repA* and downstream nucleotides (including the 199 bp region)
5	1,331	100	61,077	62,407	IS*2*	142,998	141,668	IS*2* (2)
6	1,333	99.85				3,873	5,205	IS*2* (1)
7	1,051	99.05	46,646	47,696	IS*100*	21,273	20,223	IS*100* (1)
8	1,051	99.05				80,619	81,669	IS*100* (2)
9	748	91.71	20,098	20,835	*klcA* and upstream (not ORF)	168,655	169,396	*klcA* and upstream (not ORF)
10	397	91.94	62,015	62,407	Part of IS*2*	564	180	Part of IS*2*
11	361	90.03	27,083	27,435	Putative anti-restriction protein	168,696	169,049	Putative anti-restriction protein
12	275	93.46	31,855	32,128	Part of *istB*	10,812	10,549	Putative part of *istB*
13	211	93.37	46,257	46,466	Putative IS*21*-family IS*100* ORF	145,178	144,968	Putative IS*21*-family IS*100* ORF

Table 4.2: List of sites of homology between *S. sonnei* 53G pINV and CS6 pRES.
The sites of homology in pINV and pRES of > 199 bp and > 85% nucleotide identity were detected using blastn. Reference sequence, *S. sonnei* 53G pINV (accession no. NC_016833); query sequence, *S. sonnei* CS6 pRES.

Overall, these results show that hybrid pINV/pRES plasmids in transconjugants have junctions at sites of sequence identity, which led to different phenotypes of colonies on CRA. Due to the formation of hybrid plasmids which might or might not be maintained in the bacterial host, the detection of the plasmid replicon by PCR as a means to determine the presence of plasmids in transconjugants may be inaccurate. As the hybrid plasmids in transconjugants contain junctions at sites of homology, it was hypothesised that RecA-dependent HR is involved in the formation of hybrid plasmids prior to pINV transfer.

4.3 Mechanisms of pRES-mediated transfer of pINV

Hybrid pINV/pRES plasmids were detected in transconjugants when investigating pINV transfer by pRES. These hybrid plasmids harbour parts from both pINV and pRES, joining at different junctions. For transconjugants harbouring wild-type pINV, some of them lack pRES which is essential for pINV transfer (5 of 11). Earlier bioinformatic analysis revealed that TraI and TraD from 53G pINV are likely to be non-functional and that pINV lacks an *oriT* (Figure 3.1.1A). However, it is possible that an *oriT*-like sequence on 53G pINV could be recognised by the relaxase encoded by CS6 pRES, enabling mobilisation of pINV into the recipient without pRES. Therefore, matings were performed using genetically modified donors with the pINV$^-$ pRES$^-$ *aroG::aph(3')-Ia S. sonnei* 53G as the recipient to investigate the mechanisms of pRES-mediated pINV transfer.

4.3.1 *oriT*, NikB and RecA contribute to pINV transfer by pRES

It was hypothesised that if pINV could be transferred into the recipient without pRES, it would harbour an *oriT*-like sequence which could be recognised by pRES relaxase NikB as pINV TraI is likely non-functional. To test this hypothesis, the 5'-end of *nikB* (2,048 bp) encoding for the pRES relaxase (345) was deleted in the donor to generate a markerless mutant, so that the nearby *oriT* site and *nikA*, which NikA is required for relaxosome formation with NikB (346), were unaffected. As the C-terminal end of NikB was shown to be not required for DNA relaxation (345), 223 codons at the C-terminal was not removed to not affect the downstream gene *tolA*. The loss of NikB should prevent pRES transfer, but not the assembly of the T4SS. Conjugation assays were performed using a donor with pRES Δ*nikB* and pINV, with the 53G recipient lacking pINV and pRES at 1:1 D:R ratio. Results show that deleting *nikB* leads to undetectable pRES transfer (Figure 4.3.1A). Of note, pINV transfer was also abolished, indicating that pRES NikB is

required for transfer of both pINV and pRES. This suggests that pINV transfer cannot be mediated by the putative TraI relaxases encoded on the *S. sonnei* 53G pINV and chromosome.

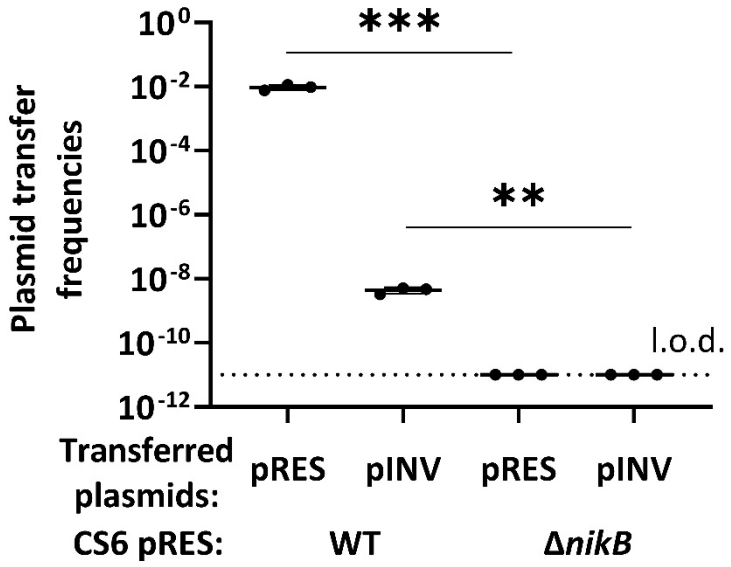


Figure 4.3.1A: Deletion of *nikB* reduces pRES and pINV transfer. Each dot represents three conjugations in a replicate. Unpaired t-tests were performed to compare pRES or pINV transfer by WT *vs*. Δ*nikB* donors. **, p = 0.0015; ***, p = 0.0008. (N=3)

Next, the *oriT* on pRES was deleted to establish whether pRES transfer is required for pINV transfer, or there is an *oriT*-like sequence on pINV as *oriT* needs to be on the molecule that is transferred and cannot be provided in *trans* like NikB (345). Theoretically, deletion of *oriT* should not affect the function of the conjugation system on pRES, but should inhibit pRES transfer. 40 bp of the *oriT* that includes the nick site for relaxase recognition and cleavage was removed without affecting *nikA* which is 65 bp downstream of the site of deletion. Conjugation assays were performed using a pRES Δ*oriT* donor and the recipient at 1:1 D:R ratio. As expected, deletion of *oriT* prevents pRES transfer. In addition, pINV transfer was also abolished (Figure 4.3.1B), suggesting that there is no functional *oriT*-like sequence on pINV which can be recognised by the pRES relaxase, and that pINV is transferred with pRES as a single molecule. Together with the earlier detection of hybrid plasmids in transconjugants (proposed to be formed by HR and/or IS transposition), it was hypothesised that pINV is transferred with pRES as a single fusion molecule.

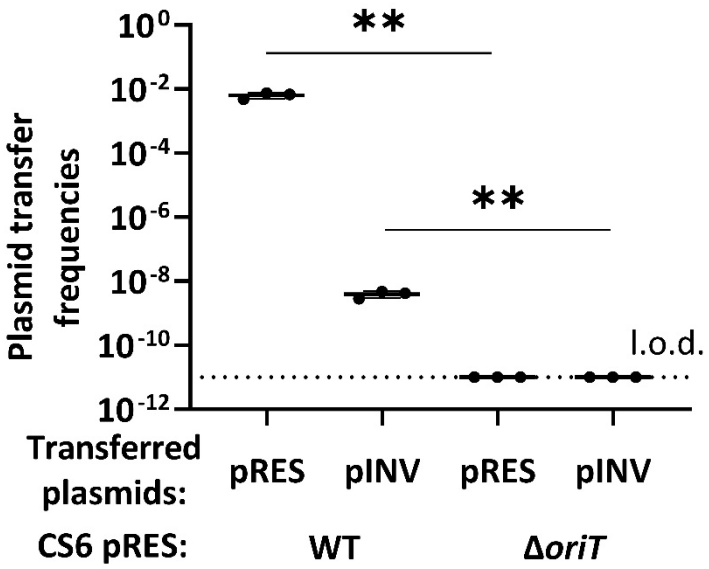


Figure 4.3.1B: Deletion of pRES *oriT* abolished pRES and pINV transfer.
Each dot represents three conjugations in a replicate. Unpaired t-tests were performed to compare pRES or pINV transfer by WT *vs.* Δ*oriT* donors. **, p = 0.0013 or 0.0022. (N=3)

RecA is a multi-functional protein which is involved in HR (347). As the formation of the single molecule could result from HR at sites of sequence identity on pINV and pRES (Table 4.1), *recA* was deleted from the donor strain and used in conjugations with the recipient at a 1:1 D:R ratio. Results show that deletion of *recA* from the donor did not affect pRES transfer, but pINV transfer was below the limit of detection (10^{-11}, Figure 4.3.1C). This demonstrates the importance of RecA in the donor for pINV transfer. RecA could be involved in the majority of the HR events that precede pINV transfer. This is consistent with the hypothesis that pINV and pRES form a single molecule by HR prior to transfer by conjugation.

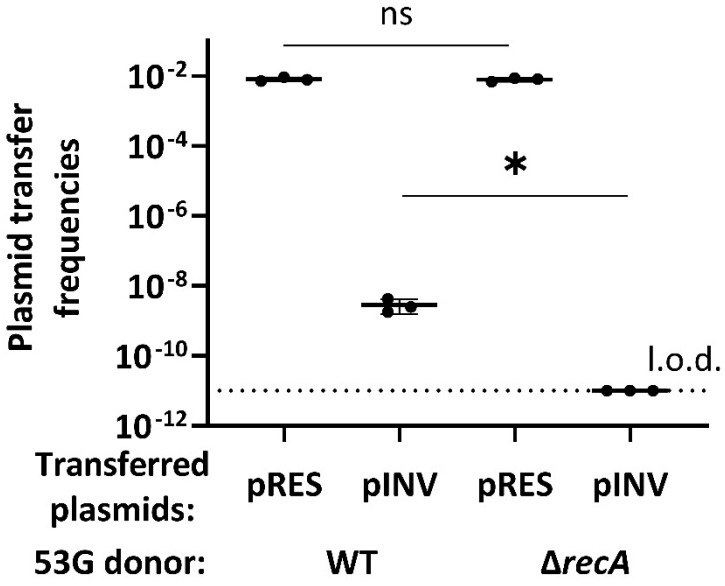


Figure 4.3.1C: Deletion of *recA* in the donor inhibits pINV transfer, but did not affect pRES transfer.
Each dot represents three conjugations in a replicate. Unpaired t-tests were performed to compare pRES transfer or pINV transfer with WT *vs.* Δ*recA* donors. *, p = 0.0183; ns, not significant. (N=3)

Taken together, these findings indicate that pINV transfer occurs after pINV forms a single molecule with pRES *via* HR, with pRES *oriT* and its relaxase required to initiate the transfer of the fusion molecule. However, my results do not exclude the possibility that rarely IS transposition can be responsible for formation of a single molecule consisting of pRES and pINV. Plasmids consisting of regions of pINV and pRES were hypothesised to form in the donor mainly due to RecA-dependent HR before the transfer, given the importance of RecA in the donor. Furthermore, the detection of hybrid plasmids without pRES *oriT* in recipients suggests that HR could occur in the recipient to resolve the fusion plasmid into smaller hybrid plasmids.

4.3.2 IS*21* plays a significant role in pINV transfer

Given the hypothesis that a single fusion plasmid forms in the donor by RecA-mediated HR prior to conjugation, RecA in the recipient could contribute to the resolution of the fusion plasmid into hybrid plasmids by HR. Therefore, *recA* in the recipient was deleted in an attempt to prevent the resolution of a fusion plasmid, and thereby increase the likelihood of detecting a fusion plasmid in the recipient, providing further evidence to support the working model for pINV transfer.

Conjugation was performed between the wild-type donor and a Δ*recA* recipient at a 1:1 D:R ratio. Given that a fusion plasmid would be large (> 300 kbp) and impose significant fitness costs, conjugations were performed with a modified step for selecting transconjugants to improve the chance of detecting a fusion plasmid. Therefore, after four-hour conjugation at 37°C, bacteria were incubated at 30°C for approximately 48 hours to potentially minimise the fitness cost for any transconjugants carrying a large plasmid.

Initially, the transfer of pRES to the Δ*recA* recipient was quantified and results demonstrated no significant difference in pRES transfer frequency to the Δ*recA* recipient compared with the wild-

type recipient, at frequencies of 6.0 x 10^{-3} and 9.7 x 10^{-3}, respectively ($p > 0.05$, Figure 4.3.2A). This suggests that RecA in the recipient does not affect the conjugation frequency of pRES.

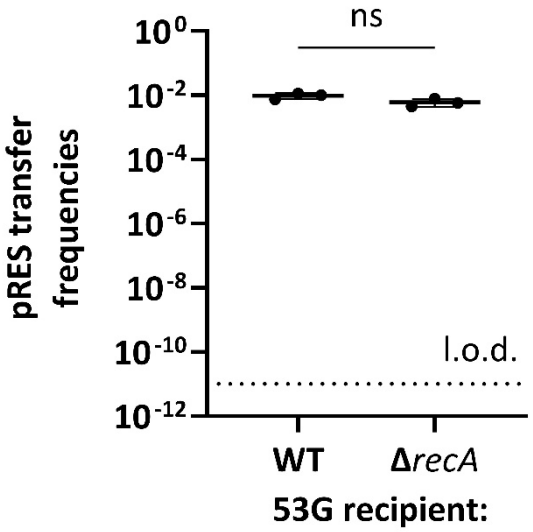


Figure 4.3.2A: pRES transfer frequency was reduced following deletion of *recA* in the recipient.
Each dot represents three conjugations in a replicate. Unpaired t-test was performed to assess the significance of the differences in pRES transfer frequencies. ns, not significant. (N=3)

Transconjugants with evidence they had acquired pINV and pRES were selected on CRA with kanamycin, chloramphenicol and carbenicillin. DNA was extracted from three Δ*recA* transconjugants and subjected to ONT and Illumina sequencing (MicrobesNG) to characterise their plasmids. Sequence data were assembled by the hybrid assembler Unicycler, analysed and annotated manually according to alignments using BLAST. Sequencing revealed the presence of hybrid plasmids in two out of three Δ*recA* transconjugants. The hybrid plasmids had junctions at the 199 bp region but with different IS*21*$_{pINV}$-IS*21*$_{pRES}$ junctions; both plasmids were identical to those detected in wild-type recipients (Table 4.3). This indicates that RecA-independent genetic events can resolve fusion plasmids into hybrid plasmids in the recipients.

	Hybrid plasmids	No. of TCs (n= 2)
Hybrid pINV/pRES plasmids	**Type 1**: Junctions at IS*21*$_{pINV}$(1)/IS*21*$_{pRES}$ and the 199 bp pHP241: From pRES: replicon, T4SS, AMR genes; from pINV: *parAB*, *relBE*, O-Ag, T3SS PAI pHP90: From pINV: replicon, *vapBC*; from pRES: *parB*, *pndAC*, *relBE*	1
	Type 3: Junctions at IS*21*$_{pINV}$(3)/IS*21*$_{pRES}$ and the 199 bp pHP146: From pRES: replicon, *relBE*, T4SS, AMR genes; from pINV: O-Ag, *parAB* pHP186: From pINV: replicon, *vapBC*, T3SS PAI; from pRES: *relBE*, *parB*, *pndAC*	1

Table 4.3: Hybrid plasmids detected in Δ*recA* transconjugants.
These hybrid plasmids are identical to those detected in WT 53G transconjugants.

However, one of the TCs harboured a large fusion plasmid formed by IS*21*$_{pRES}$ transposition (Figure 4.3.2B). IS*21*$_{pRES}$ contains both *istA* and *istB* encoding for a transposase and ATPase, respectively, with two IRs, IRR and IRL (Figure 4.3.2B, Table 4.2) (242, 245). IS*21* integrated pRES into pINV at a site 784 bp downstream of pINV *repA* which contains a 5´-TCGG-3´ sequence, the preferred integration site for IS*21*, located within the 199 bp region near the replicon (Figure 4.2.1C) (242, 245). IS*21*$_{pRES}$ and the 4 bp GCGG sequences were duplicated and present at both ends of the insertion (Figure 4.3.2B), providing evidence for IS*21* transposition rather than HR (246). The junctions between the plasmids were verified by PCR and Sanger sequencing (Source BioScience). This suggests an alternative mechanism for the formation of a fusion plasmid which is RecA- and HR-independent. IS*21* transposition was not observed when *recA* was deleted from the donor, which could be due to the low frequency of the event under these conditions. The frequency of IS*21* transposition has only been studied in the absence of *recA* in *E. coli* using small (< 7 kbp) vector plasmids (246). Further studies are required to assess the differences in frequencies of IS transposition and RecA-dependent HR in the formation of fusion plasmids during pINV transfer.

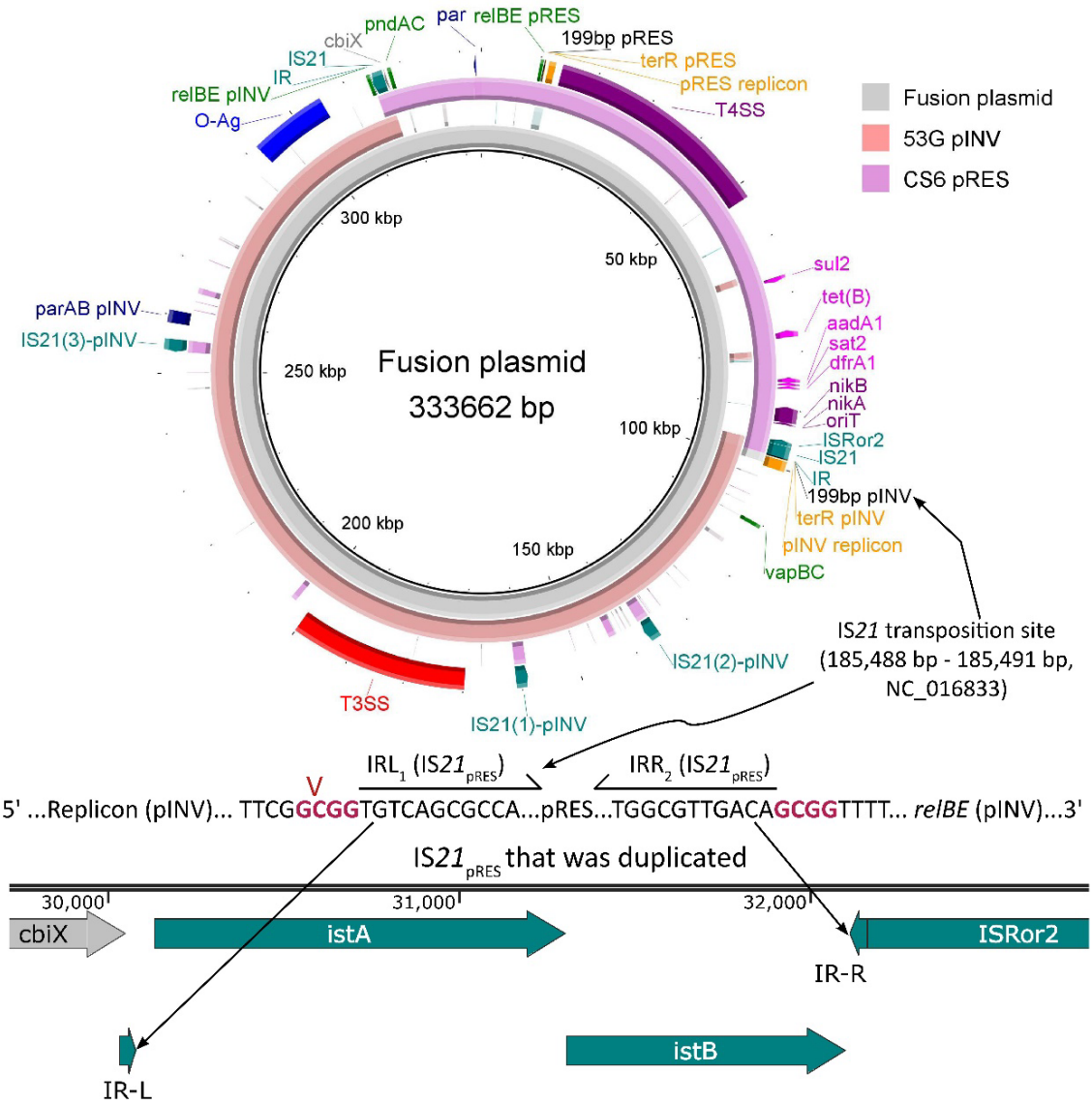


Figure 4.3.2B: Alignment of the pINV/pRES fusion plasmid receovered from a Δ*recA* transconjugants. The nucleotide sequence of the fusion plasmid was aligned against 53G pINV (coral) and CS6 pRES (purple) as reference. The site of IS*21*$_{pRES}$ insertion is expanded and the location on 53G pINV (accession no. NC_016833) labelled. The duplication of 4 bp at the site of IS*21*$_{pRES}$ insertion was labelled in magenta, with an arrow pointing at the site of transposition. IS*21* from pRES was shown on the top of the figure. Specific elements were annotated and colour-coded: replicons, yellow; pINV *vapBC* and *relBE*, pRES *relBE* and *pndAC*, green; pINV *parAB* and pRES *parB* and *parM*, pine; the T3SS and virulence genes, red; ISs, pale blue; T4SS and conjugation-related genes, purple; ORFs, grey.

As IS*21* transposition can mediate formation of fusion plasmids, I next performed bioinformatic analyses to identify this IS in *S. sonnei* plasmids. I interrogated complete closed circular sequences of a collection of 150 plasmids from *S. sonnei* including an extra 16 pINV sequences in addition to the 134 plasmids in Figure 3.1.3B. Only complete copies of IS*21* which harbour both IRs and *istAB* were considered as this means they could be functional. Results reveal that

IS*21* is present on all available examples of *S. sonnei* pINV (Figure 4.3.2C), with 15 out of 16 pINV contain three copies of IS*21*, while one pINV contains two copies.

In contrast, single copies of IS*21* were only detected in three out of 29 IncI1 plasmids, one out of 30 IncB/O/K/Z plasmids and four out of 50 IncFII plasmids. It is possible that IS-mediated HR and transposition could be mediated by other ISs. This is supported by identification of a hybrid plasmid that might have resulted from by transposition of IS*630* from pINV (Figure 4.2.2C).

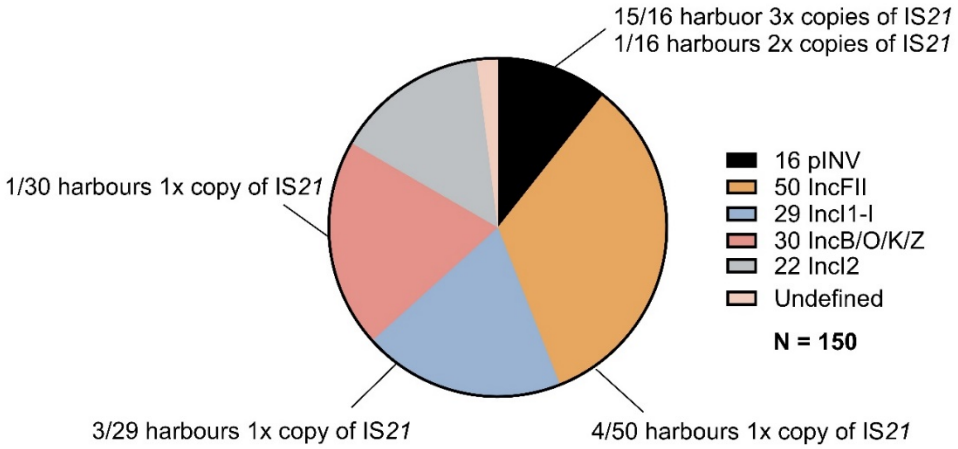
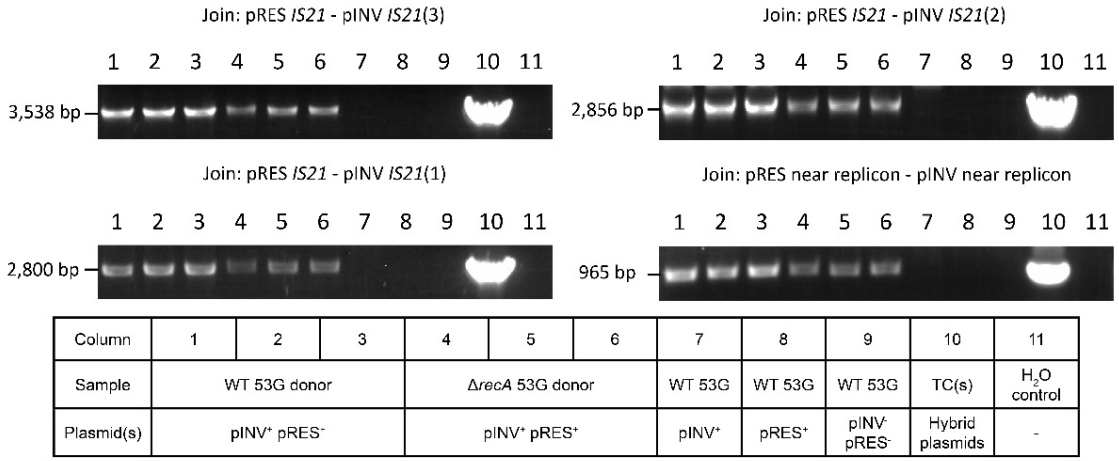


Figure 4.3.2C: IS*21* is not found on all *S. sonnei* plasmids.
Pie chart showing the proportion of plasmids belong to different Inc groups. The number of plasmids in the corresponding groups is labelled in the legend. The number of plasmids with the number of copies of IS*21* detected is annotated.

4.3.3 RecA-dependent and –independent HR could lead to formation and resolution of fusion plasmids

Interestingly, deletion of RecA in the recipient did not abolish the resolution of fusion plasmids. The detection of hybrid plasmids in Δ*recA* transconjugants indicates that RecA-independent events can resolve fusion plasmids. Therefore, similar events could also contribute to the formation of fusion plasmids in the donor. It was hypothesised that if the formation of fusion plasmids could occur spontaneously in donor strains containing pINV and pRES, the junctions between pINV and pRES should be detectable by PCR. Therefore, PCR was performed to detect the four most frequent junctions (involving copies of IS*21* and the 199 bp region) in wild-type and Δ*recA S. sonnei* donors harbouring pINV and pRES. *S. sonnei* 53G with pINV or pRES only,

plasmid-less 53G, and sequenced transconjugants with hybrid plasmids with known pINV/pRES junctions were included as controls. Results show the presence of these junctions detected by PCR in both wild-type and Δ*recA* donors, while these bands were not amplified from 53G with pINV or pRES alone, or from plasmid-less 53G (Figure 4.3.3A). These suggest that HR to form fusion plasmids can be RecA-independent. However, the frequency of fusion plasmid formation mediated by RecA-independent event is likely to be lower than RecA-dependent HR and limit the detection of transconjugants mediated by non-HR events, explaining why pINV transfer was not detected when *recA* was deleted in the donor.



Column	1	2	3	4	5	6	7	8	9	10	11
Sample	WT 53G donor			Δ*recA* 53G donor			WT 53G	WT 53G	WT 53G	TC(s)	H_2O control
Plasmid(s)	pINV$^+$ pRES$^-$			pINV$^+$ pRES$^+$			pINV$^+$	pRES$^+$	pINV$^-$ pRES$^-$	Hybrid plasmids	-

Figure 4.3.2A: Detection of junctions from hybrid plasmids in wild-type and Δ*recA* donors.
Three single CR$^+$ colonies of WT and Δ*recA S. sonnei* 53G grown on CRA with antibiotics were boiled at 95°C for five minutes, then subjected to PCR using primers to detect common pINV/pRES junctions. Long read sequenced transconjugants were used as positive controls in PCR analyses.

4.3.4 Proposed model for pINV transfer by pRES

Taken together, my results lead to a model of pINV transfer mediated by pRES. I propose that pINV transfer happens through multiple steps. First, in the donor, large pINV/pRES fusion plasmids or pINV/pRES hybrid plasmids are formed by HR and/or IS transposition. A large fusion plasmid with an intact conjugation system or a hybrid plasmid with *oriT* is transferred from the donor into the recipient *via* the T4SS and the pRES relaxase. Any hybrid plasmid formed in the donor that lacks an *oriT* would not be mobilised so not detected in the recipient. In the recipient,

further genetic events (including HR and IS transposition) occur to resolve the large fusion plasmid into pINV±pRES or hybrid plasmids. Wild-type plasmids would reform in transconjugants by recombination at the same site that generated the fusion plamid in the donor; hybrid plasmids would appear through the second event occurring at a distinct site to that involved in the formation of the fusion plasmid. Alternatively, the fusion/hybrid plasmid formed in the donor would remain unchanged after entering the recipient. Finally, wild-type or hybrid plasmids could be lost during subsequent growth and therefore not be detected.

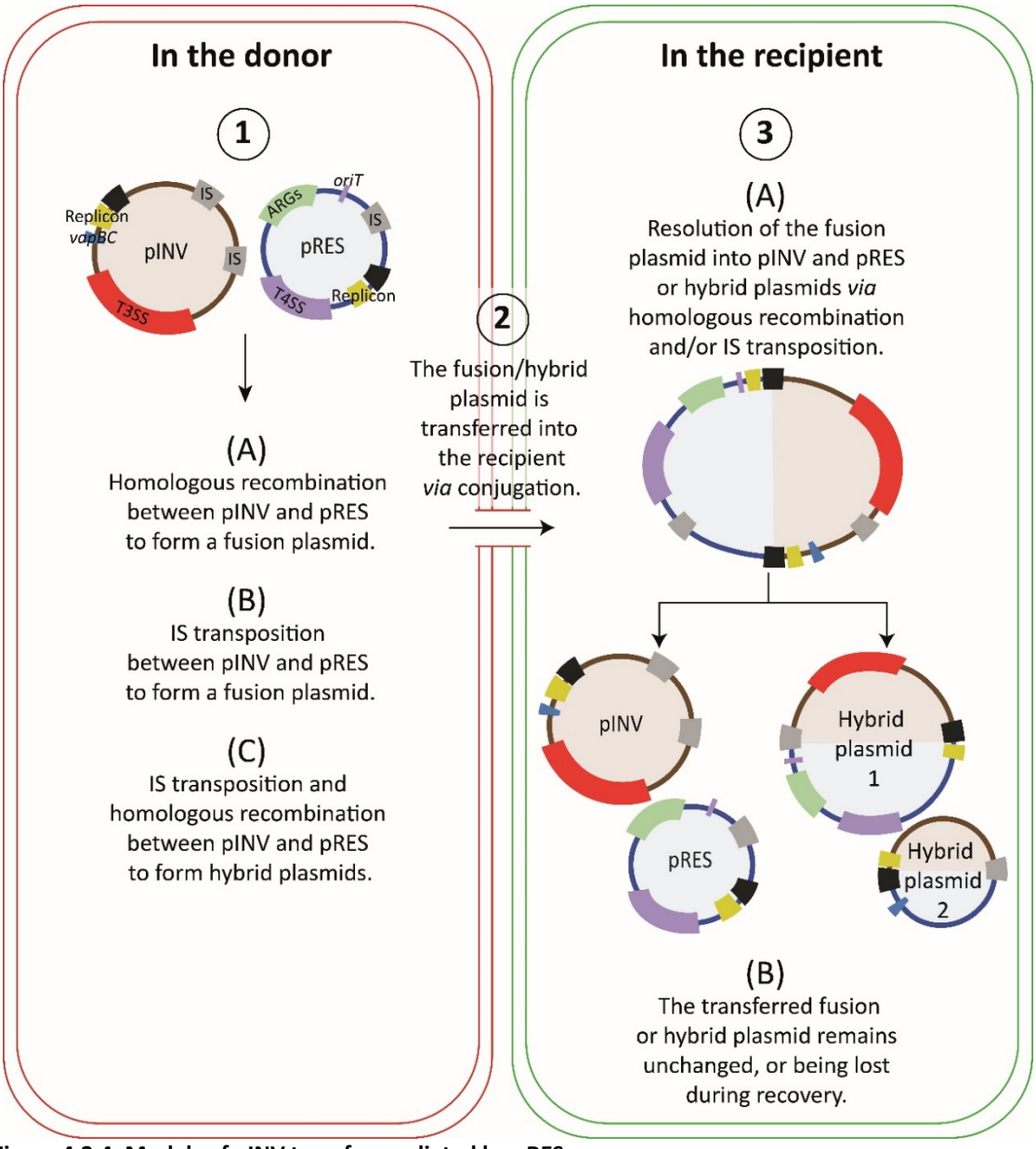


Figure 4.3.4: Models of pINV transfer mediated by pRES.
pINV transfer occurs through multiple steps: (1) Formation of a fusion or hybrid plasmid between pINV and pRES; (2) Conjugal transfer of the fusion or hybrid plasmid into the recipient; (3) The fusion plasmid resolves into pINV±pRES, or the fusion or hybrid plasmids remain in the recipient.

4.4 Summary

In this chapter, I demonstrated that pINV can be transferred into a new bacterial host in the presence of a naturally occurring IncB/O/K/Z resistance plasmid, pRES. The transfer of pINV facilitated by pRES demonstrates the potential for cross-lineage dissemination of virulence factors and antibiotic resistance determinants in *S. sonnei*. While pRES was obtained from a Lineage III strain, the *S. sonnei* background strain used for the experiments (53G) belongs to Lineage II. *S. sonnei* 53G was first isolated in Japan before 1980, and has been used in many studies (237). Interestingly, the version of the strain we have in the laboratory differs from the version that was used to determine its genome sequence. This is evidenced by the discrepancy in the small plasmids found in the two strains (Figure 3.1.4C).

I characterised the mechanisms involved in pINV transfer mediated by pRES. pINV has been characterised as a non-mobilisable plasmid (201), and it is not self-transmissible (Figure 4.1.1A) as it has no *oriT* and only harbours remnants of a MOB_F-type conjugation system encoding a relaxase TraI and coupling protein TraD, which are likely non-functional as predicted by bioinformatic analysis (Figure 3.1.1C and D). However, by forming a fusion or hybrid plasmids with pRES, part or the entire pINV can be mobilised between bacteria when the relaxase is provided in *cis* or *trans* to mobilise a molecule harbouring the *oriT* sequence (348).

I also showed that pINV transfer *via* HR between pINV and pRES leads to the formation of hybrid mosaic plasmids which harbour elements originating from the two plasmids. The emergence of these plasmids was reflected by the range of colony phenotypes when growing TCs on CRA (Figure 4.2.1A). As the resulting hybrid plasmids can lose or gain elements important for plasmid maintenance, some hybrid plasmids might not be stably maintained and would be lost overtime, while others might be successfully maintained in the bacterial population. This could lead to gain or loss of key features affecting bacterial growth and survival under particular conditions, potentially influencing inter-bacterial competition and bacterial pathogenesis, and shaping

evolution of bacteria. Further investigation into hybrid plasmids, including their maintenance, growth, and ability to be transferred horizontally, could lead to further understanding of how events around conjugation could give rise to new plasmid variants through re-assortment of genes on plasmids co-existing in a single bacterium.

My results demonstrate the importance of RecA-dependent and -independent HR events during pINV transfer. Deleting *recA* in the donor resulted in pINV transfer frequency below the limit of detection of the assay (10^{-11}), without affecting pRES transfer frequency (Figure 4.3.1F). RecA-independent recombination pathways involving RarA occur at low frequency in bacteria (213), so the low frequency of RecA-independent recombination could explain the failure to detect pINV transfer in the absence of RecA in the donor. While RecA plays a role during formation of hybrid pINV/pRES plasmids, the pathways underlying *recA*-independent recombination before plasmid transfer could be identified. This can initially be achieved by quantifying recombination frequencies in the absence of both *recA* and *rarA*, which was involved in *recA*-independent HR (213). pINV transfer frequency could be examined in the absence of RarA in combination of RecA to investigate the impact of these recombination events to pINV transfer. Further characterisation of RecA-independent pINV transfer could be carried out at single-cell level which is more sensitive than the current conjugation assay. As pINV transfer occurs at a low frequency (Figure 4.1.1A), pINV transfer frequency could be dependent on the frequency of HR and/or IS-mediated transposition.

pINV is a mosaic plasmid as demonstrated by the abundance of ISs and different GC content across the plasmid backbone (44). HR between identical copies of ISs can lead to deletion of important elements, such as the T3SS PAI (163). Similarly, HR between ISs can also lead to gain of elements, such as forming a fusion plasmid with pRES and acquisition of genes conferring the ability to be horizontally transferred. Results also show that pINV can be integrated into the chromosome by IS*1*-mediated HR (Figure 4.2.2A), similar to previous findings for pINV of *S.*

flexneri which was mediated by HR between copies of IS*1294* on pINV and the chromosome (163). Although some copies of ISs on pINV and the chromosome are incomplete, they could still act as sites for HR.

The findings also shed light on the importance of IS*21* during pINV transfer, either due to HR or transposition (Table 4.1). It is interesting that IS*21* is not always found on conjugative plasmids in *S. sonnei* despite its prominence as an element generating fusion or hybrid plasmids in my experiments (Figure 4.1.1B). There might be some important features of IS*21* that specifically contribute to this process which remain to be elucidated.

ISs play a significant role in shaping evolution of the plasmids and the pathogens, which could contribute to affecting AMR and virulence in the pathogens (39, 44, 84, 163, 255, 349, 350). Although *S. sonnei* pINV lacks an intact T4SS, it can still be horizontally transferred by forming a fusion/hybrid molecule with the conjugative pRES through HR and/or IS transposition. IS*21*, frequently detected at the pINV/pRES junctions, plays a major role in this process. Since IS*21* is also found on the *S. sonnei* chromosome, its involvement may indicate co-evolution of pINV with the bacteria or a compensatory mechanism for the loss of T4SS, allowing continued horizontal transfer of pINV. IS*630* also contributes to the formation of hybrid pINV/pRES plasmids. This suggests that different ISs might play a similar role to facilitate pINV transfer by plasmids other than pRES, as IS*21* is not always present on conjugative plasmids (Figure 4.3.2C). As ISs are found in other bacteria, the contribution of ISs in shaping the transfer of originally non-mobilisable elements *via* HR or IS transposition could be widely applicable. Furthermore, non-IS sequences, the 199 bp region is also a frequent site of pINV/pRES junctions. Therefore, plasmids with this sequence might have a higher propensity to transfer pINV. The impact of these processes on the ability of *S. sonnei* to acquire novel genetic traits and the diversity of its conjugative plasmids remains to be understood. In addition, *S. sonnei* could be an effective

model for studying plasmid evolution, providing insights applicable to other host-plasmid systems.

Altogether, results in this chapter demonstrate the importance of plasmid interactions in mediating pINV transfer. In addition to ISs, pINV/pRES junctions can occur at a 199 bp region of homology downstream of plasmid replicons (Table 4.2). This region of homology will be characterised in the next chapter.

5 Characterisation of a conserved region close to replicons of a range of plasmids

In the previous chapter, I investigated the mechanisms of the horizontal transfer of pINV mediated by the resistance plasmid, pRES. I proposed a model that pINV and pRES first form a fusion/hybrid molecule due to RecA-dependent or -independent HR or IS transposition, then the entire molecule is transferred into the recipient by conjugation (Figure 4.3.4). Sequencing results demonstrate that a 199 bp region downstream of the replicons in pINV and pRES was the site of pINV/pRES junctions in hybrid plasmids that were frequently found in transconjugants (Table 4.1 and 4.3). The sequence of the 199 bp region (designated 199R) is highly conserved between pINV and pRES which are IncFII and IncB/O/K/Z plasmids, respectively. Both these plasmids replicate unidirectionally (Figure 1.3.1A, B and C) (287, 302).

As the 199R is located close to replicons and is highly conserved in these two plasmids that belong to different Inc groups, I decided to investigate the function of the 199R. I first analysed the 199R on pINV and pRES, as well as sequences from a collection of 147 *S. sonnei* plasmids bioinformatically. I then characterised the importance of the 199R on both pINV and pRES by CRLA and PLA (Chapter 2.4 and 2.5). I also explored the potential usage of the conserved 199R in plasmid typing and evolution analysis using command lines and RStudio.

5.1 Characterisation of the 199R from pINV and pRES

5.1.1 53G pINV and CS6 pRES harbour a conserved 199 bp region downstream their replicons

Initially, I examined the genetic context of the 199R from pINV and pRES in relation to sequences that are known to be involved in the replication of plasmids. The 199R is located 593 bp and 582 bp downstream of *repA* stop codon on 53G pINV and CS6 pRES, respectively (Figure 5.1.1A). A

terR sequence, with the core region (5´- NNGNNTGTTGTAACTANN -3´) which is recognised by the Tus protein, is located 34 bp and 40 bp upstream of the 199R on pINV and pRES, respectively (351). Binding of Tus to *ter* sites block the replication fork and lead to the termination of replication (352). In addition, there is a *terL* sequence identified on pINV, 113 bp upstream of the *terR* site, which is not found on pRES. The *terL* site is not involved in termination of plasmid replication in the closely related R1 plasmid (296).

The R1 plasmid is an IncFII plasmid that belongs to the same Inc group as pINV (286). The replication and termination of replication of the R1 plasmid has been extensively studied (287, 296). In both IncFII and IncB/O/K/Z plasmids, the 199R overlaps with a single stranded initiation site (*ssi*), where the DnaG primase synthesises a complementary RNA primer that allows the start of synthesis of the leading strand of DNA (294, 302); the site has been mapped by defining the junction between RNA and DNA at the start of the newly synthesised leading strand of the R1 plasmid; the *ssi* is at 140 bp of the 199R (294).

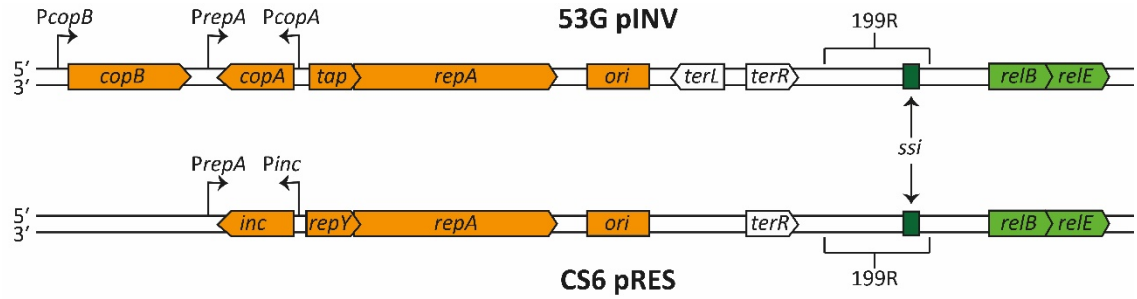


Figure 5.1.1A: Features around the 199R on *S. sonnei* 53G pINV (upper diagram) and CS6 pRES (lower). Important elements in plasmid replication and termination are annotated. The 199R overlaps with the *ssi* which is the template for synthesis of an RNA primer that marks the initiation of synthesis of the leading strand. Plasmid replication proceeds left to right in a unidirectional manner. Figure not to scale.

In the model of R1 plasmid replication proposed by Masai and Arai (Figure 5.1.1B) (294), binding of the RepA protein to the *ori* sequence leads to melting of the downstream DNA duplex, thereby exposing both strands as ssDNA. This allows replication of lagging strand of DNA (which proceeds in an anti-clockwise direction, right to left in Figure 5.1.1A and B) through the synthesis of Okazaki fragments. Exposure of the *ssi* allows synthesis of an RNA primer that is required to

initiate DNA replication of the leading strand. The replication fork progresses unidirectionally in a clockwise direction (left to right in Figure 5.1.1A and B) around the plasmid until it reaches the *terR* site.

There is a gap between the *ori* and *ssi* which *in vitro* studies indicate will only be replicated towards the end of plasmid replication and could be a slow process (294). Therefore, during plasmid replication, this gap (Figure 5.1.1B), which includes part of the 199R, could remain single stranded, and thus be recognised by recombinases such as RecA and be the site of HR (347). HR at the highly conserved 199Rs from pINV and pRES, would lead to the formation of fusion plasmids, which could then be transferred into another bacterium by conjugation. This could explain the high frequency of the 199R at pINV/pRES junctions in fusion and hybrid plasmids.

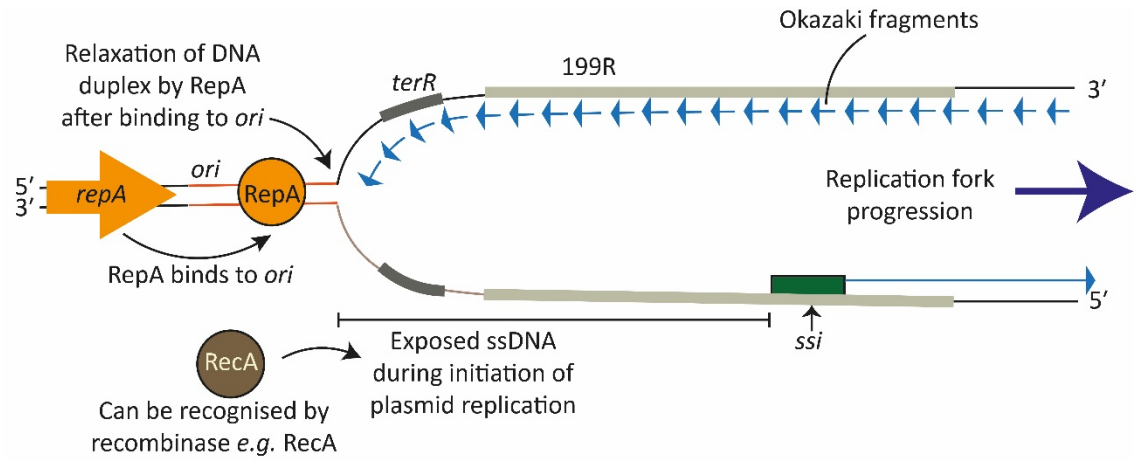


Figure 5.1.1B: Location of the 199R relative to sequences involved in plasmid replication and termination.
Model of replication of the R1 plasmid (from Masai and Arai 1989) (294), which has a replicon closely related to pINV RepA binds to the ori and unwinds DNA downstream. Replication of the leading strand starts at the *ssi* site (green box) where an RNA primer is synthesised; lagging strand replication *via* Okazaki fragments occurs on the 5′-strand when it is single stranded. Replication is unidirectional, and ends at the terR site (dark grey box) once the rest of the plasmid has been synthesised; the site of the 199R is indicated (light grey box). A ssDNA region between the *ori* and *ssi* on the 3′-strand is replicated once the rest of the plasmid has been replicated. This single stranded region could be recognised by be recognised by RecA, leading to HR with homologous sequence on a different plasmid, resulting in fusion or hybrid plasmids. Figure not to scale.

Therefore, existing models for plasmid replication indicate that the 199R includes sequences that are important for plasmid replication and maintenance as it harbours the *ssi* site (Figure

5.1.1B). Further investigation of the 199R was performed to understand the contribution of this region to plasmid maintenance.

5.1.2 The 199R is conserved in a wide range of plasmids in *S. sonnei*

As the 199R in pINV and pRES share a high level of sequence identity (99.5%) from *S. sonnei* which belong to different Inc groups, the conservation of this sequence was examined in other *S. sonnei* plasmids. Initially, sequences of plasmids from different Inc groups were investigated, including pINV from *S. sonnei* 53G and Ss046, and those belonging to IncFII (accession no. MW396858) (353), IncI1 (accession no. AB021078) (276), IncB/O/K/Z (accession no. MG299131) (181), and IncI2 (accession no. MG299138) (181), which replicate unidirectionally (303, 354, 355). Nucleotide sequences were aligned by Minimap2 (322) using the 53G pINV sequence as the reference, and output as a SAM file. SAMTOOLS (356) was used to convert the SAM file to a BAM file, which sorted and indexed the BAM file for visualisation using Integrative Genomics Viewer (IGV) (357). Alignment of the sequences demonstrates that the region of conservation in all six plasmids extends from the terR sequence to beyond the 199R (Figure 5.1.2A, Figure 7.10). IncI1 (NC_002122.1) and IncI2 (NZ_MG299138.1) plasmids share smaller region of homology with other plasmids, spanning from 1,336 - 1,706 bp and 1,323 - 1,708 bp, respectively. Interestingly, the IncFII (NZ_MW396858.1) and IncB/O/K/Z (NZ_MG299131.1) plasmids harbour *repA* genes which highly homologous to *repA* from 53G pINV (NC_016833.1), with nucleotide identities of 93.0% and 84.7%, respectively, indicating the close relationship between these plasmid groups (358). Results also demonstrate the importance of this region among plasmids from these groups as the nucleotide identity of the region including terR and the 199R was high (> 90%) compared with the region on 53G pINV.

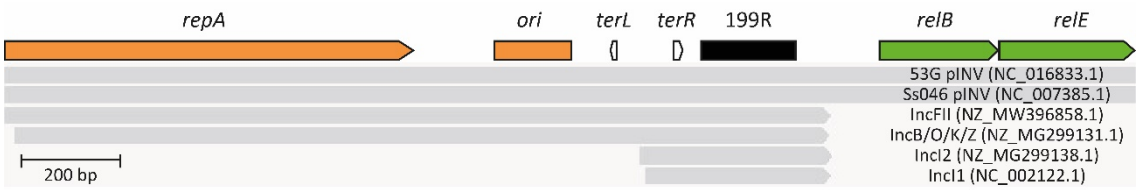


Figure 5.1.2A: Alignment of sequences around the 199R from plasmids belonging to different Inc groups. Features were annotated from *S. sonnei* 53G pINV. Alignments with nucleotide identity of > 90% are shown in grey. The accession numbers of plasmid sequences are shown in brackets. Image exported from IGV.

To investigate the conservation of this region in other plasmids, the presence of the 199R was next determined in a collection of 147 plasmids (excluding three uncharacterised plasmids from the previous collection of 150 plasmids in Figure 4.3.2C) from clinical isolates of *S. sonnei*, compiled by Dr. Arcari (La Sapienza, Rome). This collection contains assembled, closed circular long-read sequences of plasmids > 40 kb in size from different Inc groups, including 16 IncFII pINV, 50 IncFII plasmids, 29 IncI1 plasmids, 30 IncB/O/K/Z plasmids, and 22 IncI2 plasmids. Plasmids from these groups replicate unidirectionally using similar mechanisms (276, 290, 303, 354), with no bi-directionally replicating plasmid in the collection. Nucleotide sequences from the gene encoding the replication initiation protein up to the next ORF downstream of 199R were extracted and manually annotated.

Results reveal that all these unidirectionally replicating plasmids harbour sequences related to the 199R. The genetic arrangement of the region surrounding the 199R is highly conserved among the plasmids. A *terR* sequence is situated downstream of *repA* and is subsequently followed by the 199R, reflecting the genetic arrangement of *repA*-*terR*-199R (Table 7.2). However, there is significant variation downstream of the 199R in these plasmids (Table 5.1, for more details see Table 7.2). In pINV, a *relBE* TA system is located downstream of the 199R. A total of 48 out of 50 IncFII plasmids harbour a gene encoding a CPBP family intramembrane metalloprotease according to BLASTx, with two exceptions harbouring *yacAB* or IS*1294*. Plasmids from the other three Inc groups possess a wider variety of elements downstream of the 199R. In IncI1 plasmids, a *yacAB* TA system is found in 13 out of 29 plasmids, while another

13 plasmids have a gene encoding hypothetical protein located downstream of the 199R. Another IncI1 plasmid harbours a *relBE* TA system, and two plasmids carry an IS*91*-family transposase downstream of the 199R. IncB/O/K/Z group plasmids either harbour a *relBE* or *yacAB* TA system downstream of the 199R. Interestingly, IncI2 plasmids harbour different MGEs downstream of the 199R (3 of 22, IS*91*-family transposase; 8 of 22, IS*Ec44*), with the rest (11 out of 22) harbouring a *hok/sok* TA system.

Inc groups	No. of plasmids	Elements downstream of replicons
IncFII (pINV)	16	*relBE* (16/16)
IncFII	50	CPBP family intramembrane metalloprotease (48/50); *yacAB* (1/50) or IS*1294* (1/50)
IncI1	29	*yacAB* (13/29); hypothetical protein (13/29)
IncB/O/K/Z	30	*relBE* (15/30); *yacAB* (15/30)
IncI2	22	IS*91*-family transposase (3/22); IS*Ec44* (8/22); *hok/sok* (11/22)

Table 5.1: Summary of elements downstream of the 199R from plasmids of different Inc groups.

Overall, the results show that the arrangement of *repA*, *terR*, and the 199R is highly conserved, even in plasmids from different Inc groups, indicating this sequence probably has an important function in plasmids that replicate in a unidirectional manner. The conservation of sequences in and around the 199R also suggests that genetic rearrangement could occur at this site in plasmids belonging to a broad range of Inc groups, as already observed between pINV and pRES. However, elements downstream of the 199R differ even among plasmids from the same Inc group.

5.2 The contribution of the 199R in plasmid maintenance

Although the 199R is conserved in different plasmids from *S. sonnei*, this region has not been studied systematically. Previously, a 218 bp deletion (here designated as 218B) arose in the R1 plasmid downstream of the replicon following attempts to reverse the orientation of its *ori* (285, 359). The 218B deletion includes *terR* and led to increased plasmid loss (293). There is a 99 bp overlap (here designated as 99O, Figure 5.2.1A) between the 218B and 199R. As the 199R does not include the *terR* site, but encompasses the *ssi*, the site of synthesis of the RNA primer for leading strand replication, the contribution of the 199R and 99O to plasmid maintenance was investigated.

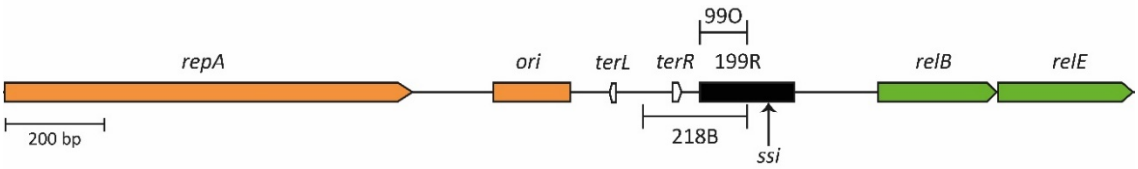


Figure 5.2.1A: A schematic of 218B, 199R and 99O from 53G pINV.
The 218B includes *terR*, but does not overlap with the *ssi*. The 218B overlaps with the 199R over 99 bp (99O), which does not include *terR* or the *ssi*.

5.2.1 The conservation of sequences extends beyond the 199R

Initially, bioinformatic analysis was performed to investigate the conservation of the sequence lost in the 218B deletion in other plasmids. The set of six plasmids from different Inc groups described in Figure 5.1.2A was examined. As the *repA* genes are not conserved in these plasmids, sequences from the end of *repA* until the nucleotide before the next ORF downstream were extracted and analysed (Figure 7.10). Results reveal that sequence conservation between the plasmids begins just after the start of the 218B that was characterised by Burger (293) (Figure 5.2.1B), with a nucleotide identity of > 92.9%.

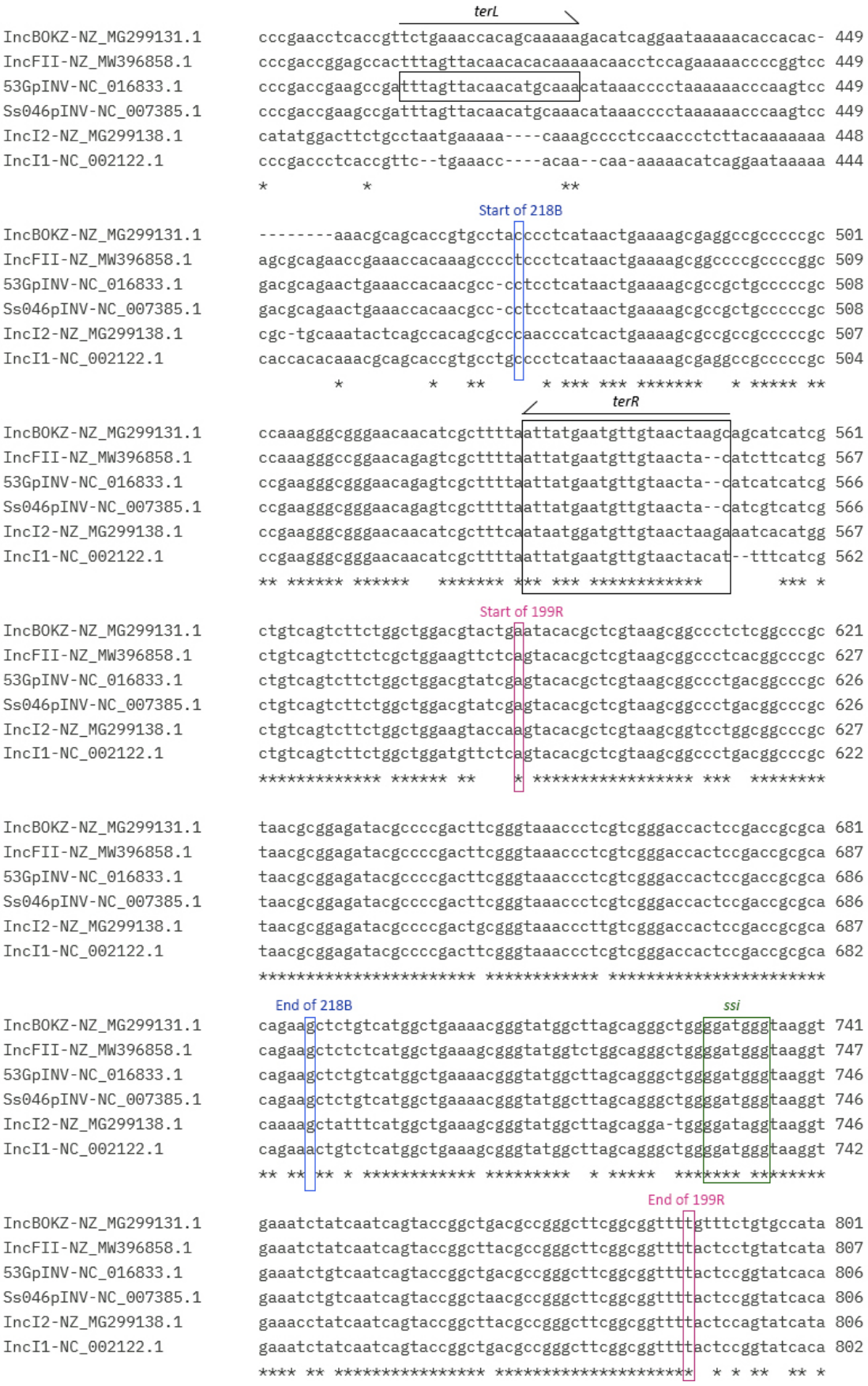


Figure 5.2.1B: Nucleotide alignment of the region downstream of *repA* in different plasmids.
Elements were annotated according to Burger 1983 (293) with *S. sonnei* 53G pINV as the reference. The alignment was generated using ClustalOmega. Asterisk (*) indicates identical nucleotides.

5.2.2 Deletion of the 199R increases loss of pINV and pRES from *S. sonnei*

To investigate the contribution of the 199R to plasmid biology, the 199R was deleted from 53G pINV and CS6 pRES, and frequency of plasmid loss assessed. As these plasmids belong to different Inc groups, this analysis would demonstrate whether the 199R affects plasmid loss irrespective of Inc group. Plasmid loss was measured using the CRLA for pINV, and a plasmid loss assay (PLA) using a *sacB::bla*$_{TEM\text{-}116}$ inserted into pRES.

A *S. sonnei* 53G strain, Δ199R, was generated containing pINV with a markerless deletion of the 199R and *cat* gene inserted downstream of *vapBC*. The presence of *cat* on pINV allowed selection of bacteria retaining the plasmid while making frozen stocks. If bacteria lack *vapBC*, they are likely to have lost pINV as CR$^-$ bacteria from *S. sonnei* usually arise through plasmid loss (163). CRLAs were performed by streaking bacteria from frozen stocks onto solid media with CR but without antibiotics, then picking single CR$^+$ colonies into PBS, then plating dilutions to media with CR to determine the proportion of CR$^+$ and CR$^-$ bacteria that were in the colony. However, a range of colony phenotypes were observed when *S. sonnei* Δ199R was grown from frozen stocks without antibiotics; most colonies of this strain appeared to be almost entirely CR$^-$, indicating a very high level of plasmid loss during the growth of the strain from frozen stocks. Therefore, the most CR positive colonies of *S. sonnei* Δ199R were selected for CRLAs. For each of the three replicates, three CR$^+$ colonies were selected for each strain, and resuspended in 100 µl PBS, serially diluted then inoculated onto on plain CRA, and incubated at 37°C for 16 hours. Colonies were scored as CR$^+$ if they had any detectable red coloration, as this would indicate that the colony arose from a CR$^+$ bacterium but then pINV was lost subsequently during cell division as the colony formed over the 16-hour incubation (Figure 5.2.2A).

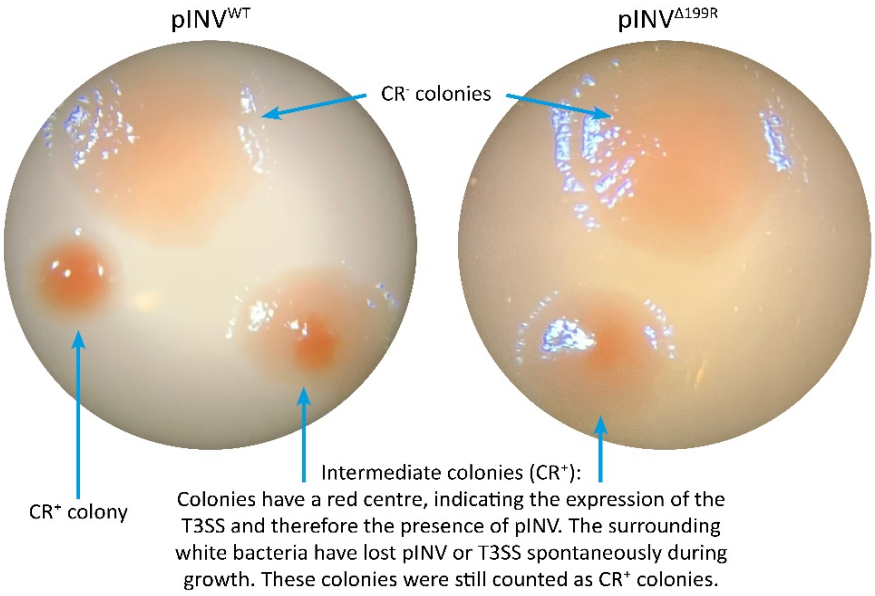


Figure 5.2.2A: CR phenotypes of 53G with pINV^{WT} (left) or pINV$^{\Delta 199R}$ (right).
The colonies were visualised by light microscopy. The shade in the figure was the edge of the lens.

Results of the CRLA demonstrate that deletion of the 199R from *S. sonnei* 53G pINV significantly increased the frequency of plasmid loss from 7.9 x 10^{-2} for the wild-type plasmid to 8.5 x 10^{-1} for pINV$^{\Delta 199R}$ ($p < 0.0001$, Figure 5.2.1C). While CR$^+$ colonies were assumed to still harbour pINV, 10 CR$^-$ colonies per plate (*i.e.* 30 colonies per replicate) were randomly selected and checked for the presence of pINV by multiplex PCR (amplifying 271 bp of *ori*, 1,105 bp of *virB*, 905 bp of *virF*, 1,458 bp of *hns*). From wild-type *S. sonnei* 53G, a total of 88 out of 90 CR$^-$ colonies had lost pINV with two harbouring the plasmid with *virB* and *virF* deletion. For *S. sonnei* Δ199R, all 90 CR$^-$ colonies had lost pINV. Therefore, the 199R is important for pINV maintenance.

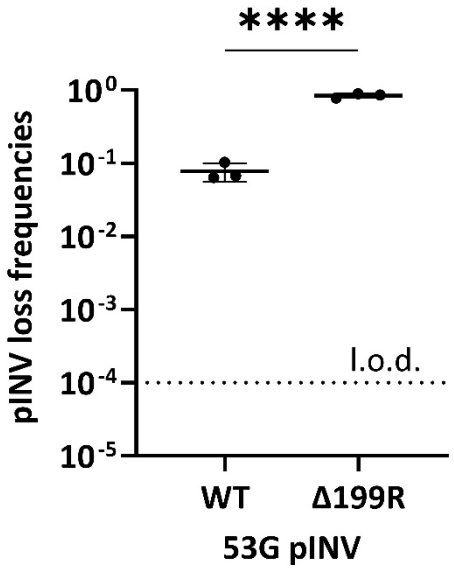


Figure 5.2.2B: pINV$^{\Delta 199R}$ is lost more frequently from *S. sonnei* than pINV^{WT}
Each dot represents results for three colonies from a replicate. Data were subjected to Shapiro-Wilk tests and unpaired t-test to assess the differences in the frequency of pINV loss which is shown as the proportion of the entire population. ****, $p < 0.0001$. (N = 3).

Next, the 199R was deleted from the IncB/O/K/Z plasmid pRES generating pRES$^{\Delta 199R}$; PLAs were performed in *S. sonnei* 53G lacking pINV but harbouring pRES containing *sacB* inserted downstream of *aadA1* along with *bla*$_{TEM\text{-}116}$ as in previous assays (Figure 3.2.1C); the presence of *sacB* allowed for the selection of bacteria that had lost the plasmid/*sacB* so can grow on media containing sucrose. For the PLA, strains were grown from the frozen stocks on plain LBA at 37°C for 16 hours. In each replicate, three single colonies were selected, diluted in PBS, then inoculated onto LBA with carbenicillin (to recover bacteria harbouring pRES), or salt-free LBA with 10% sucrose (to recover bacteria which had lost pRES). To confirm loss of pRES, 10 colonies recovered on LBA containing sucrose were re-streaked onto LBA with carbenicillin to detect the presence of pRES in these colonies. Re-streaked colonies that grew on LBA with carbenicillin were re-streaked back to LBA with sucrose to determine whether these colonies harboured pRES but were resistant to sucrose. Additionally, the presence of pRES was assessed by PCR of a 1,698 bp fragment from the pRES replicon. A total of 90 sucrose-resistant colonies (10 colonies for each of three colonies in three separate experiments) obtained from pRES$^{\Delta 199R}$ lacked pRES, while 13 out of 90 pRES^{WT} colonies still harbour pRES and sucrose resistance. pRES loss frequencies were quantified as the proportion of the bacterial population that was both sucrose resistant and carbenicillin sensitive, and were confirmed to have lost pRES by PCR. Results demonstrate that the loss of pRES was less frequent than loss of pINV, potentially due to the different repertoire of TA systems on these plasmids. The results indicate that pRES$^{\Delta 199R}$ is lost more frequently than the wild-type pRES, with loss frequencies ranging from 9.5×10^{-6} to 2.9×10^{-4} for the wild-type pRES and pRES$^{\Delta 199R}$, respectively (Figure 5.2.2C). However, despite a mean difference of 2.8×10^{-4} between the two groups, the difference was not statistically significant, likely due to the small sample size (N=3) and resulting low statistical power, which is a limitation of the study. Nevertheless, the data suggest that the 199R region plays a critical role in maintaining both pRES and pINV.

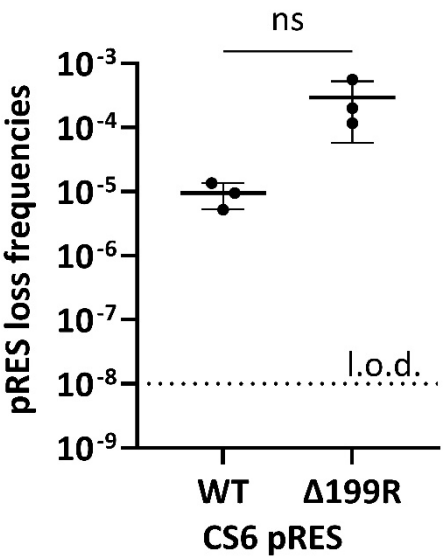


Figure 5.2.2C: Deletion of 199R from pRES increases plasmid loss
Each point represents three colonies in a replicate. Plasmid loss frequencies of pRES^{WT} in the absence of pINV also shown earlier in Figure 3.2.1C. Unpaired t-test was applied to compare pRES loss between the two strains. ns, not significant. (N=3)

5.2.3 A 99 bp sequence where the 218B overlaps with the 199R affects plasmid maintenance

The *terR* site was lost from the R1 plasmid following the 218B deletion which is therefore expected to affect the termination of plasmid replication (293). In contrast, removal of the 199R from pINV and pRES eliminates the *ssi* site so is likely to reduce the replication of the leading strand; initiation of replication can start at other sites but at different efficiency (360). Thus, the 218B and 199R might contribute to plasmid maintenance through distinct processes.

However, the 218B and 199R overlap by 99 bp which contains neither the *terR* nor *ssi* site. Therefore, to investigate whether there is also a shared mechanism by which 199R and 218B contribute to plasmid maintenance, the 99O was deleted from pINV (generating pINV$^{\Delta 99O}$) in *S. sonnei* 53G. The strain with pINV$^{\Delta 99O}$ was subjected to CRLA as above, with *S. sonnei* 53G containing pINV^{WT} or pINV$^{\Delta 199R}$ used as controls. Interestingly, deletion of 99O alone resulted in similar colony phenotypes as pINV$^{\Delta 199R}$ when grown on CRA without antibiotics, with colonies appearing mostly CR^{-} (Figure 5.2.3A). However, individual colonies of bacteria containing pINV$^{\Delta 99O}$ were of higher redness than colonies of bacteria with pINV$^{\Delta 199R}$. Therefore, a similar approach was employed to score CRLAs with colonies containing any visible redness counted as

CR^+. Plasmid loss frequencies were quantified as the proportion of $pINV^-$ bacteria within the entire population.

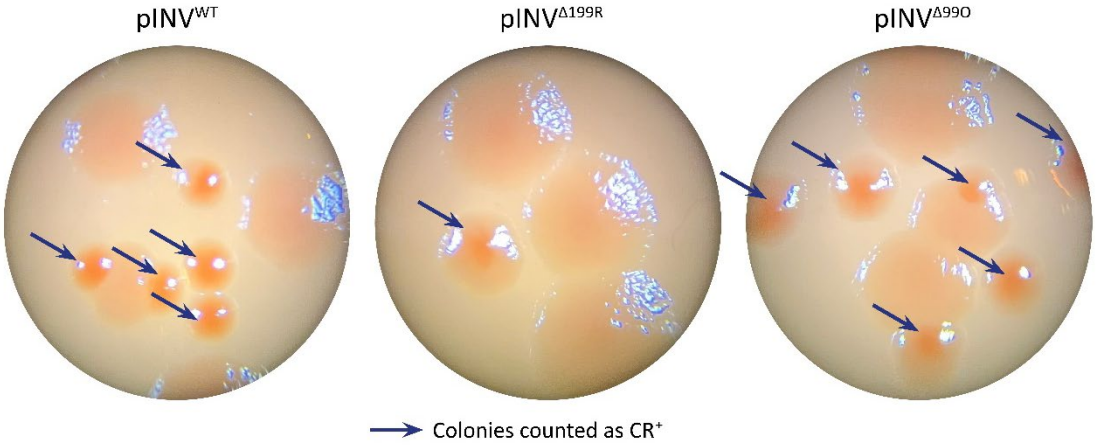


Figure 5.2.3A: Phenotypes of colonies of *S. sonnei* containing $pINV^{\Delta 99O}$, $pINV^{WT}$ or $pINV^{\Delta 199R}$.
Bacteria were grown on media without antibiotics and observed with a light microscope. Arrows indicate the colonies counted as CR^+. The shade surrounding the figures was the edge of the lens of the microscopes.

A CRLA was performed as above. The absence of pINV in CR^- colonies was checked by PCR. Results reveal that deletion of 99O from pINV increased plasmid loss compared with $pINV^{WT}$ (bacteria with plasmid loss as a proportion of the entire population, 1.2×10^{-1} to 6.2×10^{-1} for $pINV^{WT}$ and $pINV^{\Delta 99O}$, respectively, $p = 0.0001$, Figure 5.2.3B). Results also indicate that pINV loss was significantly different between $pINV^{\Delta 99O}$ and $pINV^{\Delta 199R}$ ($p = 0.0250$, Figure 5.2.3B). Although the colouration of colonies was difficult to quantify, the difference in the redness of colonies harbouring different plasmid mutants could be observed (Figure 5.2.3A). Together, these results demonstrate that the 99O contributes to pINV maintenance, even though no function has been ascribed to this sequence previously (290, 293, 294, 302).

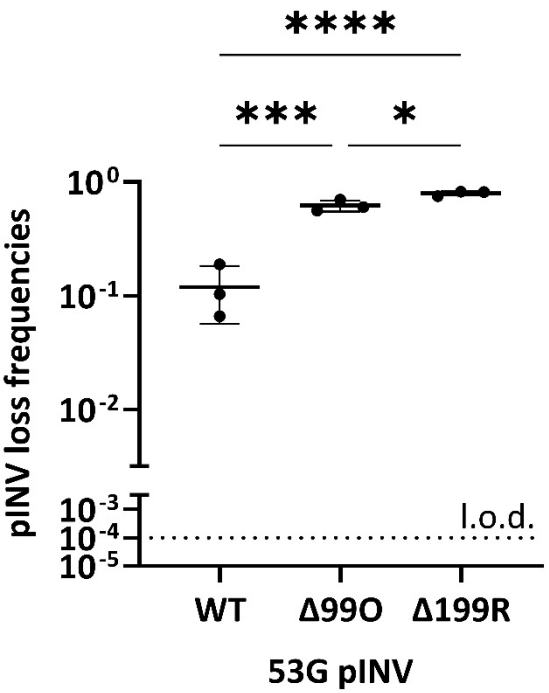


Figure 5.2.3B: Deletion of 99O increases pINV loss from *S. sonnei*.
Each dot represents pINV loss frequency in three CR^+ colonies in a replicate. pINV loss frequencies were quantified as the proportion of $pINV^-$ bacteria (*i.e.* CR^- and verified by PCR) in the population. Ordinary one-way ANOVA test was used to assess the significance of the differences between three sets of data. *, p = 0.0250; ***, p = 0.0001; ****, p < 0.0001.

Previously, it was suggested that the region containing the 99O could encode a protein that was involved in plasmid replication; however, there was no evidence supporting this hypothesis (359). Instead, another hypothesis was proposed that DNA hairpin structures in the region contribute to plasmid replication, and predicted that there were stable secondary structures embedded within 218B (293, 359). Therefore, further investigation was undertaken to test this hypothesis.

5.2.4 The involvement of predicted DNA hairpin structures in plasmid maintenance

It has been predicted that sequences in the 218B for secondary structures including at the *terR* site (293). Secondary structures could form once the region becomes ssDNA due to RepA binding the *ori*, unwinding downstream sequences. These DNA structures could influence the initiation and/or termination of replication, thereby affecting plasmid maintenance.

Therefore, bioinformatic analyses were performed to understand potential secondary structures in the 99O from *S. sonnei* 53G pINV. Tools such as RNAfold and Mfold are commonly

used to predict potential structures of nucleotides, with RNAfold found to perform slightly better than Mfold (361). Therefore, RNAfold (362, 363) was used to predict potential structures in the sequence of both DNA strands, using the built-in model optimised for ssDNA analysis. The resulting putative secondary structures were coloured according to the probability of base pairing. The free energy of thermodynamic ensemble measuring the structures' stability and its tendency to undergo spontaneous changes at 37°C was predicted. Similar analyses were done by Mfold, and predicted similar structures and free energy as RNAfold (not shown).

Results reveal likely DNA structures in the 99O, with the structures of both strands mostly mirroring each other (Figure 5.2.4A). The predicted free energy of the thermodynamic ensemble is -21.34 kcal/mol (5´-strand) and -28.47 kcal/mol (3´-strand), suggesting that the 3´-strand structure is more stable than the 5´-strand structure.

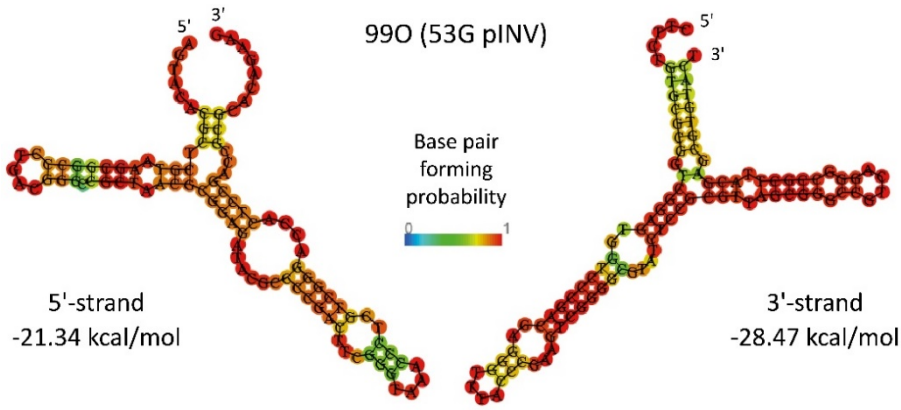


Figure 5.2.4A: DNA structures predicted by RNAfold in 99O from both strands of *S. sonnei* 53G pINV. Nucleotides were coloured according to the probability of forming base pairs, with the colour scale included in the figure. Structures from both strands were labelled with the predicted free energy of the thermodynamic ensemble in kcal/mol.

I also examined 99O from other plasmids to see whether the SNPs they contain could affect any DNA secondary structures. Results reveal that secondary structures formed by these plasmids (belonging to IncFII, IncI1, IncB/O/K/Z and IncI2) are likely conserved despite several SNPs in the region. Similar to 53G pINV, the predicted secondary structures of the 3´-strand from these plasmids are more stable than the secondary structures on the 5´-strand (Figure 5.2.4B).

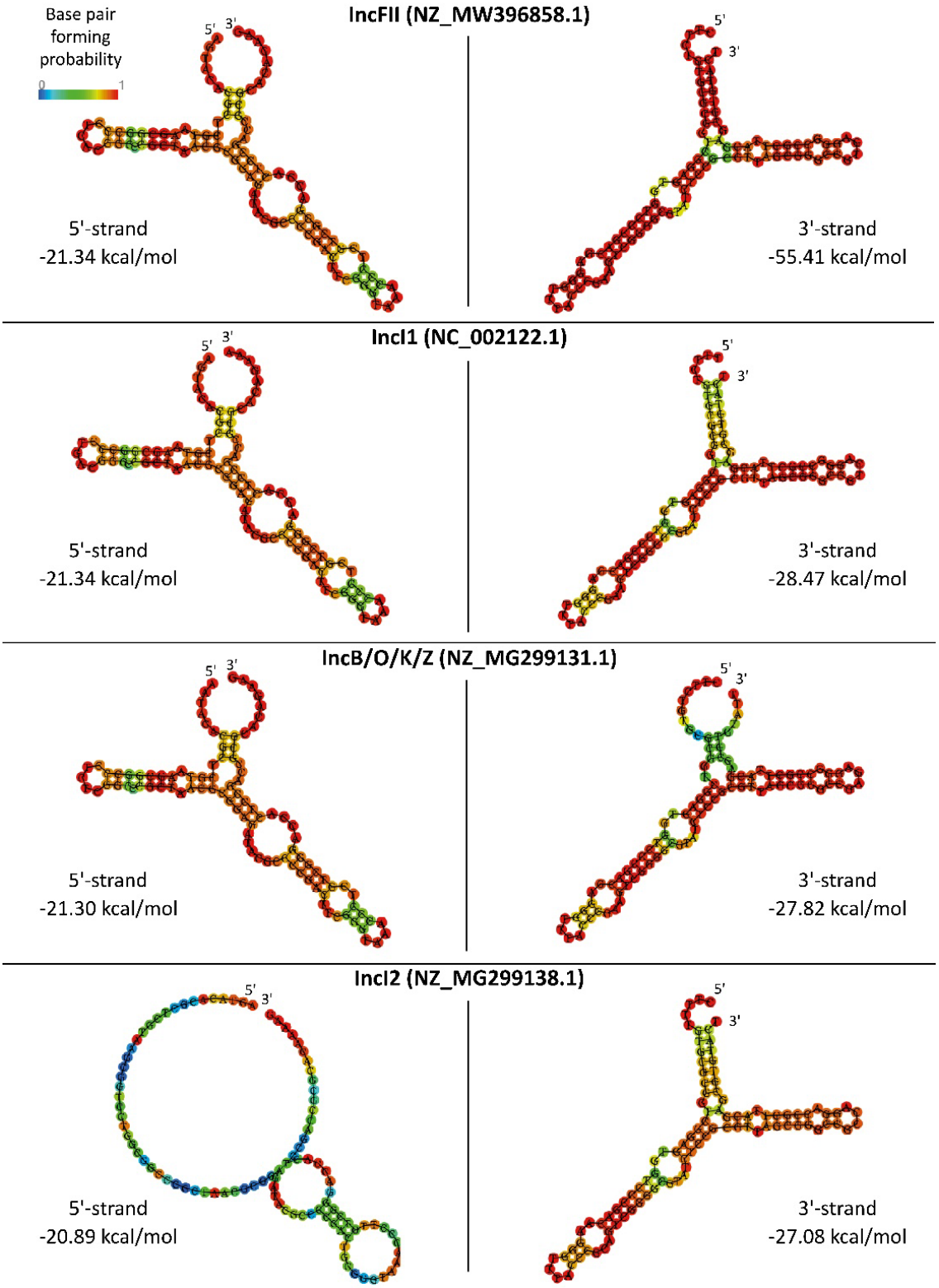


Figure 5.2.4B: Potential DNA structures in both strands of the 99O from different plasmids predicted by RNAfold.

The accession numbers and Inc groups of plasmids are shown. Nucleotides were coloured according to their probability of forming base pairs, with the colour scale included. Structures from both strands were labelled with the predicted free energy of the thermodynamic ensemble.

DNA hairpins can contribute to replication in several ways (364). DNA hairpins can act as a physical obstacle blocking the progression of the replication fork (365) or affect interactions with DNA binding proteins (*e.g.*, the DNA polymerase holoenzyme involved in lagging strand replication (366)). Surprisingly, the contribution of hairpin structures in unidirectional plasmid replication has not been studied in detail.

To assess the contribution of the possible hairpin structures in the 99O on *S. sonnei* 53G pINV, point mutations were introduced into the sequence. Two nucleotide substitutions were made in Mut1 (G59A/G61A) and Mut2 (G71A/C73T), while all four changes were included in Mut3 (Figure 5.2.4C). These changes are predicted to disrupt the potential DNA hairpin structures to different extents, leading to a reduction in the free energy required to open the 5´- and 3´-strand structures from 21.34 kcal to -15.68 kcal/mol for the 5´-strand structure (WT to Mut3), and -28.47 kcal/mol to -22.12 kcal/mol for the 3´-strand structure (WT to Mut3).

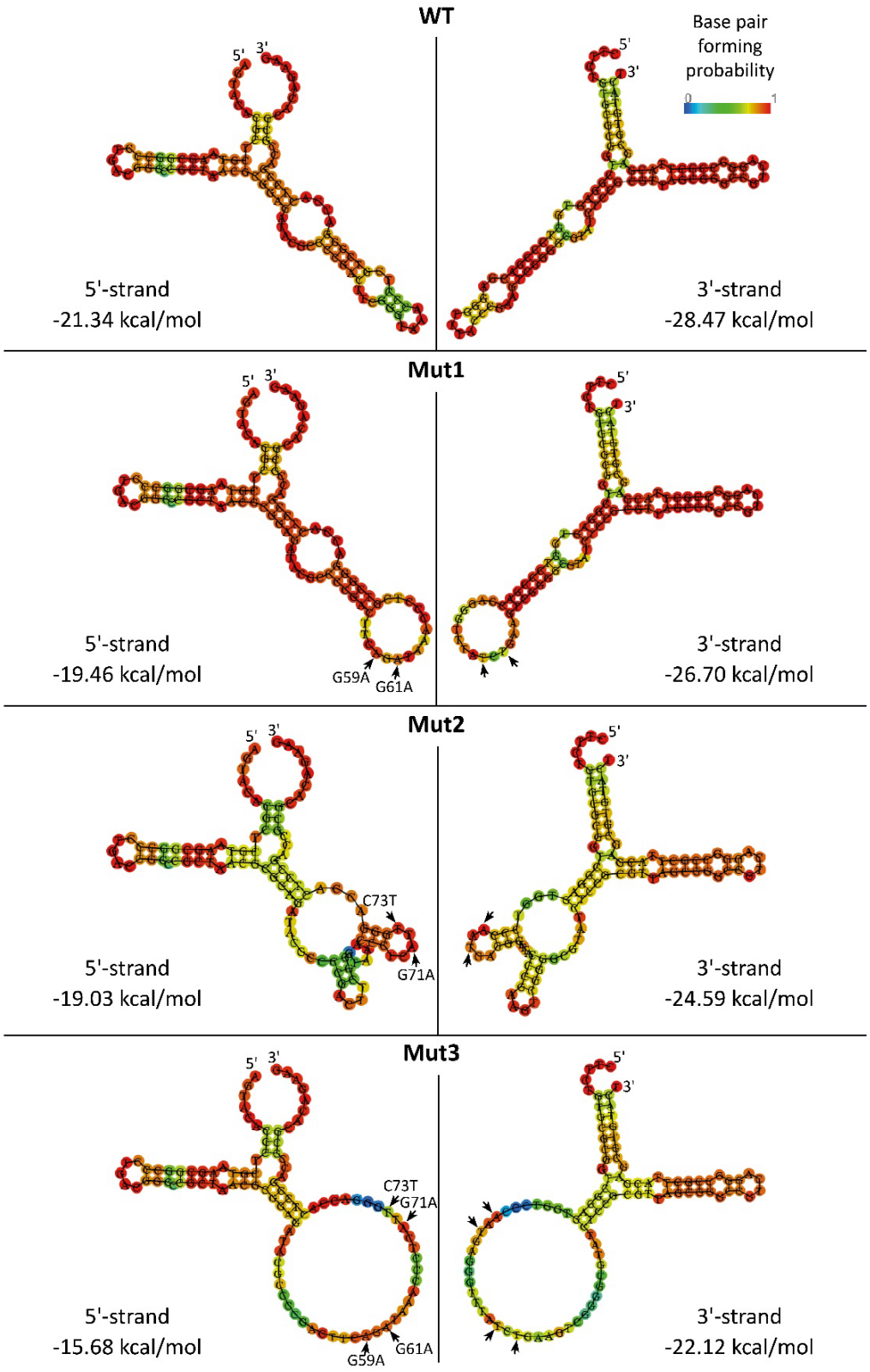


Figure 5.2.4C: The predicted secondary structures of 99O Mut1, Mut2 and Mut3.
All structures are annotated with their predicted free energy as determined by RNAfold. The mutations aimed to disrupt the largest branch of the structures. Mut1 harbours G59A and G61A mutations, leading to the disruption of the loops at the tip of the branch; Mut2 harbours G71A and C73T mutations, causing the disruption of the loops in the middle of the branch; Mut3 harbours all four changes and disrupts all loops in the branch.

CRLA were performed on *S. sonnei* 53G containing pINV^{Mut1}, pINV^{Mut2} or pINV^{Mut3}; pINV^{WT} and pINV$^{\Delta 99O}$ were included as controls. Colonies of pINV^{Mut1}, pINV^{Mut2} and pINV^{Mut3} showed different levels of red colouration following growth on plain CRA, which corresponds to the degree of disruption of predicted DNA hairpin structures (Figure 5.2.4D). The colonies of *S. sonnei* with pINV^{Mut1} were most similar to pINV^{WT} with a slight reduction in the redness of the colonies, while pINV^{Mut2} had an intermediate phenotype. In contrast, colonies of pINV^{Mut3} appeared to be largely CR^{-} with central faint red zone, similar to pINV$^{\Delta 99O}$.

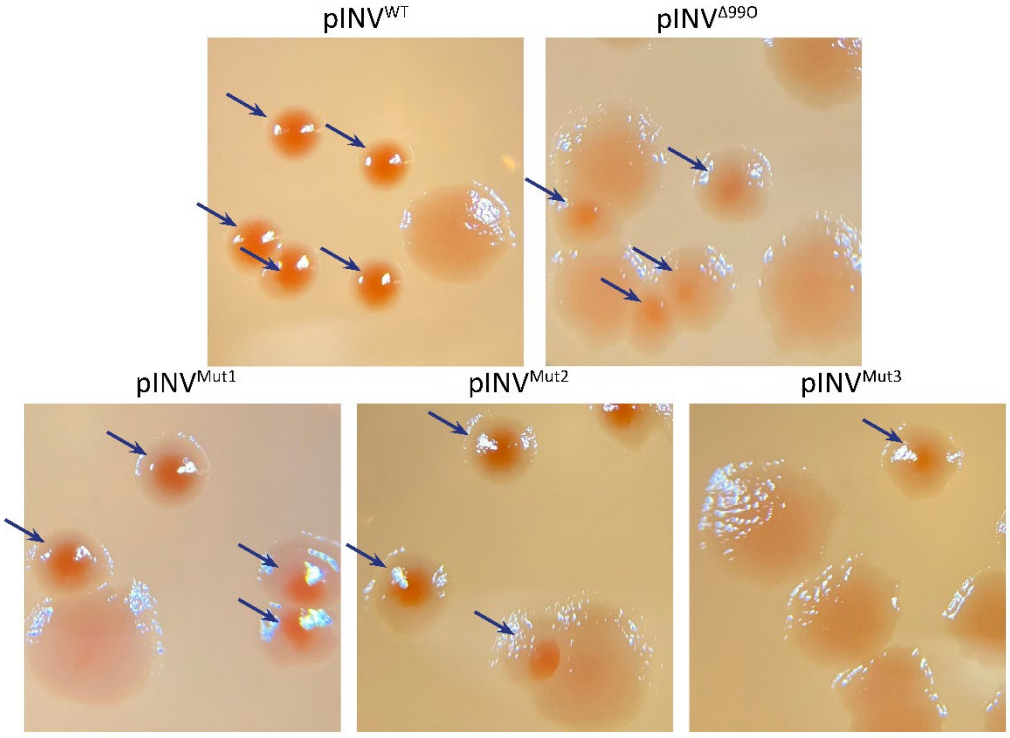


Figure 5.2.4D: Phenotypes of colonies from different mutants.
S. sonnei 53G with different versions of pINV (indicated above the panels) grown on CRA and observed with a light microscope. Navy arrows indicate colonies counted as CR^{+}. Differences in the redness of colonies was noticed but was difficult to quantify.

CRLA were performed using these strains. Results reveal that there was no significant difference between frequency of loss of pINV^{WT} and pINV^{Mut1}, which occurred at frequencies of 7.3 x 10^{-2} and 9.5 x 10^{-2} of the entire population, respectively ($p = 0.9335$, Figure 5.2.4E). In contrast, pINV^{Mut2} and pINV^{Mut3} showed a significant increase in pINV loss at frequencies of 1.7 x 10^{-1} and

8.2 x 10^{-1} of the entire population, respectively (p = 0.0448 and p < 0.0001) when compared with $pINV^{WT}$. Interestingly, pINV loss in $pINV^{Mut3}$ is even higher than $pINV^{\Delta 99O}$ (p = 0.0021). The results are consistent with the secondary structures in the 99O having a significant impact on plasmid maintenance.

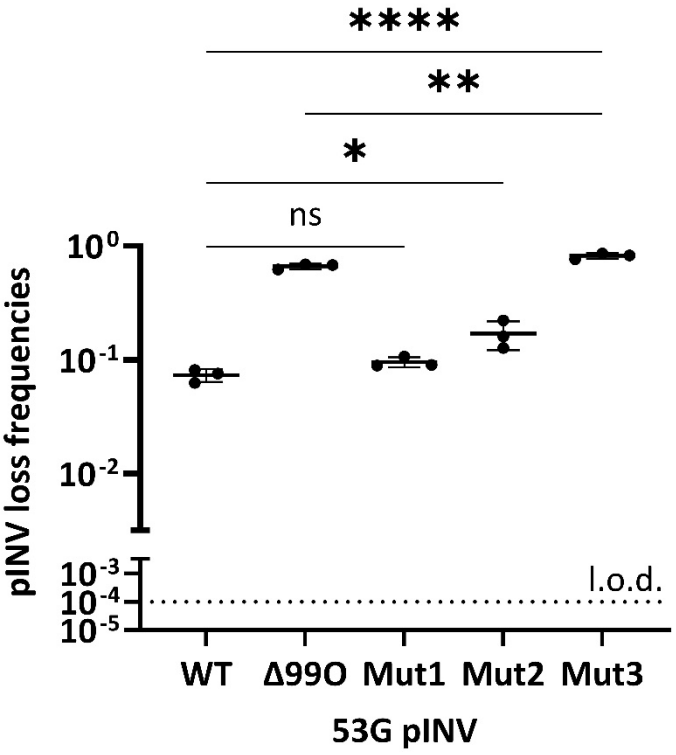


Figure 5.2.4E: pINV loss is affected by nucleotide substitutions in the 99O.
Each dot represents plasmid loss in three colonies in a replicate. Data were subjected to ordinary one-way ANOVA tests to assess significance. *, p = 0.0448; ** p = 0.0021; ****, p < 0.0001; ns, not significant. (N=3)

Overall, the results indicate that the 199R containing the putative RNA primer binding site is important for plasmid maintenance as its deletion significantly increased plasmid loss. A similar effect was observed in 218B which overlaps with 199R by the initial 99 bp and included *terR* (Figure 5.2.1A). Deletion of the overlapping sequence, 99O, also led to increased plasmid loss. As stable secondary DNA structures are predicted to form in ssDNA of the 99O, disrupting the putative structures by introducing nucleotide substitutions led to increased pINV loss, suggesting the predicted secondary structures are important in plasmid maintenance. Further investigation is required to understand the functions of these secondary structures.

5.3 Variation in sequences in and around the 199R

5.3.1 Potential usage of the 199R in plasmid classification

Given the presence of the 199R in IncFII, IncI1, IncB/O/K/Z and IncI2 plasmids, I evaluated the feasibility of utilising this sequence for plasmid classification through phylogenetic analysis. I used the 199R from 53G pINV as the reference sequence, aligning it using BLASTn with corresponding sequences from the 147 *S. sonnei* plasmids compiled by Dr. Arcari (Table 7.2). Extracted sequences were generated as FASTA files, and processed to remove spaces and incompatible syntax (such as ">") for downstream analysis. Aligned sequences of < 150 bp were removed from the analysis manually as longer sequences would provide a more comprehensive context for identifying and distinguishing between variants, reducing the probability of missing significant variations in the 199R from different plasmids. Sequences were aligned by Prank at the default setting in Ubuntu (324), aligned using RaxML-NG (325)with the GTR+G model to generate index files for analysis started with 25 random starting trees and 25 parsimony starting trees. The best tree was then bootstrapped with 1000 replicates and mapped with the branch support data containing information of accession numbers and Inc groups. The mapped best tree was visualised by FigTree (v1.4.4) and exported as a .nwk file, then annotated using a manually written script in RStudio.

The relationship of different plasmids is indicated by the length of the branches with the scale of the branches shown on Figure 5.3. The branch length indicates the presence of SNPs in the 199R. Longer branches generally indicate greater evolutionary distance or recombination between groups (325, 367). The short branch lengths in the tree demonstrate the high degree of conservation of the 199R, which means that the presence of only a few SNPs could distort the tree.

Results reveal that the tree shows relationships between the 199R generally follows the grouping of plasmids according to their Inc groups, with some discrepancies. For example, 15

out of 16 of pINV sequences (green) grouped together, and were closely related to IncB/O/K/Z plasmids (red) indicated by the location on the tree, but not the IncFII plasmids (blue). This is consistent with the previous finding that the nucleotide sequences of *repA* from IncB/O/K/Z plasmid share a high level of identity with *repA* from *S. sonnei* 53G pINV (IncFII, Figure 5.1.2A). The only exception is a pINV (accession no. CP019696) which possesses an 199R that groups with IncI2 plasmids (orange); IncI2 plasmids are all clustered together (22 out of 22).

Despite being categorised as the same Inc group as pINV, the 199R from 49 out of 50 of the IncFII plasmids (blue) showed closer relationship with IncI1 plasmids (purple) than pINV, with one exception (accession no. NC_013727) which is closely related to IncB/O/K/Z plasmids. Interestingly, there is also one exception from IncB/O/K/Z (accession no. NZ_OP038276) with a 199R being closely related to IncFII plasmids.

The 199R from IncB/O/K/Z plasmids was highly conserved, with 26 out of 30 plasmids grouped together. IncI1 plasmids (purple) showed the highest variation in their 199R, and contains plasmids with sequences related to different Inc groups. The 199R from 22 out of 29 of IncI1 plasmids grouped together, with one sequence (accession no. NZ_MG569891) grouped with IncB/O/K/Z plasmids, and one (accession no. NZ_CP010831) grouped with IncI2 plasmids. There is a small branch grouping five IncI1 and three IncB/O/K/Z plasmids together.

Interestingly, the phylogeny of these groups partially reflects the sequences downstream of the 199R (figure xxx). The IncFII exception harbours *yacAB* downstream of the 199R (accession no. NC_013727) which is closely related to IncB/O/K/Z plasmids harbouring *yacAB* at that site rather than *relBE* which is usually present in pINV. Of note, six out of eight plasmids in the stand-out branch with a mixture of IncI1 and IncB/O/K/Z plasmids harbour *yacAB*, with the remaining two plasmids (NZ_CP099772 and NZ_CP099780) belonging to the IncI1 group have IS*91* downstream of the 199R. Further discrepancies of sequences downstream of the 199R to the phylogeny is reflected by the IncI1 exception (NZ_CP010831) that harbours *relBE* downstream of the 199R as

it is closely related to IncI2 group with no *relBE* found. Similarly, the IncB/O/K/Z exception (NZ_OP038276) harbours *relBE* that is not found in the closely related IncFII plasmids.

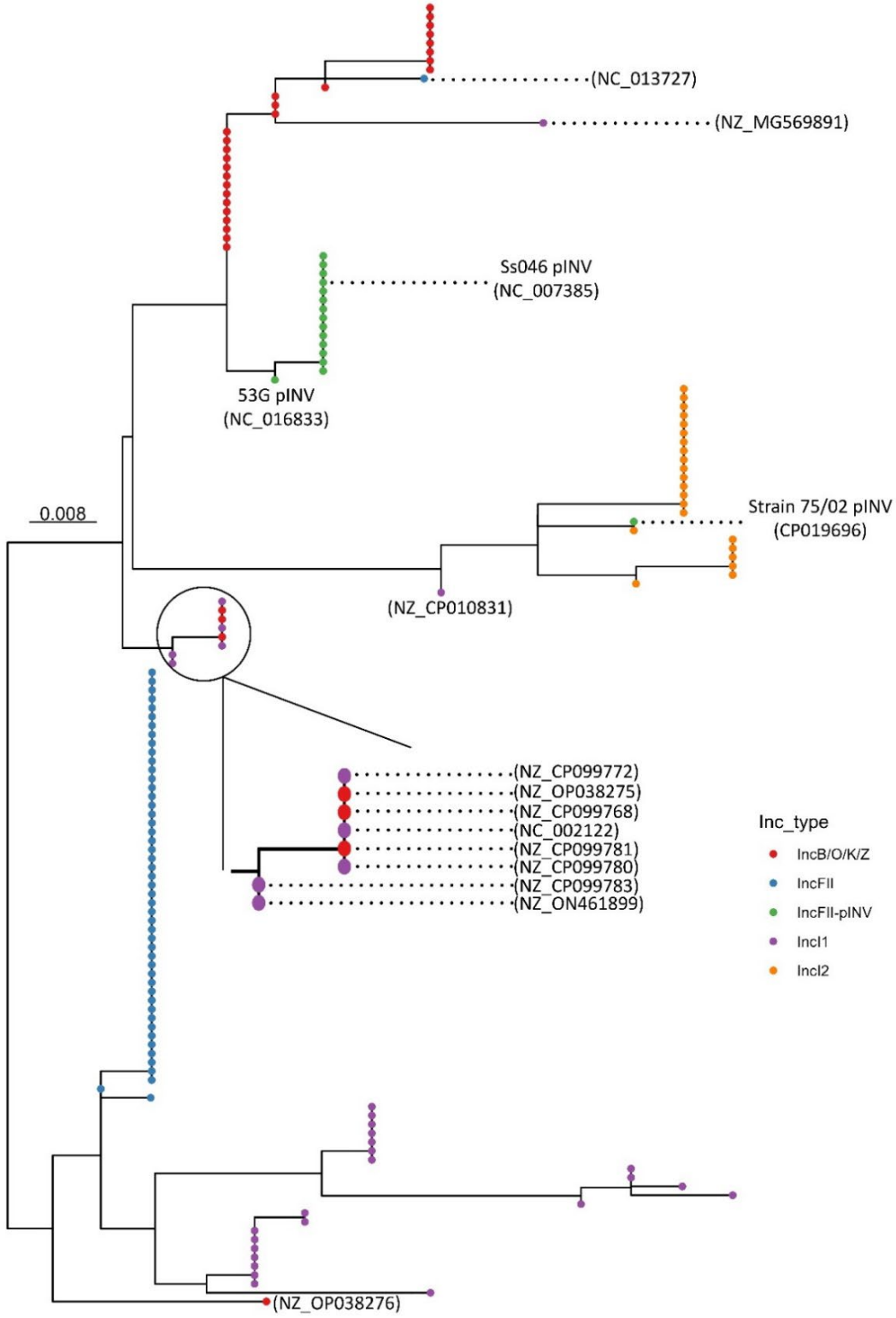


Figure 5.3: Phylogenetic tree created using alignment of 199R on plasmids
The figure was created using RStudio using the following libraries: readxl, tidyverse, ape, ggplot2, ggtree, cowplot, RColorBrewer, phytools, treeio, readr, phylobase, and tidytree. The branch tips were labelled with the corresponding Inc groups using colours from the Set 1 palette. Each dot represents a plasmid. The exceptions from different Inc groups were labelled with their accession numbers. The tree scale is labelled.

Analysis based on the 199R aligns with over three-quarters of the groupings indicated by the widely used PCR Inc typing scheme (Table 5.3). The discrepancies indicate that this region could subject to recombination and reassortment which causes variation in plasmid typing, and also reveals some evolutionary relationships among these plasmids. Therefore, while the 199R may not be appropriate for plasmid classification, phylogeny of this region could shed light into the evolutionary histories of plasmids.

Inc groups	No. of plasmids (x out of n)	Percentage (%)
pINV (IncFII)	15 out of 16	93.75
IncFII	49 out of 50	98.00
IncI1	22 out of 29	75.86
IncB/O/K/Z	26 out of 30	86.67
IncI2	22 out of 22	100.00

Table 5.3: The accuracy of grouping using the 199R of plasmids from different Inc groups.
The accuracy was determined by the grouping which is consistent to the current plasmid typing scheme according to the 199R.

In summary, there are obvious disparities between the Inc groups, which are based on sequences upstream of the *ori*, and the phylogenetic tree generated through analysis of the 199R as well as genes and ISs downstream of the 199R. These disparities suggest that there has been previous recombination between plasmids belonging to different Inc group which has occurred in or around the 199R. However, further fine mapping of potential recombination sites is required.

5.4 Summary

In this chapter, I identified and characterised a conserved 199 bp region situated downstream of the replicons in a wide range of plasmids across different Inc groups, including IncFII, IncI1, IncB/O/K/Z and IncI2. Bioinformatic analysis of plasmids that replicate unidirectionally using similar mechanisms demonstrated the conservation of this and surrounding sequences, uncovering a high level of similarity in the genetic arrangement of sequences downstream of the *ori*. This conservation suggests a close relationship between plasmids originating from these Inc groups, and could provide valuable insights into plasmid evolution.

Furthermore, my studies investigated the role of 199R in plasmid maintenance. Deleting this region led to a significant increase in plasmid loss, while putative DNA hairpin structures predominantly localised within the initial 99 bp of the 199R may have a role in replication. Disrupting these structures resulted in a significant increase in plasmid loss, emphasising the potential functional relevance of the DNA hairpins in plasmid replication. The results contribute to a hypothesis regarding the contribution of DNA hairpin structures to plasmid replication.

The 99O lies between where leading strand replication is initiated (*ssi*) and where plasmid replication ends (at the *terR* sequence in 218B). Given that this region is found in most *S. sonnei* plasmids that replicate unidirectionally using similar mechanisms, the secondary structures in the 99O could block the progression of the replication fork in the direction opposite to the leading strand replication (*i.e.* towards the *ori*). A gap identified between the R1 plasmid *ori* and the *ssi* in the late replication intermediate synthesised *in vitro* suggests that this gap could be filled at a later stage of plasmid replication, a process that may be slow. In addition, replication of the R1 plasmid's lagging strand toward the direction of *repA* has been shown to stop at the site where RepA binds to *ori in vitro* (294). Based on these results, it was earlier proposed that RepA remains bound to DNA until the end of plasmid replication, thereby blocking replication of the lagging strand (in the anti-clockwise direction) at the *ori*. I propose that secondary

structures in 99O could block inappropriate replication in the direction opposite to that of the leading strand replication (Figure 5.4). These mechanisms remain to be explored.

The Tus protein binds to dsDNA at *ter* sequences (297) to arrest the progression of the DNA replication fork *in vitro* (352). However, during plasmid replication, the dsDNA at this site becomes single-stranded upon RepA binding to the *ori*. Further investigation into the dynamics of termination of plasmid replication is required to understand how the transition from double-stranded to single-stranded DNA affects Tus binding and the overall regulation of plasmid replication termination.

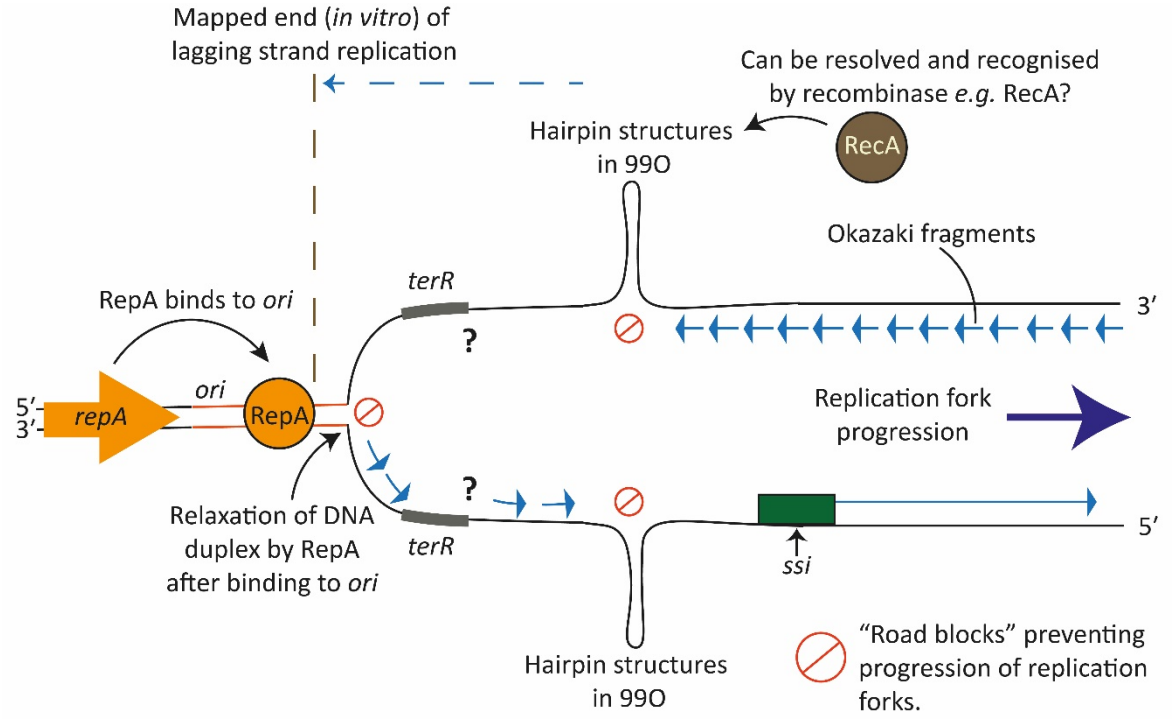


Figure 5.4: Proposed mechanism of replication affected by the 99O using 53G pINV as a model. Illustration not to scale. Important elements are annotated. The predicted secondary DNA structures of both strands were shown in Figure 5.2.3A. Replication fork progress from downstream of the *ori*. Lagging strand replication by Okazaki fragments progresses towards the *ori*; this could be blocked by obstacles formed by hairpin structures in the 99O, as well as RepA binding to *ori*, to ensure unidirectional replication. The later-stage of the replication remains to be understood.

Further exploration of the molecular mechanisms by which the 199R contributes to plasmid maintenance is necessary. Investigating potential proteins binding to the DNA sequence could provide valuable insights into the regulatory networks orchestrating plasmid replication.

Additionally, defining then disrupting the DNA secondary structures in this region *in vivo* could enhance our understanding of plasmid replication. A deeper understanding of the dynamic interplay between the 199R and other elements involved in plasmid replication should significantly advance our knowledge of plasmid biology.

I explored the potential of utilising the 199R for plasmid typing and to predict the sequences downstream of the 199R. However, results were limited due to the small number for full length sequences of plasmids in *S. sonnei*. A larger sample sizes containing more plasmids from *E. coli* and from different Inc groups could be included for a more comprehensive assessment of the suitability of the 199R for plasmid classification; findings should also be informative about the dynamics of plasmid evolution. Further analysis is required to examine the feasibility of its usage in diagnostics.

6 Discussion

6.1 Influence of horizontal gene transfer on *S. sonnei* evolution and dissemination of virulence and resistance

AMR in Enterobacterales, a family of Gram-negative bacteria that includes *E. coli*, *Klebsiella* spp., *Salmonella* spp., and *Shigella* spp., poses a significant global health challenge (368-372). AMR in these organisms is largely driven by the acquisition and dissemination of AMR genes through MGEs, such as plasmids, transposons, and integrons (373). For instance, plasmid-mediated resistance, particularly extended-spectrum β-lactamases (ESBLs) like CTX-M enzymes confer resistance to critical β-lactam antibiotics, are widely reported in *E. coli* and *K. pneumoniae* (374-376). Similarly, the emerging carbapenem-resistant Enterobacterales (CRE) render last-resort antibiotics ineffective (377). These carbapenemase resistance genes are frequently located on plasmids and other MGEs, enabling their rapid spread among bacterial populations (378-380).

Several Inc groups of plasmids have been closely associated with the spread of AMR genes by HGT in Enterobacterales, with IncF, IncI, IncA/C, IncL, IncN and IncH plasmids harbouring the greatest variety of resistance genes (204). For instance, in *E. coli*, IncFIC and IncI1 plasmids are associated with genes such as bla_{TEM} and bla_{CTX} encoding ESBLs, while aminoglycoside resistance is closely linked to IncP, IncI1 and IncA/C plasmids (381). In addition, resistance genes could co-locate with virulence factors on the same plasmids, as observed in *K. pneumoniae* and *Salmonella* spp., facilitating both persistence and pathogenicity (281, 284, 382-384). The association between specific Inc groups and AMR genes highlights the importance of understanding plasmid dynamics in combating AMR. This will contribute to developing strategies to monitor and control the dissemination of AMR genes in clinical and environmental settings.

Understanding the interplay between plasmids that can confer mobility to non-mobilisable elements is crucial. Typically, non-mobilisable genetic elements are defined by their inability to be mobilised even in the presence of a conjugative element (201). However, *S. sonnei* pINV, previously thought to be non-mobilisable (84), was transferable in the presence of a naturally-occuring, conjugative plasmid pRES; transfer involved genetic recombination. Non-mobilisable elements can exploit the conjugation apparatus, *oriT*, and relaxase of a conjugative plasmid either by forming a single co-integrated molecule with the conjugative plasmid or by acquiring mobilisation genes from the conjugative plasmid. The formation of fusion plasmids between virulence and resistance plasmids in *S. sonnei* demonstrates how co-residing plasmids can recombine to create hybrid elements, enhancing both AMR and pathogenicity within a single replicon. Such events have also been observed in other Enterobacterales systems, where conjugative plasmids play a central role in HGT. For example, the HR-mediated formation of hybrid plasmids with virulence and resistance genes was also reported in other clinically important pathogens, such as *K. pneumoniae* (203, 215, 281-284, 382) and *Yersinia pseudotuberculosis* (385). These findings suggest a common theme across Enterobacterales, where the interplay between conjugative plasmids and non-conjugative elements drives the evolution of AMR and hypervirulent pathogens, highlighting the role of plasmid-mediated recombination in amplifying the public health threat posed by these bacteria. This mode of HGT expands the potential way in which genetic traits are disseminated within and between bacterial species, underscoring the adaptability of bacteria in exploiting existing genetic resources for their evolutionary advantage.

Fusion plasmids formed in the donor could be highly unstable and costly, leading to their resolution into hybrid plasmids by further HR-dependent and/or -independent events after transfer into recipients. The precise recombination events that occur in the recipients will determine the identity and success of the hybrid plasmids, impacting the dissemination of certain genetic traits and shaping their structures and functions over time. For example, the

hybrid plasmid pHP241, which harbours both T3SS and the conjugation system T4SS, imposes a high fitness cost to the bacterial hosts, as bacteria harbouring this plasmid form smaller colonies compared to the bacteria containing wild-type pINV and/or pRES. This would negatively impact the survival of the strain in environments with limited resources, and therefore the plasmid would not be maintained in the bacterial population. This negative impact also suggests a hypothesis that the loss of the T4SS on pINV would abolish the ability of the plasmid to be horizontally transferred, but it might also reduce the fitness cost, thereby increasing the bacterial hosts' competitiveness.

I found evidence of remnants of related conjugation systems on *S. sonnei* pINV and chromosome. My work and others also described IS-mediated, chromosomal integration of pINV in *S. sonnei* and *S. flexneri* as this reduces T3SS expression, mitigating the fitness cost (163). These findings indicate that pINV might have been an integrated conjugative element (ICE), with the capacity to excise then transfer to other bacteria. However, many components necessary for successful conjugation are missing from the chromosome and pINV. Furthermore, there is no obvious advantage for the T3SS PAI to be encoded on a plasmid given its importance during infection, rather than the chromosome. It is possible that the frequent loss of virulence, through plasmid loss in *S. sonnei*, and emergence of an avirulent population of bacteria provides some benefit in host-pathogen interactions during mixed infection with virulent bacteria. Alternatively, plasmid loss is an evolutionary trade-off that is compensated by the ability of a plasmid to be transferred to a new host. This could be influenced by environmental selection pressure, such as changing niche and antimicrobial treatment, during evolution, that pINV can be easily lost to reduce fitness burden as it can be regained in the presence of resistance plasmids, which also explain the ability of *S. sonnei* to acquire resistance plasmids (62). This could further enhance the competitiveness of the bacterial hosts under extreme conditions, contributing to increase the likelihood of pINV maintenance within the bacterial population.

To mitigate the fitness costs associated with plasmid carriage, adaptive mutations can occur to enhance plasmid persistence in bacterial populations under selective pressures such as antibiotic use (301, 386). These mutations can occur in plasmid-encoded genes or regulatory regions, optimising the expression of resistance or virulence factors to reduce metabolic burdens (301). The fitness trade-offs between plasmid carriage and host survival depend on host-plasmid compatibility and environmental conditions. For instance, in *E. coli*, mutations in the plasmid replication machinery have been shown to reduce fitness costs while maintaining resistance capabilities (301, 387). While in *Pseudomonas aeruginosa*, the acquisition of certain resistance plasmids may impose significant burdens unless specific compensatory mutations arise (388, 389). These dynamics suggest that plasmid-host co-evolution is highly context-dependent. Therefore, the persistence of plasmids in host populations could be influenced by a delicate balance between selective pressures, environmental factors, and the availability of compensatory adaptations, which collectively shape the evolution of plasmid-host systems.

The evolutionary processes and genetic plasticity within bacterial genomes are observed by the detection of remnants of a conjugation system on both pINV and the *S. sonnei* chromosome. These remnants reflect the evolutionary history of these elements, indicating their origin from ancestral MGEs, such as the *E. coli* F plasmid. The presence of conjugation-related genes suggests that these elements could have acquired the capacity for HGT, contributing to the genetic diversity and adaptability of bacterial populations, which was then lost over time.

Differences in evolution can also be investigated by comparing the mobility of pINV between different *Shigella* species, the most common species, *S. sonnei* and S*. flexneri*, revealing critical differences in their genetic architecture and evolutionary trajectories. While both species harbour pINV, variations in their mobility may arise from differences in the genetic content or regulatory elements of these plasmids. Studies have demonstrated that *S. sonnei* pINV can be mobilised in the presence of a conjugative plasmid, but there is no report of *S. flexneri* pINV

being transferred horizontally in the presence of a conjugative plasmid and instead requires integration of conjugation system into the plasmid to facilitate its mobilisation (390). Additionally, pINV from both species harbours different numbers and families of ISs, indicating differential evolution, potentially caused by adaptation and dissemination in different environments (44). Environmental factors and selective pressures may influence the prevalence and dissemination of pINV within each species, shaping their respective epidemiological profiles.

Overall, the interplay between plasmids, HGT, and HR highlights the dynamic nature of bacterial genomes and their role in bacterial evolution. The detection of fusion or hybrid plasmids formed by HR and IS transposition during the mobilisation of pINV by pRES indicates the pivotal roles of HGT and HR in plasmid evolution. These processes lead to the formation of new plasmids, which can be transferred horizontally. HGT enables the acquisition of genetic material from distantly related organisms, broadening the genetic repertoire of bacterial populations (391). HR facilitates the exchange of genetic material between closely related sequences, promoting genetic diversity and adaptation (392). Understanding these processes is essential for elucidating the mechanisms driving the emergence and spread of AMR and virulence in bacterial pathogens. Continued research in this field is required for developing effective strategies to combat the evolving threat of AMR and virulent bacterial infections. Interdisciplinary approaches that integrate metagenomics and ecological modelling can enhance the prediction of plasmid-mediated AMR evolution and inform innovative applications.

6.2 The importance and implications of the conserved 199 bp region downstream of plasmid replicons

Work in Chapter 4 demonstrated that the interplay between plasmids in *S. sonnei* can contribute to the formation and spread of new plasmids through HGT and HR. A conserved 199 bp region (199R) downstream of plasmid replicons was frequently detected at pINV/pRES junctions in fusion or hybrid plasmids formed by HR or IS transposition. Therefore, I investigated the 199R in Chapter 5. The conservation of 199R across plasmids from different Inc groups provides insights into their evolutionary relationships and the functional significance of this region. Despite the vast genetic diversity among plasmids, certain regions associated with essential functions like replication and maintenance are normally conserved (209). The conservation of 199R downstream of plasmid replicons suggests that it plays a crucial role in plasmid biology and is subject to evolutionary constraints that preserve its sequence across diverse bacterial populations and diverse plasmids. The 199R region could be a hotspot for HR between related replicons or IS transposition. This phenomenon may also apply to other bacteria harbouring plasmids of these replicons that could associate with virulence and AMR genes, which this requires further investigation (204).

In *S. sonnei* 53G pINV there is a second copy of 199R (199R-2) found on the plasmid (145,915 bp - 146,111 bp, 198 bp). 199R-2 is located close to IS*21*(2) on pINV, and shares 81% nucleotide identity when compared with the 199R on pINV (Figure 7.9). At 34 bp upstream of 199R-2 is a highly mutated *terR* site lacking the core consensus sequence recognised by Tus (351), and a mutated putative *ssi* within 199R-2, which could affect their functions (Figure 6.2.1A). The 199R-2 is surrounded by a putative MFS transporter (136 bp upstream, 145,526 bp - 145,777 bp) and IS*1* (116 bp downstream, 146,228 bp - 146,995 bp). To investigate the conservation of the region, 85 bp upstream *terR* (as in the 199 bp-2 before the ORF of the MFS transporter) and 116 bp downstream (as in 199R-2 before the IS*1* IRL) of both copies of 199R (199R-1: 185,158 bp -

185,614 bp; 199R-2: 145,778 bp - 146,227 bp) were compared for their nucleotide sequences using ClustalOmega. The extended region shares 62% nucleotide identity (Figure 6.2.1A). Due to the accumulation of mutations within 199R-2, it was hypothesised that this region is likely to be derived from another replicon which is now non-functional. This could also serve as an evidence of extensive genetic rearrangement on pINV during evolution, and the multiple ISs flanking this region also further suggest the involvement of ISs during pINV evolution.

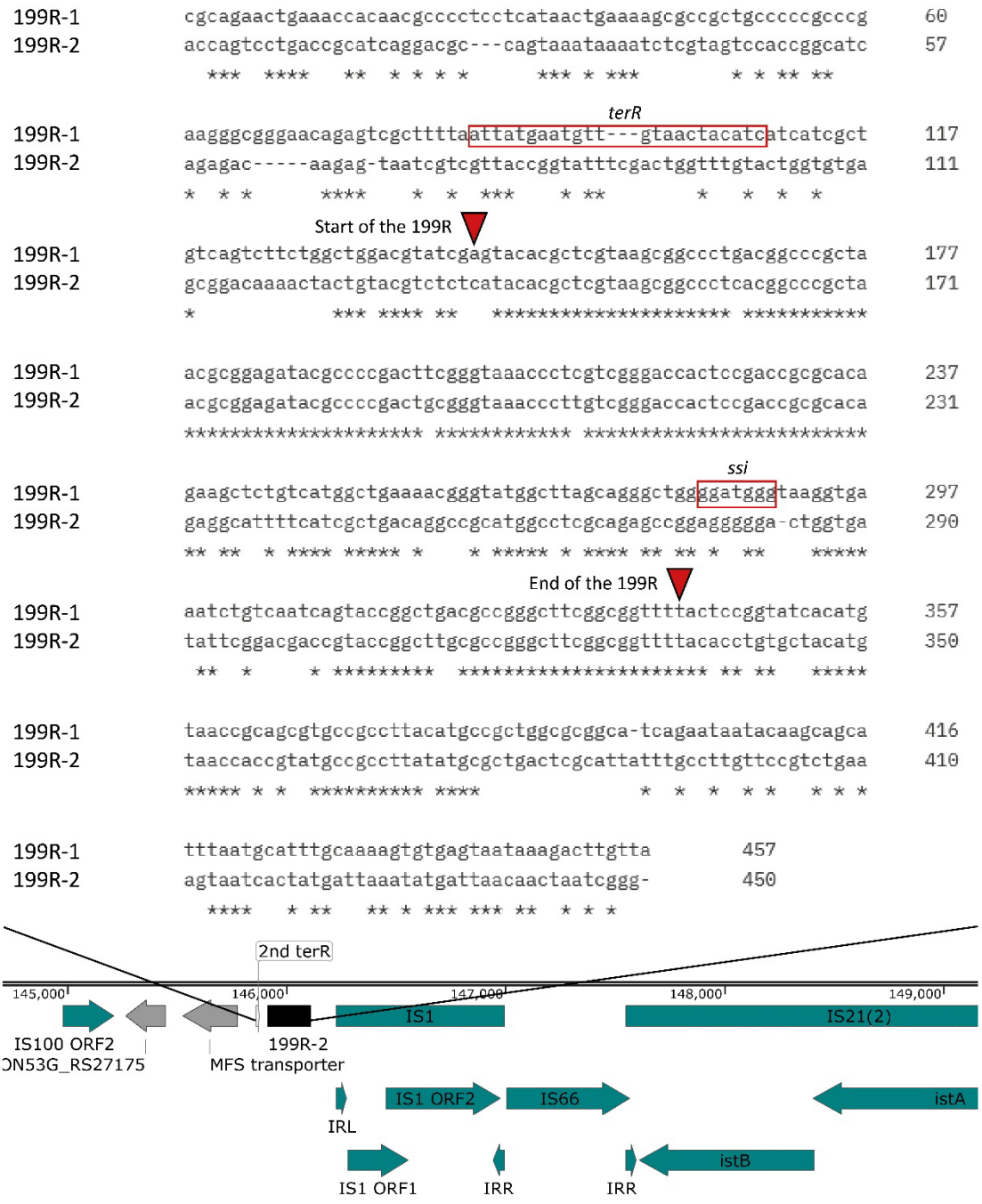


Figure 6.2.1A: The 199R-2 identified on 53G pINV and the elements surrounding.
Nucleotide alignment were done by ClustalOmega, with important elements annotated. The illustration of the gene with features annotated was exported from Snapgene.

In addition, phylogenetic analysis of plasmids from different Inc groups using sequences related to the 199R revealed genetic changes in plasmid backbones, demonstrating the genetic plasticity of plasmids. These findings indicate that 199R may be involved in events that result in changes/reassortment of sequences downstream of plasmid replicons. On many occcsaions, the genes downstream of the replicon are involved in plasmid maintenance such as TA systems. The diversity of sequences downstream of 199R suggests this region is a hotspot for genetic re-arrangements, supported by my experiments in which I observed frequent HR and IS transposition at the 199R during the mobilisation of pINV by pRES.

pINV and pRES replicate *via* Class A theta mechanism (287, 393). Class A theta plasmid replication typically involves melting of DNA duplex upon recognition of origin of replication by the replication initiation protein, leading to recruitment of host replication factors and coupled replication of both strands (394, 395). This process involves a single-stranded initiation origin (*ssi*) and termination sequences (*ter*) recognised by termination proteins to complete plasmid replication. Of note, 199R related sequences are not always found on plasmids that replicate unidirectionally by the Class A theta mechanism.

For example, broad host range IncP plasmids, such as RK2 plasmid (396), utilise the Class A theta mode to replicate unidirectionally. The genetic architecture of the plasmid replicon and the regulation of plasmid replication differ from those of IncFII and IncI-complex plasmids (168, 287, 304). The origin of replication (*oriV*) of IncP plasmids is flanked by AMR genes, and the region containing genes encoding proteins for plasmid replication regulation (*korA*-IncP1(II)-*korB*) is located downstream of *oriV* (397). Further differences may exist in the sequences for important functions like 199R and *terR* in IncFII and IncI-complex plasmids. However, studies on IncP plasmid replication are limited and further research is needed to determine whether sequences with functions similar to 199R can be found in other Class A theta replicating plasmids. Similarly, pSC101 which harbours a replicon with a different structure and replicates unidirectionally in a

different manner compared to IncFII and IncI-complex plasmids does not contain sequences related to the 199R (398). In addition, the 199R is not found in the *E. coli* F plasmid or R6K (399, 400) which replicate bi-directionally. Therefore, I hypothesise that the 199R is present in plasmids with similar replicon structures and replication mechanisms (*i.e.* those controlled by antisense RNA or with close evolutionary relationships), with its conservation indicating its importance in these processes. It would be interesting to explore if sequences containing secondary structures can be found in other plasmids where the 199R was not found.

Deletion of the 199R from plasmids across different Inc groups (such as pINV and pRES) consistently resulted in increased plasmid loss. This observation underscores the importance of the 199R in plasmid maintenance and persistence within bacterial hosts. The consistent outcome of increased plasmid loss following 199R deletion across different plasmids highlights its functional significance, suggesting it plays a fundamental role in processes essential for plasmid survival and propagation.

Deletion of the 199R included the *ssi* for leading strand synthesis, as determined previously through both *in vivo* and *in vitro* experiments (290, 294). A study by Burger (293) characterised mutant with a deletion including *terR* (218B) that overlaps with the first 99 bp (99O) of the 199R; the mutant showed a similar phenotype of increased pINV loss as plasmids lacking the 199R. It was predicted that there are hairpin structures within the 99O on the plasmids. Hairpin structures can play crucial roles in successful plasmid replication and termination of replication, ensuring accurate replication and segregation of plasmid DNA during cell division (364). For example, these hairpins can be involved in site-specific recombination, regulation of transcription, and recognised by proteins (364). Disrupting these structures in 99O increased plasmid loss, indicating their importance in plasmid maintenance. However, the effect of these secondary DNA structures on the dynamics of plasmid replication and termination remain to be elucidated.

The conservation of the 199R or 99O regions across diverse plasmids offers potential applications. Firstly, these conserved regions could inform understanding of the evolutionary history of plasmids. By comparing the sequences and functions of these regions across different plasmids, we can gain insights into the mechanisms driving plasmid evolution, including the contribution of selection and HGT.

The conservation of the 199R makes it an attractive target for novel strategies to eliminate virulence and resistance plasmids from bacteria to combat the threat of bacterial pathogens.and AMR spread by plasmids. For example, targeting the sequence could destabilise plasmid maintenance due to disrupted plasmid replication, leading to increased plasmid loss and reduced persistence within bacterial populations. Technologies such as CRISPR-Cas can specifically target and disrupt these regions, thereby inhibiting plasmid replication or maintenance. CRISPR-Cas can be delivered by phages or broad host range conjugative plasmids (401, 402). These interventions would not necessarily lead to bacterial death, so the selective pressure to evolve resistance against strategies to induce plasmid loss might be reduced compared to when antibitoics are used to treat infection. However, if plasmid segregation is affected (by disrupting plasmid replication for example) then a proportion of the bacterial population will undergo PSK.

The conserved 199R or 99O regions within plasmids can be exploited for the development of novel genetic tools and biotechnological applications. For example, by optimising the structural features of the 99O, the efficiency of plasmid replication and stability could potentially be improved, enhancing the performance of gene delivery vectors for applications, such as gene therapy or genetic engineering. In addition, the conserved region can serve as platforms for the development of biosensors capable of detecting specific environmental signals or molecular targets. Synthetic biology strategies can incorporate sensor elements, such as aptamers or protein domains, within the 199R or 99O to create genetically encoded sensors that respond to

particular stimuli. By engineering conserved regions to house ligand-binding domains, biosensors that detect small molecules or metabolites can be constructed to provide real-time monitoring of cellular dynamics or environmental conditions. Engineered plasmids with this region could be a stable platform for various biotechnological interventions and potentially used in vaccine development.

In conclusion, characterisation and exploitation of the 199R or 99O could enable the development of novel genetic tools and biotechnological applications. By understanding the functional properties of these regions, custom plasmids with tailored functions could be engineered to allow precise control of gene expression, development of biosensors, and improvement of gene delivery vectors for diverse biotechnological applications, while contributing to the current understanding of plasmid biology.

6.3 Limitations and future works

In this thesis, I studied the mechanisms of pRES-mediated pINV transfer, but several limitations remain to be addressed. There is no evidence of pINV transfer across different lineages based on analysis of a limited set of available pINV sequences (20). As pINV is lost during laboratory culture, many WGS lack pINV sequences. The lack of available pINV sequences and the high conservation of this plasmid in *S. sonnei* due to the clonal expansion make it difficult to find evidence of HGT of pINV from WGS databases. Expanding collection of the long-read sequences of *S. sonnei* isolates would significantly enhance epidemiological studies and the tracking of plasmid spread/evolution. As IS*21* is frequently found at the junctions of hybrid plasmids, analysis of a larger collection of complete plasmid sequences would also inform the prevalence of IS*21* on the plasmids. Alongside deleting the single IS*21* copy on pRES, these studies could contribute to understanding the impact of IS*21* on pINV transfer.

I performed conjugation assays using on solid media. Conjugation studies using traditional culture-based methods have provided valuable insights into plasmid transfer dynamics. However, there are several limitations. For example, these methods require compatible donor and recipient strains, optimal growth conditions, and selective media to detect transconjugants, which may not accurately represent the diversity of microbial ecosystems or the natural rates of HGT (403, 404). There are also ongoing debates regarding the quantification of conjugation frequencies. In addition, these conjugation methods often fail to capture interactions involving unculturable or slow-growing microbes, underestimating the ecological and evolutionary impacts of conjugation in complex communities such as the gut microbiome or environmental biofilms (217). Furthermore, these methods could overlook the intermediate molecules formed during conjugation, such as the potential formation of fusion plasmid by HR during conjugation in *S. sonnei*. For example, the transfer of pINV mediated by pRES occurs at low frequencies, making it uncertain whether this process happens during *in vivo* colonisation, or if it is merely a laboratory artifact.

To address these gaps, different approaches have been proposed and developed. Interdisciplinary methods have been designed to better quantify conjugation rates by also considering factors such as the impact of prophages and the growth of the bacteria during conjugation (405, 406). Exploring plasmid movements in murine models or chemostat experiments could reflect the environments encountered by *S. sonnei* in the intestinal tract, and could contribute to understanding the dynamics of plasmids between the pathogens and other bacteria in the microbiome. Another promising approach is the use of fluorescence-based reporters to track conjugation events *in situ* and in real-time, which could be challenging for pINV transfer detection as this is a rare event *in vitro* (407). Furthermore, the limitations of traditional culture-based conjugation methods are also addressed by new culture-independent methods (391). Techniques such as metagenomic sequencing and plasmidome analysis allow direct identification of plasmid transfer events by detecting plasmid sequences within microbial communities without the need for cultivation (408). Single-cell genomics has also emerged as a powerful tool to trace plasmid transfer at the cellular level, enabling the detection of HGT events between unculturable organisms (409). These methods provide a more comprehensive and accurate understanding of plasmid dynamics, offering critical insights into microbial evolution and resistance dissemination in natural ecosystems.

In addition to conjugation, the fate of pINV/pRES hybrid or fusion plasmids within transconjugants could be examined. It is possible that some of the plasmids do not impose a significant fitness burden on their hosts, and could also be conjugative or mobilisable. Further experiments could be done to investigate the maintenance and transfer of these newly arising hybrid plasmids to assess their ability to undergo vertical and horizontal transmission. Serial culture experiments should be informative about the success of combinations of particular plasmid replicons and TA and partitioning systems.

During pINV transfer mediated by pRES, 199R is frequently detected at the junctions of hybrid and fusion pINV/pRES plasmids. To better understand the prevalence, variability, and evolutionary role of 199R-containing plasmids, comprehensive genomic studies using long-read sequencing are needed. These studies would allow for detailed analyses of the arrangement, function, and evolution of 199R across diverse bacterial populations.

It was surprising that deletion of the entire 99O results in less plasmid loss than introducing the four SNPs within the 99O (*i.e.* in Mut3, Figure 5.2.4C). While I checked the sequence of the 99O and surrounding region in Mut3, it is possible that additional, unidentified mutations in the strain could lead to this outcome. To determine whether the reduced plasmid loss is solely due to the intended mutations, WGS such as Illumina sequencing should be performed to analyse the sequences of *S. sonnei* strains carrying pINV^{Mut3}.

The role of predicted 99O DNA structures in plasmid replication remains unclear. These structures may be critical in initiating or regulating replication. Future studies should aim to determine the exact DNA structures *in vivo*, which could be challenging. Techniques such as circular dichroism spectroscopy, which has been used to study conformational properties of DNA (410, 411), could be employed to detect the formation of DNA structures within the 99O. *In vitro* replication assays in controlled environments could be employed to test whether 99O DNA structures block the progression of plasmid replication forks (412). Additionally, advanced imaging techniques, such as live-cell imaging combined with fluorescent tagging of key replication and termination proteins (413), could provide real-time insights into how these DNA structures influence plasmid replication dynamics *in vivo*, providing direct evidence of their impact on replication.

7 Appendix

(A)

```
vapBC-CS6pINV       tcagctccagtcttcagttctcaggccgcccacacgttcaaattcccgggtgttattagt      60
vapBC-Ss046pINV     tcagctccagtcttcagttctcaggccgcccacacgttcaaattcccgggtgttattagt      60
vapBC-53GpINV       tcagctccagtcttcagttctcaggccgcccacacgttcaaattcccgggtgttattagt      60
                    ************************************************************

vapBC-CS6pINV       cacaataatcagcccccgactgcgggcatgacctgcgatcatttgatcaaatggcccgac      120
vapBC-Ss046pINV     cacaataatcagcccccgactgcgggcatgacctgcgatcatttgatcaaatggcccgac      120
vapBC-53GpINV       cacaataatcagcccccgactgcgggcatgacctgcgatcatttgatcaaatggcccgac      120
                    ************************************************************

vapBC-CS6pINV       agggcgtccctgacgggcaagttctgctcttatctggccggtgtgtgtggctgcagcagc      180
vapBC-Ss046pINV     agggcgtccctgacgggcaagttctgctcttatctggccggtgtgtgtggctgcagcagc      180
vapBC-53GpINV       agggcgtccctgacgggcaagttctgctcttatctggccggtgtgtgtggctgcagcagc      180
                    ************************************************************

vapBC-CS6pINV       gtcgtaatccagaacatcaatgcgggaaacaaatccctcgatcacagcgagattacgttc      240
vapBC-Ss046pINV     gtcgtaatccagaacatcaatgcgggaaacaaatccctcgatcacagcgagattacgttc      240
vapBC-53GpINV       gtcgtaatccagaacatcaatgcgggaaacaaatccctcgatcacagcgagattacgttc      240
                    ************************************************************

vapBC-CS6pINV       aggcatctggctttttttctgcaccatatatcagctccatcagggtgaccgaactgataca      300
vapBC-Ss046pINV     aggcatctggctttttttctgcaccatatatcagctccatcagggtgaccgaactgataca      300
vapBC-53GpINV       aggcatctggctttttttctgcaccatatatcagctccatcagggtgaccgaactgataca      300
                    ************************************************************

vapBC-CS6pINV       cattctcccctggttcaggttaaaacgctccctgacgctggcgggtttgttctttatcgt      360
vapBC-Ss046pINV     cattctcccctggttcaggttaaaacgctccctgacgctggcgggtttgttctttatcgt      360
vapBC-53GpINV       cattctcccctggttcaggttaaaacgctccctgacgctggcgggtttgttctttatcgt      360
                    ************************************************************

vapBC-CS6pINV       aaaaatgcagatgttggtatcgagcataaacttcagcatcagaatgactccctttcctgc      420
vapBC-Ss046pINV     aaaaatgcagatgttggtatcgagcataaacttcagcatcagaatgactccctttcctgc      420
vapBC-53GpINV       aaaaatgcagatgttggtatcgagcataaacttcagcatcagaatgactccctttcctgc      420
                    ************************************************************

vapBC-CS6pINV       atgccgggctgttccctgttatccataaaatcggcgctgacgctgtgtccgtcgaaccat      480
vapBC-Ss046pINV     atgccgggctgttccctgttatccataaaatcggcgctgacgctgtgtccgtcgaaccat      480
vapBC-53GpINV       atgccgggctgttccctgttatccataaaatcggcgctgacgctgtgtccgtcgaaccat      480
                    ************************************************************

vapBC-CS6pINV       tcgtcccatgtctctccggctggcgtaatgattctggttcgtccgacagcaatcacttca      540
vapBC-Ss046pINV     tcgtcccatgtctctccggctggcgtaatgattctggttcgtccgacagcaatcacttca      540
vapBC-53GpINV       tcgtcccatgtctctccggctggcgtaatgattctggttcgtccgacagcaatcacttca      540
                    ************************************************************

vapBC-CS6pINV       acgcgttttacgttttccggcaatgcaaccgcttttggcagtctgaccgcctggctgcgg      600
vapBC-Ss046pINV     acgcgttttacgttttccggcaatgcaaccgcttttggcagtctgaccgcctggctgcgg      600
vapBC-53GpINV       acgcgttttacgttttccggcaatgcaaccgcttttggcagtctgaccgcctggctgcgg      600
                    ************************************************************

vapBC-CS6pINV       ttgctgagaaatacggtggtttccat 626
vapBC-Ss046pINV     ttgctgagaaatacggtggtttccat 626
vapBC-53GpINV       ttgctgagaaatacggtggtttccat 626
                    ***************************
```

(B)

```
relBE-CS6pINV       atgcctaacatcattctgtctgagaccagtgcaagtgtcagtgaactgaaaaaaaatcct      60
relBE-Ss046pINV     atgcctaacatcattctgtctgagaccagtgcaagtgtcagtgaactgaaaaaaaatcct      60
relBE-53GpINV       atgcctaacatcattctgtctgagaccagtgcaagtgtcagtgaactgaaaaaaaatcct      60
                    ************************************************************

relBE-CS6pINV       atggctacagtcagtgccggagatggttaccctgtcgccattcttaaccggaatcagccg      120
relBE-Ss046pINV     atggctacagtcagtgccggagatggttaccctgtcgccattcttaaccggaatcagccg      120
relBE-53GpINV       atggctacagtcagtgccggagatggttaccctgtcgccattcttaaccggaatcagccg      120
                    ************************************************************

relBE-CS6pINV       gcattctactgtgtgcctgctgaactgtatgaaaggatgcttgatgctctggatgaccag      180
relBE-Ss046pINV     gcattctactgtgtgcctgctgaactgtatgaaaggatgcttgatgctctggatgaccag      180
relBE-53GpINV       gcattctactgtgtgcctgctgaactgtatgaaaggatgcttgatgctctggatgaccag      180
                    ************************************************************

relBE-CS6pINV       gaactggtgaaactggttactgaacgcagcaatcagccactgcatgatgtggacctggac      240
relBE-Ss046pINV     gaactggtgaaactggttactgaacgcagcaatcagccactgcatgatgtggacctggac      240
relBE-53GpINV       gaactggtgaaactggttactgaacgcagcaatcagccactgcatgatgtggacctggac      240
                    ************************************************************

relBE-CS6pINV       agttatttatgacttatacggtaaaattcagggatgatgcgctcaaggaatggctgaaac      300
relBE-Ss046pINV     agttatttatgacttatacggtaaaattcagggatgatgcgctcaaggaatggctgaaac      300
relBE-53GpINV       agttatttatgacttatacggtaaaattcagggatgatgcgctcaaggaatggctgaaac      300
                    ************************************************************

relBE-CS6pINV       tggataagaccattcaacaacagtttgtgaaaaaactaaaaaaatgcagcgaaaatcctc      360
relBE-Ss046pINV     tggataagaccattcaacaacagtttgtgaaaaaactaaaaaaatgcagcgaaaatcctc      360
relBE-53GpINV       tggataagaccattcaacaacagtttgtgaaaaaactaaaaaaatgcagcgaaaatcctc      360
                    ************************************************************

relBE-CS6pINV       atataccgtcggcaaaactaagagggctaaaggactgctacaaaattaaactacgtgcat      420
relBE-Ss046pINV     atataccgtcggcaaaactaagagggctaaaggactgctacaaaattaaactacgtgcat      420
relBE-53GpINV       atataccgtcggcaaaactaagagggctaaaggactgctacaaaattaaactacgtgcat      420
                    ************************************************************

relBE-CS6pINV       cgggatttcgtctggtttatcaggtgattgacgatatgttaattattgcggttgttgcgg      480
relBE-Ss046pINV     cgggatttcgtctggtttatcaggtgattgacgatatgttaattattgcggttgttgcgg      480
relBE-53GpINV       cgggatttcgtctggtttatcaggtgattgacgatatgttaattattgcggttgttgcgg      480
                    ************************************************************

relBE-CS6pINV       taggaaaacgagaacgtagtaacgtttataatctcgccagtgagagaatgagatag   536
relBE-Ss046pINV     taggaaaacgagaacgtagtaacgtttataatctcgccagtgagagaatgagatag   536
relBE-53GpINV       taggaaaacgagaacgtagtaacgtttataatctcgccagtgagagaatgagatag   536
                    ***********************************************************
```

(C)

```
parAB-CS6pINV       atgactagttttgagcagttaagtaaggttgcgcagcgtgcagataaaatgttgctcgcc      60
parAB-53GpINV       atgactagttttgagcagttaagtaaggttgcgcagcgtgcagataaaatgttgctcgcc      60
parAB-Ss046pINV     atgactagttttgagcagttaagtaaggttgcgcagcgtgcagataaaatgttgctcgcc      60
                    ************************************************************

parAB-CS6pINV       ctaacaaaacagattcaggagcaaaagcaggagtttcaggctgatgttttttatcaagtc      120
parAB-53GpINV       ctaacaaaacagattcaggagcaaaagcaggagtttcaggctgatgttttttatcaagtc      120
parAB-Ss046pINV     ctaacaaaacagattcaggagcaaaagcaggagtttcaggctgatgttttttatcaagtc      120
                    ************************************************************
```

```
parAB-CS6pINV       tattcaaagtctgctgtagctaagcttccaaaattaacgcgcgcaagcgtggatggtgca      180
parAB-53GpINV       tattcaaagtctgctgtagctaagcttccaaaattaacgcgcgcaagcgtggatggtgca      180
parAB-Ss046pINV     tattcaaagtctgctgtagctaagcttccaaaattaacgcgcgcaagcgtggatggtgca      180
                    ************************************************************

parAB-CS6pINV       gtcggtgagatggaggctcagggctatcagtttgaaaaacgcccagcaggaaccgcgaca      240
parAB-53GpINV       gtcggtgagatggaggctcagggctatcagtttgaaaaacgcccagcaggaaccgcgaca      240
parAB-Ss046pINV     gtcggtgagatggaggctcagggctatcagtttgaaaaacgcccagcaggaaccgcgaca      240
                    ************************************************************

parAB-CS6pINV       aaatatgcattaaccatccagaacattatcgatatctatgcacatcgtggcattcctaaa      300
parAB-53GpINV       aaatatgcattaaccatccagaacattatcgatatctatgcacatcgtggcattcctaaa      300
parAB-Ss046pINV     aaatatgcattaaccatccagaacattatcgatatctatgcacatcgtggcattcctaaa      300
                    ************************************************************

parAB-CS6pINV       tatcgcgatcgttatagtgaggcatattccatattcatcggtagtctaaaaggcggtgtg      360
parAB-53GpINV       tatcgcgatcgttatagtgaggcatattccatattcatcggtagtctaaaaggcggtgtg      360
parAB-Ss046pINV     tatcgcgatcgttatagtgaggcatattccatattcatcggtagtctaaaaggcggtgtg      360
                    ************************************************************

parAB-CS6pINV       tctaaaaccgtgagtagtgtttccgtagctcacgccctgcgtgctcatcctcacctgctg      420
parAB-53GpINV       tctaaaaccgtgagtagtgtttccgtagctcacgccctgcgtgctcatcctcacctgctg      420
parAB-Ss046pINV     tctaaaaccgtgagtagtgtttccgtagctcacgccctgcgtgctcatcctcacctgctg      420
                    ************************************************************

parAB-CS6pINV       tcagaagatttacgtattctgctccttgatttagatccccagtcatcagccactatgttt      480
parAB-53GpINV       tcagaagatttacgtattctgctccttgatttagatccccagtcatcagccactatgttt      480
parAB-Ss046pINV     tcagaagatttacgtattctgctccttgatttagatccccagtcatcagccactatgttt      480
                    ************************************************************

parAB-CS6pINV       ctgaactacttgcatgcggtggggctggttgatacaaccgcaccgcaggctatgttgcag      540
parAB-53GpINV       ctgaactacttgcatgcggtggggctggttgatacaaccgcaccgcaggctatgttgcag      540
parAB-Ss046pINV     ctgaactacttgcatgcggtggggctggttgatacaaccgcaccgcaggctatgttgcag      540
                    ************************************************************

parAB-CS6pINV       aacgtgagtcgtgaggaactactggaagattttatcgttccgtcagttatccctggcgtt      600
parAB-53GpINV       aacgtgagtcgtgaggaactactggaagattttatcgttccgtcagttatccctggcgtt      600
parAB-Ss046pINV     aacgtgagtcgtgaggaactactggaagattttatcgttccgtcagttatccctggcgtt      600
                    ************************************************************

parAB-CS6pINV       tatgttatgccggcttctattgacgacgcctttattgcatcaaactgggatactctgtgc      660
parAB-53GpINV       tatgttatgccggcttctattgacgacgcctttattgcatcaaactgggatactctgtgc      660
parAB-Ss046pINV     tatgttatgccggcttctattgacgacgcctttattgcatcaaactgggatactctgtgc      660
                    ************************************************************

parAB-CS6pINV       gaggagcatctattaggacaaaacaaacacgcgatattgcgtgagaatattattgataag      720
parAB-53GpINV       gaggagcatctattaggacaaaacaaacacgcgatattgcgtgagaatattattgataag      720
parAB-Ss046pINV     gaggagcatctattaggacaaaacaaacacgcgatattgcgtgagaatattattgataag      720
                    ************************************************************

parAB-CS6pINV       ctgaagcatgatttcgacttcatccttattgatacaggcccgcatctggacgcatttttg      780
parAB-53GpINV       ctgaagcatgatttcgacttcatccttattgatacaggcccgcatctggacgcatttttg      780
parAB-Ss046pINV     ctgaagcatgatttcgacttcatccttattgatacaggcccgcatctggacgcatttttg      780
                    ************************************************************
```

```
parAB-CS6pINV       aaaaatgccattgctgccgcagatatcatgtttacaccggtgccgccagcacaagttgat      840
parAB-53GpINV       aaaaatgccattgctgccgcagatatcatgtttacaccggtgccgccagcacaagttgat      840
parAB-Ss046pINV     aaaaatgccattgctgccgcagatatcatgtttacaccggtgccgccagcacaagttgat      840
                    ************************************************************

parAB-CS6pINV       ttccattccacgcttaaatatttggctcgtttacctgaacttgtacagattattgaacag      900
parAB-53GpINV       ttccattccacgcttaaatatttggctcgtttacctgaacttgtacagattattgaacag      900
parAB-Ss046pINV     ttccattccacgcttaaatatttggctcgtttacctgaacttgtacagattattgaacag      900
                    ************************************************************

parAB-CS6pINV       gatggatgttcatgccgtttgcaggcaaatattggcttcatgtcgaaactggcgaataag      960
parAB-53GpINV       gatggatgttcatgccgtttgcaggcaaatattggcttcatgtcgaaactggcgaataag      960
parAB-Ss046pINV     gatggatgttcatgccgtttgcaggcaaatattggcttcatgtcgaaactggcgaataag      960
                    ************************************************************

parAB-CS6pINV       tcagatcataaatactgccatagcctgacaaaagagattttcggtggtgacatgcttgat     1020
parAB-53GpINV       tcagatcataaatactgccatagcctgacaaaagagattttcggtggtgacatgcttgat     1020
parAB-Ss046pINV     tcagatcataaatactgccatagcctgacaaaagagattttcggtggtgacatgcttgat     1020
                    ************************************************************

parAB-CS6pINV       gtttcaatgccacgtcttgatggttttgagagaagcggcgagtcgtttgataccgtgata     1080
parAB-53GpINV       gtttcaatgccacgtcttgatggttttgagagaagcggcgagtcgtttgataccgtgata     1080
parAB-Ss046pINV     gtttcaatgccacgtcttgatggttttgagagaagcggcgagtcgtttgataccgtgata     1080
                    ************************************************************

parAB-CS6pINV       tctgcgaatcctgttacttatgttggcagtggtgaagctttgaaaaatgcacgtatggca     1140
parAB-53GpINV       tctgcgaatcctgttacttatgttggcagtggtgaagctttgaaaaatgcacgtatggca     1140
parAB-Ss046pINV     tctgcgaatcctgttacttatgttggcagtggtgaagctttgaaaaatgcacgtatggca     1140
                    ************************************************************

parAB-CS6pINV       gcagaagatttcgctaaagcagtttttgatcgtattgaatttataagggccaactactga     1200
parAB-53GpINV       gcagaagatttcgctaaagcagtttttgatcgtattgaatttataagggccaactactga     1200
parAB-Ss046pINV     gcagaagatttcgctaaagcagtttttgatcgtattgaatttataagggccaactactga     1200
                    ************************************************************

parAB-CS6pINV       tggaaaacagaaagcaccgaccgactattggccgtacgctgaatacaaatatcctgaata     1260
parAB-53GpINV       tggaaaacagaaagcaccgaccgactattggccgtacgctgaatacaaatatcctgaata     1260
parAB-Ss046pINV     tggaaaacagaaagcaccgaccgactattggccgtacgctgaatacaaatatcctgaata     1260
                    ************************************************************

parAB-CS6pINV       atactgaagaaatatcggctccagttcatgtatttaccctcaatactggccgaaaggcaa     1320
parAB-53GpINV       atactgaagaaatatcggctccagttcatgtatttaccctcaatactggccgaaaggcaa     1320
parAB-Ss046pINV     atactgaagaaatatcggctccagttcatgtatttaccctcaatactggccgaaaggcaa     1320
                    ************************************************************

parAB-CS6pINV       aatttactgagatcaaggttgatcatgacaaagtggacactcaaacttttgtggttgaag     1380
parAB-53GpINV       aatttactgagatcaaggttgatcatgacaaagtggacactcaaacttttgtggttgaag     1380
parAB-Ss046pINV     aatttactgagatcaaggttgatcatgacaaagtggacactcaaacttttgtggttgaag     1380
                    ************************************************************

parAB-CS6pINV       aggtgaatggccgcgaacagacagccctgacgccggactcactaaaagacatcaccagaa     1440
parAB-53GpINV       aggtgaatggccgcgaacagacagccctgacgccggactcactaaaagacatcaccagaa     1440
parAB-Ss046pINV     aggtgaatggccgcgaacagacagccctgacgccggactcactaaaagacatcaccagaa     1440
                    ************************************************************
```

```
parAB-CS6pINV       ctatacatttacaacaattctatccatgtattgggattcggactggcgatcttattgaga      1500
parAB-53GpINV       ctatacatttacaacaattctatccatgtattgggattcggactggcgatcttattgaga      1500
parAB-Ss046pINV     ctatacatttacaacaattctatccatgtattgggattcggactggcgatcttattgaga      1500
                    ************************************************************

parAB-CS6pINV       ttttggatggttcccgtcgtcgggcagcagctttgctatgcaaggttggattacgtgtac      1560
parAB-53GpINV       ttttggatggttcccgtcgtcgggcagcagctttgctatgcaaggttggattacgtgtac      1560
parAB-Ss046pINV     ttttggatggttcccgtcgtcgggcagcagctttgctatgcaaggttggattacgtgtac      1560
                    ************************************************************

parAB-CS6pINV       ttgtaacggatgatgagcttactgtttctgaagctcagcatctggcgaaagatttgcaga      1620
parAB-53GpINV       ttgtaacggatgatgagcttactgtttctgaagctcagcatctggcgaaagatttgcaga      1620
parAB-Ss046pINV     ttgtaacggatgatgagcttactgtttctgaagctcagcatctggcgaaagatttgcaga      1620
                    ************************************************************

parAB-CS6pINV       cgtcacttgaacataatattcgtgaaattgggttgcgtttagtacgtctcaaagaggccg      1680
parAB-53GpINV       cgtcacttgaacataatattcgtgaaattgggttgcgtttagtacgtctcaaagaggccg      1680
parAB-Ss046pINV     cgtcacttgaacataatattcgtgaaattgggttgcgtttagtacgtctcaaagaggccg      1680
                    ************************************************************

parAB-CS6pINV       gtatgaaccagaagcaaattgcggaaagagagggattgtcggcagctaaggtgactcgtg      1740
parAB-53GpINV       gtatgaaccagaagcaaattgcggaaagagagggattgtcggcagctaaggtgactcgtg      1740
parAB-Ss046pINV     gtatgaaccagaagcaaattgcggaaagagagggattgtcggcagctaaggtgactcgtg      1740
                    ************************************************************

parAB-CS6pINV       ccctccaagcagcgtccgtacctaaagattttgtgtctcttttcctgttcagtctgagc      1800
parAB-53GpINV       ccctccaagcagcgtccgtacctaaagattttgtgtctcttttcctgttcagtctgagc      1800
parAB-Ss046pINV     ccctccaagcagcgtccgtacctaaagattttgtgtctcttttcctgttcagtctgagc      1800
                    ************************************************************

parAB-CS6pINV       tgacatacgccgattacaggcaactggctgaattaagtgaacgactgaggctgggtgata      1860
parAB-53GpINV       tgacatacgccgattacaggcaactggctgaattaagtgaacgactgaggctgggtgata      1860
parAB-Ss046pINV     tgacatacgccgattacaggcaactggctgaattaagtgaacgactgaggctgggtgata      1860
                    ************************************************************

parAB-CS6pINV       tatctattgatgaagttgtgaagaatatctcgccgagtattgagcttattactgctgatg      1920
parAB-53GpINV       tatctattgatgaagttgtgaagaatatctcgccgagtattgagcttattactgctgatg      1920
parAB-Ss046pINV     tatctattgatgaagttgtgaagaatatctcgccgagtattgagcttattactgctgatg      1920
                    ************************************************************

parAB-CS6pINV       acaatctttctgaagacgaggttaaaaacagtattatgcggcttataaccaaggaaatgt      1980
parAB-53GpINV       acaatctttctgaagacgaggttaaaaacagtattatgcggcttataaccaaggaaatgt      1980
parAB-Ss046pINV     acaatctttctgaagacgaggttaaaaacagtattatgcggcttataaccaaggaaatgt      1980
                    ************************************************************

parAB-CS6pINV       cttcattgcttgattcaggagtaaaggataaagcggttgtaactttgttatggaaatttg      2040
parAB-53GpINV       cttcattgcttgattcaggagtaaaggataaagcggttgtaactttgttatggaaatttg      2040
parAB-Ss046pINV     cttcattgcttgattcaggagtaaaggataaagcggttgtaactttgttatggaaatttg      2040
                    ************************************************************

parAB-CS6pINV       atagtaaagataaatttgctcgaaaacgagtaaagggacgtacgttctcatatgaattcg      2100
parAB-53GpINV       atagtaaagataaatttgctcgaaaacgagtaaagggacgtacgttctcatatgaattcg      2100
parAB-Ss046pINV     atagtaaagataaatttgctcgaaaacgagtaaagggacgtacgttctcatatgaattcg      2100
                    ************************************************************
```

```
parAB-CS6pINV       gcagattaccgttggaagttcaggataagttagatcgtatgattgctcttgtactaaaag      2160
parAB-53GpINV       gcagattaccgttggaagttcaggataagttagatcgtatgattgctcttgtactaaaag      2160
parAB-Ss046pINV     gcagattaccgttggaagttcaggataagttagatcgtatgattgctcttgtactaaaag      2160
                    ************************************************************

parAB-CS6pINV       ataatctgaattcgctgtga      2180
parAB-53GpINV       ataatctgaattcgctgtga      2180
parAB-Ss046pINV     ataatctgaattcgctgtga      2180
                    ********************
```

Figure 7.1: Nucleotide alignment of vapBC (A), relBE (B) and parAB (C) from pINV of *S. sonnei* CS6, 53G and Ss046.

The alignment was created by ClustalOmega. Astericks (*) indicate identical nucleotides of all sequences in the comparison.

```
ISSso4-53GpINV      tgcaaaatggcatctgaatcgatcgcagtttggcattcgaatcgatcatgatttggcatc      60
ISSso4-CS6pINV      tgcaaaatggcatctgaatcgatcgcagtttggcattcgaatcgatcatgatttggcatc      60
ISSso4-Ss046pINV    tgcaaaatggcatctgaatcgatcgcagtttggcattcgaatcgatcatgatttggcatc      60
                    ************************************************************

ISSso4-53GpINV      catctccgatcacagtttggcatcacaatcgatcacgatttggcatacttccgatcattg      120
ISSso4-CS6pINV      catctccgatcacagtttggcatcacaatcgatcacgatttggcatacttccgatcattg      120
ISSso4-Ss046pINV    catctccgatcacagtttggcatcacaatcgatcacgatttggcatacttccgatcattg      120
                    ************************************************************

ISSso4-53GpINV      attggaattcgccagtcactccgggaattaactcttttcaccacagtctccattgccgtg      180
ISSso4-CS6pINV      attggaattcgccagtcactccgggaattaactcttttcaccacagtctccattgccgtg      180
ISSso4-Ss046pINV    attggaattcgccagtcactccgggaattaactcttttcaccacagtctccattgccgtg      180
                    ************************************************************

ISSso4-53GpINV      tttaaaccagtggagacgtcaatgtccaaaaagagaaaatccaggagcactatggaaact      240
ISSso4-CS6pINV      tttaaaccagtggagacgtcaatgtccaaaaagagaaaatccaggagcactatggaaact      240
ISSso4-Ss046pINV    tttaaaccagtggagacgtcaatgtccaaaaagagaaaatccaggagcactatggaaact      240
                    ************************************************************

ISSso4-53GpINV      tgttttaagatccttcagctcaagttcgataagaagctgactaaccgttgtatcggactg      300
ISSso4-CS6pINV      tgttttaagatccttcagctcaagttcgataagaagctgactaaccgttgtatcggactg      300
ISSso4-Ss046pINV    tgttttaagatccttcagctcaagttcgataagaagctgactaaccgttgtatcggactg      300
                    ************************************************************

ISSso4-53GpINV      acgttacacatcagcgcctcaaccgtttttgaggttcttgcccgcttcaaagcctcttct      360
ISSso4-CS6pINV      acgttacacatcagcgcctcaaccgtttttgaggttcttgcccgcttcaaagcctcttct      360
ISSso4-Ss046pINV    acgttacacatcagcgcctcaaccgtttttgaggttcttgcccgcttcaaagcctcttct      360
                    ************************************************************

ISSso4-53GpINV      ctgtcatggccattacctgcggatatatcccatgacacgctggagaaactcattttcccg      420
ISSso4-CS6pINV      ctgtcatggccattacctgcggatatatcccatgacacgctggagaaactcattttcccg      420
ISSso4-Ss046pINV    ctgtcatggccattacctgcggatatatcccatgacacgctggagaaactcattttcccg      420
                    ************************************************************
```

```
ISSso4-53GpINV       cccaaagacacctctgcgtcagagcttgttatgcctgacatgctgtacttcgatacggaa     480
ISSso4-CS6pINV       cccaaagacacctctgcgtcagagcttgttatgcctgacatgctgtacttcgatacggaa     480
ISSso4-Ss046pINV     cccaaagacacctctgcgtcagagcttgttatgcctgacatgctgtacttcgatacggaa     480
                     ************************************************************

ISSso4-53GpINV       atgagaaaacctggcgtgacgcgccaactgttgtggatggaatacaaggcacaggcgggt     540
ISSso4-CS6pINV       atgagaaaacctggcgtgacgcgccaactgttgtggatggaatacaaggcacaggcgggt     540
ISSso4-Ss046pINV     atgagaaaacctggcgtgacgcgccaactgttgtggatggaatacaaggcacaggcgggt     540
                     ************************************************************

ISSso4-53GpINV       gacaaagcgatgggatactcacacttctgccgttgctaccggaagtggaaaaaaacgcga     600
ISSso4-CS6pINV       gacaaagcgatgggatactcacacttctgccgttgctaccggaagtggaaaaaaacgcga     600
ISSso4-Ss046pINV     gacaaagcgatgggatactcacacttctgccgttgctaccggaagtggaaaaaaacgcga     600
                     ************************************************************

ISSso4-53GpINV       cgcctctccatgcgacaggagcatcgtgctggcgaaaagctctttatcgacttctgtggt     660
ISSso4-CS6pINV       cgcctctccatgcgacaggagcatcgtgctggcgaaaagctctttatcgacttctgtggt     660
ISSso4-Ss046pINV     cgcctctccatgcgacaggagcatcgtgctggcgaaaagctctttatcgacttctgtggt     660
                     ************************************************************

ISSso4-53GpINV       cctacggtccctgttattaatccggacaccggagagatacgccgcgtggcaatcttcgtg     720
ISSso4-CS6pINV       cctacggtccctgttattaatccggacaccggagagatacgccgcgtggcaatcttcgtg     720
ISSso4-Ss046pINV     cctacggtccctgttattaatccggacaccggagagatacgccgcgtggcaatcttcgtg     720
                     ************************************************************

ISSso4-53GpINV       gctgtaatgggggcgtccaactatacctatgttgaagcgtgcgaaggtcaggacatgatg     780
ISSso4-CS6pINV       gctgtaatgggggcgtccaactatacctatgttgaagcgtgcgaaggtcaggacatgatg     780
ISSso4-Ss046pINV     gctgtaatgggggcgtccaactatacctatgttgaagcgtgcgaaggtcaggacatgatg     780
                     ************************************************************

ISSso4-53GpINV       tcatggttaaatgcccacagccgttgcctgaccttcctgggcggcgtaccgaagctgctc     840
ISSso4-CS6pINV       tcatggttaaatgcccacagccgttgcctgaccttcctgggcggcgtaccgaagctgctc     840
ISSso4-Ss046pINV     tcatggttaaatgcccacagccgttgcctgaccttcctgggcggcgtaccgaagctgctc     840
                     ************************************************************

ISSso4-53GpINV       atccccgataacctcagaagcgcagtgaaaaaagccgatcgctatgagcctgtcatcaac     900
ISSso4-CS6pINV       atccccgataacctcagaagcgcagtgaaaaaagccgatcgctatgagcctgtcatcaac     900
ISSso4-Ss046pINV     atccccgataacctcagaagcgcagtgaaaaaagccgatcgctatgagcctgtcatcaac     900
                     ************************************************************

ISSso4-53GpINV       gacagttatcaggcgctggctgagcactatggcacggttattatcccggcccgcccgcgc     960
ISSso4-CS6pINV       gacagttatcaggcgctggttgagcactatggcacggttattatcccggcccgcccgcgc     960
ISSso4-Ss046pINV     gacagttatcaggcgctggttgagcactatggcacggttattatcccggcccgcccgcgc     960
                     ****************** ****************************************

ISSso4-53GpINV       aaacctaaagataaaccaaaagcagaaaacggcgtcctcatcgtcgagctatggcttctt     1020
ISSso4-CS6pINV       aaacctaaagataaaccaaaagcagaaaacggcgtcctcatcgtcgagcgatggcttctt     1020
ISSso4-Ss046pINV     aaacctaaagataaaccaaaagcagaaaacggcgtcctcatcgtcgagcgatggcttctt     1020
                     ********************************************** **********

ISSso4-53GpINV       gcccgcatccgtaatgaaaccttccacacgctcagggcgctgaatgcacgccttcgggag     1080
ISSso4-CS6pINV       gcccgcatccgtaatgaaaccttccacacgctcagggcgctgaatgcacgccttcgggag     1080
ISSso4-Ss046pINV     gcccgcatccgtaatgaaaccttccacacgctcagggcgctgaatgcacgccttcgggag     1080
                     ************************************************************
```

```
ISSso4-53GpINV       ttgctgacggacatgaataaccgaccaatgaagggatacggtaatcaaacccgcgcagaa     1140
ISSso4-CS6pINV       ttgctgacggacatgaataaccgaccaatgaagggatacggtaatcaaacccgcgcagaa     1140
ISSso4-Ss046pINV     ttgctgacggacatgaataaccgaccaatgaagggatacggtaatcaaacccgcgcagaa     1140
                     ************************************************************

ISSso4-53GpINV       cgcttccgcatgctggatgcaccagcactctctccgctgcctctggagccttatgaatac     1200
ISSso4-CS6pINV       cgcttccgcatgctggatgcaccagcactctctccgctgcctctggagccttatgaatac     1200
ISSso4-Ss046pINV     cgcttccgcatgctggatgcaccagcactctctccgctgcctctggagccttatgaatac     1200
                     ************************************************************

ISSso4-53GpINV       acggaatacaaagccgtgaaggttggccctgactaccacgttgaatatgcccgtcactgg     1260
ISSso4-CS6pINV       acggaatacaaagccgtgaaggttggccctgactaccacgttgaatatgcccgtcactgg     1260
ISSso4-Ss046pINV     acggaatacaaagccgtgaaggttggccctgactaccacgttgaatatgcccgtcactgg     1260
                     ************************************************************

ISSso4-53GpINV       tactccgttccacacgaactggtcggtcagcgtctgtcactgaaggtggggcagtcggtg     1320
ISSso4-CS6pINV       tactccgttccacacgaactggtcggtcagcgtctgtcactgaaggtggggcagtcggtg     1320
ISSso4-Ss046pINV     tactccgttccacacgaactggtcggtcagcgtctgtcactgaaggtggggcagtcggtg     1320
                     ************************************************************

ISSso4-53GpINV       gtacagctctggcacaaggggcagtgtgtggcgcaacatccgcgcagcacacatgaatat     1380
ISSso4-CS6pINV       gtacagctctggcacaaggggcagtgtgtggcgcaacatccgcgcagcacacatgaatat     1380
ISSso4-Ss046pINV     gtacagctctggcacaaggggcagtgtgtggcgcaacatccgcgcagcacacatgaatat     1380
                     ************************************************************

ISSso4-53GpINV       aaacacacaaccaatccgcttcacatgccagaacgccaccggcggcatgggacatggaca     1440
ISSso4-CS6pINV       aaacacacaaccaatccgcttcacatgccagaacgccaccggcggcatgggacatggaca     1440
ISSso4-Ss046pINV     aaacacacaaccaatccgcttcacatgccagaacgccaccggcggcatgggacatggaca     1440
                     ************************************************************

ISSso4-53GpINV       ccggaacgccttatcgagcagggaaaccgtacaggtccgtccacaggccgggtggttgaa     1500
ISSso4-CS6pINV       ccggaacgccttatcgagcagggaaaccgtacaggtccgtccacaggccgggtggttgaa     1500
ISSso4-Ss046pINV     ccggaacgccttatcgagcagggaaaccgtacaggtccgtccacaggccgggtggttgaa     1500
                     ************************************************************

ISSso4-53GpINV       tccatgctgaaagccaaaccccatccggagctggcttaccgtgcagtcctgggtctgctg     1560
ISSso4-CS6pINV       tccatgctgaaagccaaactccatccggagctggcttaccgtgcagtcctgggtctgctg     1560
ISSso4-Ss046pINV     tccatgctgaaagccaaaccccatccggagctggcttaccgtgcagtcctgggtctgctg     1560
                     ******************* ****************************************

ISSso4-53GpINV       gcacttcagaaaaaatacggcccggaacggctggaaaaggcctgttacgttgccctgcac     1620
ISSso4-CS6pINV       gcacttcagaaaaaatacggcccggaacggctggaaaaggcctgttacgttgccctgcac     1620
ISSso4-Ss046pINV     gcacttcagaaaaaatacggcccggaacggctggaaaaggcctgttacgttgccctgcac     1620
                     ************************************************************

ISSso4-53GpINV       tataacgccccggacagacgctttatcgacaatctgctgcgtcatcatcgggataacgtt     1680
ISSso4-CS6pINV       tataacgccccggacagacgctttatcgacaatctgctgcgtcatcatcgggataacgtt     1680
ISSso4-Ss046pINV     tataacgccccggacagacgctttatcgacaatctgctgcgtcatcatcgggataacgtt     1680
                     ************************************************************

ISSso4-53GpINV       gagcttccgttatcccgccagggcgaacaacatccggcatatgcttcagagcatgaaaac     1740
ISSso4-CS6pINV       gagcttccgttatcccgccagggcgaacaacatccggcatatgcttcagagcatgaaaac     1740
ISSso4-Ss046pINV     gagcttccgttatcccgcctgggcgaacaacatccggcatatgcttcagagcatgaaaac     1740
                     ******************* ****************************************
```

```
ISSso4-53GpINV       ctgcgtggtcctggctactaccactgacagaaggaacaaccgtatgagtgataatctgtt     1800
ISSso4-CS6pINV       ctgcgtggtcctggctactaccactgacagaaggaacaaccgtatgagtgataatctgtt     1800
ISSso4-Ss046pINV     ctgcgtggtcctggctactaccactgacagaaggaacaaccgtatgagtgataatctgtt     1800
                     ************************************************************

ISSso4-53GpINV       aaacaaactgacccaactgaaactcccggcgatggccgggagcctgatacggcaacggga     1860
ISSso4-CS6pINV       aaacaaactgacccaactgaaactcccggcgatggccgggagcctgatacggcaacggga     1860
ISSso4-Ss046pINV     aaacaaactgacccaactgaaactcccggcgatggccgggagcctgatacggcaacggga     1860
                     ************************************************************

ISSso4-53GpINV       aacgcctcagacttacgacgaactgtcgttcgaggaaaggctgacactgctggttgatga     1920
ISSso4-CS6pINV       aacgcctcagacttacgacgaactgtcgttcgaggaaaggctgacactgctggttgatga     1920
ISSso4-Ss046pINV     aacgcctcagacttacgacgaactgtcgttcgaggaaaggctgacactgctggttgatga     1920
                     ************************************************************

ISSso4-53GpINV       cgagcttctgagcagagaaaacagtcgtgtggcacggttgagaaaaaatgcctgcctgaa     1980
ISSso4-CS6pINV       cgagcttctgagcagagaaaacagtcgtgtggcacggttgagaaaaaatgcctgcctgaa     1980
ISSso4-Ss046pINV     cgagcttctgagcagagaaaacagtcgtgtggcacggttgagaaaaaatgcctgcctgaa     1980
                     ************************************************************

ISSso4-53GpINV       atatcaggcaacaccggaagggctgcgttatccggcctcaagaggtctgcgggctgaaca     2040
ISSso4-CS6pINV       atatcaggcaacaccggaagggctgcgttatccggcctcaagaggtctgcgggctgaaca     2040
ISSso4-Ss046pINV     atatcaggcaacaccggaagggctgcgttatccggcctcaagaggtctgcgggctgaaca     2040
                     ************************************************************

ISSso4-53GpINV       gatgagggaactgctgaacggccactacataatccacaggaaaaacctactgataaccgg     2100
ISSso4-CS6pINV       gatgagggaactgctgaacggccactacataatccacaggaaaaacctactgataaccgg     2100
ISSso4-Ss046pINV     gatgagggaactgctgaacggctactacataatccacaggaaaaacctactgataaccgg     2100
                     ********************** ***********************************

ISSso4-53GpINV       tccgacaggatgcggtaagagctggatagccaatgcgctgggcgaacaggcatgtcgaca     2160
ISSso4-CS6pINV       tccgacaggatgcggtaagagctggatagccaatgcgctgggcgaacaggcatgtcgaca     2160
ISSso4-Ss046pINV     tccgacaggatgcggtaagagctggatagccaatgcgctgggcgaacaggcatgtcgaca     2160
                     ************************************************************

ISSso4-53GpINV       gaaatacagcgtccggtactgtagaaccgggcgattactggaacagttggctcagggacg     2220
ISSso4-CS6pINV       gaaatacagcgtccggtactgtagaaccgggcgattactggaacagttggctcagggacg     2220
ISSso4-Ss046pINV     gaaatacagcgtccggtactgtagaaccgggcgattactggaacagttggctcagggacg     2220
                     ************************************************************

ISSso4-53GpINV       tgttgatggcagttggctgaaatatctgaaacagttgcaaaaaatacaggttctaatcct     2280
ISSso4-CS6pINV       tgttgatggcagttggctgaaatatctgaaacagttgcaaaaaatacaggttctaatcct     2280
ISSso4-Ss046pINV     tgttgatggcagttggctgaaatatctgaaacagttgcaaaaaatacaggttctaatcct     2280
                     ************************************************************

ISSso4-53GpINV       ggacgatctggggctggaacagttaagtaacgcccagtgtaatgaccttctggagatcac     2340
ISSso4-CS6pINV       ggacgatctggggctggaacagttaagtaacgcccagtgtaatgaccttctggagatcac     2340
ISSso4-Ss046pINV     ggacgatctggggctggaacagttaagtaacgcccagtgtaatgaccttctggagatcac     2340
                     ************************************************************

ISSso4-53GpINV       agaggatcgttacggacaaagcagcacgatagtggtaagccagttcccggtagataaatg     2400
ISSso4-CS6pINV       agaggatcgttacggacaaagcagcacgatagtggtaagccagttcccggtagataaatg     2400
ISSso4-Ss046pINV     agaggatcgttacggacaaagcagcacgatagtggtaagccagttcccggtagataaatg     2400
                     ************************************************************
```

```
ISSso4-53GpINV       gcatggtctgatggagaacccaacaacagcagatgccatcctggacagactggtacataa     2460
ISSso4-CS6pINV       gcatggtctgatggagaacccaacaacagcagatgccatcctggacagactggtacataa     2460
ISSso4-Ss046pINV     gcatggtctgatggagaacccaacaacagcagatgccatcctggacagactggtacataa     2460
                     ************************************************************

ISSso4-53GpINV       ctcacacagagtggtccttcagggagagtccctgaggaaaaatccccctacagtggaaag     2520
ISSso4-CS6pINV       ctcacacagagtggtccttcagggagagtccctgaggaaaaatccccgacagtggaaag     2520
ISSso4-Ss046pINV     ctcacacagagtggtccttcagggagagtccctgaggaaaaatccccgacagtggaaag     2520
                     ********************************************** ***********

ISSso4-53GpINV       cagcgaaaaaacgagttaaaaagcggtagaaaaagttaagttcccggttgatcgaatgga     2580
ISSso4-CS6pINV       cagcgaaaaaacgagttaaaaagcggtagaaaaagttaagttcccggttgatcgaatgga     2580
ISSso4-Ss046pINV     cagcgaaaaaacgagttaaaaagcggtagaaaaagttaagttcccggttgatcgaatgga     2580
                     ************************************************************

ISSso4-53GpINV       tgcgaatcgctgatcgaacttcgatgccaatccgtgatcgaatagagcggaatacgca       2638
ISSso4-CS6pINV       tgcgaatcgctgatcgaacttcgatgccaatccgtgatcgaatagagcggaatacgca       2638
ISSso4-Ss046pINV     tgcgaatcgctgatcgaacttcgatgccaatccgtgatcgaatagagcggaatacgca       2638
                     ************************************************************
```

Figure 7.2: Nucleotide alignment of IS*Sso4* from pINV of *S. sonnei* CS6, 53G and Ss046.
The alignment was created by ClustalOmega. Astericks (*) indicate identical nucleotides of all sequences in the comparison. The nucleotide identities are 99.85% and 99.89% for 53G and Ss046 pINV, respectively, when compared to IS*Sso4* from CS6 pINV.

(A)

```
TraD-F               MSFNAKDMTQGGQIASMRIRMFSQIANIMLYCLFIFFWILVGLVLWIKISWQTFVNGCIY       60
TraD-CS6pINV         --------------------MFSQIANIMLYCLFIFFWILVGLVLWVKISWQTFVNGCIY       40
TraD-Ss046pINV       --------------------MFSQIANIMLYCLFIFFWILVGLVLWVKISWQTFVNGCIY       40
TraD-53GpINV         --------------------MFSQIANIMLYCLFIFFWILVGLVLWVKISWQTFVNGCIY       40
                                         *************************:*************

TraD-F               WWCTTLEGMRDLIKSQPVYEIQYYGKTFRMNAAQVLHDKYMIWCSEQLWSAFVLAAVVAL       120
TraD-CS6pINV         WWCTTLEGMRDLIKSQPVYEIQYYGKTFRMNAAQVLHDKYMIWCGEQLWSAFVLATVVAL       100
TraD-Ss046pINV       WWCTTLEGMRDLIKSQPVYEIQYYGKTFRMNAAQVLHDKYMIWCGEQLWSAFVLATVVAL       100
TraD-53GpINV         WWCTTLEGMRDLIKSQPVYEIQYYGKTFRMNAAQVLHDKYMIWCGEQLWSAFVLATVVAL       100
                     ****************************************.**********:****

TraD-F               VICLITFFVVSWILGRQGKQQSENEVTGGRQLTDNPKDVARMLKKDGKDSDIRIGDLPII       180
TraD-CS6pINV         VICLITFFVVSWILGRQGKQQSENEVTGGRQLTDNPKDVARMLKKDGRDSDIRIGDLPII       160
TraD-Ss046pINV       VICLITFFVVSWILGRQGKQQSENEVTGGRQLTDNPKDVARMLKKDGRDSDIRIGDLPII       160
TraD-53GpINV         VICLITFFVVSWILGRQGKQQSENEVTGGRQLTDNPKDVARMLKKDGRDSDIRIGDLPII       160
                     ********************************************:*************

TraD-F               RDSEIQNFCLHGTVGAGKSEVIRRLANYARQRGDMVVIYDRSGEFVKSYYDPSIDKILNP       240
TraD-CS6pINV         RDSEIQNFCLHGTVGAGKSEVIRRLANYARQRGDMVVIYDRSGEFVKSYYDPSIDKILNP       220
TraD-Ss046pINV       RDSEIQNFCLHGTVGAGKSEVIRRLANYARQRGDMVVIYDRSGEFVKSYYDPSIDKILNP       220
TraD-53GpINV         RDSEIQNFCLHGTVGAGKSEVIRRLANYARQRGDMVVIYDRSGEFVKSYYDPSIDKILNP       220
                     ************************************************************

TraD-F               LDARCAAWDLWKECLTQPDFDNTANTLIPMGTKEDPFWQGSGRTIFAEAAYLMRNDPNRS       300
TraD-CS6pINV         LDARCAAWDLWKECLTQPDFDNTANTLIPMGTKEDPFWQGSGRTIFAEAAYLMRNDPNRS       280
TraD-Ss046pINV       LDARCAAWDLWKECLTQPDFDNTANTLIPMGTKEDPFWQGSGRTIFAEAAYLMRNDPNRS       280
TraD-53GpINV         LDARCAAWDLWKECLTQPDFDNTANTLIPMGTKEDPFWQGSGRTIFAEAAYLMRNDPNRS       280
                     ************************************************************
```

```
TraD-F          YSKLVDTLLSIKIEKLRTYLRNSPAANLVEEKIEKTAISIRAVLTNYVKAIRYLQGIEHN   360
TraD-CS6pINV    YSKLVDTLLSIKIEKLRTYLRNSPAANLVEEKIEKTAISIRAVLTNYVKAIRYLQGIEHN   340
TraD-Ss046pINV  YSKLVDTLLSIKIEKLRTYLRNSPAANLVEEKIEKTAISIRAVLTNYVKAIRYLQGIEHN   340
TraD-53GpINV    YSKLVDTLLSIKIEKLRTYLRNSPAANLVEEKIEKTAISIRAVLTNYVKAIRYLQGIEHN   340
                ************************************************************

TraD-F          GEPFTIRDWMRGVREDQKNGWLFISSNADTHASLKPVISMWLSIAIRGLLAMGENRNRRV   420
TraD-CS6pINV    GEPFTIRDWMRGVREDQKNGWLFISSNADTHASLKPVISMWLSIAIRGLLAMGENRNRRV   400
TraD-Ss046pINV  GEPFTIRDWMRGVREDQKNGWLFISSNADTHASLKPVISMWLSIAIRGLLAMGENRNRRV   400
TraD-53GpINV    GEPFTIRDWMRGVREDQKNGWLFISSNADTHASLKPVISMWLSIAIRGLLAMGENRNRRV   400
                ************************************************************

TraD-F          WFFCDELPTLHKLPDLVEILPEARKFGGCYVFGIQSYAQLEDIYGEKAAASLFDVMNTRA   480
TraD-CS6pINV    WFFCDELPTLHKLPDLVEILPEARKFGGCYVFGIQSYAQLEDIYGEKAAATLFDVMNTRA   460
TraD-Ss046pINV  WFFCDELPTLHKLPDLVEILPEARKFGGCYVFGIQSYAQLEDIYGEKAAATLFDVMNTRA   460
TraD-53GpINV    WFFCDELPTLHKLPDLVEILPEARKFGGCYVFGIQSYAQLEDIYGEKAAATLFDVMNTRA   460
                ************************************************:***********

TraD-F          FFRSPSHKIAEFAAGEIGEKEHLKASEQYSYGADPVRDGVSTGKDMERQTLVSYSDIQSL   540
TraD-CS6pINV    FFRSPSHKIAEFAVGEIGEKEHLKASEQYSYGADPVRDGVSTGKDMERQTLVSYSDIQSL   520
TraD-Ss046pINV  FFRSPSHKIAEFAVGEIGEKEHLKASEQYSYGADPVRDGVSTGKDMERQTLVSYSDIQSL   520
TraD-53GpINV    FFRSPSHKIAEFAVGEIGEKEHLKASEQYSYGADPVRDGVSTGKDMERQTLVSYSDIQSL   520
                *************.************************************************

TraD-F          PDLTCYVTLPGPYPAVKLSLKYQTRPKVAPEFIPRDINPEMENRLSAVLAAREAEGRQMA   600
TraD-CS6pINV    PDLTCYVTLPGPYPAVKLSLKYQERPKVAPEFIPRDINPEMENRLSAVLAAREAEGRQMA   580
TraD-Ss046pINV  PDLTCYVTLPGPYPAVKLSLKYQERPKVAPEFIPRDINPEMENRLSAVLAAREAEGRQMA   580
TraD-53GpINV    PDLTCYVTLPGPYPAVKLSLKYQERPKVAPEFIPRDINPEMENRLSAVLAAREAEGRQMA   580
                ********************** ***********************************

TraD-F          SLFEPDVPEVVSGEDVTQAEQP----------------------------------QQ   624
TraD-CS6pINV    SLFEPDVPEVVSGEDVTQAEQPQQPQQPQQPQQPQ---------QPQQPQQPQQPQQPQQ   631
TraD-Ss046pINV  SLFEPDVPEVVSGEDVTQAEQPQQPQQPQQPQQPQQPQQPQQPQQPQQPQQPQQPQQPQQ   640
TraD-53GpINV    SLFEPDVPEVVSGEDVTQAEQPQQPQQPQQPQQP-----------QQPQQPQQPQQPQQ   628
                **********************                                    **
                                        C-terminal tail of TraD essential for conjugation
TraD-F          PVSPAINDKKSDSGVNVPAGGIEQELKMKPEEEMEQQLPPGISESGEVVDMAAYEAWQQE   684
TraD-CS6pINV    PVSPVINDKKSDSGVNVPAGGIEQELKMKPEEEMEQQLPPGISESGEVVDMAAYEAWQQE   691
TraD-Ss046pINV  PVSPVINDKKSDSGVNVPAGGIEQELKMKPEEEMEQQLPPGISESGEVVDMAAYEAWQQE   700
TraD-53GpINV    PVSPVINDKKSDSGVNVPAGGIEQELKMKPEEEMEQQLPPGISESGEVVDMAAYEAWQQE   688
                ****.********************************************************

TraD-F          NHPDIQQQMQRREEVNINVHRERGEDVEPGDDF*  717
TraD-CS6pINV    NHPDTWQQMQRREEVNINVHRERGEDVEPGDDF*  724
TraD-Ss046pINV  NHPDTWQQMQRREEVNINVHRERGEDVEPGDDF*  733
TraD-53GpINV    NHPDTWQQMQRREEVNINVHRERGEDVEPGDDF*  721
                ****  ****************************
```

(B)

```
TraI-CS6pINV    MMSIAQVRSAGSADNYYTDKDNYYVLGSMGERWAGQGAEQLGLQGSVDKDVFTRLLEGRL   60
TraI-Ss046pINV  MMSIAQVRSAGSADNYYTDKDNYYVLGSMGERWAGQGAEQLGLQGSVDKDVFTRLLEGRL   60
TraI-F          MMSIAQVRSAGSAGNYYTDKDNYYVLGSMGERWAGRGAEQLGLQGSVDKDVFTRLLEGRL   60
TraI-53GpINV    MMSIAQVRSAGSADNYYTDKDNYYVLGSMGERWAGQGAEQLGLQGSVDKDVFTRLLEGRL   60
                *************.********************:************************
                                                    Relaxase domain
```

```
TraI-CS6pINV        PDGADLSRMQDGSNRHRPGYDLTFSAPKSVSMMAMLGGDKRLIEAHNQAVDFAVRQVEAL       120
TraI-Ss046pINV      PDGADLSRMQDGSNRHRPGYDLTFSAPKSVSMMAMLGGDKRLIEAHNQAVDFAVRQVEAL       120
TraI-F              PDGADLSRMQDGSNRHRPGYDLTFSAPKSVSMMAMLGGDKRLIDAHNQAVDFAVRQVEAL       120
TraI-53GpINV        PDGADLSRMQDGSNRHRPGYDLTFSAPKSVSMMAMLGGDKRLIEAHNQAVDFAVRQVEAL       120
                    ****************************************:****************

TraI-CS6pINV        ASTRVMTDGQSETVLTGNLVMALFNHDTSRDQEPQLHTHAVVTNVTQHNGEWKTLSSDKV       180
TraI-Ss046pINV      ASTRVMTDGQSETVLTGNLVMALFNHDTSRDQEPQLHTHAVVTNVTQHNGEWKTLSSDKV       180
TraI-F              ASTRVMTDGQSETVLTGNLVMALFNHDTSRDQEPQLHTHAVVANVTQHNGEWKTLSSDKV       180
TraI-53GpINV        ASTRVMTDGQSETVLTGNLVMALFNHDTSRDQEPQLHTHAVVTNVTQHNGEWKTLSSDKV       180
                    ****************************************:****************

TraI-CS6pINV        GKTGFIENVYANQIAFGRLYREKRKEQVEALGYETEVVGKHGMWEMPGVPVEAFSGRSQT       240
TraI-Ss046pINV      GKTGFIENVYANQIAFGRLYREKRKEQVEALGYETEVVGKHGMWEMPGVPVEAFSGRSQT       240
TraI-F              GKTGFIENVYANQIAFGRLYREKLKEQVEALGYETEVVGKHGMWEMPGVPVEAFSGRSQT       240
TraI-53GpINV        GKTGFIENVYANQIAFGRLYREKRKEQVEALGYETEVVGKHGMWEMPGVPVEAFSGRSQT       240
                    ********************** ***********************************

TraI-CS6pINV        IREAVGEDASLKSRDVAALDTRKSKQHVDPEVRMAEWMQTLKETGFDIRAYRDAAEQRAY       300
TraI-Ss046pINV      IREAVGEDASLKSRDVAALDTRKSKQHVDPEVRMAEWMQTLKETGFDIRAYRDAAEQRAY       300
TraI-F              IREAVGEDASLKSRDVAALDTRKSKQHVDPEIKMAEWMQTLKETGFDIRAYRDAADQRAD       300
TraI-53GpINV        IREAVGEDASLKSRDVAALDTRKSKQHVDPEVRMAEWMQTLKETGFDIRAYRDAAEQRAY       300
                    ********************************::**********************:***

TraI-CS6pINV        TRTQTPGPASQDGPDVQQAVTQAIAGLSERKVQFMYTDLLARTVGILPPENGVIERARAG       360
TraI-Ss046pINV      TRTQTPGPASQDGPDVQQAVTQAIAGLSERKVQFMYTDLLARTVGILPPENGVIERARAG       360
TraI-F              LRTLTPGPASQDGPDVQQAVTQAIAGLSERKVQFTYTDVLARTVGILPPENGVIERARAG       360
TraI-53GpINV        TRTQTPGPASQDGPDVQQAVTQAIAGLSERKVQFMYTDLLARTVGILPPENGVIERARAG       360
                     ** ****************************** ***:**********************
                                                        Helicase domain

TraI-CS6pINV        IDEAISREQLIPLDREKGLFTSGIHMLDELSVRALSRDIMKQNRVTVHPEKSVPRTAGYS       420
TraI-Ss046pINV      IDEAISREQLIPLDREKGLFTSGIHMLDELSVRALSRDIMKQNRVTVHPEKSVPRTAGYS       420
TraI-F              IDEAISREQLIPLDREKGLFTSGIHVLDELSVRALSRDIMKQNRVTVHPEKSVPRTAGYS       420
TraI-53GpINV        IDEAISREQLIPLDREKGLFTSGIHMLDELSVRALSRDIMKQNRVTVHPEKSVPRTAGYS       420
                    ************************:**********************************

TraI-CS6pINV        DAVSVLAQDRPSLAIVSGQGGAAGQRERVAELVMMAREQGREVQIIAADRRSQMNLKQDE       480
TraI-Ss046pINV      DAVSVLAQDRPSLAIVSGQGGAAGQRERVAELVMMAREQGREVQIIAADRRSQMNLKQDE       480
TraI-F              DAVSVLAQDRPSLAIVSGQGGAAGQRERVAELVMMAREQGREVQIIAADRRSQMNMKQDE       480
TraI-53GpINV        DAVSVLAQDRPSLAIVSGQGGAAGQRERVAELVMMAREQGREVQIIAADRRSQMNLKQDE       480
                    ******************************************************:****

TraI-CS6pINV        RLSGELITGRRQLLEGMAFTPGSTVIVDQGEKLSLKETLTLLDGAARHNVQVLITDSGQR       540
TraI-Ss046pINV      RLSGELITGRRQLLEGMAFTPGSTVIVDQGEKLSLKETLTLLDGAARHNVQVLITDSGQR       540
TraI-F              RLSGELITGRRQLLEGMAFTPGSTVIVDQGEKLSLKETLTLLDGAARHNVQVLITDSGQR       540
TraI-53GpINV        RLSGELITGRRQLLEGMAFTPGSTVIVDQGEKLSLKETLTLLDGAARHNVQVLITDSGQR       540
                    ************************************************************

TraI-CS6pINV        TGTGSALMAMKDAGVNTYRWQGGEQRPATIISEPDRNVRYARLAGDFAASVKAGEESVAQ       600
TraI-Ss046pINV      TGTGSALMAMKDAGVNTYRWQGGEQRPATIISEPDRNVRYARLAGDFAASVKAGEESVAQ       600
TraI-F              TGTGSALMAMKDAGVNTYRWQGGEQRPATIISEPDRNVRYARLAGDFAASVKAGEESVAQ       600
TraI-53GpINV        TGTGSALMAMKDAGVNTYRWQGGEQRPATIISEPDRNVRYARLAGDFAASVKAGEESVAQ       600
                    ************************************************************

TraI-CS6pINV        VSGVREQAILTQAIRSELKTQGVLGHPEVTMTALSPVWLDSRSRYLRDMYRPGMVMEQWN       660
TraI-Ss046pINV      VSGVREQAILTQAIRSELKTQGVLGHPEVTMTALSPVWLDSRSRYLRDMYRPGMVMEQWN       660
TraI-F              VSGVREQAILTQAIRSELKTQGVLGLPEVTMTALSPVWLDSRSRYLRDMYRPGMVMEQWN       660
TraI-53GpINV        VSGVREQAILTQAIRSELKTQGVLGHPEVTMTALSPVWLDSRSRYLRDMYRPGMVMEQWN       660
                    ************************ **********************************
```

```
TraI-CS6pINV        PETRSHDRYVTERVTAQSHSLTLRNAQGETQVVRISSLDSSWSLFRPEKMPVADGERLRV        720
TraI-Ss046pINV      PETRSHDRYVTERVTAQSHSLTLRNAQGETQVVRISSLDSSWSLFRPEKMPVADGERLRV        720
TraI-F              PETRSHDRYVIDRVTAQSHSLTLRDAQGETQVVRISSLDSSWSLFRPEKMPVADGERLRV        720
TraI-53GpINV        PETRSHDRYVTERVTAQSHSLTLRNAQGETQVVRISSLDSSWSLFRPEKMPVADGERLRV        720
                    ********** :************:***********************************

TraI-CS6pINV        TGKIPGLRVSGGDRLQVASVSEDAMTVVVPGRAEPATLPVSDSPFTALKLENGWVETPGH        780
TraI-Ss046pINV      TGKIPGLRVSGGDRLQVASVSEDAMTVVVPGRAEPATLPVSDSPFTALKLENGWVETPGH        780
TraI-F              TGKIPGLRVSGGDRLQVASVSEDAMTVVVPGRAEPATLPVSDSPFTALKLENGWVETPGH        780
TraI-53GpINV        TGKIPGLRVSGGDRLQVASVSEDAMTVVVPGRAEPATLPVSDSPFTALKLENGWVETPGH        780
                    ************************************************************

TraI-CS6pINV        SVSDSATVFASVTQMAMDNATLNGLARSGRDVRLYSSLDETRTAEKLARHPSFTVVSEQI        840
TraI-Ss046pINV      SVSDSATVFASVTQMAMDNATLNGLARSGRDVRLYSSLDETRTAEKLARHPSFTVVSEQI        840
TraI-F              SVSDSATVFASVTQMAMDNATLNGLARSGRDVRLYSSLDETRTAEKLARHPSFTVVSEQI        840
TraI-53GpINV        SVSDSATVFASVTQMAMDNATLNGLARSGRDVRLYSSLDETRTAEKLARHPSFTVVSEQI        840
                    ************************************************************

TraI-CS6pINV        KARAGETLLETAISLQKAGLHTPAQQAIHLALPVLESKNLAFSMVDLLTEAKSFAAEGTG        900
TraI-Ss046pINV      KARAGETLLETAISLQKAGLHTPAQQAIHLALPVLESKNLAFSMVDLLTEAKSFAAEGTG        900
TraI-F              KTRAGETSLETAISHQKSALHTPAQQAIHLALPVVESKKLAFSMVDLLTEAKSFAAEGTG        900
TraI-53GpINV        KARAGETLLETAISLQKAGLHTPAQQAIHLALPVLESKNLAFSMVDLLTEAKSFAAEGTG        900
                    *:***** ****** **:.***************:***:**********************

TraI-CS6pINV        FADLGGEINAQIKRGDLLYVDVAKGYGTGLLVSRASYEAEKSILRHILEGKEAVTPLMER        960
TraI-Ss046pINV      FADLGGEINAQIKRGDLLYVDVAKGYGTGLLVSRASYEAEKSILRHILEGKEAVTPLMER        960
TraI-F              FTELGGEINAQIKRGDLLYVDVAKGYGTGLLVSRASYEAEKSILRHILEGKEAVMPLMER        960
TraI-53GpINV        FADLGGEINAQIKRGDLLYVDVAKGYGTGLLVSRASYEAEKSILRHILEGKEAVTPLMER        960
                    *::************************************************ *****

TraI-CS6pINV        VPGELMEKLTSGQRAATRMILETSDRFTVVQGYAGVGKTTQFRAVMSAVNMLPESERPRV        1020
TraI-Ss046pINV      VPGELMEKLTSGQRAATRMILETSDRFTVVQGYAGVGKTTQFRAVMSAVNMLPESERPRV        1020
TraI-F              VPGELMEKLTSGQRAATRMILETSDRFTVVQGYAGVGKTTQFRAVMSAVNMLPESERPRV        1020
TraI-53GpINV        VPGELMEKLTSGQRAATRMILETSDRFTVVQGYAGVGKTTQFRAVMSAVNMLPESERPRV        1020
                    ************************************************************

TraI-CS6pINV        VGLGPTHRAVGEMRSAGVDAQTLASFLHDTQLQQRSGETPDFSNTLFLLDESSMVGNTDM        1080
TraI-Ss046pINV      VGLGPTHRAVGEMRSAGVDAQTLASFLHDTQLQQRSGETPDFSNTLFLLDESSMVGNTDM        1080
TraI-F              VGLGPTHRAVGEMRSAGVDAQTLASFLHDTQLQQRSGETPDFSNTLFLLDESSMVGNTDM        1080
TraI-53GpINV        VGLGPTHRAVGEMRSAGVDAQTLASFLHDTQLQQRSGETPDFSNTLFLLDESSMVGNTDM        1080
                    ************************************************************

TraI-CS6pINV        ARAYALIAVGGGRAVASGDTDQLQAIAPGQPFRLQQTRSAADVVIMKEIVRQTPELREAV        1140
TraI-Ss046pINV      ARAYALIAVGGGRAVASGDTDQLQAIAPGQPFRLQQTRSAADVVIMKEIVRQTPELREAV        1140
TraI-F              ARAYALIAAGGGRAVASGDTDQLQAIAPGQPFRLQQTRSAADVAIMKEIVRQTPELREAV        1140
TraI-53GpINV        ARAYALIAVGGGRAVASGDTDQLQAIAPGQPFRLQQTRSAADVVIMKEIVRQTPELREAV        1140
                    ********.***********************************.*****************

TraI-CS6pINV        YSLINRDVERALSGLERVKPSQVPRLEGAWAPEHSVTEFSHSQEAKLAEAQQKAMLKGEA        1200
TraI-Ss046pINV      YSLINRDVERALSGLERVKPSQVPRLEGAWAPEHSVTEFSHSQEAKLAEAQQKAMLKGEA        1200
TraI-F              YSLINRDVERALSGLESVKPSQVPRQEGAWAPEHSVTEFSHSQEAKLAEAQQKAMLKGEA        1200
TraI-53GpINV        YSLINRDVERALSGLERVKPSQVPRLEGAWAPEHSVTEFSHSQEAKLAEAQQKAMLKGEA        1200
                    *************** ******** **********************************

TraI-CS6pINV        FPDVPMTLYEAIVRDYTGRTPEAREQTLIVTHLNEDRRVLNSMIHDAREKAGELGQVQVM        1260
TraI-Ss046pINV      FPDVPMTLYEAIVRDYTGRTPEAREQTLIVTHLNEDRRVLNSMIHDAREKAGELGQVQVM        1260
TraI-F              FPDVPMTLYEAIVRDYTGRTPEAREQTLIVTHLNEDRRVLNSMIHDVREKAGELGKEQVM        1260
TraI-53GpINV        FPDVPMTLYEAIVRDYTGRTPEAREQTLIVTHLNEDRRVLNSMIHDAREKAGELGQVQVM        1260
                    ********************************************.********: ***
```

```
TraI-CS6pINV        VPVLNTANIRDGELRRLSTWENNPDALALVDNVYHRIAGISKDDGLITLQDAEGNTRLIS        1320
TraI-Ss046pINV      VPVLNTANIRDGELRRLSTWENNPDALALVDNVYHRIAGISKDDGLITLQDAEGNTRLIS        1320
TraI-F              VPVLNTANIRDGELRRLSTWETHRDALVLVDNVYHRIAGISKDDGLITLQDAEGNTRLIS        1320
TraI-53GpINV        VPVLNTANIRDGELRRLSTWENNPDALALVDNVYHRIAGISKDDGLITLQDAEGNTRLIS        1320
                    **********************.: ***.*********************************

TraI-CS6pINV        PREAVAEGVTLYTPDTIRVGTGDRIRFTKSDRERGYVANSVWTVTAVSGDSVTLSDGQQT        1380
TraI-Ss046pINV      PREAVAEGVTLYTPDTIRVGTGDRIRFTKSDRERGYVANSVWTVTAVSGDSVTLSDGQQT        1380
TraI-F              PREAVAEGVTLYTPDTIRVGTGDRMRFTKSDRERGYVANSVWTVTAVSGDSVTLSDGQQT        1380
TraI-53GpINV        PREAVAEGVTLYTPDTIRVGTGDRIRFTKSDRERGYVANSVWTVTAVSGDSVTLSDGQQT        1380
                    ************************:*********************************

TraI-CS6pINV        RVIRPGQERAEQHIDLAYAITAHGAQGASETFAIALEGTEGNRKLMR*PEFNGHSFLK*N        1438
TraI-Ss046pINV      RVIRPGQERAEQHIDLAYAITAHGAQGASETFAIALEGTEGNRKLMR*PEFNGHSFLK*N        1438
TraI-F              REIRPGQEQAEQHIDLAYAITAHGAQGASETFAIALEGTEGNRKLMAGFESAYVALSRMK        1440
TraI-53GpINV        RVIRPGQERAEQHIDLAYAITAHGAQGASETFAIALEGTEGNRKLMAGFESAYVALSRMK        1440
                    * ******:************************************   *    :: : :

TraI-CS6pINV        GIR-------RS***A----EKPNVTLKSSKPKLSERFL--KINFRSVKALPDYLFLK-A        1481
TraI-Ss046pINV      GIR-------RS***A----EKPNVTLKSSKPKLSERFL--KINFRSVKALPDYLFLK-A        1481
TraI-F              QHVQVYTDNRQGWTDAINNAVQKGTAHDVFEPKPDREVMNAERLFSTARELRDVAAGRAV        1500
TraI-53GpINV        QHVQVYTDNRQGWTDAINNAVQKGTAHDVFEPKPDREVMNAERLFSTARELRDVAAGRAV        1500
                              :.   *     : ..: .  :** ....:  :  * :.: * *    : .
                    ----------C-terminal domain essential for conjugative DNA transfer

TraI-CS6pINV        L*DNGLPPPEKGSVLLVPARWLNWNLKFCNCVRR*MKLALSEIY*KKQQRILHRSR*KIR        1537
TraI-Ss046pINV      L*DNGLPPPEKGSVLLVPARWLNWNLKFCNCVRR*MKLALSEIY*KKQQRILHRSR*KIR        1537
TraI-F              LRQAGLAGGDSPARFIAPGRK------YPQ-----------------------------        1524
TraI-53GpINV        LRQAGLAGGDSPARFIAPGRK------YPQ-----------------------------        1524
                    * : **   :. : ::.*.*       : :

TraI-CS6pINV        VNRTMATTISH*SDVSGIWCIQERLL---------------QLGTA*TL-----------        1569
TraI-Ss046pINV      VNRTMATTISH*SDVSGIWCIQERLL---------------QLGTA*TL-----------        1569
TraI-F              -PYVALPAFDRNGKSAGIWLNPLTTDDGNGLRGFSGEGRVKGSGDAQFVALQGSRNGESL        1583
TraI-53GpINV        -PYVALPAFDRNGKSAGIWLNPLTTDDGNGLRGFSGEGRVKGSGDAQFVALQGSRNGESL        1583
                       .   ::.: .. :***                        * *  :

TraI-CS6pINV        -RQK---TK**AAKTGD*G--GTYPHSRNIWNPAAPDGAGREWHHRWS*PTGTSS*GA--        1616
TraI-Ss046pINV      -RQK---TK**AAKTGD*G--GTYPHSRNIWNPAAPDGAGREWHHRWS*PTGTSS*GA--        1616
TraI-F              LADNMQDGVRIARDNPDSGVVVRIAGEGRPWNPGAITG-GRVWGDIPD---NSVQPGAGN        1639
TraI-53GpINV        LADNMQDGVRIARDNPDSGVVVSIAGEGRPWNPGAITG-GRVWGDIPD---NSVQPGAGN        1639
                      ::       * .. * *       . . ***.*  * ** * .  .   .: . **

TraI-CS6pINV        KATL*AETQVQ--SDYELEPQSASCAKSAEPDVRSYS-TKSGLGGGPDVCCHTGGMVVPR        1672
TraI-Ss046pINV      KATL*AETQVQ--SDYELEPQSASCAKSAEPDVRSYS-TKSGLGGGPDVCCHTGGMVVPR        1672
TraI-F              GEPVTAEVLAQRQAEEAIRRET---ERRADEIVRKMAENKPDLPDGKTEQ----------        1686
TraI-53GpINV        GEPVTAEVLAQRQAEEAIRRET---ERRADEIVRKMAENKPDLPDGKTEL----------        1686
                       : **. .*  ::  :. ::    : *:  **. : .* .* .*

TraI-CS6pINV        WHQRCLYVRNCRLRHGRAHDKRADR*SPVYGAQEPAPTCRANPPL*SRFTVLRIRLPGHT        1730
TraI-Ss046pINV      WHQRCLYVRNCRLRHGRAHDKRADR*SPVYGAQEPAPTCRANPPL*SRFTVLRIRLPGHT        1730
TraI-F              ---------AVREIAGQERDRAA-----ITEREAALPEGVLREPQRVREAVREIA----R        1728
TraI-53GpINV        ---------AVREIAGQERDRAA-----ITEREAALPESVLRESQREREAVREVA----R        1728
                             *   *: :*: *     :   :   *    .     * :* .:

TraI-CS6pINV        GAVWSENINVA*R*LLRQRSDGKLLGNAEK*EPEPLSF**PG*SHLSNTGIH*DFLQSSA        1783
TraI-Ss046pINV      GAVWSENINVA*R*LLRQRSDGKLLGNAEK*EPEPLSF**PG*SHLSNTGIH*DFLQSSA        1783
TraI-F              ENLLQERLQQMERDMVRDLQKEKTLGGD*------------------------------        1756
TraI-53GpINV        ENLLQERLQQMERDMVRDLQKEKTPGGD*------------------------------        1756
                      : .*.::   * ::*: .. *  *.
```

(C)

```
TraX-F              MTTDNTNTTRNDSLVARTDTWLQSLLVWSPGQRDIIKTVALVLMVLDHANRILHLDQSWM        60
TraX-CS6pINV        MTTDNTNTTRNDSLAARTDTWLQSFLVWSPGQRDIIKTVALVLMVLDHINLIFQLKQEWM        60
TraX-Ss046pINV      MTTDNTNTTRNDSLAARTDTWLQSFLVWSPGQRDIIKTVALVLMVLDHINLIFQLKQEWM        60
TraX-53GpINV        MTTDNTNTTRNDSLAARTDTWLQSFLVWSPGQRDIIKTVALVLMVLDHINLIFQLKQEWM        60
                    **************.*********:********************** * *::*.*.**

TraX-F              FLAGRGAFPLFALVWGLNLSRHAHIRQLAINRLWGWAVFSQFWYYFAGFPWYEGNILFAF        120
TraX-CS6pINV        FLAGRGAFPLFALVWGLNLSRHAHIRQPAINRLWGWGIIAQFAYYLAGFPWYEGNILFAF        120
TraX-Ss046pINV      FLAGRGAFPLFALVWGLNLSRHAHIRQPAINRLWGWGIIAQFAYYLAGFPWYEGNILFAF        120
TraX-53GpINV        FLAGRGAFPLFALVWGLNLSRHAHIRQPAINRLWGWGIIAQFAYYLAGFPWYEGNILFAF        120
                    ************************** ********.:::** **:***************

TraX-F              AVAAQVLTWCETRSGWRTAAAILMMALWGPLSGTSYGIAGLLMLAVSHRLYRAEDRMERL        180
TraX-CS6pINV        AVAAQVLTWCETRSGWRTAAAILLMALLGAFVRHQLRHCR------------AADAGSQ-        167
TraX-Ss046pINV      AVAAQVLTWCETRSGWRTAAAILLMALLGAFVRHQLRHCR------------AADAGSQ-        167
TraX-53GpINV        AVAAQVLTWCETRSGWRTAAAILLMALLGAFVRHQLRHCR------------AADAGSQ-        167
                    ***********************:*** * :   .   .            * *  .:

TraX-F              VLVACLLAVIPALNLATSDAAAVAGLVMTVLTVGLVSCAGKLLSRFWPGDFFPTFYACHL        240
TraX-CS6pINV        ------PPTVPGGRQ---SG---------------TSGAGRLPAGCYSGA*---------        193
TraX-Ss046pINV      ------PPTVPGGRQ---SG---------------TSGAGRLPAGCYSGA*---------        193
TraX-53GpINV        ------PPTVPGGRQ---SG---------------TSGAGRLPAGCYSGA*---------        193
                           .:*. .    ..              .* **:* :  : *

TraX-F              AVLGVLAL*   248
TraX-CS6pINV        ---------   193
TraX-Ss046pINV      ---------   193
TraX-53GpINV        ---------   193
```

Figure 7.3: Amino acid alignment of TraD (A), TraI (B) and TraX (C) from pINV of *S. sonnei* CS6, 53G and Ss046, as well as *E. coli* F plasmid.

The alignment was created by ClustalOmega. Astericks (*) indicate identical nucleotides of all sequences in the comparison. The domains of important functions for TraD and TraI are annotated according to those from F plasmids (221, 330, 331). TraI on CS6, Ss046 and 53G pINV share 82.70%, 82.70% and 97.15% amino acid identity, respectively, comparing to F plasmid TraI.

(A)

```
relBE-CS6pINV       atgcc-----------------taacatcattctgtctgagaccagtgcaagtgtcagtg 43
relBE-CS6pRES       atgaaaaacaataccgcacaggcaacaaaagtcattaccgcgcatgtgccattacctatg 60
                    ***                   ****  * **  *      *  **** * *  *  **

relBE-CS6pINV       aactgaaaaa-aaatcctatggctacagtcagtgccggagatggttaccctgtcgccatt 102
relBE-CS6pRES       gctgataaagtcgaccagatggccgcc------agactggaacgctcccggggctggatt 114
                          ***    * *  *****  *            **  * * **  * *   ***

relBE-CS6pINV       cttaaccggaatcagccggcattctactgtgtgcctgctgaactgtatgaaaggatgctt 162
relBE-CS6pRES       atcaaacaggcgctttctgcatg---------gcttgcccaggaggaggagcgtaatcgc 165
                     * ** * *   *   * ****         ** ***  *   * * **  * *  *

relBE-CS6pINV       gatgctctggatgaccaggaactggtgaaactggttactgaacgcagcaatcagccactg 222
relBE-CS6pRES       ctgacgctggaagccctggacgatgtgacatccggacaggttatcgaccatcaggctgta 225
                        * ***** * ** ***    **** *   *     *    *  * ***** *  *
```

```
relBE-CS6pINV      catgatgtggacctggacagttatttatgacttatacggtaaaattcagggatgatgcgc 282
relBE-CS6pRES      caggcctggg-cggacagcctcagtactgacaa-t-------ccgttaccggtgccacgc 276
                   ** *    ** *    *   * * *  ****   *          * *  * **   ***

relBE-CS6pINV      tcaaggaatggctgaaactggataagaccattca---------------acaacagtttg 327
relBE-CS6pRES      tgatggaactgaagtggaccagtaaggcgctttctgatttggcgcggttatatgattttc 336
                   * * ****  *  *       **** *  **                * *  * ***

relBE-CS6pINV      tgaa--aaaactaaaaaaatgcagcgaaaatcctcatataccgtcggcaaaactaagagg 385
relBE-CS6pRES      tggtgctggccagtaaacctgcggccgccagaacgg--tgcagtccct-----gacacag 389
                   **       *   ***  *** **    *       * * ***        *    *

relBE-CS6pINV      gctaaaggactgctacaaaattaaactacgtgcatcgggatttcgtctggtttatcaggt 445
relBE-CS6pRES      gctccggtcattctgtta-actcatccacgtatgggggaaca------gttatttca--- 439
                   ***   *   * **   * * * * * ****    ** *       * * * ***

relBE-CS6pINV      gattgacgatatgttaattattgcggttgttgcggtaggaaaacgagaacgtagtaacgt 505
relBE-CS6pRES      gtttgaacccagggaagtcagacggatttttgctggcgagtacgaaatccgttacgaact 499
                   * ****    * *  * * *    * ** **** *  *   *   *   ***    *  *

relBE-CS6pINV      ttat----------------aatctcgccagtgagagaa-tgag-------atag      536
relBE-CS6pRES      tactggccagactatttatgtattgcgcctgtggcacacacgagaaaacaggtag      554
                   *  *               **  **** ***  * *   ***       ***
```

(B)

```
relB-CS6pINV      MPN-------IILSETSASVSELKKNPMATVSAGDGYPVAILNRNQPAFYCVPAELYERM  53
relB-CS6pRES      MKNNTAQATKVITAHVPLPMADKVDQMAARLERSRGWII---KQALSAWLAQEEERNRLT  57
                  * *       :* :..   :::  .:  * :. . *: :   ::   *: .   *  .

relB-CS6pINV      LDALDDQELVKLVTERSNQPLHDVDLDSYL*------------  83
relB-CS6pRES      LEALDDVTSGQVI--------DHQAVQAWADSLSTDNPLPVPR*  92
                  *:****    :::        ..  ::::

relE-CS6pINV      MTYTVKFRDDALKEWLKLDKTIQ----QQFVKKLKKCSENPHIPSAKLRGLKDCYKI---  53
relE-CS6pRES      --MELKWTSKALSDLARLYDFLVLASKPAAARTVQSLTQAPVILLTHPRMGEQLFQFEPR  58
                      :*: ..**.:  :* . :        .:.::. :: * *  :: *  :: :::

relE-CS6pINV      ---KLRASGFRLVYQVIDDMLIIAVVAVGKRERSNVYNLASERMR*        95
relE-CS6pRES      EVRRIFAGEYEIRYELTGQTIYVLRLW-HTRENR*-----------        91
                     :: *. :.: *:: .: : :  :   .**.
```

Figure 7.4: Nucleotide (A) and amino acid (B) alignment of RelBE on *S. sonnei* CS6 pINV and pRES. The alignments were created by ClustalOmega. Astericks (*) indicate identical nucleotides and amino acids of all sequences in the comparison. The two RelBE systems share 46.06% nucleotide identity. The RelB and RelE share 16.67% and 13.58% amino acid identity.

(A)

```
Kil-pKY1            MRKRFFVGIFAINLLVGCQANYIRDVQGGTVAPSSSSKLTGISVQ*     45
Lys-CS6pColE1       MRKRFFVGIFAINLLVGCQANYIRDVQGGTVAPSSSSKLTGISVQ*     45
Lys-Ss046pColE1     MRKRFFVGIFAINLLVGCQANYIRDVQGGTVAPSSSSKLTGISVQ*     45
                    *********************************************
```

(B)

```
Cea-pKY1            METAVAYYKDGVPYDDKGEVIITLLNGNPDGSGSGGGGGTGGSKSESSAAIHATAKWSTA      60
Cea-CS6pColE1       METAVAYYKDGVPYDDKGQVIITLLNGNPDGSGSGGGGGTGGSKSESSAAIHATAKWSTA      60
Cea-Ss046pColE1     METAVAYYKDGVPYDDKGQVIITLLNGNPDGSGSGGGGGTGGSKSESSAAIHATAKWSTA      60
                    ******************:*******************************************

Cea-pKY1            QLKKTQAEQAARAKAAAEAQAKAKANRDALTQHLKDIVNEALRHNSTHPEVIDLLMPIMQ      120
Cea-CS6pColE1       QLKKTQAEQAARAKAAAEAQAKAKANRDALTQHLKDIVNEALRHNSTHPEVIDLAHANNA      120
Cea-Ss046pColE1     QLKKTQAEQAARAKAAAEAQAKAKANRDALTQHLKDIVNEALRHNSTHPEVIDLAHANNA      120
                    **************************************************

Cea-pKY1            RCRQKQSGCALQKQKKKPVKKRKRAEKSFQEAEQRRKEIEKEQAETERQLKLAEDEEKRL      180
Cea-CS6pColE1       AMQAEAERLRLAKAEEKARKEAEAAEKAFQEAEQRRKEIEKEQAETERQLKLAEDEEKRL      180
Cea-Ss046pColE1     AMQAEAERLRLAKAEEKARKEAEAAEKAFQEAEQRRKEIEKEQAETERQLKLAEDEEKRL      180
                      : : .   * * ::*  *: : ***:*********************************

Cea-pKY1            AALSEEARAVEVAQKNLAAAQSELAKVDEEINTLNTRLSSSIHARDAETNTLSGKRNELD      240
Cea-CS6pColE1       AALSEEARAVEVAQKNLAAAQSELAKVDEEINTLNTRLSSSIHARDAETNTLSGKRNELD      240
Cea-Ss046pColE1     AALSEEARAVEVAQKNLAAAQSELAKVDEEINTLNTRLSSSIHARDAETNTLSGKRNELD      240
                    ************************************************************

Cea-pKY1            QASAKYKELDERVKLLSPRANDPLQSRPFFEATRLRARRGDEMEEKQKQVTATETRLNQI      300
Cea-CS6pColE1       QASAKYKELDERVKLLSPRANDPLQSRPFFEATRLRARAGDEMEEKQKQVTASETRLNQI      300
Cea-Ss046pColE1     QASAKYKELDERVKLLSPRANDPLQSRPFFEATRLRARAGDEMEEKQKQVTASETRLNQI      300
                    ************************************ *************:*******

Cea-pKY1            SSEINGIQEAISQANNKRSTAVSRIHDAEDNLKTAQTNLLNSQIKDAVDATVSFYQTLSE      360
Cea-CS6pColE1       SSEINGIQEAISQANNKRSTAVSRIHDAEDNLKTVQTNLLNSQIKDAVDATVSFYQTLSE      360
Cea-Ss046pColE1     SSEINGIQEAISQANNKRSTAVSRIHDAEDNLKTVQTNLLNSQIKDAVDATVSFYQTLSE      360
                    **********************************.*************************

Cea-pKY1            KYGEKYSKMAQELADKSKGKKISNVNEALAAFEKYKDVLNKKFSKADRDAIFNALEAVKY      420
Cea-CS6pColE1       KYGEKYSKMAQELADKSKGKKISNVNEALAAFEKYKDVLNKKFSKADRDAIFNALEAVKY      420
Cea-Ss046pColE1     KYGEKYSKMAQELADKSKGKKISNVNEALAAFEKYKDVLNKKFSKADRDAIFNALEAVKY      420
                    ************************************************************

Cea-pKY1            EDWAKHLDQFAKYLKITGHVSFGYDVVSDILKIKDTGDWKPLFLTLEKKAVDAGVSYVVV      480
Cea-CS6pColE1       EDWAKHLDQFAKYLKITGHVSFGYDVVSDILKIKDTGDWKPLFLTLEKKAVDAGVSYVVV      480
Cea-Ss046pColE1     EDWAKHLDQFAKYLKITGHVSFGYDVVSDILKIKDTGDWKPLFLTLEKKAVDAGVSYVVV      480
                    ************************************************************

Cea-pKY1            LLFSVLAGTTLGIWGIAIVTGILCAFIDKNKLNTINEVLGI* 521
Cea-CS6pColE1       LLFSVLAGTTLGIWGIAIVTGILCAFIDKNKLNTINEVLGI* 521
Cea-Ss046pColE1     LLFSVLAGTTLGIWGIAIVTGILCAFIDKNKLNTINEVLGI* 521
                    ****************************************
```

(C)

```
Imm-CS6         MSLRYYIKNILFGLYCALIYIYLITKNNEGYYFLASDKMLYAIVISTILCPYSKYAIEHI     60
Imm-Ss046       MSLRYYIKNILFGLYCALIYIYLITKNNEGYYFLASDKMLYAIVISTILCPYSKYAIEHI     60
Imm-pKY1        MSLRYYIKNILFGLYCALIYIYLITKNNEGYYFLASDKMLYAIVISTILCPYSKYAIEHI     60
                ************************************************************

Imm-CS6         FFKFIKKDFFRKRKNLNNAPVAKLNLF-MLYNLLCLVLAIPFGLLGLFISIKNN*  113
Imm-Ss046       FFKFIKKDFFRKRKNLNNAPVAKLNLF-MLYNLLCLVLAIPFGLLGLFISIKNN*  113
Imm-pKY1        FFKFIKKDFFRKRKNLNKCPRGKIKPYLCVYNLLCLVLAIPFGLLGLVYINKE*-  113
                *****************:.* .*:: :  :*****************.   *:
```

Figure 7.5: Amino acid alignments of Lys and Kil (A), ColE1 encoded by *cea* (B), and Imm (C) on *S. sonnei* CS6 and Ss046 pColE1 and pKY1.

The alignments were created by ClustalOmega. Astericks (*) indicate identical amino acids of all sequences in the comparison.

```
cea-pKY1            atggaaacagctgtagcgtactataaagatggtgttccttatgatgataaggggaggta      60
cea-CS6pColE1       atggaaacagctgtagcgtactataaagatggtgttccttatgatgataaggggcaggta      60
cea-Ss046pColE1     atggaaacagctgtagcgtactataaagatggtgttccttatgatgataaggggcaggta      60
cea-53GpColE1       atggaaacagctgtagcgtactataaagatggtgttccttatgatgataaggggcaggta      60
                    ************************************************** *****

cea-pKY1            atcattactcttttgaatggtaatccagacgggagtggctctggcggtggtggtggaact      120
cea-CS6pColE1       atcattactcttttgaatggtaatccagacgggagtggctctggcggtggtggtggaact      120
cea-Ss046pColE1     atcattactcttttgaatggtaatccagacgggagtggctctggcggtggtggtggaact      120
cea-53GpColE1       atcattactcttttgaatggtaatccagacgggagtggctctggcggtggtggtggaact      120
                    ************************************************************

cea-pKY1            ggaggtagcaaaagtgaaagttctgcagccattcatgccacagctaaatggtctactgct      180
cea-CS6pColE1       ggaggtagcaaaagtgaaagttctgcagccattcatgccacagctaaatggtctactgct      180
cea-Ss046pColE1     ggaggtagcaaaagtgaaagttctgcagccattcatgccacagctaaatggtctactgct      180
cea-53GpColE1       ggaggtagcaaaagtgaaagttctgcagccattcatgccacagctaaatggtctactgct      180
                    ************************************************************

cea-pKY1            caattgaagaaaacgcaggcagaacaggctgcccgagcaaaagctgccgcagaagcacag      240
cea-CS6pColE1       caattgaagaaaacgcaggcagaacaggctgcccgagcaaaagctgccgcagaagcacag      240
cea-Ss046pColE1     caattgaagaaaacgcaggcagaacaggctgcccgagcaaaagctgccgcagaagcacag      240
cea-53GpColE1       caattgaagaaaacgcaggcagaacaggctgcccgagcaaaagctgccgcagaagcacag      240
                    ************************************************************

cea-pKY1            gctaaagcaaaagcaaaccgggatgcgctgactcaacatctgaaggatattgtgaatgag      300
cea-CS6pColE1       gctaaagcaaaagcaaaccgggatgcgctgactcaacatctgaaggatattgtgaatgag      300
cea-Ss046pColE1     gctaaagcaaaagcaaaccgggatgcgctgactcaacatctgaaggatattgtgaatgag      300
cea-53GpColE1       gctaaagcaaaagcaaaccgggatgcgctgactcaacatctgaaggatattgtgaatgag      300
                    ************************************************************

cea-pKY1            gcgcttcgccataattccactcatccggaggttattgacct-gctcatgccaataatgca      359
cea-CS6pColE1       gcgcttcgccataattccactcatccggaggttattgacctggctcatgccaataatgca      360
cea-Ss046pColE1     gcgcttcgccataattccactcatccggaggttattgacctggctcatgccaataatgca      360
cea-53GpColE1       gcgcttcgccataattccactcatccggaggttattgacctggctcatgccaataatgca      360
                    ***************************************** ******************
```

```
cea-pKY1            gcgatgcaggcagaagcagagcggttgcgccttgcaaaagcagaagaaaaagcccgtaaa      419
cea-CS6pColE1       gcgatgcaggcagaagcagagcggttgcgccttgcaaaagcagaagaaaaagcccgtaaa      420
cea-Ss046pColE1     gcgatgcaggcagaagcagagcggttgcgccttgcaaaagcagaagaaaaagcccgtaaa      420
cea-53GpColE1       gcgatgcaggcagaagcagagcggttgcgccttgcaaaagcagaagaaaaagcccgtaaa      420
                    ************************************************************

cea-pKY1            gaagcggaagcgtgcggaaaagtcttttcaggaagcagaacaacgacgtaaagagataga      479
cea-CS6pColE1       gaagcggaagc-tgcggaaaaggcttttcaggaagcagaacaacgacgtaaagagataga      479
cea-Ss046pColE1     gaagcggaagc-tgcggaaaaggcttttcaggaagcagaacaacgacgtaaagagataga      479
cea-53GpColE1       gaagcggaagc-tgcggaaaagtcttttcaggaagcagaacaacgacgtaaagagataga      479
                    *********** ********** ***********************************

cea-pKY1            gaaggagcaggctgaaacagaacgccaattgaaactggctgaggatgaagagaaacgcct      539
cea-CS6pColE1       gaaggagcaggctgaaacagaacgccaattgaaactggctgaggatgaagagaaacgcct      539
cea-Ss046pColE1     gaaggagcaggctgaaacagaacgccaattgaaactggctgaggatgaagagaaacgcct      539
cea-53GpColE1       gaaggagcaggctgaaacagaacgccaattgaaactggctgaggatgaagagaaacgcct      539
                    ************************************************************

cea-pKY1            ggcagcattgagtgaagaggcccgggctgtggaggtggcacaaaaaaatcttgctgctgc      599
cea-CS6pColE1       ggcagcattgagtgaagaggcccgggctgtggaggtggcacaaaaaaatcttgctgctgc      599
cea-Ss046pColE1     ggcagcattgagtgaagaggcccgggctgtggaggtggcacaaaaaaatcttgctgctgc      599
cea-53GpColE1       ggcagcattgagtgaagaggcccgggctgtggaggtggcacaaaaaaatcttgctgctgc      599
                    ************************************************************

cea-pKY1            acaatctgagctggcgaaagtggatgaagagattaatacgctcaatacccgtttaagctc      659
cea-CS6pColE1       acaatctgagctggcgaaagtggatgaagagattaatacgctcaatacccgtttaagctc      659
cea-Ss046pColE1     acaatctgagctggcgaaagtggatgaagagattaatacgctcaatacccgtttaagctc      659
cea-53GpColE1       acaatctgagctggcgaaagtggatgaagagattaatacgctcaatacccgtttaagctc      659
                    ************************************************************

cea-pKY1            cagtattcatgcccgtgatgcagaaacgaatacgctgtccggaaaacgaaatgagttgga      719
cea-CS6pColE1       cagtattcatgcccgtgatgcagaaacgaatacgctgtccggaaaacgaaatgagttgga      719
cea-Ss046pColE1     cagtattcatgcccgtgatgcagaaacgaatacgctgtccggaaaacgaaatgagttgga      719
cea-53GpColE1       cagtattcatgcccgtgatgcagaaacgaatacgctgtccggaaaacgaaatgagttgga      719
                    ************************************************************

cea-pKY1            tcaggcatctgctaaatataaagaactggatgaaagggttaaacttttatctcctagggc      779
cea-CS6pColE1       tcaggcatctgctaaatataaagaactggatgaaagggttaaacttttatctcctagggc      779
cea-Ss046pColE1     tcaggcatctgctaaatataaagaactggatgaaagggttaaacttttatctcctagggc      779
cea-53GpColE1       tcaggcatctgctaaatataaagaactggatgaaagggttaaacttttatctcctagggc      779
                    ************************************************************

cea-pKY1            gaatgatccgcttcagagtcgtccttttttttgaggccaccagactacgggcgagacgtgg      839
cea-CS6pColE1       gaatgatccgcttcagagtcgtccttttttttgaggccaccagactacgggcgagagctgg      839
cea-Ss046pColE1     gaatgatccgcttcagagtcgtccttttttttgaggccaccagactacgggcgagagctgg      839
cea-53GpColE1       gaatgatccgcttcagagtcgtccttttttttgaggccaccagactacgggcgagagctgg      839
                    ********************************************************  ***

cea-pKY1            tgatgagatggaggaaaaacaaaagcaggtaacagctacagaaacgcgtcttaaccagat      899
cea-CS6pColE1       tgatgagatggaggaaaaacaaaagcaggtaacagcatcagaaacgcgtcttaaccagat      899
cea-Ss046pColE1     tgatgagatggaggaaaaacaaaagcaggtaacagcatcagaaacgcgtcttaaccagat      899
cea-53GpColE1       tgatgagatggaggaaaaacaaaagcaggtaacagcatcagaaacgcgtcttaaccagat      899
                    ***********************************  **********************

cea-pKY1            tagctctgagataaatggaatccaagaggctatttctcaggctaataataagcgaagtac      959
cea-CS6pColE1       tagctctgagataaatggaatccaagaggctatttctcaggctaataataagcgaagtac      959
cea-Ss046pColE1     tagctctgagataaatggaatccaagaggctatttctcaggctaataataagcgaagtac      959
cea-53GpColE1       tagctctgagataaatggaatccaagaggctatttctcaggctaataataagcgaagtac      959
                    ************************************************************
```

```
cea-pKY1            agcagtttcacgtattcatgatgctgaagataatttgaaaacagcacagactaatctcct      1019
cea-CS6pColE1       agcagtttcacgtattcatgatgctgaagataatttgaaaacagtacagactaatctcct      1019
cea-Ss046pColE1     agcagtttcacgtattcatgatgctgaagataatttgaaaacagtacagactaatctcct      1019
cea-53GpColE1       agcagtttcacgtattcatgatgctgaagataatttgaaaacagcacagactaatctcct      1019
                    ****************************************** ***************

cea-pKY1            gaactcgcagattaaggatgctgtggatgcaacagttagcttttatcaaacgctatctga      1079
cea-CS6pColE1       gaactcgcagattaaggatgctgtggatgcaacagttagcttttatcaaacgctatctga      1079
cea-Ss046pColE1     gaactcgcagattaaggatgctgtggatgcaacagttagcttttatcaaacgctatctga      1079
cea-53GpColE1       gaactcgcagattaaggatgctgtggatgcaacagttagcttttatcaaacgctatctga      1079
                    ************************************************************

cea-pKY1            aaaatatggtgaaaaatattcaaaaatggcacaggaacttgctgataaatctaaaggtaa      1139
cea-CS6pColE1       aaaatatggtgaaaaatattcaaaaatggcacaggaacttgctgataaatctaaaggtaa      1139
cea-Ss046pColE1     aaaatatggtgaaaaatattcaaaaatggcacaggaacttgctgataaatctaaaggtaa      1139
cea-53GpColE1       aaaatatggtgaaaaatattcaaaaatggcacaggaacttgctgataaatctaaaggtaa      1139
                    ************************************************************

cea-pKY1            gaaaatcagcaatgtgaatgaagctctagctgcttttgagaaatacaaggatgtattaaa      1199
cea-CS6pColE1       gaaaatcagcaatgtgaatgaagctctagctgcttttgagaaatacaaggatgtattaaa      1199
cea-Ss046pColE1     gaaaatcagcaatgtgaatgaagctctagctgcttttgagaaatacaaggatgtattaaa      1199
cea-53GpColE1       gaaaatcagcaatgtgaatgaagctctagctgcttttgagaaatacaaggatgtattaaa      1199
                    ************************************************************

cea-pKY1            taagaaattcagcaaagcagaccgtgatgcaattttcaatgcactggaggcggttaagta      1259
cea-CS6pColE1       taagaaattcagcaaagcagaccgtgatgcaattttcaatgcactggaggcggttaagta      1259
cea-Ss046pColE1     taagaaattcagcaaagcagaccgtgatgcaattttcaatgcactggaggcggttaagta      1259
cea-53GpColE1       taagaaattcagcaaagcagaccgtgatgcaattttcaatgcactggaggcggttaagta      1259
                    ************************************************************

cea-pKY1            tgaagactgggctaagcatttagatcagtttgccaagtacttgaagattacgggacatgt      1319
cea-CS6pColE1       tgaagactgggctaagcatttagatcagtttgccaagtacttgaagattacgggacatgt      1319
cea-Ss046pColE1     tgaagactgggctaagcatttagatcagtttgccaagtacttgaagattacgggacatgt      1319
cea-53GpColE1       tgaagactgggctaagcatttagaacagtttgccaagtacttgaagattacgggacatgt      1319
                    ********************** **********************************

cea-pKY1            ttcttttggatatgatgtggtatctgatatcctaaaaattaaggatacaggtgactggaa      1379
cea-CS6pColE1       ttcttttggatatgatgtggtatctgatatcctaaaaattaaggatacaggtgactggaa      1379
cea-Ss046pColE1     ttcttttggatatgatgtggtatctgatatcctaaaaattaaggatacaggtgactggaa      1379
cea-53GpColE1       ttcttttggatatgatgtggtatctgatatcctaaaaattaaggatacaggtgactggaa      1379
                    ************************************************************

cea-pKY1            gccgctatttcttacattagagaagaaagctgtagatgctggagttagttatgttgttgt      1439
cea-CS6pColE1       gccgctatttcttacattagagaagaaagctgtagatgctggagttagttatgttgttgt      1439
cea-Ss046pColE1     gccgctatttcttacattagagaagaaagctgtagatgctggagttagttatgttgttgt      1439
cea-53GpColE1       gccgctatttcttacattagagaagaaagctgtagatgctggagttagttatgttgttgt      1439
                    ************************************************************

cea-pKY1            tttacttttagtgtgcttgctggaactacattaggtatctgggggattgctattgttac      1499
cea-CS6pColE1       tttacttttagtgtgcttgctggaactacattaggtatctgggggattgctattgttac      1499
cea-Ss046pColE1     tttacttttagtgtgcttgctggaactacattaggtatctgggggattgctattgttac      1499
cea-53GpColE1       tttacttttagtgtgcttgctggaactacattaggtatctgggggattgctattgttac      1499
                    ************************************************************

cea-pKY1            aggcattctatgtgcctttattgataagaataaacttaatactataaatgaggtgttggg      1559
cea-CS6pColE1       aggcattctatgtgcctttattgataagaataaacttaatactataaatgaggtgttggg      1559
cea-Ss046pColE1     aggcattctatgtgcctttattgataagaataaacttaatactataaatgaggtgttggg      1559
cea-53GpColE1       aggcattctatgtgcctttattgataagaataaacttaatactataaatgaggtgttggg      1559
                    ************************************************************
```

```
cea-pKY1            tatttaa    1566
cea-CS6pColE1       tatttaa    1566
cea-Ss046pColE1     tatttaa    1566
cea-53GpColE1       tatttaa    1566
                    *******
```

Figure 7.6: Nucleotide alignment of *cea* on pColE1 from *S. sonnei* CS6 and 53G and Ss046, and pKY1. The alignments were created by ClustalOmega. Astericks (*) indicate identical nucleotides of all sequences in the comparison.

```
pKYM-NC_001378      aacactctgcgagtggtacattttccccggattatcgtcctgagcctgccgctggtctct        60
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      ttctaccgcctcgctttgctcgttgctcaacgcctcacagacacggattaaaatccgcat        120
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      ccgttcaccgttttttaaagtccgttaaaagcatgatgccatctccgagagttaatctcg        180
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      tcaaatgctaaatcgtgggggtcctttggggttccgatttagtgattgcgacaccaccga        240
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      ttaaaaaacttatgcggggtggatggtttcacgaagtgaggccatccacctgtaagacag        300
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      ggttttgtttttattccctgttttggtgatcgggtgtgtggaaaaggttggggtaagccg        360
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      ttcgggggtgcttgttttgggggggttaaaattgtggttatttttttgcgcaattctcgcgc        420
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      gtgatccttgtatttatacttaagggataaatggcggatatgaaatagtggtttagccca        480
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      gtaatgacgaggctttgagtgggttttgacaggtcaaagaaaatggagcagaattgaggc        540
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      gtttttaatcggcgttggggagtgcgtcaacactccccaacatttcgaatgtgtcacctc        600
HE616531            ------------------------------------------------------------        0

pKYM-NC_001378      agcggcaaactctggtgacatgtactggctcgcaatgcacaggtacgtgatgaatatacc        660
HE616531            ------------------------------------------------------------        0
```

```
pKYM-NC_001378      acatcaaatcacagcctgcccagatcggagcaggcttaatgtcagaagataaattccttt        720
HE616531            -----------------------------------atgtcagaagataaattccttt         22
                                                       **********************

pKYM-NC_001378      cggactacagcccccgtgatgcagtttgggatacccagcgcacgcttaccgattctgtcg        780
HE616531            cggactacagcccccgtgatgcagtttgggatacccagcgcacgcttaccgattctgtcg         82
                    ************************************************************

pKYM-NC_001378      ggggtatctaccagactgctggtcaattcgagcgctatgcactccgtatggcctcctgta        840
HE616531            ggggtatctaccagactgctgctgaattcgagcgctatgcactccgtatggcctcctgta        142
                    ********************* * ***********************************

pKYM-NC_001378      gcggtttgttacgttttggttggtctaccatcatggaaaccggagaaacgcgcctacggc        900
HE616531            gcggtttgttacgttttggttggtctaccatcatggaaaccggagaaacgcgcctacggc        202
                    ************************************************************

pKYM-NC_001378      ttcgtagtgcgcaattttgccgtgtccgtcattgccctgtctgccagtggagaagaaccc        960
HE616531            ttcgtagtgcgcaattttgccgtgtccgtcattgccctgtctgccagtggagaagaaccc        262
                    ************************************************************

pKYM-NC_001378      tcatgtggcaagcccgtttttatcaggctctaccgaaaatcgttgtggattacccgtctt       1020
HE616531            tcatgtggcaagcccgtttttatcaggctctaccgaaaatcgttgtggattacccgtctt        322
                    ************************************************************

pKYM-NC_001378      cccgatggttgtttctgacgttaactgtcaggaactgcgagataggtgaacttggaacag       1080
HE616531            cccgatggttgtttctgacgttaactgtcaggaactgcgagataggtgaacttggaacag        382
                    ************************************************************

pKYM-NC_001378      tccttacagcaatgaatgcggcgtttaagcgaatggaaaagcgaaaggagctatcacctg       1140
HE616531            tccttacagcaatgaatgcggcgtttaagcgaatggaaaagcgaaaggagctatcacctg        442
                    ************************************************************

pKYM-NC_001378      ttcaggggtggatcagggctacggaggtgacgcgaggtaaggatggcagcgcacatccgc       1200
HE616531            ttcaggggtggatcagggctacggaggtgacgcgaggtaaggatggcagcgcacatccgc        502
                    ************************************************************

pKYM-NC_001378      attttcactgtctgctgatggtgcaaccttcttggtttaaagggaagaactacgttaagc       1260
HE616531            attttcactgtctgctgatggtgcaaccttcttggtttaaagggaagaactacgttaagc        562
                    ************************************************************

pKYM-NC_001378      acgaacgttgggtagaactctggcgcgattgcttgcgggtgaactatgagccgaatatcg       1320
HE616531            acgaacgttgggtagaactctggcgcgattgcttgcgggtgaactatgagccgaatatcg        622
                    ************************************************************

pKYM-NC_001378      atattcgggcagtaaaaactaagacaggtgaggttgtggccaacgttgccgagcaactgc       1380
HE616531            atattcgggcagtaaaaactaagacaggtgaggttgtggccaacgttgccgagcaactgc        682
                    ************************************************************

pKYM-NC_001378      aaagcgcggttgctgaaacgctgaaatactccgttaaaccggaagatatggcaaacgatc       1440
HE616531            aaagcgcggttgctgaaacgctgaaatactccgttaaaccggaagatatggcaaacgatc        742
                    ************************************************************

pKYM-NC_001378      ctgagtggtttcttgacgtgacgcggcagcttcacaagcgccgttttatctcgaccggtg       1500
HE616531            ctgagtggtttcttgagctgacgcggcagcttcacaagcgccgttttatctcgaccggtg        802
                    ****************  ***************************************

pKYM-NC_001378      gggcgctaaaaaacgtcctccagttggatcgagaaaccaatgaggatcttgtcattgccg       1560
HE616531            gggcgctaaaaaacgtcctccagttggatcgagaaaccaatgaggatcttgtcattgccg        862
                    ************************************************************
```

```
pKYM-NC_001378      acgatgtaggggatggcactgatgacgggaagcggacggcgtttgtctgggattcaggta      1620
HE616531            acgatgtaggggatggcactgatgacgggaagcggacggcgtttgtctgggattcaggta      922
                    ************************************************************

pKYM-NC_001378      aacggcgttacaaacgcgcccctgagaaggataaatcggattaacgtatgaatattaata      1680
HE616531            aacggcgttacaaacgcgcccctgagaaggataaatcggattaacgtatgaatattaata      982
                    ************************************************************

pKYM-NC_001378      ttgaatacctgaatggaaataagactattggtttatttttttaagaagtgaagcgtgat      1740
HE616531            ttgaatacctgaatggaaataagactattggtttatttttttaagaagtgaagcggtga      1042
                    ********************************************************

pKYM-NC_001378      tc--ctgacagtttaaaaaccttattttgcttattgatggattaagttttggcacatttg      1798
HE616531            ttcctgacaggtttaaaaaccttattttgcttattgatggattaagttttggcacatttg      1102
                    *         ************************************************

pKYM-NC_001378      gttttcatccgcacgaaggttttgaggatgaattaattttatatattcagaaaacaaacg      1858
HE616531            gttttcatccgcacgaaggttttgaggatgaattaattttatatattcagaaaacaaacg      1162
                    ************************************************************

pKYM-NC_001378      agagggtaaaaactcttttgtgaaaatagatttaaacggtatcaagtttggatttttaa      1918
HE616531            agagggtaaaaactcttttgtgaaaatagatttaaacggtatcaagtttggatttttaa      1222
                    ************************************************************

pKYM-NC_001378      gaacgcattcttagttctggaaaagagccagcggcaggctgaggtgataggtacgagatt      1978
HE616531            gaacgcattcttagttctggaaaagagccagcggcaggctgaggtgataggtacgagatt      1282
                    ************************************************************

pKYM-NC_001378      gcatgcaatctctagtgctctgtctatcctgcattatcctcagcattatcctcagccttg      2038
HE616531            gcatgcaatctctagtgctctgtctatcctgcattatcctcagcattatcctcagccttg      1342
                    ************************************************************

pKYM-NC_001378      ccaactcgacaccaatgcaggatagacaatccgatgtcaaatgtt---------------      2083
HE616531            ccaactcgacaccaatgcaggatagacaatccgatgtcaaatgttaacactctgcgagtg      1402
                    *******************************************

pKYM-NC_001378      ------------------------------------------------------------      2083
HE616531            gtacattttccccggattatcgtcctgagcctgccgctggctctctttctaccgcctcgc      1462

pKYM-NC_001378      ------------------------------------------------------------      2083
HE616531            tttgctcgttgctcaacgcctcacagacacggattaaaatccgcatccgttcaccgtttt      1522

pKYM-NC_001378      ------------------------------------------------------------      2083
HE616531            ttaaagtccgttaaaagcatgatgccatctccgagagttaatctcgtcaaatgctaaatc      1582

pKYM-NC_001378      ------------------------------------------------------------      2083
HE616531            gtgggggtccccttttggggttccgatttagtgattgacgacaccaccgattaaaaaactt      1642

pKYM-NC_001378      ------------------------------------------------------------      2083
HE616531            atgcggggtggatggtttcacgaagtgaggccatccacctgtaagacagggttttgtttt      1702

pKYM-NC_001378      ------------------------------------------------------------      2083
HE616531            tattccctgttttggtgatcgggtgtgtggaaaaggttggggtaagccgttcgggggtgc      1762
```

```
pKYM-NC_001378      ------------------------------------------------------------        2083
HE616531            ttgttttgggggttaaaattgtggttatttttgcgcaattctcgcgcgtgatccttgt        1822

pKYM-NC_001378      ------------------------------------------------------------        2083
HE616531            atttatacttaagggataaatggcggatatgaaatagtggtttagcccagtaatgacgag        1882

pKYM-NC_001378      ------------------------------------------------------------        2083
HE616531            gctttgagtgggttttgacaggtcaaagaaaatggagcagaattgaggcgtttttaatcg        1942

pKYM-NC_001378      ------------------------------------------------------------        2083
HE616531            gcgttggggagtgcgtcaacactccccaacatttcgaatgtgtcacctcagcggcaaact        2002

pKYM-NC_001378      ------------------------------------------------------------        2083
HE616531            ctggtgacatgtactggctcgcaatgcacaggtacgtgatgaatataccacatcaaatca        2062

pKYM-NC_001378      --------------------------- 2083
HE616531            cagcctgcccagatcggagcaggctta 2089
```

Figure 7.7: Nucleotide alignment of pKYM (NC_001378) and 2,083 bp plasmid from 53G (HE616531). The alignments were created by ClustalOmega. Astericks (*) indicate identical nucleotides of all sequences in the comparison.

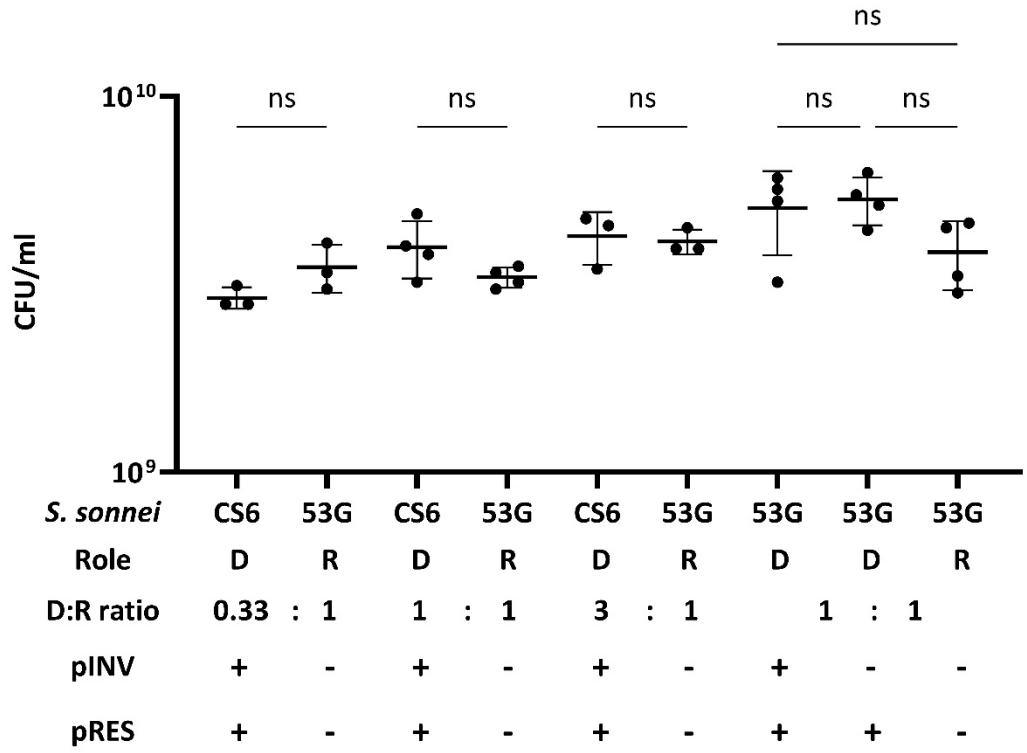


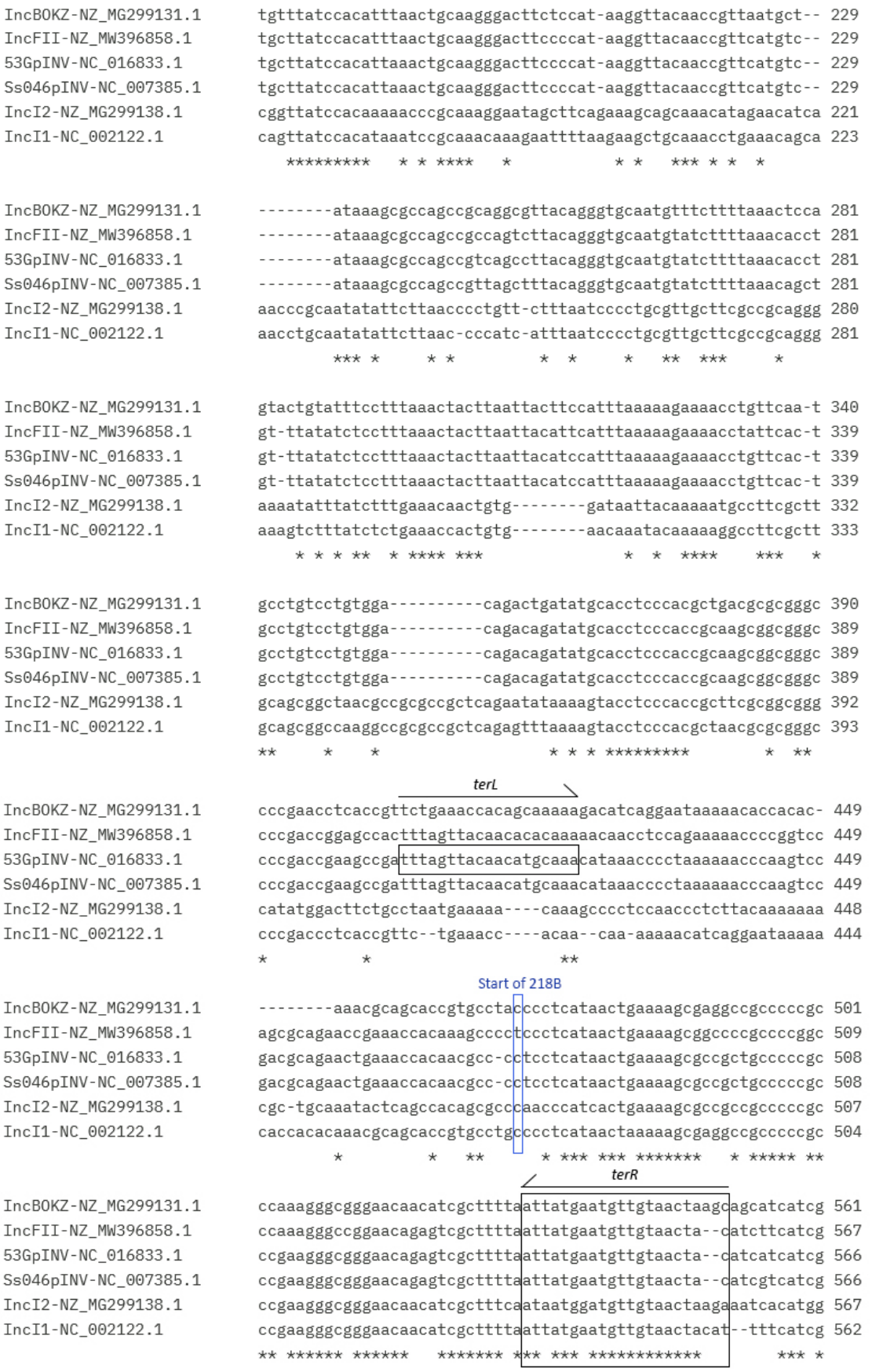
Figure 7.8: CFU/ml of *S. sonnei* donors and recipients before conjugation at different conditions. Ordinary one-way ANOVA tests were performed to investigate the difference between concentration of *S. sonnei* before conjugation. ns, not significant.

```
IS630-pINV      ctacacttttgccagtccatgtttgcctccggggaatgggctgacggtttccataaaatg   60
ISVsa5-pRES     ----------------------------------------tcataatttccccaaagcg   19
                                                       * *   *****  ***  *

IS630-pINV      gcgaacttttttcaacagtt---gccacattgagctgcactgatgattacgcgttattgt   117
ISVsa5-pRES     taaccatgtgtgaataaattttgagctagtagggttgcagccacgagtaagtcttccctt   79
                     * * *  *  * **    *   * * * ****   * ** ** *  **    *

IS630-pINV      gtcgtgaagtgcctgccatagccgttcaacatgattcacccatggcgagtaaaccggctg   177
ISVsa5-pRES     gttattgtgtagccagaatgccg--------caaaacttccatgcctaag---cgaactg   128
                **  *   **  *    **  *          *  *  ***** * *    *   ***

IS630-pINV      ataaatgaccctgaacttcgggttctccttcagccagctctgtgtttcccggcttttgtg   237
ISVsa5-pRES     tt----gagagtacgtttcgatttctgac-----------tgtgttagcctgg-------   166
                 *    **   *    ****  ****             ******  ** *

IS630-pINV      gataatgtagttgtccac--gatcagcgtgatggttttcg----cccgacggtatgtcgc   291
ISVsa5-pRES     ----aagtgcttgtcccaaccttgtttctgagcatgaacgcccgcaagccaacatgttag   222
                    * **  ******     *    ***   *   **    *  * *   ****

IS630-pINV      tttaagccgcttcagcaggctgatgaacagcgccgaacttttgctgttgccgcccacaca   351
ISVsa5-pRES     ttgaa------gcatcagggcgattagcagcatga---------tatcaaaacgctctga   267
                ** **       ** ****  *** * ****            * *    * * *  *

IS630-pINV      gctgactttacctgtcccgctgtgcagcgctccggccagataatatttttcattctgtcc   411
ISVsa5-pRES     gctgctc------gttcggctatggcgtaggc----------------------------   293
                ****         ** * *** **  *    *

IS630-pINV      cggcgtgaccacccgtttttgctgtccgcgcagttgccagtccgc-accgattttgggat   470
ISVsa5-pRES     --------------ctagtccgtaggcaggacttttcaagtctcggaaggtttcttcaat   339
                             *  *   *   *  *   ** * ****    *  * ** *   **

IS630-pINV      taagatggatatccacttcatcttcataaaagacctgatgttctgcgctgcattcgtcca   530
ISVsa5-pRES     ctgcattcgcttcgaatagatattaaca----agttgtttgggtgttcgaattt---caa   392
                    **    ** * *  ** ** * *    *  ** *    **  *    **   * *

IS630-pINV      gtgctttatggattgctgccatcttttcatctttatgcgggtcacggatacgcagagttg   590
ISVsa5-pRES     caggtaagttagttgctagaacccatggctcctttgccgacgc-----tgagtagatttt   447
                  * *   *   *****   * *  *   ** **   **   *     *  * *** **

IS630-pINV      gcgcagcccttcgccacacaatccccgcagacggcaaccagcggcgaacggttccggcat   650
ISVsa5-pRES     aggtgacgggtggtgacaatg-------agtccgtgtcgagcgctgattttt-tcggcct   499
                  *   *   * *  ***         ** * *   * ****  **    *  **** *

IS630-pINV      ttaactggca---------accggttatctcattgatttttattgccagca--------g   693
ISVsa5-pRES     ttagagcgagatttatacaatagaatttggcatgagattggattgcttttagtcagcctc   559
                ***    *          *  *  * *  ***    **  *****    *

IS630-pINV      ttctgtactccagcgtgaacgctggtagccaaa-------gtcgccgggagaatgtttta   746
ISVsa5-pRES     ttatagcctaaagtctttgagtgactagatgacatatcatgtaagttgctgataggtttc   619
                ** *   **  **  *    *    ***   *       **     *  **  * ***
```

```
IS630-pINV       ccagctcacgtaac--agtgtgcagatatgctcaaacggccagcgacgggcacgcccgg-   803
ISVsa5-pRES      cagttttccgctcctaggtctgcatatgtgctataa--atcagctcattcgagtacggaa   677
                 *    *  **   *   ** **** ** ****  **    ****       *   * *

IS630-pINV       caggtaatgatttcagtccctcaacacccgactgcgtgaaccagttaatccagcgtccaa   863
ISVsa5-pRES      caggaaagtttctcat-----------------------------------------------   693
                 **** **   * ***

IS630-pINV       cagaggagcgggcgcagcagagcgttctggcaacgtcgctgacacggtcgccccggtgca   923
ISVsa5-pRES      ------------------------------------------------------------   693


IS630-pINV       gcatcagcatggcagtcagtctgcgggcataatttttatcgtgtgttttatggatggctt   983
ISVsa5-pRES      ------------------------------------------------------------   693


IS630-pINV       tctgcatcaggcgtcgttcgtcacgggaaattggtgctatgatcggcat      1032
ISVsa5-pRES      -------------------------------------------------      693
```

Figure 7.9: Nucleotide alignment of IS*630*$_{pINV}$ and IS*Vsa5*$_{pRES}$.
The alignments were created by ClustalOmega. Astericks (*) indicate identical nucleotides of all sequences in the comparison.

```
IncBOKZ-NZ_MG299131.1     gtttgacctcatctgaataatccggcctgcgccggaggcttccgcacgtctgaagccc-g 59
IncFII-NZ_MW396858.1      aagtgacctcctctgaataatccggcctgcgccggaggcttccgcacgtctgaagccc-g 59
53GpINV-NC_016833.1       aagtgacctcctcagaataatccggcacgcgccggaggcattcgcacgcctgaagccc-g 59
Ss046pINV-NC_007385.1     aagtgacctcctcagaataatccggcacgcgccggaggcattcgcacgcctgaagccc-g 59
IncI2-NZ_MG299138.1       -------tttagctgcctgatacaatcggc-----actctcttcagccccttcggggctt 48
IncI1-NC_002122.1         -----------acagcatcattccctcagcactgaatcatcgccagcccctccggggctt 49
                                     * *  * ** *    **    *         *  **   *  *

IncBOKZ-NZ_MG299131.1     gcgacgcaca--aaaaagcagcacagcacacaaaaaacagcctcatcatctgccttccgg 117
IncFII-NZ_MW396858.1      acagcgcaca--aaaaatcagcaccacatacaaaaaacaacctcatcatccagcttctgg 117
53GpINV-NC_016833.1       ccagtgcaca--aaaaatcagcgctacatacaaaaaacaacctcatcatccaccttctgg 117
Ss046pINV-NC_007385.1     ccagtgcaca--aaaaatcagcgctacatacaaaaaacaacctcatcatccaccttctgg 117
IncI2-NZ_MG299138.1       ttcgtgttgcctgcatttcatgaaattcctcatatttgacgaccgtttttcatttt---g 105
IncI1-NC_002122.1         tcggcgcagattccgctca-gcccaaaatccttagtagtca-ccttaaatcccctc---a 104
                               *                     *  *        * *       *

IncBOKZ-NZ_MG299131.1     ctcatcatgctccccctgtttttc--acgcaaa---acacgtctcacagacgggtaattc 172
IncFII-NZ_MW396858.1      tgcatccggccccccctgttttcg--atacaaa---acacgcctcacagacggggaattt 172
53GpINV-NC_016833.1       tgcatccggtccccccctgtttttg--atacaaa---acacgcctcacagacggggaattt 172
Ss046pINV-NC_007385.1     tgcatccggtccccccctgtttttg--atacaaa---acacgcctcacagacggggaattt 172
IncI2-NZ_MG299138.1       ----tcccgctctgacagttttttgatggccgcaatagggtatcggagtaaaaacgcctt 161
IncI1-NC_002122.1         gaggggcatatctgccc-ataaaaccacgcatcagtcatcagaacgtggccacgttgttt 163
                                    *   *   *        *                          *
```


IncBOKZ-NZ_MG299131.1 tgtttatccacatttaactgcaagggacttctccat-aaggttacaaccgttaatgct-- 229
IncFII-NZ_MW396858.1 tgcttatccacatttaactgcaagggacttccccat-aaggttacaaccgttcatgtc-- 229
53GpINV-NC_016833.1 tgcttatccacattaaactgcaagggacttccccat-aaggttacaaccgttcatgtc-- 229
Ss046pINV-NC_007385.1 tgcttatccacattaaactgcaagggacttccccat-aaggttacaaccgttcatgtc-- 229
IncI2-NZ_MG299138.1 cggttatccacaaaaacccgcaaaggaatagcttcagaaagcagcaaacatagaacatca 221
IncI1-NC_002122.1 cagttatccacataaatccgcaaacaaagaattttaagaagctgcaaacctgaaacagca 223
IncBOKZ-NZ_MG299131.1 --------ataaagcgccagccgcaggcgttacagggtgcaatgtttcttttaaactcca 281
IncFII-NZ_MW396858.1 --------ataaagcgccagccgccagtcttacagggtgcaatgtatcttttaaacacct 281
53GpINV-NC_016833.1 --------ataaagcgccagccgtcagccttacagggtgcaatgtatcttttaaacacct 281
Ss046pINV-NC_007385.1 --------ataaagcgccagccgttagctttacagggtgcaatgtatcttttaaacagct 281
IncI2-NZ_MG299138.1 aacccgcaatatattcttaacccctgtt-ctttaatcccctgcgttgcttcgccgcaggg 280
IncI1-NC_002122.1 aacctgcaatatattcttaac-cccatc-atttaatcccctgcgttgcttcgccgcaggg 281
IncBOKZ-NZ_MG299131.1 gtactgtatttcctttaaactacttaattacttccatttaaaaagaaaacctgttcaa-t 340
IncFII-NZ_MW396858.1 gt-ttatatctcctttaaactacttaattacattcatttaaaaagaaaacctattcac-t 339
53GpINV-NC_016833.1 gt-ttatatctcctttaaactacttaattacatccatttaaaaagaaaacctgttcac-t 339
Ss046pINV-NC_007385.1 gt-ttatatctcctttaaactacttaattacatccatttaaaaagaaaacctgttcac-t 339
IncI2-NZ_MG299138.1 aaaatatttatctttgaaacaactgtg--------gataattacaaaaatgccttcgctt 332
IncI1-NC_002122.1 aaagtctttatctctgaaaccactgtg--------aacaaatacaaaaaggccttcgctt 333
IncBOKZ-NZ_MG299131.1 gcctgtcctgtgga----------cagactgatatgcacctcccacgctgacgcgcgggc 390
IncFII-NZ_MW396858.1 gcctgtcctgtgga----------cagacagatatgcacctcccaccgcaagcggcgggc 389
53GpINV-NC_016833.1 gcctgtcctgtgga----------cagacagatatgcacctcccaccgcaagcggcgggc 389
Ss046pINV-NC_007385.1 gcctgtcctgtgga----------cagacagatatgcacctcccaccgcaagcggcgggc 389
IncI2-NZ_MG299138.1 gcagcggctaacgccgcgccgctcagaatataaaagtacctcccaccgcttcgcggcggg 392
IncI1-NC_002122.1 gcagcggccaaggccgcgccgctcagagtttaaaagtacctcccacgctaacgcgcgggc 393
terL
IncBOKZ-NZ_MG299131.1 cccgaacctcaccgttctgaaaccacagcaaaaagacatcaggaataaaaacaccacac- 449
IncFII-NZ_MW396858.1 cccgaccggagccactttagttacaacacacaaaaacaacctccagaaaaaccccggtcc 449
53GpINV-NC_016833.1 cccgaccgaagccgatttagttacaacatgcaaacataaaccccctaaaaaacccaagtcc 449
Ss046pINV-NC_007385.1 cccgaccgaagccgatttagttacaacatgcaaacataaaccccctaaaaaacccaagtcc 449
IncI2-NZ_MG299138.1 catatggacttctgcctaatgaaaaa----caaagcccctccaaccctcttacaaaaaaa 448
IncI1-NC_002122.1 cccgaccctcaccgttc--tgaaacc----acaa--caa-aaaaacatcaggaataaaaa 444
Start of 218B
IncBOKZ-NZ_MG299131.1 --------aaacgcagcaccgtgcctaccctcataactgaaaagcgaggccgcccccgc 501
IncFII-NZ_MW396858.1 agcgcagaaccgaaaccacaaagccccctccctcataactgaaaagcggccccgccccggc 509
53GpINV-NC_016833.1 gacgcagaactgaaaccacaacgcc-cctcctcataactgaaaagcgccgctgcccccgc 508
Ss046pINV-NC_007385.1 gacgcagaactgaaaccacaacgcc-cctcctcataactgaaaagcgccgctgcccccgc 508
IncI2-NZ_MG299138.1 cgc-tgcaaatactcagccacagcgcccaacccatcactgaaaagcgccgccgcccccgc 507
IncI1-NC_002122.1 caccacacaaacgcagcaccgtgcctgcccctcataactaaaaagcgaggccgcccccgc 504
terR
IncBOKZ-NZ_MG299131.1 ccaaagggcgggaacaacatcgctttaattatgaatgttgtaactaagcagcatcatcg 561
IncFII-NZ_MW396858.1 ccaaagggccggaacagagtcgctttaattatgaatgttgtaacta--catcttcatcg 567
53GpINV-NC_016833.1 ccgaagggcgggaacagagtcgctttaattatgaatgttgtaacta--catcatcatcg 566
Ss046pINV-NC_007385.1 ccgaagggcgggaacagagtcgctttaattatgaatgttgtaacta--catcgtcatcg 566
IncI2-NZ_MG299138.1 ccgaagggcgggaacaacatcgctttcaataatggatgttgtaactaagaaatcacatgg 567
IncI1-NC_002122.1 ccgaagggcgggaacaacatcgctttaattatgaatgttgtaactacat--tttcatcg 562

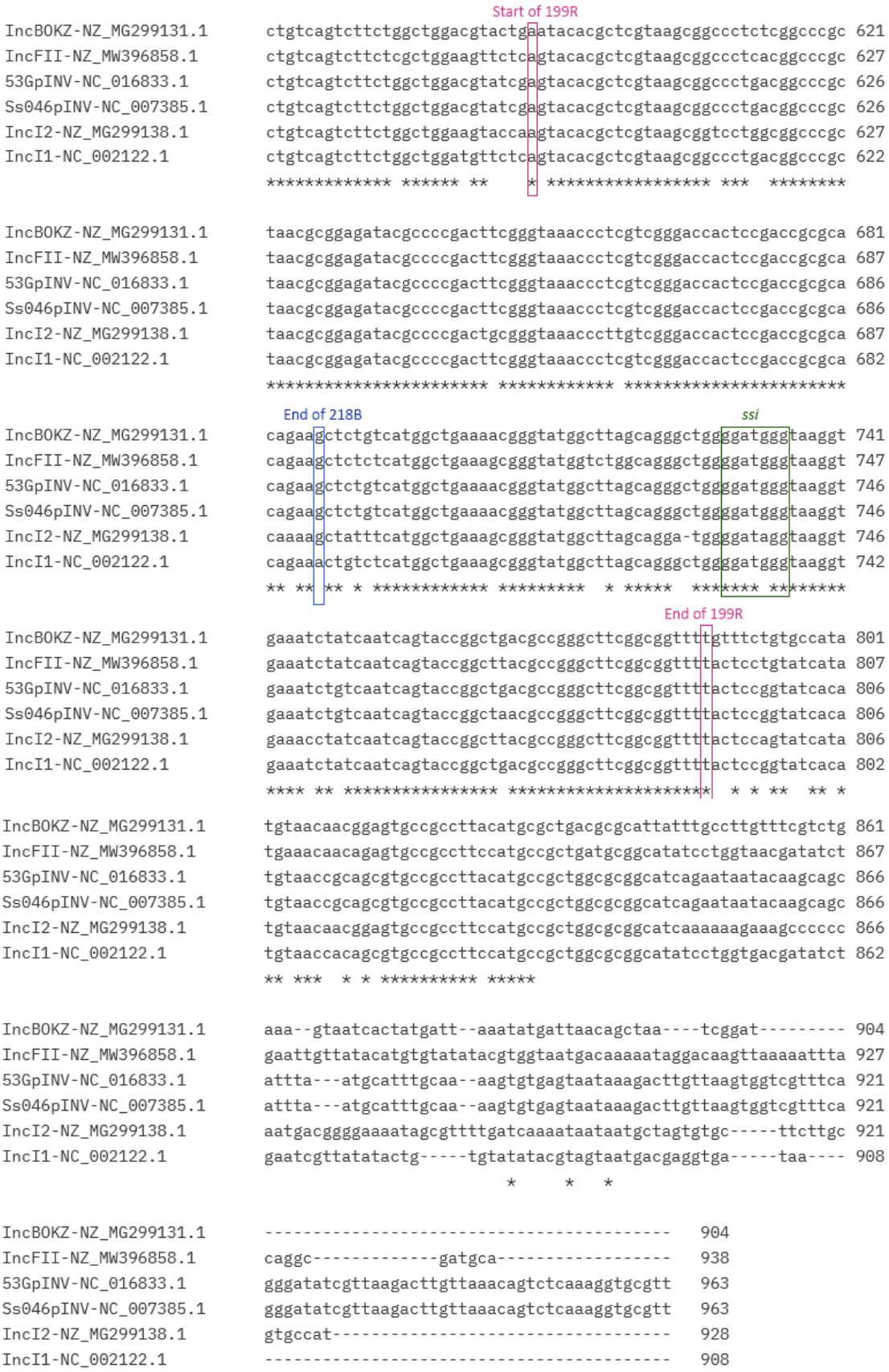


Figure 7.10: Nucleotide alignment of the sequence between *repA* and the next ORF on different plasmids.

The alignments were created by ClustalOmega. Astericks (*) indicate identical nucleotides of all sequences in the comparison.

Antibiotic classes	Mechanism of resistance	Gene	Location	Example(s) of MGE associated
Quinolones	DNA gyrase and topoisomerase IV	*gyrA*	C	Quinolone resistance-determining regions (QRDRs) of DNA gyrase and topoisomerase IV
		gyrB	C	
		parC	C	
		parE	C	
	Plasmid-mediated quinolone resistance (PMQR) gene	*qnrA1*	I, P	IS*CR1* (414, 415)
		qnrB	I, P	—
		qnrB4	I, P	IS*CR1* (415)
		qnrB6	I, P	IS*CR1* (415)
		qnrB19	I, P	IS*Ecp1*, IS*26* (415)
		qnrC	I, P	—
		qnrS1	I, P	pAH0376 (416)
		qnrS2	I, P	—
	Acetyltransferase	*aac-(60)-Ib-cr*	P	IS*26* (415)
	Efflux pumps	*qepA1*	P	IS*26*, IS*CR3C* (415)
β-lactam	Class A β-lactamases	bla_{SHV-2}	P	—
		bla_{SHV-11}	C, P	—
		bla_{SHV-12}	P	—
		bla_{PER-2}	—	—
		bla_{TEM-1}	I, P	IncR pKSR100 (417), IncI1 ST16, IncI2 plasmids (418), IS*Ecp1* (418), Tn*3* (419)

	bla$_{TEM-1a}$ (420)	P (420)	Minor MSMA sublineage plasmid (420), Tn*3*, IncB/O/K/Z plasmid p866 (197)
	bla$_{TEM-1b}$	I, P	IncR pKSR100 (417), IncFII pSF470 (419)
	bla$_{TEM-15}$	P (417)	IncR pKSR100 (417)
	bla$_{TEM-17}$	P (417)	IncR pKSR100 (417)
	bla$_{TEM-19}$	P (417)	IncR pKSR100 (417)
	bla$_{TEM-20}$	P (417)	IncR pKSR100 (417)
	bla$_{TEM-52}$	P	IncR pKSR100 (417)
	bla$_{CTX-M-1}$	P	—
	bla$_{CTX-M-2}$	P	—
	bla$_{CTX-M-3}$	P	—
	bla$_{CTX-M-14}$	P	IncF plasmids (418), IS*Ecp1*, IS*903D* (418)
	bla$_{CTX-M-15}$	C, P	IncI1 plasmid pKHSB1, pSH4469, IncI1 ST16, ST31 plasmids (418), IS*Ecp1* (418)
	bla$_{CTX-M-22}$	P	—
	bla$_{CTX-M-24}$	P	IS*Ecp1*
	bla$_{CTX-M-27}$	P	IncI plasmids, IS*Ecp1*, IS*903D* (418)
	bla$_{CTX-M-28}$	–	—
	bla$_{CTX-M-39}$	–	—
	bla$_{CTX-M-55}$	P	IncI1, IncA/C and IncZ plasmids (421), IncI2 plasmids (418), IS*Ecp1*, IS*26*-IS*Ecp1* and IS*Ecp1*-IS*5* (421)
	bla$_{CTX-M-57}$	–	—

		bla$_{CTX\text{-}M\text{-}64}$	P	—
		bla$_{CTX\text{-}M\text{-}65}$	P	—
		bla$_{CTX\text{-}M\text{-}79}$	P	—
	Hybrid of CTX-M-9 and CTX-M-1 β-lactamases	*bla*$_{CTX\text{-}M\text{-}123}$	P	IncHI2 and IncF conjugative plasmids
	Class B β-lactamases	*bla-IMP-like*	P	—
		bla$_{KPC}$	–	—
		bla$_{IMP\text{-}3}$	P	—
		bla-VIM-like	–	—
	Class C β-lactamases	*bla*$_{CMY\text{-}2}$	C, P	pSHH21, pSHH22, IS*Ecp1* (422)
		bla$_{CMY\text{-}59}$	C	—
		bla$_{DHA\text{-}1}$	C, I, P	—
	Class D β-lactamases	*bla*$_{OXA\text{-}1}$*-like*	I, P, C	In*1*, Tn*2603*, *Shigella* resistance locus multidrug resistance element (417), SRL PAI (197, 420)
		bla$_{OXA\text{-}2}$*-like*	I	In*1*
		bla$_{OXA\text{-}5}$*-like*	I, P	In*1*
		bla$_{OXA\text{-}30}$*-like*	I, P	In*1*, Tn*1409*
Macrolide	Enzymatic inactivation (macrolide 2'-phosphotransferase)	*mphA*	P	IS*26*, IS*6100*, p2246–CTX-M plasmid, IncR pKSR100, Minor MSMA sublineage plasmid, MSMA sublineage 3 plasmid (420)

	rRNA methylase	*ermB*	P	IncFI and IncFII plasmids, IncR pKSR100, IS*CR3*
Colistin	Phosphatidylethanolamine transferase	*mcr-1*	P	IncI2, IncFI, IncHI2, IncFIB, IncP, IncY, and IncX4, IS*Apl1*, Tn*6390*
Aminoglycosides	Streptomycin phosphotransferase	*strA*	P	IncFII and pNV-Y394, pCERC1 (R-plasmid) (417), pSs046_spA, pSFxv_3 (420), small MDR plasmid spA (20)
		strB	P	IncFII plasmids and pNV-Y394, pCERC1 (R-plasmid) (417), pSs046_spA, pABC-3, pSFxv_3 (420), small MDR plasmid spA (20)
	Adenyl transferase	*aadA1*	I, P	*Shigella* resistance locus multidrug resistance element (417), In*1*/Tn*3* & In*2*/Tn*7* (20, 420)
		aadA2	I, P	—
		aadA5	I, P	IncR pKSR100 (420)
Tetracycline	Efflux pumps	*tetA*	C, P	pSs046_spA, Minor MSMA sublineage plasmid (420), small MDR plasmid spA (20)
		tetB	C, P	*Shigella* resistance locus multidrug resistance element (417), chromosomal Islands SRL-MDRE (420)
		tet(C)	—	—
		tet(D)	—	—
		tetG	C, P	—
		tet(M)	—	—

Phenicols	Chloramphenicol acetyltransferase	*catA-like*	P	*Shigella* resistance locus multidrug resistance element (417), chromosomal Islands SRL-MDRE (420)
		catP	P	—
	Efflux pumps	*cmlA1*	I, P	—
Sulphonamides	Plasmid-borne resistance (sulphonamide resistant dihydropteroate synthase)	*sul1*	I, P	In*1*, non conjugative R plasmid, IncR pKSR100, MSMA sublineage 3 plasmid (420)
		sul2	I, P	Large transmissible plasmids or small non-conjugative plasmids, pCERC1 (R-plasmid) (417), pSs046_spA, pABC-3, pSFxv_3 (420), small MDR plasmid spA (20)
		sul3	I, P	In*1*
Streptothricin	Streptothricin acetyltransferase (423)	*sat2*	I, P	*In*2/*Tn7* (20)
Trimethoprim	Dihydrofolate reductases	*dfrA1*	I, P, C	In*1* & *In*2/*Tn7*, MSMA sublineage 3 plasmid (20, 420)
		dfrA5	I, P	In*1*
		dfrA7	I, P	In*1*
		dfrA8	P	In*1*
		dfrA12	I, P	Conjugative R plasmids, In*1*
		dfrA13	P	In*1*
		dfrA14	P	A small non-conjugative plasmid, pCERC1 (R-plasmid) (417), pABC-3 (420)

	dfrA14-like	P	In*1*
	dfrA15	I, P	In*1*
	dfrA16	P	In*1*
	dfrA17	I, P	In*1*, pKSR100 conjugative R-plasmid (420)
	dfrV	I, P	—

Table 7.1: Summary of transfer mechanisms associated with resistance genes detected in *Shigella* spp., from selected publications.
Unless indicated otherwise, the information in Table 7.1 is adapted from Ranjbar et al., 2019, Table 1 (424). Abbreviations: C, chromosome; P, plasmid; I, integron; -, unknown; MGE, mobile genetic element; MSMA, men who have sex with men-associated; IS, insertion sequence; Tn, transposon; In*1*, class 1 integrons; In*2*, class 2 integrons.

Inc groups	Accession number	RepA		*terR*		199R		Downstream		
		Start (bp)	End (bp)	Start (bp)	End (bp)	Start (bp)	End (bp)	Element	Start (bp)	End (bp)
IncFII, pINV	NC_016833.1	183,849	184,706	185,243	185,265	185,300	185,498	*relBE*	185,670	186,205
	CP102113.1	97,491	98,348	96,932	96,954	96,699	96,897	*relBE*	95,992	96,527
	CP104409.1	91,006	91,863	92,400	92,422	92,457	92,655	*relBE*	92,827	93,362
	CP104417.1	8,755	9,612	8,196	8,218	7,963	8,161	*relBE*	7,256	7,791
	CP104421.1	6,741	7,598	8,135	8,157	8,192	8,390	*relBE*	8,562	9,097
	CP104427.1	187,883	188,740	189,277	189,299	189,334	189,532	*relBE*	189,704	190,239
	CP104430.1	154,624	155,481	156,018	156,040	156,075	156,273	*relBE*	156,445	156,980
	NC_007385.1	180,760	181,617	182,154	182,176	182,211	182,409	*relBE*	182,581	183,116
	NZ_CP010830.1	88,165	89,022	87,606	87,628	87,373	87,571	*relBE*	86,666	87,255
	NZ_CP019696.1	88,168	89,018	87,736	87,758	87,368	87,565	*relBE*	86,661	87,196
	NZ_CP023646.1	14,975	15,832	16,369	16,391	16,426	16,624	*relBE*	16,796	17,331
	NZ_CP046285.1	30,272	31,129	31,666	31,688	31,723	31,921	*relBE*	32,093	32,628
	NZ_CP053752.1	21,938	22,795	21,379	21,401	21,146	21,344	*relBE*	20,439	20,974
	NZ_CP109776.1	1	858	1,395	1,417	1,452	1,650	*relBE*	1,822	2,356
	NZ_MW396859.1	1	858	1,395	1,417	1,452	1,650	*relBE*	1,822	2,356
	CP104425.1	6,743	7,600	8,137	8,159	8,194	8,392	*relBE*	8,564	9,099
IncFII	NZ_CP049168.1	9,228	10,097	8,668	8,690	8,435	8,633	CPBP family intramembrane metalloprotease	7,636	8,289
	NZ_KX008967.1	12,878	13,747	14,285	14,307	14,342	14,540	CPBP family intramembrane metalloprotease	14,686	15,339
	CP022460.1	43,687	44,556	43,127	43,149	42,894	43,092	CPBP family intramembrane metalloprotease	42,095	42,748
	CP102116.1	15,015	15,884	14,455	14,477	14,222	14,420	CPBP family intramembrane metalloprotease	13,423	14,076
	CP104412.1	19,413	20,282	20,820	20,842	20,877	21,075	CPBP family intramembrane metalloprotease	21,221	21,874
	CP104416.1	65,820	66,689	67,227	67,249	67,284	67,482	CPBP family intramembrane metalloprotease	67,628	68,281
	CP104420.1	13,287	14,156	12,727	12,749	12,494	12,692	CPBP family intramembrane metalloprotease	11,695	12,348

CP104422.1	51,895	52,764	53,302	53,324	53,359	53,557	CPBP family intramembrane metalloprotease	53,703	54,356
CP104426.1	52,340	53,209	53,747	53,769	53,804	54,002	CPBP family intramembrane metalloprotease	54,148	54,801
CP104433.1	74,064	74,933	75,471	75,493	75,528	75,726	CPBP family intramembrane metalloprotease	75,872	76,525
CP104437.1	45,460	46,329	46,867	46,889	46,924	47,122	CPBP family intramembrane metalloprotease	47,268	47,921
NC_013727.1	784	1,653	2,191	2,213	2,248	2,446	*yacAB*	2,567	3,114
NZ_CP045525.1	69,065	69,934	68,505	68,527	68,272	68,470	CPBP family intramembrane metalloprotease	67,473	68,126
NZ_CP049166.1	14,678	15,547	14,118	14,140	13,885	14,083	CPBP family intramembrane metalloprotease	13,097	13,739
NZ_CP049170.1	40,049	40,918	41,456	41,478	41,513	41,711	CPBP family intramembrane metalloprotease	41,857	42,510
NZ_CP049172.1	21,583	22,452	21,023	21,045	20,790	20,988	CPBP family intramembrane metalloprotease	19,991	20,644
NZ_CP049174.1	50,340	51,209	49,780	49,802	49,547	49,745	CPBP family intramembrane metalloprotease	48,748	49,401
NZ_CP049186.1	5,758	6,627	7,165	7,187	7,222	7,420	CPBP family intramembrane metalloprotease	7,566	8,219
NZ_CP061363.1	76,474	77,343	77,881	77,903	77,938	78,136	CPBP family intramembrane metalloprotease	78,282	78,935
NZ_CP099767.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_CP099770.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_CP099771.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_CP099773.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_CP099776.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_CP101110.2	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450

NZ_CP115395.1	548	1,417	1,955	1,977	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_MW396858.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OP038267.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OP038274.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OP038278.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OP038279.1	1	858	1,396	1,415	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OP038287.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OP038290.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OP038295.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OP038298.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OQ291176.1	1	858	1,369	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_OQ291178.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
NZ_ON461901.1	27,080	27,949	28,488	28,506	28,547	28,745	CPBP family intramembrane metalloprotease	28,891	29,544
NZ_OP038277.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
CP049164.1	4,444	5,313	5,850	5,872	5,907	6,105	CPBP family intramembrane metalloprotease	6,251	6,904
LR213461.1	24,219	25,088	23,659	23,681	23,426	23,624	CPBP family intramembrane metalloprotease	22,627	23,280
NZ_CP045934.1	1	858	1,395	1,417	1,452	1,650	CPBP family intramembrane metalloprotease	1,796	2,449

	NZ_CP053755.1	68,956	69,825	68,396	68,418	68,163	68,361	CPBP family intramembrane metalloprotease	67,364	68,017
	NZ_LT174531.1	547	1,418	1,956	1,978	2,013	2,211	(Partial) IS1294	2,275	2,754
	NZ_OP038282.1	14,208	15,077	15,615	15,637	15,672	15,870	CPBP family intramembrane metalloprotease	16,016	16,669
	NZ_OP038284.1	1	858	1,395	1,417	1,452	1,650	CPBP family intramembrane metalloprotease	1,796	2,449
	NZ_OP038293.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
	NZ_OP038301.1	1	858	1,396	1,418	1,453	1,651	CPBP family intramembrane metalloprotease	1,797	2,450
	NZ_OQ230386.1	4,166	5,188	5,737	5,759	5,794	5,992	CPBP family intramembrane metalloprotease	6,138	6,791
	NZ_OQ230388.1	54,639	55,661	54,068	54,090	53,835	54,033	CPBP family intramembrane metalloprotease	53,036	53,689
IncI1	NZ_OP038300.1	53,018	54,049	54,582	54,600	54,641	54,839	*yacAB*	54,960	55,486
	NZ_OP038292.1	46	1,077	1,611	1,633	1,670	1,866	*yacAB*	1,987	2,534
	NZ_OQ230390.1	16,664	17,695	18,229	18,251	18,287	18,484	Hypothetical protein	18,765	19,274
	NZ_OR237795.1	1	1,032	1,565	1,587	1,623	1,820	Hypothetical protein	2,101	2,610
	NZ_OR237799.1	1	1,032	1,565	1,587	1,623	1,820	Hypothetical protein	2,101	2,610
	NZ_OR237801.1	46	1,077	1,610	1,628	1,669	1,867	*yacAB*	1,988	2,514
	NC_020991.1	36,225	37,256	35,669	35,691	35,436	35,633	Hypothetical protein	34,646	35,155
	NZ_CP022456.1	32,569	33,600	32,013	32,035	31,780	31,976	Hypothetical protein	30,990	31,499
	NZ_CP099772.1	46	1,076	1,610	1,632	1,667	1,865	IS91-family transposase	2,100	3,308
	NZ_ON461899.1	42,465	43,496	44,029	44,051	44,086	44,248	*yacAB*	44,405	44,952
	CP035009.1	1	1,032	1,565	1,583	1,624	1,822	*yacA*	1,943	2,212
	NC_002122.1	455	1,486	2,019	2,041	2,076	2,274	*yacAB*	2,395	2,942
	NZ_CP010831.1	65,587	66,618	65,033	65,055	64,801	64,996	*relBE*	64,124	64,683
	NZ_CP014096.2	57,354	58,385	56,802	56,822	56,565	56,763	*yacA*	56,331	56,444
	NZ_CP032524.1	92,719	93,750	94,284	94,306	94,343	94,537	Hypothetical protein	283	792
	NZ_CP045527.1	60,549	61,580	59,993	60,015	59,760	59,957	Hypothetical protein	58,970	59,479
	NZ_CP049176.1	60,477	61,508	59,926	59,944	59,687	59,885	*yacAB*	59,040	59,566
	NZ_CP049178.1	6,145	7,176	5,594	5,612	5,355	5,553	*yacAB*	4,708	5,234

	NZ_CP049180.1	51,211	52,242	50,660	50,678	50,421	50,619	*yacAB*	49,774	50,300
	NZ_CP053754.1	7,218	8,249	6,663	6,685	6,430	6,628	Hypothetical protein	5,640	6,149
	NZ_CP099778.1	46	1,077	1,610	1,632	1,667	1,865	Hypothetical protein	2,146	2,655
	NZ_CP099779.1	46	1,077	1,611	1,633	1,670	1,866	Hypothetical protein	2,147	2,656
	NZ_CP099780.1	46	1,076	1,610	1,632	1,667	1,865	IS91-family transposase	2,100	3,308
	NZ_CP099782.1	1	1,032	1,565	1,587	1,623	1,820	Hypothetical protein	2,101	2,610
	NZ_CP099783.1	46	1,077	1,610	1,632	1,667	1,865	*yacAB*	1,986	2,533
	NZ_KJ406378.1	455	1,486	2,020	2,042	2,078	2,275	Hypothetical protein	2,556	3,065
	NZ_MG569891.1	5,926	6,957	5,375	5,393	5,137	5,334	*yacAB*	4,469	5,016
	NZ_ON461902.1	59,912	60,943	61,476	61,494	61,535	61,733	Hypothetical protein	62,014	62,523
	NZ_OP038281.1	46	1,077	1,610	1,628	1,669	1,867	*yacAB*	1,988	2,514
	NZ_OQ230387.1	21,039	21,917	22,441	22,459	22,500	22,698	*yacAB*	22,819	23,276
	NZ_OQ230385.1	60,239	61,117	59,698	59,716	59,459	59,657	*yacAB*	58,881	59,338
	NZ_CP099769.1	1	879	1,403	1,421	1,464	1,660	*relBE*	1,778	2,337
	NZ_KY471628.1	1	879	1,403	1,421	1,464	1,660	*relBE*	1,778	2,337
	NZ_KY471629.1	1	879	1,403	1,421	1,464	1,660	*relBE*	1,778	2,337
	NZ_MG299131.1	1	879	1,403	1,421	1,464	1,660	*relBE*	1,778	2,337
	NZ_MG299139.1	1	879	1,403	1,421	1,464	1,660	*relBE*	1,778	2,337
	NZ_MG299143.1	1	879	1,403	1,421	1,464	1,660	*relBE*	1,778	2,337
	NZ_MG299147.1	1	879	1,403	1,421	1,464	1,660	*relBE*	1,778	2,337
	NZ_MG299151.1	1	879	1,403	1,421	1,464	1,660	*relBE*	1,778	2,337
IncB/O/K/Z	NZ_ON461900.1	60,284	61,162	61,686	61,704	61,745	61,943	*relBE*	62,061	62,620
	NZ_OP038269.1	1	879	1,403	1,421	1,462	1,660	*relBE*	1,778	2,337
	NZ_OP038272.1	1	879	1,403	1,421	1,462	1,660	*relBE*	1,778	2,337
	NZ_OP038276.1	65,105	65,983	66,507	66,525	66,566	66,764	*relBE*	66,882	67,441
	CP104418.1	70,672	71,703	70,121	70,139	69,882	70,080	*yacAB*	69,304	69,761
	CP102114.1	70,672	71,703	70,121	70,139	69,882	70,080	*yacAB*	69,304	69,761
	CP104411.1	60,371	61,402	59,820	59,838	59,581	59,779	*yacAB*	59,003	59,460
	CP104414.1	78,550	79,581	77,999	78,017	77,760	77,958	*yacAB*	77,182	77,639
	CP104424.1	72,392	73,423	73,956	73,974	74,015	74,213	*yacAB*	74,334	74,791
	CP104429.1	65,553	66,584	67,117	67,135	67,176	67,374	*yacAB*	67,495	67,952
	CP104432.1	82,784	83,815	82,233	82,251	81,994	82,192	*yacAB*	81,416	81,873

	CP104436.1	82,797	83,828	82,246	82,264	82,007	82,205	*yacAB*	81,429	81,886
	LR213460.1	46	1,077	1,610	1,628	1,669	1,866	*relBE*	1,985	2,544
	NZ_CP022673.1	1	1,032	112,534	112,552	112,295	112,493	*relBE*	111,685	112,177
	NZ_MW396864.1	1	1,032	86,319	86,337	86,080	86,278	*yacAB*	85,502	85,959
	NZ_OP038275.1	46	1,077	1,611	1,629	1,668	1,866	*yacAB*	1,987	2,444
	NZ_CP099768.1	46	1,077	1,611	1,629	1,668	1,866	*yacAB*	1,987	2,444
	NZ_CP099774.1	46	1,077	1,610	1,628	1,669	1,867	*yacAB*	1,988	2,443
	NZ_CP099781.1	46	1,077	1,611	1,629	1,668	1,866	*yacAB*	1,987	2,444
	NZ_OP038271.1	46	1,077	1,610	1,628	1,669	1,867	*relBE*	1,985	2,544
IncI2	NZ_KY363994.1	496	1,524	2,060	2,084	2,121	2,316	*hok/sok*	2,993	3,280
	NZ_KY363995.1	496	1,524	2,060	2,084	2,121	2,316	*hok/sok*	2,993	3,280
	NZ_KY363996.1	496	1,524	2,060	2,084	2,121	2,316	IS91-family transposase	3,134	4,333
	NZ_KY363997.1	496	1,524	2,060	2,084	2,121	2,316	IS91-family transposase	2,822	4,021
	NZ_KY363998.1	496	1,524	2,060	2,084	2,121	2,316	*hok/sok*	2,993	3,280
	NZ_KY363999.1	496	1,524	2,060	2,084	2,121	2,316	IS91-family transposase	2,822	4,021
	NZ_LT174530.1	496	1,524	2,060	2,084	2,121	2,316	*hok/sok*	2,993	3,280
	NZ_KJ460501.1	38	1,066	1,598	1,623	1,657	1,854	*hok/sok*	2,340	2,627
	NZ_KX827311.1	38	1,066	1,598	1,623	1,657	1,854	*hok/sok*	2,340	2,627
	NZ_KY174331.1	38	1,066	1,598	1,623	1,657	1,854	*hok/sok*	2,340	2,627
	NZ_LC477293.1	496	1,524	2,056	2,081	2,115	2,312	*hok/sok*	2,798	3,085
	NZ_OQ230389.1	44,383	45,411	43,826	43,851	43,595	43,792	*hok/sok*	42,822	43,109
	NZ_CP019690.1	54,803	55,831	56,367	56,392	56,428	56,622	*hok/sok*	57,111	57,398
	NZ_MG299138.1	488	1,516	2,052	2,077	2,113	2,307	ISEc44	2,445	4,262
	NZ_MG299140.1	488	1,516	2,052	2,077	2,113	2,307	ISEc44	2,445	4,262
	NZ_MG299142.1	488	1,516	2,052	2,077	2,113	2,307	ISEc44	2,445	4,262
	NZ_MG299144.1	488	1,516	2,052	2,077	2,113	2,307	ISEc44	2,445	4,262
	NZ_MG299146.1	488	1,516	2,052	2,077	2,113	2,307	ISEc44	2,445	4,262
	NZ_MG299148.1	488	1,516	2,052	2,077	2,113	2,307	ISEc44	2,445	4,262
	NZ_MG299150.1	488	1,516	2,052	2,077	2,113	2,307	ISEc44	2,445	4,262
	NZ_MG299152.1	488	1,516	2,052	2,077	2,113	2,307	ISEc44	2,445	4,262
	NZ_CP045935.1	1	1,029	1,561	1,586	1,620	1,817	*hok/sok*	2,303	2,590

Table 7.2: Summary of elements surrounding 199R on plasmids from different Inc groups.
The Inc groups, accession numbers and the locations of different elements were listed.

8 References

1. Shiga K. Ueber den Erreger der Dysenterie in Japan. Cent. Bakteriol. Parasiteukunde Infektionskrank. 1898;23:599-600.
2. Lampel KA, Formal SB, Maurelli AT. A Brief History of Shigella. EcoSal Plus. 2018;8(1).
3. Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AKM. Shigellosis. Lancet. 2018;391(10122):801-12.
4. Kothary MH, Babu US. INFECTIVE DOSE OF FOODBORNE PATHOGENS IN VOLUNTEERS: A REVIEW. Journal of Food Safety. 2001;21(1):49-68.
5. Dodd CE, Jones D. A numerical taxonomic study of the genus Shigella. J Gen Microbiol. 1982;128(9):1933-57.
6. Nakamura M, Clausen CR. Lactose fermentation by Shigella sonnei. Nature. 1962;196:44-6.
7. Pupo GM, Lan R, Reeves PR. Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics. Proc Natl Acad Sci U S A. 2000;97(19):10567-72.
8. Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SyN, Wang Q, et al. Structure and genetics of Shigella O antigens. FEMS Microbiology Reviews. 2008;32(4):627-53.
9. Hale TL. Genetic basis of virulence in Shigella species. Microbiol Rev. 1991;55(2):206-24.
10. The HC, Thanh DP, Holt KE, Thomson NR, Baker S. The genomic signatures of Shigella evolution, adaptation and geographical spread. Nature Reviews Microbiology. 2016;14(4):235-50.
11. Niyogi SK. Shigellosis. J Microbiol. 2005;43(2):133-43.
12. Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, Farag TH, van Eijk A, et al. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. Clin Infect Dis. 2012;55 Suppl 4(Suppl 4):S232-45.
13. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet. 2013;382(9888):209-22.
14. Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Swerdlow DL, Sansonetti PJ, et al. Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. Bull World Health Organ. 1999;77(8):651-66.
15. Organization WH. Diarrhoeal disease 2024 [updated 7th March 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease.
16. Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, et al. Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990-2016. Lancet Infect Dis. 2018;18(11):1229-40.
17. Baker S, Scott TA. Antimicrobial-resistant Shigella: where do we go next? Nature Reviews Microbiology. 2023;21(7):409-10.
18. Khalil I, Troeger CE, Blacker BF, Reiner RC, Jr. Capturing the true burden of Shigella and ETEC: The way forward. Vaccine. 2019;37(34):4784-6.
19. Gu B, Cao Y, Pan S, Zhuang L, Yu R, Peng Z, et al. Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of Shigella between Europe-America and Asia-Africa from 1998 to 2009. Int J Antimicrob Agents. 2012;40(1):9-17.
20. Holt KE, Baker S, Weill FX, Holmes EC, Kitchen A, Yu J, et al. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat Genet. 2012;44(9):1056-9.

21. Lefèvre S, Njamkepo E, Feldman S, Ruckly C, Carle I, Lejay-Collin M, et al. Rapid emergence of extensively drug-resistant Shigella sonnei in France. Nature Communications. 2023;14(1):462.
22. Bardsley M, Jenkins C, Mitchell HD, Mikhail AFW, Baker KS, Foster K, et al. Persistent Transmission of Shigellosis in England Is Associated with a Recently Emerged Multidrug-Resistant Strain of Shigella sonnei. J Clin Microbiol. 2020;58(4).
23. Mason LCE, Greig DR, Cowley LA, Partridge SR, Martinez E, Blackwell GA, et al. The evolution and international spread of extensively drug resistant Shigella sonnei. Nature Communications. 2023;14(1):1983.
24. Melton-Celsa AR. Shiga Toxin (Stx) Classification, Structure, and Function. Microbiol Spectr. 2014;2(4):Ehec-0024-2013.
25. Levine MM, Kotloff KL, Barry EM, Pasetti MF, Sztein MB. Clinical trials of Shigella vaccines: two steps forward and one step back on a long, hard road. Nat Rev Microbiol. 2007;5(7):540-53.
26. Williams PCM, Berkley JA. Guidelines for the treatment of dysentery (shigellosis): a systematic review of the evidence. Paediatr Int Child Health. 2018;38(sup1):S50-s65.
27. Das JK, Ali A, Salam RA, Bhutta ZA. Antibiotics for the treatment of Cholera, Shigella and Cryptosporidium in children. BMC Public Health. 2013;13 Suppl 3(Suppl 3):S10.
28. Traa BS, Walker CL, Munos M, Black RE. Antibiotics for the treatment of dysentery in children. Int J Epidemiol. 2010;39 Suppl 1(Suppl 1):i70-4.
29. von Seidlein L, Kim DR, Ali M, Lee H, Wang X, Thiem VD, et al. A multicentre study of Shigella diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. PLoS Med. 2006;3(9):e353.
30. Das SK, Klontz EH, Azmi IJ, Ud-Din AI, Chisti MJ, Afrad MH, et al. Characteristics of Multidrug Resistant Shigella and Vibrio cholerae O1 Infections in Patients Treated at an Urban and a Rural Hospital in Bangladesh. ISRN Microbiol. 2013;2013:213915.
31. Shoja S, Ghasemi S, Dastranj M, Shamseddin J, Ebrahimi N, Alizade H, et al. Characterization of genotypes and antimicrobial resistance profiles of clinical isolates of Shigella from patients in the southern region of Iran. Eur J Med Res. 2023;28(1):611.
32. Kasumba IN, Badji H, Powell H, Hossain MJ, Omore R, Sow SO, et al. Shigella in Africa: New Insights From the Vaccine Impact on Diarrhea in Africa (VIDA) Study. Clin Infect Dis. 2023;76(76 Suppl1):S66-s76.
33. Thompson CN, Thieu NT, Vinh PV, Duc AN, Wolbers M, Vinh H, et al. Clinical implications of reduced susceptibility to fluoroquinolones in paediatric Shigella sonnei and Shigella flexneri infections. J Antimicrob Chemother. 2016;71(3):807-15.
34. Bose P, Chowdhury G, Halder G, Ghosh D, Deb AK, Kitahara K, et al. Prevalence and changing antimicrobial resistance profiles of Shigella spp. isolated from diarrheal patients in Kolkata during 2011-2019. PLoS Negl Trop Dis. 2024;18(2):e0011964.
35. van den Beld MJC, Reubsaet FAG. Differentiation between Shigella, enteroinvasive Escherichia coli (EIEC) and noninvasive Escherichia coli. European Journal of Clinical Microbiology & Infectious Diseases. 2012;31(6):899-904.
36. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY, Veeraraghavan B. Accurate differentiation of Escherichia coli and Shigella serogroups: challenges and strategies. New Microbes and New Infections. 2018;21:58-62.
37. Yang F, Yang J, Zhang X, Chen L, Jiang Y, Yan Y, et al. Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. Nucleic Acids Res. 2005;33(19):6445-58.
38. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 2014;38(5):865-91.
39. Hawkey J, Monk JM, Billman-Jacobe H, Palsson B, Holt KE. Impact of insertion sequences on convergent evolution of Shigella species. PLOS Genetics. 2020;16(7):e1008931.

40. Tillier ER, Collins RA. Genome rearrangement by replication-directed translocation. Nat Genet. 2000;26(2):195-7.
41. Haidar-Ahmad N, Manigat FO, Silué N, Pontier SM, Campbell-Valois FX. A Tale about Shigella: Evolution, Plasmid, and Virulence. Microorganisms. 2023;11(7).
42. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Front Microbiol. 2015;6:242.
43. Novick RP. Plasmid incompatibility. Microbiol Rev. 1987;51(4):381-95.
44. Martyn JE, Pilla G, Hollingshead S, Winther KS, Lea S, McVicker G, et al. Maintenance of the Shigella sonnei Virulence Plasmid Is Dependent on Its Repertoire and Amino Acid Sequence of Toxin-Antitoxin Systems. J Bacteriol. 2022;204(3):e0051921.
45. Ito H, Kido N, Arakawa Y, Ohta M, Sugiyama T, Kato N. Possible mechanisms underlying the slow lactose fermentation phenotype in Shigella spp. Appl Environ Microbiol. 1991;57(10):2912-7.
46. Prosseda G, Di Martino ML, Campilongo R, Fioravanti R, Micheli G, Casalino M, et al. Shedding of genes that interfere with the pathogenic lifestyle: the Shigella model. Res Microbiol. 2012;163(6-7):399-406.
47. Prunier AL, Schuch R, Fernández RE, Mumy KL, Kohler H, McCormick BA, et al. nadA and nadB of Shigella flexneri 5a are antivirulence loci responsible for the synthesis of quinolinate, a small molecule inhibitor of Shigella pathogenicity. Microbiology (Reading). 2007;153(Pt 7):2363-72.
48. Maurelli AT, Fernández RE, Bloch CA, Rode CK, Fasano A. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli. Proc Natl Acad Sci U S A. 1998;95(7):3943-8.
49. Barbagallo M, Di Martino ML, Marcocci L, Pietrangeli P, De Carolis E, Casalino M, et al. A new piece of the Shigella Pathogenicity puzzle: spermidine accumulation by silencing of the speG gene [corrected]. PLoS One. 2011;6(11):e27226.
50. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, et al. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis. 2011;17(1):7-15.
51. Mohle-Boetani JC, Stapleton M, Finger R, Bean NH, Poundstone J, Blake PA, et al. Communitywide shigellosis: control of an outbreak and risk factors in child day-care centers. Am J Public Health. 1995;85(6):812-6.
52. Black RE, Craun GF, Blake PA. Epidemiology of common-source outbreaks of shigellosis in the United States, 1961-1975. Am J Epidemiol. 1978;108(1):47-52.
53. Farag TH, Faruque AS, Wu Y, Das SK, Hossain A, Ahmed S, et al. Housefly population density correlates with shigellosis among children in Mirzapur, Bangladesh: a time series analysis. PLoS Negl Trop Dis. 2013;7(6):e2280.
54. Cohen D, Green M, Block C, Slepon R, Ambar R, Wasserman SS, et al. Reduction of transmission of shigellosis by control of houseflies (Musca domestica). Lancet. 1991;337(8748):993-7.
55. Shah N, DuPont HL, Ramsey DJ. Global etiology of travelers' diarrhea: systematic review from 1973 to the present. Am J Trop Med Hyg. 2009;80(4):609-14.
56. Kendall ME, Crim S, Fullerton K, Han PV, Cronquist AB, Shiferaw B, et al. Travel-associated enteric infections diagnosed after return to the United States, Foodborne Diseases Active Surveillance Network (FoodNet), 2004-2009. Clin Infect Dis. 2012;54 Suppl 5:S480-7.
57. Olson S, Hall A, Riddle MS, Porter CK. Travelers' diarrhea: update on the incidence, etiology and risk in military and similar populations - 1990-2005 versus 2005-2015, does a decade make a difference? Trop Dis Travel Med Vaccines. 2019;5:1.
58. Chandani Y, Ghosh V, Suresh V, Singh V, Asumah MN, Padhi BK. An outbreak of shigellosis in European travellers returning from Cape Verde. New Microbes New Infect. 2023;52:101091.

59. Bowen A, Hurd J, Hoover C, Khachadourian Y, Traphagen E, Harvey E, et al. Importation and domestic transmission of Shigella sonnei resistant to ciprofloxacin - United States, May 2014-February 2015. MMWR Morb Mortal Wkly Rep. 2015;64(12):318-20.
60. Bardhan P, Faruque AS, Naheed A, Sack DA. Decrease in shigellosis-related deaths without Shigella spp.-specific interventions, Asia. Emerg Infect Dis. 2010;16(11):1718-23.
61. Marder EP, Cieslak PR, Cronquist AB, Dunn J, Lathrop S, Rabatsky-Ehr T, et al. Incidence and Trends of Infections with Pathogens Transmitted Commonly Through Food and the Effect of Increasing Use of Culture-Independent Diagnostic Tests on Surveillance - Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2013-2016. MMWR Morb Mortal Wkly Rep. 2017;66(15):397-403.
62. Thompson CN, Duy PT, Baker S. The Rising Dominance of Shigella sonnei: An Intercontinental Shift in the Etiology of Bacillary Dysentery. PLoS Negl Trop Dis. 2015;9(6):e0003708.
63. Puzari M, Sharma M, Chetia P. Emergence of antibiotic resistant Shigella species: A matter of concern. J Infect Public Health. 2018;11(4):451-4.
64. Vinh H, Nhu NT, Nga TV, Duy PT, Campbell JI, Hoang NV, et al. A changing picture of shigellosis in southern Vietnam: shifting species dominance, antimicrobial susceptibility and clinical presentation. BMC Infect Dis. 2009;9:204.
65. Bangtrakulnonth A, Vieira AR, Lo Fo Wong DM, Pornreongwong S, Pulsrikarn C, Sawanpanyalert P, et al. Shigella from humans in Thailand during 1993 to 2006: spatial-time trends in species and serotype distribution. Foodborne Pathog Dis. 2008;5(6):773-84.
66. Ud-Din AI, Wahid SU, Latif HA, Shahnaij M, Akter M, Azmi IJ, et al. Changing trends in the prevalence of Shigella species: emergence of multi-drug resistant Shigella sonnei biotype g in Bangladesh. PLoS One. 2013;8(12):e82601.
67. Qiu S, Xu X, Yang C, Wang J, Liang B, Li P, et al. Shift in serotype distribution of Shigella species in China, 2003-2013. Clin Microbiol Infect. 2015;21(3):252.e5-8.
68. Sack DA, Hoque AT, Huq A, Etheridge M. Is protection against shigellosis induced by natural infection with Plesiomonas shigelloides? Lancet. 1994;343(8910):1413-5.
69. Shepherd JG, Wang L, Reeves PR. Comparison of O-antigen gene clusters of Escherichia coli (Shigella) sonnei and Plesiomonas shigelloides O17: sonnei gained its current plasmid-borne O-antigen genes from P. shigelloides in a recent event. Infect Immun. 2000;68(10):6056-61.
70. Jeong HJ, Jang ES, Han BI, Lee KH, Ock MS, Kong HH, et al. Acanthamoeba: could it be an environmental host of Shigella? Exp Parasitol. 2007;115(2):181-6.
71. Saeed A, Johansson D, Sandström G, Abd H. Temperature Depended Role of Shigella flexneri Invasion Plasmid on the Interaction with Acanthamoeba castellanii. Int J Microbiol. 2012;2012:917031.
72. Siddiq M, O'Flanagan H, Richardson D, Llewellyn CD. Factors associated with sexually transmitted shigella in men who have sex with men: a systematic review. Sex Transm Infect. 2023;99(1):58-63.
73. Simms I, Field N, Jenkins C, Childs T, Gilbart VL, Dallman TJ, et al. Intensified shigellosis epidemic associated with sexual transmission in men who have sex with men--Shigella flexneri and S. sonnei in England, 2004 to end of February 2015. Euro Surveill. 2015;20(15).
74. Gaufin T, Blumenthal J, Ramirez-Sanchez C, Mehta S, Pride DT, Fierer J, et al. Antimicrobial-Resistant Shigella spp. in San Diego, California, USA, 2017-2020. Emerg Infect Dis. 2022;28(6):1110-6.
75. Mitchell HD, Whitlock G, Zdravkov J, Olsson J, Silalang P, Bardsley M, et al. Prevalence and risk factors of bacterial enteric pathogens in men who have sex with men: A cross-sectional study at the UK's largest sexual health service. J Infect. 2023;86(1):33-40.
76. Zayet S, Klopfenstein T, Pierron A, Royer PY, Toko L, Garnier P, et al. Shigella sonnei, an emerging multidrug-resistant sexually transmitted pathogen in Franche-Comté, France. Emerg Microbes Infect. 2021;10(1):1702-5.

77. Fischer N, Maex M, Mattheus W, Van den Bossche A, Van Cauteren D, Laisnez V, et al. Genomic epidemiology of persistently circulating MDR Shigella sonnei strains associated with men who have sex with men (MSM) in Belgium (2013-19). J Antimicrob Chemother. 2021;77(1):89-97.
78. Braam JF, Bruisten SM, Hoogeland M, de Vries HJC, Schim van der Loeff MF, van Dam AP. Shigella is common in symptomatic and asymptomatic men who have sex with men visiting a sexual health clinic in Amsterdam. Sex Transm Infect. 2022;98(8):564-9.
79. Moreno-Mingorance A, Mir-Cros A, Goterris L, Rodriguez-Garrido V, Sulleiro E, Barberà MJ, et al. Increasing trend of antimicrobial resistance in Shigella associated with MSM transmission in Barcelona, 2020-21: outbreak of XRD Shigella sonnei and dissemination of ESBL-producing Shigella flexneri. J Antimicrob Chemother. 2023;78(4):975-82.
80. Ingle DJ, Easton M, Valcanis M, Seemann T, Kwong JC, Stephens N, et al. Co-circulation of Multidrug-resistant Shigella Among Men Who Have Sex With Men in Australia. Clin Infect Dis. 2019;69(9):1535-44.
81. Matanza XM, Clements A. Pathogenicity and virulence of Shigella sonnei: A highly drug-resistant pathogen of increasing prevalence. Virulence. 2023;14(1):2280838.
82. Pakbin B, Brück WM, Brück TB. Molecular Mechanisms of Shigella Pathogenesis; Recent Advances. International Journal of Molecular Sciences [Internet]. 2023; 24(3).
83. Ashida H, Suzuki T, Sasakawa C. Shigella infection and host cell death: a double-edged sword for the host and pathogen survival. Curr Opin Microbiol. 2021;59:1-7.
84. Jiang Y, Yang F, Zhang X, Yang J, Chen L, Yan Y, et al. The complete sequence and analysis of the large virulence plasmid pSS of Shigella sonnei. Plasmid. 2005;54(2):149-59.
85. Muthuramalingam M, Whittier SK, Picking WL, Picking WD. The Shigella Type III Secretion System: An Overview from Top to Bottom. Microorganisms. 2021;9(2).
86. Bajunaid W, Haidar-Ahmad N, Kottarampatel AH, Ourida Manigat F, Silué N, Tchagang CF, et al. The T3SS of Shigella: Expression, Structure, Function, and Role in Vacuole Escape. Microorganisms. 2020;8(12).
87. Tachiyama S, Skaar R, Chang Y, Carroll BL, Muthuramalingam M, Whittier SK, et al. Composition and Biophysical Properties of the Sorting Platform Pods in the Shigella Type III Secretion System. Frontiers in Cellular and Infection Microbiology. 2021;11.
88. Ferrari ML, Charova SN, Sansonetti PJ, Mylonas E, Gazi AD. Structural Insights of Shigella Translocator IpaB and Its Chaperone IpgC in Solution. Frontiers in Cellular and Infection Microbiology. 2021;11.
89. Buchrieser C, Glaser P, Rusniok C, Nedjari H, D'Hauteville H, Kunst F, et al. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri. Molecular Microbiology. 2000;38(4):760-71.
90. Picker MA, Wing HJ. H-NS, Its Family Members and Their Regulation of Virulence Genes in Shigella Species. Genes (Basel). 2016;7(12).
91. Tobe T, Yoshikawa M, Mizuno T, Sasakawa C. Transcriptional control of the invasion regulatory gene virB of Shigella flexneri: activation by virF and repression by H-NS. J Bacteriol. 1993;175(19):6142-9.
92. Beloin C, Dorman CJ. An extended role for the nucleoid structuring protein H-NS in the virulence gene regulatory cascade of Shigella flexneri. Mol Microbiol. 2003;47(3):825-38.
93. Falconi M, Colonna B, Prosseda G, Micheli G, Gualerzi CO. Thermoregulation of Shigella and Escherichia coli EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of virF promoter to transcriptional repressor H-NS. Embo j. 1998;17(23):7033-43.
94. Di Martino ML, Falconi M, Micheli G, Colonna B, Prosseda G. The Multifaceted Activity of the VirF Regulatory Protein in the Shigella Lifestyle. Front Mol Biosci. 2016;3:61.
95. Anderson MJ, den Hartigh AB, Fink SL. Molecular Mechanisms of Pyroptosis. In: Fink SL, editor. Pyroptosis: Methods and Protocols. New York, NY: Springer US; 2023. p. 1-16.

96. Simonian HP, Vo L, Doma S, Fisher RS, Parkman HP. Regional postprandial differences in pH within the stomach and gastroesophageal junction. Dig Dis Sci. 2005;50(12):2276-85.
97. Evans DF, Pye G, Bramley R, Clark AG, Dyson TJ, Hardcastle JD. Measurement of gastrointestinal pH profiles in normal ambulant human subjects. Gut. 1988;29(8):1035-41.
98. Bliven KA, Lampel KA. Shigella: Virulence Factors and Pathogenicity. In: Gurtler JB, Doyle MP, Kornacki JL, editors. Foodborne Pathogens: Virulence Factors and Host Susceptibility. Cham: Springer International Publishing; 2017. p. 169-208.
99. Gorden J, Small PL. Acid resistance in enteric bacteria. Infect Immun. 1993;61(1):364-7.
100. Waterman SR, Small PL. Identification of sigma S-dependent genes associated with the stationary-phase acid-resistance phenotype of Shigella flexneri. Mol Microbiol. 1996;21(5):925-40.
101. Chiang IL, Wang Y, Fujii S, Muegge BD, Lu Q, Tarr PI, et al. Biofilm Formation and Virulence of Shigella flexneri Are Modulated by pH of Gastrointestinal Tract. Infect Immun. 2021;89(11):e0038721.
102. Nickerson KP, Chanin RB, Sistrunk JR, Rasko DA, Fink PJ, Barry EM, et al. Analysis of Shigella flexneri Resistance, Biofilm Formation, and Transcriptional Profile in Response to Bile Salts. Infect Immun. 2017;85(6).
103. Brotcke Zumsteg A, Goosmann C, Brinkmann V, Morona R, Zychlinsky A. IcsA is a Shigella flexneri adhesin regulated by the type III secretion system and required for pathogenesis. Cell Host Microbe. 2014;15(4):435-45.
104. Köseoğlu VK, Hall CP, Rodríguez-López EM, Agaisse H. The Autotransporter IcsA Promotes Shigella flexneri Biofilm Formation in the Presence of Bile Salts. Infect Immun. 2019;87(7).
105. Dickenson NE, Zhang L, Epler CR, Adam PR, Picking WL, Picking WD. Conformational changes in IpaD from Shigella flexneri upon binding bile salts provide insight into the second step of type III secretion. Biochemistry. 2011;50(2):172-80.
106. Bernard AR, Jessop TC, Kumar P, Dickenson NE. Deoxycholate-Enhanced Shigella Virulence Is Regulated by a Rare π-Helix in the Type Three Secretion System Tip Protein IpaD. Biochemistry. 2017;56(49):6503-14.
107. Mahmoud RY, Stones DH, Li W, Emara M, El-Domany RA, Wang D, et al. The Multivalent Adhesion Molecule SSO1327 plays a key role in Shigella sonnei pathogenesis. Mol Microbiol. 2016;99(4):658-73.
108. Willer EdM, Lima RdL, Giugliano LG. In vitro adhesion and invasion inhibition of Shigella dysenteriae, Shigella flexneri and Shigella sonnei clinical strains by human milk proteins. BMC Microbiology. 2004;4(1):18.
109. Zhu Z, Wang W, Cao M, Zhu Q, Ma T, Zhang Y, et al. Virulence factors and molecular characteristics of Shigella flexneri isolated from calves with diarrhea. BMC Microbiology. 2021;21(1):214.
110. Nutten S, Sansonetti P, Huet G, Bourdon-Bisiaux C, Meresse B, Colombel JF, et al. Epithelial inflammation response induced by Shigella flexneri depends on mucin gene expression. Microbes Infect. 2002;4(11):1121-4.
111. Marteyn B, West NP, Browning DF, Cole JA, Shaw JG, Palm F, et al. Modulation of Shigella virulence in response to available oxygen in vivo. Nature. 2010;465(7296):355-8.
112. Haider K, Hossain A, Wanke C, Qadri F, Ali S, Nahar S. Production of mucinase and neuraminidase and binding of Shigella to intestinal mucin. J Diarrhoeal Dis Res. 1993;11(2):88-92.
113. Prizont R. Degradation of intestinal glycoproteins by pathogenic Shigella flexneri. Infect Immun. 1982;36(2):615-20.
114. Sansonetti PJ, Arondel J, Cantey JR, Prévost MC, Huerre M. Infection of rabbit Peyer's patches by Shigella flexneri: effect of adhesive or invasive bacterial phenotypes on follicle-associated epithelium. Infect Immun. 1996;64(7):2752-64.

115. Mabbott NA, Donaldson DS, Ohno H, Williams IR, Mahajan A. Microfold (M) cells: important immunosurveillance posts in the intestinal epithelium. Mucosal Immunol. 2013;6(4):666-77.
116. Rey C, Chang YY, Latour-Lambert P, Varet H, Proux C, Legendre R, et al. Transcytosis subversion by M cell-to-enterocyte spread promotes Shigella flexneri and Listeria monocytogenes intracellular bacterial dissemination. PLoS Pathog. 2020;16(4):e1008446.
117. Mounier J, Vasselon T, Hellio R, Lesourd M, Sansonetti PJ. Shigella flexneri enters human colonic Caco-2 epithelial cells through the basolateral pole. Infect Immun. 1992;60(1):237-48.
118. High N, Mounier J, Prévost MC, Sansonetti PJ. IpaB of Shigella flexneri causes entry into epithelial cells and escape from the phagocytic vacuole. Embo j. 1992;11(5):1991-9.
119. Zychlinsky A, Kenny B, Ménard R, Prévost MC, Holland IB, Sansonetti PJ. IpaB mediates macrophage apoptosis induced by Shigella flexneri. Mol Microbiol. 1994;11(4):619-27.
120. Senerovic L, Tsunoda SP, Goosmann C, Brinkmann V, Zychlinsky A, Meissner F, et al. Spontaneous formation of IpaB ion channels in host cell membranes reveals how Shigella induces pyroptosis in macrophages. Cell Death Dis. 2012;3(9):e384.
121. Ashida H, Mimuro H, Sasakawa C. Shigella Manipulates Host Immune Responses by Delivering Effector Proteins with Specific Roles. Frontiers in Immunology. 2015;6.
122. Guichon A, Zychlinsky A. Clinical isolates of Shigella species induce apoptosis in macrophages. J Infect Dis. 1997;175(2):470-3.
123. Watson JL, Sanchez-Garrido J, Goddard PJ, Torraca V, Mostowy S, Shenoy AR, et al. Shigella sonnei O-Antigen Inhibits Internalization, Vacuole Escape, and Inflammasome Activation. mBio. 2019;10(6).
124. Reeves P. Role of O-antigen variation in the immune response. Trends Microbiol. 1995;3(10):381-6.
125. Caboni M, Pédron T, Rossi O, Goulding D, Pickard D, Citiulo F, et al. An O antigen capsule modulates bacterial pathogenesis in Shigella sonnei. PLoS Pathog. 2015;11(3):e1004749.
126. Miao EA, Mao DP, Yudkovsky N, Bonneau R, Lorang CG, Warren SE, et al. Innate immune detection of the type III secretion apparatus through the NLRC4 inflammasome. Proc Natl Acad Sci U S A. 2010;107(7):3076-80.
127. Mostowy S, Boucontet L, Mazon Moya MJ, Sirianni A, Boudinot P, Hollinshead M, et al. The zebrafish as a new model for the in vivo study of Shigella flexneri interaction with phagocytes and bacterial autophagy. PLoS Pathog. 2013;9(9):e1003588.
128. Riddle RB, Jennbacken K, Hansson KM, Harper MT. Endothelial inflammation and neutrophil transmigration are modulated by extracellular matrix composition in an inflammation-on-a-chip model. Scientific Reports. 2022;12(1):6855.
129. Salvermoser M, Begandt D, Alon R, Walzog B. Nuclear Deformation During Neutrophil Migration at Sites of Inflammation. Frontiers in Immunology. 2018;9.
130. Raqib R, Ekberg C, Sharkar P, Bardhan PK, Zychlinsky A, Sansonetti PJ, et al. Apoptosis in acute shigellosis is associated with increased production of Fas/Fas ligand, perforin, caspase-1, and caspase-3 but reduced production of Bcl-2 and interleukin-2. Infect Immun. 2002;70(6):3199-207.
131. Weinrauch Y, Drujan D, Shapiro SD, Weiss J, Zychlinsky A. Neutrophil elastase targets virulence factors of enterobacteria. Nature. 2002;417(6884):91-4.
132. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil extracellular traps kill bacteria. Science. 2004;303(5663):1532-5.
133. Sansonetti PJ, Tran Van Nhieu G, Egile C. Rupture of the intestinal epithelial barrier and mucosal invasion by Shigella flexneri. Clin Infect Dis. 1999;28(3):466-75.
134. Mandic-Mulec I, Weiss J, Zychlinsky A. Shigella flexneri is trapped in polymorphonuclear leukocyte vacuoles and efficiently killed. Infect Immun. 1997;65(1):110-5.

135. François M, Le Cabec V, Dupont MA, Sansonetti PJ, Maridonneau-Parini I. Induction of necrosis in human neutrophils by Shigella flexneri requires type III secretion, IpaB and IpaC invasins, and actin polymerization. Infect Immun. 2000;68(3):1289-96.
136. Torraca V, Kaforou M, Watson J, Duggan GM, Guerrero-Gutierrez H, Krokowski S, et al. Shigella sonnei infection of zebrafish reveals that O-antigen mediates neutrophil tolerance and dysentery incidence. PLoS Pathog. 2019;15(12):e1008006.
137. Dickenson NE, Choudhari SP, Adam PR, Kramer RM, Joshi SB, Middaugh CR, et al. Oligomeric states of the Shigella translocator protein IpaB provide structural insights into formation of the type III secretion translocon. Protein Sci. 2013;22(5):614-27.
138. Russo BC, Stamm LM, Raaben M, Kim CM, Kahoud E, Robinson LR, et al. Intermediate filaments enable pathogen docking to trigger type 3 effector translocation. Nat Microbiol. 2016;1:16025.
139. Tran Van Nhieu G, Caron E, Hall A, Sansonetti PJ. IpaC induces actin polymerization and filopodia formation during Shigella entry into epithelial cells. Embo j. 1999;18(12):3249-62.
140. Demali KA, Jue AL, Burridge K. IpaA targets beta1 integrins and rho to promote actin cytoskeleton rearrangements necessary for Shigella entry. J Biol Chem. 2006;281(51):39534-41.
141. Mellouk N, Enninga J. Cytosolic Access of Intracellular Bacterial Pathogens: The Shigella Paradigm. Front Cell Infect Microbiol. 2016;6:35.
142. Ogawa M, Yoshimori T, Suzuki T, Sagara H, Mizushima N, Sasakawa C. Escape of intracellular Shigella from autophagy. Science. 2005;307(5710):727-31.
143. Campbell-Valois FX, Sachse M, Sansonetti PJ, Parsot C. Escape of Actively Secreting Shigella flexneri from ATG8/LC3-Positive Vacuoles Formed during Cell-To-Cell Spread Is Facilitated by IcsB and VirA. mBio. 2015;6(3):e02567-14.
144. Pieper R, Fisher CR, Suh MJ, Huang ST, Parmar P, Payne SM. Analysis of the proteome of intracellular Shigella flexneri reveals pathways important for intracellular growth. Infect Immun. 2013;81(12):4635-48.
145. Egile C, Loisel TP, Laurent V, Li R, Pantaloni D, Sansonetti PJ, et al. Activation of the CDC42 effector N-WASP by the Shigella flexneri IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. J Cell Biol. 1999;146(6):1319-32.
146. Fukumatsu M, Ogawa M, Arakawa S, Suzuki M, Nakayama K, Shimizu S, et al. Shigella targets epithelial tricellular junctions and uses a noncanonical clathrin-dependent endocytic pathway to spread between cells. Cell Host Microbe. 2012;11(4):325-36.
147. Teh MY, Morona R. Identification of Shigella flexneri IcsA Residues Affecting Interaction with N-WASP, and Evidence for IcsA-IcsA Co-Operative Interaction. PLOS ONE. 2013;8(2):e55152.
148. Shibata T, Takeshima F, Chen F, Alt FW, Snapper SB. Cdc42 facilitates invasion but not the actin-based motility of Shigella. Curr Biol. 2002;12(4):341-5.
149. Wright SS, Vanaja SK. Shigella "Osp"pression of innate immunity. Cell. 2022;185(13):2205-7.
150. Ashida H, Sasakawa C, Suzuki T. A unique bacterial tactic to circumvent the cell death crosstalk induced by blockade of caspase-8. Embo j. 2020;39(17):e104469.
151. Carl S. The Bacteriology of Atoxic Dysentery Bacilli (Paradysentery Bacilli). Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene, Abt I (Originale). 1915;75(5/6):408–56 pp.
152. Zhu Z, Wang L, Qian H, Gu F, Li Y, Zhang H, et al. Comparative genome analysis of 12 Shigella sonnei strains: virulence, resistance, and their interactions. International Microbiology. 2021;24(1):83-91.
153. Hawkey J, Paranagama K, Baker KS, Bengtsson RJ, Weill F-X, Thomson NR, et al. Global population structure and genotyping framework for genomic surveillance of the major dysentery pathogen, Shigella sonnei. Nature Communications. 2021;12(1):2684.

154. Karaolis DK, Lan R, Reeves PR. Sequence variation in Shigella sonnei (Sonnei), a pathogenic clone of Escherichia coli, over four continents and 41 years. J Clin Microbiol. 1994;32(3):796-802.
155. Ke X, Gu B, Pan S, Tong M. Epidemiology and molecular mechanism of integron-mediated antibiotic resistance in Shigella. Arch Microbiol. 2011;193(11):767-74.
156. McVicker G, Tang CM. Deletion of toxin–antitoxin systems in the evolution of Shigella sonnei as a host-adapted pathogen. Nature Microbiology. 2016;2(2):16204.
157. Pilla G, Tang CM. Going around in circles: virulence plasmids in enteric pathogens. Nature Reviews Microbiology. 2018;16(8):484-95.
158. Xu DQ, Cisar JO, Ambulos N, Jr., Burr DH, Kopecko DJ. Molecular cloning and characterization of genes for Shigella sonnei form I O polysaccharide: proposed biosynthetic pathway and stable expression in a live salmonella vaccine vector. Infect Immun. 2002;70(8):4414-23.
159. Brahmbhatt HN, Lindberg AA, Timmis KN. Shigella lipopolysaccharide: structure, genetics, and vaccine development. Curr Top Microbiol Immunol. 1992;180:45-64.
160. Singh G, Yadav M, Ghosh C, Rathore JS. Bacterial toxin-antitoxin modules: classification, functions, and association with persistence. Current Research in Microbial Sciences. 2021;2:100047.
161. Sonika S, Singh S, Mishra S, Verma S. Toxin-antitoxin systems in bacterial pathogenesis. Heliyon. 2023;9(4):e14220.
162. Sengupta M, Austin S. Prevalence and significance of plasmid maintenance functions in the virulence plasmids of pathogenic bacteria. Infect Immun. 2011;79(7):2502-9.
163. Pilla G, McVicker G, Tang CM. Genetic plasticity of the Shigella virulence plasmid is mediated by intra- and inter-molecular events between insertion sequences. PLOS Genetics. 2017;13(9):e1007014.
164. Sayeed S, Reaves L, Radnedge L, Austin S. The stability region of the large virulence plasmid of Shigella flexneri encodes an efficient postsegregational killing system. J Bacteriol. 2000;182(9):2416-21.
165. Winther KS, Gerdes K. Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. Proc Natl Acad Sci U S A. 2011;108(18):7403-7.
166. Nordström K, Molin S, Light J. Control of replication of bacterial plasmids: genetics, molecular biology, and physiology of the plasmid R1 system. Plasmid. 1984;12(2):71-90.
167. Nordström K, Molin S, Aagaard-Hansen H. Partitioning of plasmid R1 in Escherichia coli,: I. Kinetics of loss of plasmid derivatives deleted of the par region. Plasmid. 1980;4(2):215-27.
168. Kolatka K, Kubik S, Rajewska M, Konieczny I. Replication and partitioning of the broad-host-range plasmid RK2. Plasmid. 2010;64(3):119-34.
169. Sansonetti PJ, Kopecko DJ, Formal SB. Shigella sonnei plasmids: evidence that a large plasmid is necessary for virulence. Infect Immun. 1981;34(1):75-83.
170. Maneewannakul K, Maneewannakul S, Ippen-Ihler K. Characterization of traX, the F plasmid locus required for acetylation of F-pilin subunits. J Bacteriol. 1995;177(11):2957-64.
171. Yoshioka Y, Ohtsubo H, Ohtsubo E. Repressor gene finO in plasmids R100 and F: constitutive transfer of plasmid F is caused by insertion of IS3 into F finO. J Bacteriol. 1987;169(2):619-23.
172. Dostál L, Shao S, Schildbach JF. Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. Nucleic Acids Res. 2011;39(7):2658-70.
173. Escobar-Páramo P, Giudicelli C, Parsot C, Denamur E. The Evolutionary History of Shigella and Enteroinvasive Escherichia coli Revised. Journal of Molecular Evolution. 2003;57(2):140-8.
174. Organization WH. Disease Outbreak News; Extensively drug-resistant Shigella sonnei infections - Europe. . 2022.

175. Muthuirulandi Sethuvel DP, Devanga Ragupathi NK, Anandan S, Veeraraghavan B. Update on: Shigella new serogroups/serotypes and their antimicrobial resistance. Lett Appl Microbiol. 2017;64(1):8-18.
176. Thanh Duy P, Thi Nguyen TN, Vu Thuy D, Chung The H, Alcock F, Boinett C, et al. Commensal Escherichia coli are a reservoir for the transfer of XDR plasmids into epidemic fluoroquinolone-resistant Shigella sonnei. Nat Microbiol. 2020;5(2):256-64.
177. Pakbin B, Didban A, Brück WM, Alizadeh M. Phylogenetic analysis and antibiotic resistance of Shigella sonnei isolates. FEMS Microbiology Letters. 2022;369(1):fnac042.
178. Chung The H, Baker S. Out of Asia: the independent rise and global spread of fluoroquinolone-resistant Shigella. Microbial Genomics. 2018;4(4).
179. Toro CS, Salazar JC, Montero DA, Ugalde JA, Díaz J, Cádiz LA, et al. Antimicrobial Resistance Dynamics in Chilean Shigella sonnei Strains Within Two Decades: Role of Shigella Resistance Locus Pathogenicity Island and Class 1 and Class 2 Integrons. Front Microbiol. 2021;12:794470.
180. Charles H, Prochazka M, Thorley K, Crewdson A, Greig DR, Jenkins C, et al. Outbreak of sexually transmitted, extensively drug-resistant Shigella sonnei in the UK, 2021-22: a descriptive epidemiological study. Lancet Infect Dis. 2022;22(10):1503-10.
181. Qiu S, Liu K, Yang C, Xiang Y, Min K, Zhu K, et al. A Shigella sonnei clone with extensive drug resistance associated with waterborne outbreaks in China. Nature Communications. 2022;13(1):7365.
182. Ling Z, Yin W, Shen Z, Wang Y, Shen J, Walsh TR. Epidemiology of mobile colistin resistance genes mcr-1 to mcr-9. Journal of Antimicrobial Chemotherapy. 2020;75(11):3087-95.
183. Byrne MK, Miellet S, McGlinn A, Fish J, Meedya S, Reynolds N, et al. The drivers of antibiotic use and misuse: the development and investigation of a theory driven community measure. BMC Public Health. 2019;19(1):1425.
184. Gaurav A, Bakht P, Saini M, Pandey S, Pathania R. Role of bacterial efflux pumps in antibiotic resistance, virulence, and strategies to discover novel efflux pump inhibitors. Microbiology (Reading). 2023;169(5).
185. Bobate S, Mahalle S, Dafale NA, Bajaj A. Emergence of environmental antibiotic resistance: Mechanism, monitoring and management. Environmental Advances. 2023;13:100409.
186. Shad AA, Shad WA. Shigella sonnei: virulence and antibiotic resistance. Arch Microbiol. 2021;203(1):45-58.
187. Chung The H, Boinett C, Pham Thanh D, Jenkins C, Weill F-X, Howden BP, et al. Dissecting the molecular evolution of fluoroquinolone-resistant Shigella sonnei. Nature Communications. 2019;10(1):4828.
188. Azmi IJ, Khajanchi BK, Akter F, Hasan TN, Shahnaij M, Akter M, et al. Fluoroquinolone Resistance Mechanisms of Shigella flexneri Isolated in Bangladesh. PLOS ONE. 2014;9(7):e102533.
189. Miranda A, Ávila B, Díaz P, Rivas L, Bravo K, Astudillo J, et al. Emergence of Plasmid-Borne dfrA14 Trimethoprim Resistance Gene in Shigella sonnei. Frontiers in Cellular and Infection Microbiology. 2016;6.
190. Collis CM, Hall RM. Expression of antibiotic resistance genes in the integrated cassettes of integrons. Antimicrob Agents Chemother. 1995;39(1):155-62.
191. Souque C, Escudero JA, MacLean RC. Integron activity accelerates the evolution of antibiotic resistance. eLife. 2021;10:e62474.
192. Mazel D. Integrons: agents of bacterial evolution. Nat Rev Microbiol. 2006;4(8):608-20.
193. Gillings Michael R. Integrons: Past, Present, and Future. Microbiology and Molecular Biology Reviews. 2014;78(2):257-77.
194. Domingues S, da Silva GJ, Nielsen KM. Integrons: Vehicles and pathways for horizontal dissemination in bacteria. Mob Genet Elements. 2012;2(5):211-23.
195. Fluit AC, Schmitz FJ. Resistance integrons and super-integrons. Clin Microbiol Infect. 2004;10(4):272-88.

196. Deng Y, Bao X, Ji L, Chen L, Liu J, Miao J, et al. Resistance integrons: class 1, 2 and 3 integrons. Ann Clin Microbiol Antimicrob. 2015;14:45.
197. Allué-Guardia A, Koenig SSK, Quirós P, Muniesa M, Bono JL, Eppinger M. Closed Genome and Comparative Phylogenetic Analysis of the Clinical Multidrug Resistant Shigella sonnei Strain 866. Genome Biology and Evolution. 2018;10(9):2241-7.
198. Rodríguez-Beltrán J, DelaFuente J, León-Sampedro R, MacLean RC, San Millán Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. Nature Reviews Microbiology. 2021;19(6):347-59.
199. Coluzzi C, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. Evolution of Plasmid Mobility: Origin and Fate of Conjugative and Nonconjugative Plasmids. Mol Biol Evol. 2022;39(6).
200. Garcillán-Barcia MP, Alvarado A, de la Cruz F. Identification of bacterial plasmids based on mobility and plasmid population biology. FEMS Microbiology Reviews. 2011;35(5):936-56.
201. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EP, de la Cruz F. Mobility of plasmids. Microbiol Mol Biol Rev. 2010;74(3):434-52.
202. Ares-Arroyo M, Coluzzi C, Rocha EPC. Origins of transfer establish networks of functional dependencies for plasmid transfer by conjugation. Nucleic Acids Res. 2023;51(7):3001-16.
203. Zhang J, Xu Y, Wang M, Li X, Liu Z, Kuang D, et al. Mobilizable plasmids drive the spread of antimicrobial resistance genes and virulence genes in Klebsiella pneumoniae. Genome Medicine. 2023;15(1):106.
204. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. Journal of Antimicrobial Chemotherapy. 2018;73(5):1121-37.
205. Orlek A, Phan H, Sheppard AE, Doumith M, Ellington M, Peto T, et al. Ordering the mob: Insights into replicon and MOB typing schemes from analysis of a curated dataset of publicly available plasmids. Plasmid. 2017;91:42-52.
206. Shen Z, Tang CM, Liu G-Y. Towards a better understanding of antimicrobial resistance dissemination: what can be learnt from studying model conjugative plasmids? Military Medical Research. 2022;9(1):3.
207. Amadesi S, Ambretti S, Lazzarotto T, Gaibani P. Complete genome sequence and antimicrobial resistance analysis of ESBL-producing Shigella sonnei carrying small cryptic plasmids isolated in northern Italy. Journal of Global Antimicrobial Resistance. 2023;33:152-4.
208. Choi H, Navarathna D, Harston B, Hwang M, Corona B, San Juan MR, et al. Case of Extensively Drug-Resistant Shigella sonnei Infection, United States. Emerging Infectious Disease journal. 2023;29(8):1708.
209. Teixeira M, Pillay S, Urhan A, Abeel T. SHIP: identifying antimicrobial resistance gene transfer between plasmids. Bioinformatics. 2023;39(10):btad612.
210. Wigley DB. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. Nature Reviews Microbiology. 2013;11(1):9-13.
211. Del Val E, Nasser W, Abaibou H, Reverchon S. RecA and DNA recombination: a review of molecular mechanisms. Biochem Soc Trans. 2019;47(5):1511-31.
212. Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehrauer WM. Biochemistry of homologous recombination in Escherichia coli. Microbiol Rev. 1994;58(3):401-65.
213. Jain K, Wood EA, Romero ZJ, Cox MM. RecA-independent recombination: Dependence on the Escherichia coli RarA protein. Mol Microbiol. 2021;115(6):1122-37.
214. Yang X, Ye L, Chan Edward W-C, Zhang R, Chen S. Tracking Recombination Events That Occur in Conjugative Virulence Plasmid p15WZ-82_Vir during the Transmission Process. mSystems. 2020;5(4):10.1128/msystems.00140-20.
215. Yang X, Xie M, Xu Q, Ye L, Yang C, Dong N, et al. Transmission of pLVPK-like virulence plasmid in Klebsiella pneumoniae mediated by an IncI1 conjugative helper plasmid. iScience. 2022;25(6):104428.

216. McMillan EA, Jackson CR, Frye JG. Transferable Plasmids of Salmonella enterica Associated With Antibiotic Resistance Genes. Front Microbiol. 2020;11:562181.
217. Virolle C, Goldlust K, Djermoun S, Bigot S, Lesterlin C. Plasmid Transfer by Conjugation in Gram-Negative Bacteria: From the Cellular to the Community Level. Genes (Basel). 2020;11(11).
218. Sun D. Pull in and Push Out: Mechanisms of Horizontal Gene Transfer in Bacteria. Front Microbiol. 2018;9:2154.
219. Hasegawa H, Suzuki E, Maeda S. Horizontal Plasmid Transfer by Transformation in Escherichia coli: Environmental Factors and Possible Mechanisms. Front Microbiol. 2018;9:2365.
220. Fillol-Salom A, Alsaadi A, Sousa JAM, Zhong L, Foster KR, Rocha EPC, et al. Bacteriophages benefit from generalized transduction. PLoS Pathog. 2019;15(7):e1007888.
221. Lu J, Wong JJW, Edwards RA, Manchak J, Frost LS, Glover JNM. Structural basis of specific TraD–TraM recognition during F plasmid-mediated bacterial conjugation. Molecular Microbiology. 2008;70(1):89-99.
222. Byrd DR, Matson SW. Nicking by transesterification: the reaction catalysed by a relaxase. Mol Microbiol. 1997;25(6):1011-22.
223. Alvarez-Martinez CE, Christie PJ. Biological diversity of prokaryotic type IV secretion systems. Microbiol Mol Biol Rev. 2009;73(4):775-808.
224. Lawley TD, Klimke WA, Gubbins MJ, Frost LS. F factor conjugation is a true type IV secretion system. FEMS Microbiology Letters. 2003;224(1):1-15.
225. Garcillán-Barcia MP, de la Cruz F. Why is entry exclusion an essential feature of conjugative plasmids? Plasmid. 2008;60(1):1-18.
226. Chan PT, Ohmori H, Tomizawa J, Lebowitz J. Nucleotide sequence and gene organization of ColE1 DNA. J Biol Chem. 1985;260(15):8925-35.
227. Jerome LJ, van Biesen T, Frost LS. Degradation of FinP antisense RNA from F-like plasmids: the RNA-binding protein, FinO, protects FinP from ribonuclease E. J Mol Biol. 1999;285(4):1457-73.
228. Timmis KN, Andrés I, Achtman M. Fertility repression of F-like conjugative plasmids: physical mapping of the R6--5 finO and finP cistrons and identification of the finO protein. Proc Natl Acad Sci U S A. 1978;75(12):5836-40.
229. van Biesen T, Frost LS. The FinO protein of IncF plasmids binds FinP antisense RNA and its target, traJ mRNA, and promotes duplex formation. Mol Microbiol. 1994;14(3):427-36.
230. Koraimann G, Teferle K, Markolin G, Woger W, Högenauer G. The FinOP repressor system of plasmid R1: analysis of the antisense RNA control of traJ expression and conjugative DNA transfer. Mol Microbiol. 1996;21(4):811-21.
231. Frost L, Lee S, Yanchar N, Paranchych W. finP and fisO mutations in FinP anti-sense RNA suggest a model for FinOP action in the repression of bacterial conjugation by the Flac plasmid JCFL0. Mol Gen Genet. 1989;218(1):152-60.
232. Will WR, Frost LS. Characterization of the opposing roles of H-NS and TraJ in transcriptional regulation of the F-plasmid tra operon. J Bacteriol. 2006;188(2):507-14.
233. Alvarado A, Garcillán-Barcia MP, de la Cruz F. A degenerate primer MOB typing (DPMT) method to classify gamma-proteobacterial plasmids in clinical and environmental settings. PLoS One. 2012;7(7):e40438.
234. Garcillán-Barcia MP, Francia MV, de la Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev. 2009;33(3):657-87.
235. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903.
236. Carattoli A, Hasman H. PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). In: de la Cruz F, editor. Horizontal Gene Transfer: Methods and Protocols. New York, NY: Springer US; 2020. p. 285-94.

237. Kopecko DJ, Washington O, Formal SB. Genetic and physical evidence for plasmid control of Shigella sonnei form I cell surface antigen. Infect Immun. 1980;29(1):207-14.
238. Ross K, Varani AM, Snesrud E, Huang H, Alvarenga DO, Zhang J, et al. TnCentral: a Prokaryotic Transposable Element Database and Web Portal for Transposon Analysis. mBio. 2021;12(5):e0206021.
239. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34(Database issue):D32-6.
240. Tenzen T, Ohtsubo E. Preferential transposition of an IS630-associated composite transposon to TA in the 5'-CTAG-3' sequence. J Bacteriol. 1991;173(19):6207-12.
241. Tenzen T, Matsutani S, Ohtsubo E. Site-specific transposition of insertion sequence IS630. J Bacteriol. 1990;172(7):3830-6.
242. Reimmann C, Haas D. The istA gene of insertion sequence IS21 is essential for cleavage at the inner 3' ends of tandemly repeated IS21 elements in vitro. Embo j. 1990;9(12):4055-63.
243. Schmid S, Seitz T, Haas D. Cointegrase, a naturally occurring, truncated form of IS21 transposase, catalyzes replicon fusion rather than simple insertion of IS21. J Mol Biol. 1998;282(3):571-83.
244. Schmid S, Berger B, Haas D. Target joining of duplicated insertion sequence IS21 is assisted by IstB protein in vitro. J Bacteriol. 1999;181(7):2286-9.
245. Reimmann C, Moore R, Little S, Savioz A, Willetts NS, Haas D. Genetic structure, function and regulation of the transposable element IS21. Molecular and General Genetics MGG. 1989;215(3):416-24.
246. Reimmann C, Haas D. Mode of replicon fusion mediated by the duplicated insertion sequence IS21 in Escherichia coli. Genetics. 1987;115(4):619-25.
247. Gamas P, Galas D, Chandler M. DNA sequence at the end of IS1 required for transposition. Nature. 1985;317(6036):458-60.
248. Ton-Hoang B, Turlan C, Chandler M. Functional domains of the IS1 transposase: analysis in vivo and in vitro. Mol Microbiol. 2004;53(5):1529-43.
249. Ohtsubo H, Ohtsubo E. Nucleotide sequence of an insertion element, IS1. Proc Natl Acad Sci U S A. 1978;75(2):615-9.
250. Tavakoli N, Comanducci A, Dodd HM, Lett M-C, Albiger B, Bennett P. IS1294, a DNA Element That Transposes by RC Transposition. Plasmid. 2000;44(1):66-84.
251. Mendiola MV, Bernales I, de la Cruz F. Differential roles of the transposon termini in IS91 transposition. Proc Natl Acad Sci U S A. 1994;91(5):1922-6.
252. Mendiola MV, de la Cruz F. Specificity of insertion of IS91, an insertion sequence present in alpha-haemolysin plasmids of Escherichia coli. Mol Microbiol. 1989;3(7):979-84.
253. Garcillán-Barcia MP, Bernales I, Mendiola MV, de la Cruz F. IS 91 Rolling-Circle Transposition. Mobile DNA II2007. p. 889-904.
254. Toleman MA, Bennett PM, Walsh TR. ISCR elements: novel gene-capturing systems of the 21st century? Microbiol Mol Biol Rev. 2006;70(2):296-316.
255. Garcillán-Barcia MP, de la Cruz F. Distribution of IS91 family insertion sequences in bacterial genomes: evolutionary implications. FEMS Microbiology Ecology. 2002;42(2):303-13.
256. Muñoz-López M, García-Pérez JL. DNA transposons: nature and applications in genomics. Curr Genomics. 2010;11(2):115-28.
257. Wu S, Tian P, Tan T. Genomic landscapes of bacterial transposons and their applications in strain improvement. Applied Microbiology and Biotechnology. 2022;106(19):6383-96.
258. Hall RM, Brookes DE, Stokes HW. Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point. Mol Microbiol. 1991;5(8):1941-59.
259. Hansson K, Sundström L, Pelletier A, Roy PH. IntI2 integron integrase in Tn7. J Bacteriol. 2002;184(6):1712-21.

260. Muthuirulandi Sethuvel DP, Veeraraghavan B, Vasudevan K, Devanga Ragupathi NK, Murugan D, Walia K, et al. Complete genome analysis of clinical Shigella strains reveals plasmid pSS1653 with resistance determinants: a triumph of hybrid approach. Gut Pathogens. 2019;11(1):55.
261. Anderson MC, Vonaesch P, Saffarian A, Marteyn BS, Sansonetti PJ. Shigella sonnei Encodes a Functional T6SS Used for Interbacterial Competition and Niche Occupancy. Cell Host & Microbe. 2017;21(6):769-76.e3.
262. Hood RD, Singh P, Hsu F, Güvener T, Carl MA, Trinidad RR, et al. A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. Cell Host Microbe. 2010;7(1):25-37.
263. Bingle LEH, Bailey CM, Pallen MJ. Type VI secretion: a beginner's guide. Current Opinion in Microbiology. 2008;11(1):3-8.
264. Kapitein N, Mogk A. Deadly syringes: type VI secretion system activities in pathogenicity and interbacterial competition. Curr Opin Microbiol. 2013;16(1):52-8.
265. Leung PB, Matanza XM, Roche B, Ha KP, Cheung HC, Appleyard S, et al. Shigella sonnei utilises colicins during inter-bacterial competition. Microbiology (Reading). 2024;170(2).
266. De Silva PM, Bennett RJ, Kuhn L, Ngondo P, Debande L, Njamkepo E, et al. Escherichia coli killing by epidemiologically successful sublineages of Shigella sonnei is mediated by colicins. EBioMedicine. 2023;97:104822.
267. Riley MA, Wertz JE. Bacteriocin diversity: ecological and evolutionary perspectives. Biochimie. 2002;84(5-6):357-64.
268. Calcuttawala F, Hariharan C, Pazhani GP, Ghosh S, Ramamurthy T. Activity spectrum of colicins produced by Shigella sonnei and genetic mechanism of colicin resistance in conspecific S. sonnei strains and Escherichia coli. Antimicrob Agents Chemother. 2015;59(1):152-8.
269. Cascales E, Buchanan SK, Duché D, Kleanthous C, Lloubès R, Postle K, et al. Colicin biology. Microbiol Mol Biol Rev. 2007;71(1):158-229.
270. Weaver CA, Redborg AH, Konisky J. Plasmid-determined immunity of Escherichia coli K-12 to colicin Ia Is mediated by a plasmid-encoded membrane protein. J Bacteriol. 1981;148(3):817-28.
271. Sidikaro J, Nomura M. E3 immunity substance. A protein from e3-colicinogenic cells that accounts for their immunity to colicin E3. J Biol Chem. 1974;249(2):445-53.
272. Calcuttawala F, Pal A, Nath P, Kar R, Hazra D, Pal R. Structural and functional insights into colicin: a new paradigm in drug discovery. Archives of Microbiology. 2021;204(1):37.
273. Mader A, von Bronk B, Ewald B, Kesel S, Schnetz K, Frey E, et al. Amount of Colicin Release in Escherichia coli Is Regulated by Lysis Gene Expression of the Colicin E2 Operon. PLOS ONE. 2015;10(3):e0119124.
274. Hardy KG, Meynell GG, Dowman JE, Spratt BG. Two major groups of colicin factors: their evolutionary significance. Mol Gen Genet. 1973;125(3):217-30.
275. Hershfield V, Boyer HW, Yanofsky C, Lovett MA, Helinski DR. Plasmid ColEl as a molecular vehicle for cloning and amplification of DNA. Proc Natl Acad Sci U S A. 1974;71(9):3455-9.
276. Hama C, Takizawa T, Moriwaki H, Urasaki Y, Mizobuchi K. Organization of the replication control region of plasmid ColIb-P9. J Bacteriol. 1990;172(4):1983-91.
277. Mankovich JA, Hsu CH, Konisky J. DNA and amino acid sequence analysis of structural and immunity genes of colicins Ia and Ib. J Bacteriol. 1986;168(1):228-36.
278. Asano K, Mizobuchi K. Structural analysis of late intermediate complex formed between plasmid ColIb-P9 Inc RNA and its target RNA. How does a single antisense RNA repress translation of two genes at different rates? J Biol Chem. 2000;275(2):1269-74.
279. Reller LB. Colicin typing as an epidemiological tool in the investigation of outbreaks of Shigella sonnei. Appl Microbiol. 1971;21(1):21-6.
280. Morris GK, Wells JG. Colicin typing of Shigella sonnei. Appl Microbiol. 1974;27(2):312-6.

281. Yang X, Dong N, Chan EW-C, Zhang R, Chen S. Carbapenem Resistance-Encoding and Virulence-Encoding Conjugative Plasmids in Klebsiella pneumoniae. Trends in Microbiology. 2021;29(1):65-83.
282. Xie M, Chen K, Ye L, Yang X, Xu Q, Yang C, et al. Conjugation of Virulence Plasmid in Clinical Klebsiella pneumoniae Strains through Formation of a Fusion Plasmid. Advanced Biosystems. 2020;4(4):1900239.
283. Yang X, Wai-Chi Chan E, Zhang R, Chen S. A conjugative plasmid that augments virulence in Klebsiella pneumoniae. Nature Microbiology. 2019;4(12):2039-43.
284. Liu C, Du P, Yang P, Lu M, Shen N. Fusion plasmid enhanced the endemic extensively drug resistant Klebsiella pneumoniae clone ST147 harbored bla(OXA-48) to acquire the hypervirulence and cause fatal infection. Ann Clin Microbiol Antimicrob. 2023;22(1):11.
285. Kollek R, Oertel W, Goebel W. Isolation and characterization of the minimal fragment required for autonomous replication ("Basic replicon") of a copy mutant (pKN102) of the antibiotic resistance factor R1. Molecular and General Genetics MGG. 1978;162(1):51-7.
286. Silva RM, Saadi S, Maas WK. A basic replicon of virulence-associated plasmids of Shigella spp. and enteroinvasive Escherichia coli is homologous with a basic replicon in plasmids of IncF groups. Infection and immunity. 1988;56(4):836-42.
287. Nordström K. Plasmid R1—Replication and its control. Plasmid. 2006;55(1):1-26.
288. Blomberg P, Nordström K, Wagner EG. Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation. Embo j. 1992;11(7):2675-83.
289. Masai H, Arai K. RepA and DnaA proteins are required for initiation of R1 plasmid replication in vitro and interact with the oriR sequence. Proceedings of the National Academy of Sciences. 1987;84(14):4781-5.
290. Bernander R, Krabbe M, Nordström K. Mapping of the in vivo start site for leading strand DNA synthesis in plasmid R1. Embo j. 1992;11(12):4481-7.
291. Liriano ML, Dilworth RJ, Beuning PJ. DNA Recognition/Processing | DNA Polymerase III, Bacterial☆. In: Jez J, editor. Encyclopedia of Biological Chemistry III (Third Edition). Oxford: Elsevier; 2021. p. 460-71.
292. Łazowski K, Woodgate R, Fijalkowska IJ. Escherichia coli DNA replication: the old model organism still holds many surprises. FEMS Microbiology Reviews. 2024;48(4):fuae018.
293. Burger KJ. Removal of symmetrical sequences from the origin region of plasmid R1 reduces its replication efficiency. Embo j. 1983;2(5):657-62.
294. Masai H, Arai K. Leading strand synthesis of R1 plasmid replication in vitro is primed by primase alone at a specific site downstream of oriR. J Biol Chem. 1989;264(14):8082-90.
295. Hill TM, Pelletier AJ, Tecklenburg ML, Kuempel PL. Identification of the DNA sequence from the E. coli terminus region that halts replication forks. Cell. 1988;55(3):459-66.
296. Krabbe M, Zabielski J, Bernander R, Nordström K. Inactivation of the replication-termination system affects the replication mode and causes unstable maintenance of plasmid R1. Molecular Microbiology. 1997;24(4):723-35.
297. Jameson KH, Rudolph CJ, Hawkins M. Termination of DNA replication at Tus-ter barriers results in under-replication of template DNA. J Biol Chem. 2021;297(6):101409.
298. Light J, Molin S. Post‐transcriptional control of expression of the repA gene of plasmid R1 mediated by a small RNA molecule. The EMBO Journal. 1983;2(1):93-8-8.
299. Riise E, Stougaard P, Bindslev B, Nordström K, Molin S. Molecular cloning and functional characterization of a copy number control gene (copB) of plasmid R1. Journal of Bacteriology. 1982;151(3):1136-45.
300. Riise E, Molin S. Purification and characterization of the CopB replication control protein, and precise mapping of its target site in the R1 plasmid. Plasmid. 1986;15(3):163-71.
301. San Millan A, MacLean RC. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. Microbiol Spectr. 2017;5(5).

302. Tanaka K, Sakai T, Honda Y, Hiasa H, Sakai H, Komano T. Plasmid colIb contains an ssi signal close to the replication origin. Plasmid. 1991;25(2):125-30.
303. Tanaka K, Sakai H, Honda Y, Nakamura T, Higashi A, Komano T. Structural and functional features of cis-acting sequences in the basic replicon of plasmid ColIb-P9. Nucleic Acids Res. 1992;20(11):2705-10.
304. Asano K, Hama C, Inoue S, Moriwaki H, Mizobuchi K. The plasmid ColIb-P9 antisense Inc RNA controls expression of the RepZ replication protein and its positive regulator repY with different mechanisms. J Biol Chem. 1999;274(25):17924-33.
305. Watanabe T, Nishida H, Ogata C, Arai T, Sato S. EPISOME-MEDIATED TRANSFER OF DRUG RESISTANCE IN ENTEROBACTERIACEAE VII. Journal of Bacteriology. 1964;88(3):716-26.
306. Novick RP, Clowes RC, Cohen SN, Curtiss R, 3rd, Datta N, Falkow S. Uniform nomenclature for bacterial plasmids: a proposal. Bacteriol Rev. 1976;40(1):168-89.
307. Johnson TJ, Nolan LK. Plasmid replicon typing. Methods Mol Biol. 2009;551:27-35.
308. Couturier M, Bex F, Bergquist PL, Maas WK. Identification and classification of bacterial plasmids. Microbiol Rev. 1988;52(3):375-95.
309. Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ. Identification of plasmids by PCR-based replicon typing. Journal of Microbiological Methods. 2005;63(3):219-28.
310. Villa L, Carattoli A. Plasmid Typing and Classification. Methods Mol Biol. 2020;2075:309-21.
311. Payne SM, Finkelstein RA. Detection and differentiation of iron-responsive avirulent mutants on Congo red agar. Infect Immun. 1977;18(1):94-8.
312. Liu G, Beaton SE, Grieve AG, Evans R, Rogers M, Strisovsky K, et al. Bacterial rhomboid proteases mediate quality control of orphan membrane proteins. The EMBO Journal. 2020;39(10):e102922.
313. Simon R, Priefer U, Pühler A. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. Bio/Technology. 1983;1(9):784-91.
314. Pelicic V, Reyrat JM, Gicquel B. Expression of the Bacillus subtilis sacB gene confers sucrose sensitivity on mycobacteria. J Bacteriol. 1996;178(4):1197-9.
315. Gay P, Le Coq D, Steinmetz M, Ferrari E, Hoch JA. Cloning structural gene sacB, which codes for exoenzyme levansucrase of Bacillus subtilis: expression of the gene in Escherichia coli. J Bacteriol. 1983;153(3):1424-31.
316. Zhang Y, Buchholz F, Muyrers JP, Stewart AF. A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet. 1998;20(2):123-8.
317. Heringa S, Monroe J, Herrick J. A Simple, Rapid Method for Extracting Large Plasmid DNA from Bacteria. Nature Precedings. 2007.
318. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nature Biotechnology. 2021;39(11):1348-65.
319. Li X, Xie Y, Liu M, Tai C, Sun J, Deng Z, et al. oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. Nucleic Acids Research. 2018;46(W1):W229-W34.
320. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom. 2018;4(8).
321. Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, et al. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. Nucleic acids research. 2024;52(W1):W521-W5.
322. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094-100.
323. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92.

324. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. 2014;1079:155-70.
325. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35(21):4453-5.
326. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology. 2019;37(5):540-6.
327. Chen Z, Erickson DL, Meng J. Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. BMC Genomics. 2020;21(1):631.
328. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
329. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12(1):402.
330. Matson SW, Ragonese H. The F-plasmid TraI protein contains three functional domains required for conjugative DNA strand transfer. J Bacteriol. 2005;187(2):697-706.
331. Cheng Y, McNamara DE, Miley MJ, Nash RP, Redinbo MR. Functional Characterization of the Multidomain F Plasmid TraI Relaxase-Helicase*. Journal of Biological Chemistry. 2011;286(14):12670-82.
332. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27(7):1009-10.
333. Florensa AF, Kaas RS, Clausen P, Aytan-Aktug D, Aarestrup FM. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. Microb Genom. 2022;8(1).
334. Ramírez MS, Quiroga C, Centrón D. Novel rearrangement of a class 2 integron in two non-epidemiologically related isolates of Acinetobacter baumannii. Antimicrob Agents Chemother. 2005;49(12):5179-81.
335. Galata V, Fehlmann T, Backes C, Keller A. PLSDB: a resource of complete bacterial plasmids. Nucleic Acids Res. 2019;47(D1):D195-d202.
336. Schmartz GP, Hartung A, Hirsch P, Kern F, Fehlmann T, Müller R, et al. PLSDB: advancing a comprehensive database of bacterial plasmids. Nucleic Acids Res. 2022;50(D1):D273-d8.
337. Lee H-T, Hase T, Miyoshi J, Shimada K, Masamune Y, Takagi Y. Characterization of a ColE1-Like Plasmid Isolated from Shigella sonnei. The Journal of Biochemistry. 1982;91(4):1147-54.
338. Tomizawa J-i, Som T. Control of cole 1 plasmid replication: Enhancement of binding of RNA I to the primer transcript by the rom protein. Cell. 1984;38(3):871-8.
339. Castagnoli L, Scarpa M, Kokkinidis M, Banner DW, Tsernoglou D, Cesareni G. Genetic and structural analysis of the ColE1 Rop (Rom) protein. Embo j. 1989;8(2):621-9.
340. Hirose T, Shibata H, Higashi M, Hase T, Yamaguchi K, Masamune Y. STRUCTURE AND REPLICATION MECHANISM OF PLASMID pKYM. Journal of pharmacobio-dynamics. 1987;10.
341. Lau BT, Malkus P, Paulsson J. New quantitative methods for measuring plasmid loss rates reveal unexpected stability. Plasmid. 2013;70(3):353-61.
342. Lian ZJ, Phan M-D, Hancock SJ, Nhu NTK, Paterson DL, Schembri MA. Genetic basis of I-complex plasmid stability and conjugation. PLOS Genetics. 2023;19(6):e1010773.
343. Ayele B, Mihret A, Mekonnen Z, Sisay Tessema T, Melaku K, Nassir MF, et al. Whole genome sequencing and antimicrobial resistance among clinical isolates of Shigella sonnei in Addis Ababa, Ethiopia. PLOS ONE. 2024;19(11):e0313310.
344. Bzymek M, Lovett ST. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc Natl Acad Sci U S A. 2001;98(15):8319-25.
345. Furuya N, Nisioka T, Komano T. Nucleotide sequence and functions of the oriT operon in IncI1 plasmid R64. J Bacteriol. 1991;173(7):2231-7.

346. Furuya N, Komano T. Initiation and termination of DNA transfer during conjugation of IncI1 plasmid R64: roles of two sets of inverted repeat sequences within oriT in termination of R64 transfer. J Bacteriol. 2000;182(11):3191-6.
347. Cox MM. The bacterial RecA protein: structure, function, and regulation. In: Aguilera A, Rothstein R, editors. Molecular Genetics of Recombination. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 53-94.
348. Ramsay JP, Kwong SM, Murphy RJ, Yui Eto K, Price KJ, Nguyen QT, et al. An updated view of plasmid conjugation and mobilization in Staphylococcus. Mob Genet Elements. 2016;6(4):e1208317.
349. Kanai Y, Tsuru S, Furusawa C. Insertion sequences: Simple mobile elements with rich ecological and evolutionary structures. Current Opinion in Systems Biology. 2023;36:100481.
350. Babiker A, Lohsen S, Van Riel J, Hjort K, Weiss DS, Andersson DI, et al. Heteroresistance to piperacillin/tazobactam in Klebsiella pneumoniae is mediated by increased copy number of multiple β-lactamase genes. JAC-Antimicrobial Resistance. 2024;6(2):dlae057.
351. Gottlieb PA, Wu S, Zhang X, Tecklenburg M, Kuempel P, Hill TM. Equilibrium, kinetic, and footprinting studies of the Tus-Ter protein-DNA interaction. Journal of Biological Chemistry. 1992;267(11):7434-43.
352. Hill TM, Marians KJ. Escherichia coli Tus protein acts to arrest the progression of DNA replication forks in vitro. Proc Natl Acad Sci U S A. 1990;87(7):2481-5.
353. Locke RK, Greig DR, Jenkins C, Dallman TJ, Cowley LA. Acquisition and loss of CTX-M plasmids in Shigella species associated with MSM transmission in the UK. Microbial Genomics. 2021;7(8).
354. Diaz R, Staudenbauer WL. Origin and direction of mini-R1 plasmid DNA replication in cell extracts of Escherichia coli. Journal of bacteriology. 1982;150(3):1077-84.
355. Goodall DJ, Warecka D, Hawkins M, Rudolph CJ. Interplay between chromosomal architecture and termination of DNA replication in bacteria. Frontiers in Microbiology. 2023;14.
356. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
357. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature Biotechnology. 2011;29(1):24-6.
358. Zhang D, Zhao Y, Feng J, Hu L, Jiang X, Zhan Z, et al. Replicon-Based Typing of IncI-Complex Plasmids, and Comparative Genomics Analysis of IncIγ/K1 Plasmids. Front Microbiol. 2019;10:48.
359. Kollek R, Oertel W, Goebel W. Site-specific deletion at the replication origin of the antibiotic resistance factor R1. Mol Gen Genet. 1980;177(3):413-9.
360. Bruand C, Ehrlich SD, Jannière L. Primosome assembly site in Bacillus subtilis. Embo j. 1995;14(11):2642-50.
361. Binet T, Padiolleau-Lefèvre S, Octave S, Avalle B, Maffucci I. Comparative Study of Single-stranded Oligonucleotides Secondary Structure Prediction Tools. BMC Bioinformatics. 2023;24(1):422.
362. Lorenz R, Wolfinger MT, Tanzer A, Hofacker IL. Predicting RNA secondary structures from sequence and probing data. Methods. 2016;103:86-98.
363. Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:26.
364. Bikard D, Loot C, Baharoglu Z, Mazel D. Folded DNA in action: hairpin formation and biological functions in prokaryotes. Microbiology and molecular biology reviews : MMBR. 2010;74(4):570-88.
365. Voineagu I, Narayanan V, Lobachev KS, Mirkin SM. Replication stalling at unstable inverted repeats: Interplay between DNA hairpins and fork stabilizing proteins. Proceedings of the National Academy of Sciences. 2008;105(29):9936-41.

366. Hacker KJ, Alberts BM. The rapid dissociation of the T4 DNA polymerase holoenzyme when stopped by a DNA hairpin helix. A model for polymerase release following the termination of each Okazaki fragment. Journal of Biological Chemistry. 1994;269(39):24221-8.
367. Lyons-Weiler J, Takahashi K. Branch length heterogeneity leads to nonindependent branch length estimates and can decrease the efficiency of methods of phylogenetic inference. J Mol Evol. 1999;49(3):392-405.
368. Kowalski M, Minka Obama B, Catho G, Dewez JE, Merglen A, Ruef M, et al. Antimicrobial resistance in Enterobacterales infections among children in sub-Saharan Africa: a systematic review and meta-analysis. eClinicalMedicine. 2024;70.
369. Seijas-Pereda L, Rescalvo-Casas C, Hernando-Gozalo M, Angmorkie-Eshun V, Agyei E, Adu-Gyamfi V, et al. The Antimicrobial Resistance (AMR) Rates of Enterobacterales in a Rural Hospital from the Eastern Region, Ghana: A Retrospective Study, 2022. Antibiotics (Basel). 2023;12(8).
370. Tompkins K, Juliano JJ, van Duin D. Antimicrobial Resistance in Enterobacterales and Its Contribution to Sepsis in Sub-saharan Africa. Front Med (Lausanne). 2021;8:615649.
371. Kumar CPG, Bhatnagar T, Sathya Narayanan G, Swathi SS, Sindhuja V, Siromany VA, et al. High-level Colonization With Antibiotic-Resistant Enterobacterales Among Individuals in a Semi-Urban Setting in South India: An Antibiotic Resistance in Communities and Hospitals (ARCH) Study. Clinical Infectious Diseases. 2023;77(Supplement_1):S111-S7.
372. Auguet OT, Niehus R, Gweon HS, Berkley JA, Waichungo J, Njim T, et al. Population-level faecal metagenomic profiling as a tool to predict antimicrobial resistance in Enterobacterales isolates causing invasive infections: An exploratory study across Cambodia, Kenya, and the UK. eClinicalMedicine. 2021;36.
373. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev. 2018;31(4).
374. Castanheira M, Simner PJ, Bradford PA. Extended-spectrum β-lactamases: an update on their characteristics, epidemiology and detection. JAC-Antimicrobial Resistance. 2021;3(3):dlab092.
375. Husna A, Rahman MM, Badruzzaman ATM, Sikder MH, Islam MR, Rahman MT, et al. Extended-Spectrum β-Lactamases (ESBL): Challenges and Opportunities. Biomedicines [Internet]. 2023; 11(11).
376. Li Q, Chang W, Zhang H, Hu D, Wang X. The Role of Plasmids in the Multiple Antibiotic Resistance Transfer in ESBLs-Producing Escherichia coli Isolated From Wastewater Treatment Plants. Frontiers in Microbiology. 2019;10.
377. Shropshire WC, Konovalova A, McDaneld P, Gohel M, Strope B, Sahasrabhojane P, et al. Systematic Analysis of Mobile Genetic Elements Mediating β-Lactamase Gene Amplification in Noncarbapenemase-Producing Carbapenem-Resistant Enterobacterales Bloodstream Infections. mSystems. 2022;7(5):e00476-22.
378. Ramsamy Y, Mlisana KP, Amoako DG, Abia ALK, Ismail A, Allam M, et al. Mobile genetic elements-mediated Enterobacterales-associated carbapenemase antibiotic resistance genes propagation between the environment and humans: A One Health South African study. Science of The Total Environment. 2022;806:150641.
379. Acman M, Wang R, van Dorp L, Shaw LP, Wang Q, Luhmann N, et al. Role of mobile genetic elements in the global dissemination of the carbapenem resistance gene blaNDM. Nature Communications. 2022;13(1):1131.
380. Walas N, Slown S, Amato HK, Lloyd T, Bender M, Varghese V, et al. The role of plasmids in carbapenem resistant E. coli in Alameda County, California. BMC Microbiology. 2023;23(1):147.
381. Tohmaz M, Askari Badouei M, Kalateh Rahmani H, Hashemi Tabar G. Antimicrobial resistance, virulence associated genes and phylogenetic background versus plasmid replicon types: the possible associations in avian pathogenic Escherichia coli (APEC). BMC Veterinary Research. 2022;18(1):421.

382. Xu Y, Zhang J, Wang M, Liu M, Liu G, Qu H, et al. Mobilization of the nonconjugative virulence plasmid from hypervirulent Klebsiella pneumoniae. Genome Medicine. 2021;13(1):119.
383. Ramirez Maria S, Traglia German M, Lin David L, Tran T, Tolmasky Marcelo E. Plasmid-Mediated Antibiotic Resistance and Virulence in Gram-Negatives: the Klebsiella pneumoniae Paradigm. Microbiology Spectrum. 2014;2(5):10.1128/microbiolspec.plas-0016-2013.
384. Khajanchi BK, Foley SL. Antimicrobial Resistance and Increased Virulence of Salmonella. Microorganisms. 2022;10(9).
385. Lesic B, Zouine M, Ducos-Galand M, Huon C, Rosso ML, Prévost MC, et al. A natural system of chromosome transfer in Yersinia pseudotuberculosis. PLoS Genet. 2012;8(3):e1002529.
386. Hall JPJ, Wright RCT, Harrison E, Muddiman KJ, Wood AJ, Paterson S, et al. Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by compensatory mutation. PLOS Biology. 2021;19(10):e3001225.
387. Nair Ramith R, Andersson Dan I, Warsi Omar M. Antibiotic resistance begets more resistance: chromosomal resistance mutations mitigate fitness costs conferred by multi-resistant clinical plasmids. Microbiology Spectrum. 2024;12(5):e04206-23.
388. Elfadadny A, Ragab RF, AlHarbi M, Badshah F, Ibáñez-Arancibia E, Farag A, et al. Antimicrobial resistance of Pseudomonas aeruginosa: navigating clinical impacts, current resistance trends, and innovations in breaking therapies. Frontiers in Microbiology. 2024;15.
389. Zhao Y, Xu H, Wang H, Wang P, Chen S. Multidrug resistance in Pseudomonas aeruginosa: genetic control mechanisms and therapeutic advances. Molecular Biomedicine. 2024;5(1):62.
390. Sansonetti PJ, Kopecko DJ, Formal SB. Involvement of a plasmid in the invasive ability of Shigella flexneri. Infect Immun. 1982;35(3):852-60.
391. Brito IL. Examining horizontal gene transfer in microbial communities. Nature Reviews Microbiology. 2021;19(7):442-53.
392. Didelot X, Maiden MC. Impact of recombination on bacterial evolution. Trends Microbiol. 2010;18(7):315-22.
393. Rozwandowicz M, Hordijk J, Bossers A, Zomer AL, Wagenaar JA, Mevius DJ, et al. Incompatibility and phylogenetic relationship of I-complex plasmids. Plasmid. 2020;109:102502.
394. Lilly J, Camps M. Mechanisms of Theta Plasmid Replication. Microbiology Spectrum. 2015;3(1):10.1128/microbiolspec.plas-0029-2014.
395. Kim JW, Bugata V, Cortés-Cortés G, Quevedo-Martínez G, Camps M. Mechanisms of Theta Plasmid Replication in Enterobacteria and Implications for Adaptation to Its Host. EcoSal Plus. 2020;9(1).
396. Meyer RJ, Helinski DR. Unidirectional replication of the P-group plasmid RK2. Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis. 1977;478(1):109-13.
397. Schmidhauser TJ, Helinski DR. Regions of broad-host-range plasmid RK2 involved in replication and stable maintenance in nine species of gram-negative bacteria. J Bacteriol. 1985;164(1):446-55.
398. Manen D, Caro L. The replication of plasmid pSC101. Mol Microbiol. 1991;5(2):233-7.
399. Lovett MA, Sparks RB, Helinski DR. Bidirectional replication of plasmid R6K DNA in Escherichia coli; correspondence between origin of replication and position of single-strand break in relaxed complex. Proceedings of the National Academy of Sciences. 1975;72(8):2905-9.
400. Rakowski SA, Filutowicz M. Plasmid R6K replication control. Plasmid. 2013;69(3):231-42.
401. Tao S, Chen H, Li N, Liang W. The Application of the CRISPR-Cas System in Antibiotic Resistance. Infect Drug Resist. 2022;15:4155-68.
402. Derollez E, Lesterlin C, Bigot S. Design, potential and limitations of conjugation-based antibacterial strategies. Microbial Biotechnology. 2024;17(11):e70050.
403. Šubrtová Salmonová H, Bunesova V. Methods of Studying Diversity of Bacterial Comunities: A Review. Scientia Agriculturae Bohemica. 2017;48.

404. Alderliesten JB, Duxbury SJN, Zwart MP, de Visser JAGM, Stegeman A, Fischer EAJ. Effect of donor-recipient relatedness on the plasmid conjugation frequency: a meta-analysis. BMC Microbiology. 2020;20(1):135.
405. Huisman JS, Benz F, Duxbury SJN, de Visser JAGM, Hall AR, Fischer EAJ, et al. Estimating plasmid conjugation rates: A new computational tool and a critical comparison of methods. Plasmid. 2022;121:102627.
406. Igler C, Schwyter L, Gehrig D, Wendling CC. Conjugative plasmid transfer is limited by prophages but can be overcome by high conjugation rates. Philosophical Transactions of the Royal Society B: Biological Sciences. 2021;377(1842):20200470.
407. Couturier A, Virolle C, Goldlust K, Berne-Dedieu A, Reuter A, Nolivos S, et al. Real-time visualisation of the intracellular dynamics of conjugative plasmid transfer. Nat Commun. 2023;14(1):294.
408. Stockdale SR, Hill C. Incorporating plasmid biology and metagenomics into a holistic model of the human gut microbiome. Current Opinion in Microbiology. 2023;73:102307.
409. Cyriaque V, Ibarra-Chávez R, Kuchina A, Seelig G, Nesme J, Madsen JS. Single-cell RNA sequencing reveals plasmid constrains bacterial population heterogeneity and identifies a non-conjugating subpopulation. Nature Communications. 2024;15(1):5853.
410. Kypr J, Kejnovská I, Renciuk D, Vorlícková M. Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Res. 2009;37(6):1713-25.
411. Vorlíčková M, Kejnovská I, Bednářová K, Renčiuk D, Kypr J. Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. Chirality. 2012;24(9):691-8.
412. Kowalczyk L, Rajewska M, Konieczny I. Positioning and the specific sequence of each 13-mer motif are critical for activity of the plasmid RK2 replication origin. Molecular Microbiology. 2005;57(5):1439-49.
413. Crivat G, Taraska JW. Imaging proteins inside cells with fluorescent tags. Trends Biotechnol. 2012;30(1):8-16.
414. Al-Khafaji NSK, Almjalawi BSA, Ewadh RMJ, Al-Dahmoshi HOM, Abed SY, Nasrolahi A, et al. Prevalence of plasmid-mediated quinolone resistance genes and biofilm formation in different species of quinolone-resistant clinical Shigella isolates: a cross-sectional study. European Journal of Medical Research. 2024;29(1):419.
415. Jacoby George A, Strahilevitz J, Hooper David C. Plasmid-Mediated Quinolone Resistance. Microbiology Spectrum. 2014;2(5):10.1128/microbiolspec.plas-0006-2013.
416. Hata M, Suzuki M, Matsumoto M, Takahashi M, Sato K, Ibe S, et al. Cloning of a novel gene for quinolone resistance from a transferable plasmid in Shigella flexneri 2b. Antimicrob Agents Chemother. 2005;49(2):801-3.
417. Baker KS, Dallman TJ, Ashton PM, Day M, Hughes G, Crook PD, et al. Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: A cross-sectional study. The Lancet Infectious Diseases. 2015;15(8):913-21.
418. Kim S, Park AK, Kim JS, Park J, Shin E, Jung HJ, et al. The role of international travellers in the spread of CTX-M-15-producing Shigella sonnei in the Republic of Korea. Journal of Global Antimicrobial Resistance. 2019;18:298-303.
419. Sethuvel DPM, Anandan S, Ragupathi NKD, Gajendiran R, Kuroda M, Shibayama K, et al. IncFII plasmid carrying antimicrobial resistance genes in Shigella flexneri: Vehicle for dissemination. Journal of Global Antimicrobial Resistance. 2019;16:215-9.
420. Baker KS, Dallman TJ, Field N, Childs T, Mitchell H, Day M, et al. Genomic epidemiology of Shigella in the United Kingdom shows transmission of pathogen sublineages and determinants of antimicrobial resistance. Scientific Reports. 2018;8(1):7389-.
421. Kim JS, Kim S, Park J, Shin E, Yun YS, yong Lee D, et al. Plasmid-mediated transfer of CTX-M-55 extended-spectrum beta-lactamase among different strains of Salmonella and Shigella spp. in the Republic of Korea. Diagnostic Microbiology and Infectious Disease. 2017;89(1):86-8.

422. Chiu CH, Lee JJ, Wang MH, Chu C. Genetic analysis and plasmid-mediated blaCMY-2 in Salmonella and Shigella and the Ceftriaxone Susceptibility regulated by the ISEcp-1 tnpA-blaCMY-2-blc-sugE. Journal of Microbiology, Immunology and Infection. 2020.
423. Tietze E, Brevet J. Nucleotide sequence of the streptothricin-acetyl-transferase gene sat-2. Nucleic Acids Res. 1990;18(5):1283.
424. Ranjbar R, Farahani A. Shigella: Antibiotic-resistance mechanisms and new horizons for treatment. Dove Medical Press Ltd.; 2019. p. 3137-67.