Plasticity and maintenance of the virulence plasmid in *Shigella flexneri* and *Shigella sonnei*



Giulia Pilla

Sir William Dunn School of Pathology
Lincoln College
University of Oxford

A thesis in partial fulfilment of the requirements for the degree of Doctor of Philosophy
Trinity Term 2020

Plasticity and maintenance of the virulence plasmid in *Shigella flexneri* and *Shigella sonnei*

Giulia Pilla, DPhil Pathology, Trinity Term 2020, Lincoln College

Abstract

Shigella is the major cause of human bacillary dysentery worldwide. *Shigella flexneri* and *Shigella sonnei* account for >90% of shigellosis globally, and evolved from commensal *Escherichia coli* following acquisition of a large plasmid, pINV. This plasmid is essential for virulence as it harbours genes encoding a Type three secretion system (T3SS) on a 32 kb pathogenicity island (T3SS PAI). pINV is present in the bacterium in low copy number, and possesses multiple partitioning systems and toxin-antitoxin (TA) systems whose role in plasmid maintenance is unclear. Furthermore, during growth in the laboratory both *S. flexneri* and *S. sonnei* spontaneously lose T3SS expression, becoming avirulent.

In this work I evaluated how insertion sequences (IS), TA systems and partitioning systems influence the maintenance of virulence in *S. flexneri* and *S. sonnei*. I demonstrate that in *S. flexneri* avirulence mainly results from discrete deletions involving the T3SS PAI mediated by insertion sequences (IS). Repositioning of *vapBC*, which encodes a TA system essential for pINV maintenance, near the T3SS PAI prevents loss of the PAI, indicating that VapBC maintains the entire plasmid as well as local sequences. IS also mediate reversible integration of pINV into the chromosome, which results in down-regulation of T3SS expression, and reversible avirulence. In contrast, *S. sonnei* becomes avirulent by mainly losing pINV, which I found is caused by the absence of the two TA systems present in *S. flexneri*, CcdAB and GmvAT, as well as amino acid substitutions in VapBC. The contribution of the partitioning system ParAB on pINV maintenance is influenced by bacterial growth conditions and the presence of TA systems. However, the function of the second partitioning system StbAB is unknown. The absence of this system in *S. sonnei* pINV suggests that the two partitioning systems may be redundant.

Finally, I have designed and evaluated CRISPR-Cas to induce the loss of virulence or reduced viability in *Shigella* by targeting T3SS genes and *vapBC*, thus, validating a novel approach to treat shigellosis.

Acknowledgements

The list of people who I would like to acknowledge is probably very long and words are not enough to thank all of them, but I will try to be concise.

First of all, I would like to express my infinite gratitude to my supervisor as well as mentor Prof. Christoph Tang. Five years ago, I would have never imagined that a six-month placement would have become a stepping-stone for a DPhil in such prestigious lab and university. I would not have been here without his guidance and precious advices, inside and outside the lab. He made my dream come true.

I am also thankful to all members (past and present) of the Tang lab, who became my second family. Dr Gareth McVicker, who started part of this work before I joined the group and has been an amazing teacher during the first two years of my DPhil. Dr Guangyu Liu, who has been an incredible source of cloning tricks. Dr Jessica Martyn, Dr Emeline Lawarée, and Dr Mariya Lobanovska, who became my sisters and have been inestimable colleagues.

Outside the lab, I would like to thank several people, who assisted me in different ways. Lucia, the perfect housemate, always supportive and caring. Federico, for the routine-breaking coffee chats and his inspiring DPhil/career advices. Valentina, my favourite gym friend and confidant. Lusiana, the companion on last-minute dinners. And Chiara, who has always supported me from distance with energy and affection.

I am especially thankful to my family, who has taught me that there is no goal impossible to be achieved. Last but not least, I would like to thank my better half, Lorenzo, who has always been there for me despite the distances and has been an incredible support especially in the last few difficult months.

Abbreviations

bp	Base pair
CFU	Colony forming unit
CR	Congo red
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
fMet	Formylated methionine
gDNA	Genomic deoxyribonucleic acid
GFP	Green fluorescent protein
kb	Kilobase
nt	Nucleotide
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
pINV	Invasion plasmid
OD600	Optical density measured at absorbance of 600 nm
RNA	Ribonucleic acid
RNase	Ribonuclease
tRNA	Transfer ribonucleic acid
T3SS	Type 3 Secretion System

Table of Contents

1 LIST OF FIGURES IX

2 LIST OF TABLES XI

3 LIST OF APPENDICES XI

1 INTRODUCTION 1

1.1 *Shigella* 2

1.1.1 Microbiology 2

1.1.2 Epidemiology 3

1.1.2.1 The curious case of *S. sonnei* 4

1.1.3 Evolution 5

1.1.4 Symptoms 9

1.1.5 Pathogenesis 9

1.1.5.1 Gaining access to the intestinal epithelium 10

1.1.5.2 Crossing the intestinal epithelium and survival inside the phagocytes 11

1.1.5.3 Invasion, replication and spread across the epithelium 12

1.1.5.4 Role of polymorphonuclear neutrophils 15

1.1.5.5 Manipulation of host immune responses 16

1.1.6 Treatment of shigellosis, antimicrobial resistance and *Shigella* vaccines 17

1.2 *Shigella*'s virulence 19

1.2.1 The large virulence plasmid in *Shigella* spp 19

1.2.1.1 Evolution and phylogenesis 20

1.2.1.2 Virulence determinants of pINV 22

1.2.1.3 Maintenance of pINV 24

1.2.1.3.1 Replication of pINV 25

1.2.1.3.2 Partitioning systems 27

1.2.1.3.3 Toxin-antitoxin systems 32

1.2.1.4 Insertion Sequences 40

1.2.2 Chromosomal virulence genes 47

1.2.3 Regulation of virulence 47

1.3 CRISPR-Cas systems and *Shigella* 51

1.3.1 Activity of CRISPR-Cas systems 51

1.3.2 CRISPR-Cas systems in *Shigella* 53

1.4 Aims 54

2 MATERIALS AND METHODS 56

2.1 Growth media and strains 56

2.2 Construction of strains 58

2.3 Tri-parental mating 66

2.4 CR-binding assay 67

2.5 Virulence plasmid stability assay 68

2.6 Multiplex PCR 68

2.7 Genome sequencing 69

2.8 Purification and visualisation of pINV 69

2.9 Protein Secretion 70

2.10 RNA extraction and quantitative real-time PCR 70

2.11 Quantitative PCR analysis to measure plasmid copy number 71

2.12 Viability assay 72

2.13 Statistical analyses 72

3 RESULTS: ANALYSIS OF THE STABILITY OF PINV IN *S. FLEXNERI* 74

3.1 The role of TA systems 74

3.1.1 VapBC efficiently maintains pINV 75

3.1.2 VapBC maintains nearby sequences 78

3.2 The role of Insertion Sequences 80

3.2.1 IS and homologous recombination pathways are involved in the deletion of the T3SS PAI 81

3.2.2 IS mediate reversible integration of pINV into the chromosome 85

3.3 The role of partitioning systems in pINV maintenance 89

3.3.1 Contribution of single partitioning systems on pINV maintenance 90

3.3.2 Investigating pINV copy number as a potential factor influencing the effect of ParAB 94

3.4 Summary 96

4 RESULTS: ANALYSIS OF THE STABILITY OF PINV IN *S. SONNEI* 99

4.1 The role of TA systems 99

4.1.1 The absence of *ccdAB* and *gmvAT* contributes to pINV loss 100

4.1.2 Amino acid substitutions in VapBC contribute to pINV stability 102

4.2 The role of Insertion Sequences 107

4.2.1 IS are involved in the deletion of the T3SS PAI in *S. sonnei* 107

4.3 Summary 110

5 RESULTS: EVALUATION OF CRISPR-CAS GENE EDITING SYSTEM TO INDUCE LOSS OF VIRULENCE IN *SHIGELLA* 113

5.1 Testing *virB* and *vapBC* as potential CRISPR-Cas9 targets to induce loss of virulence in *Shigella* 113

5.1.1 Targeting *virB* by CRISPR-Cas9 induces deletion of the T3SS PAI 115

5.1.2 Targeting of *vapBC* by CRISPR-Cas9 leads to distinct events 118

5.2 Inducible CRISPR-Cas9 targeting *vapBC* 121

5.3 Summary 124

6 DISCUSSION 129

6.1 TA systems as plasmid guardians and potential clues of plasmid evolution 130

6.2 The double-edged sword of IS on pINV 133

6.3 Partitioning systems: two better than one? 137

6.4 CRISPR-Cas9 as a novel tool to treat shigellosis and investigate PSK 140

7 APPENDIX 143

8 BIBLIOGRAPHY 156

1 List of Figures

Figure 1.1 Summary of the evolution of *Shigella* spp. 6

Figure 1.2 *Shigella* pathogenesis. 10

Figure 1.3 *S. flexneri* virulence plasmid pWR501. 20

Figure 1.4 Genetic map of the T3SS PAI of the *S. flexneri* pWR501. 22

Figure 1.5 Architecture of T3SS in *Shigella* 23

Figure 1.6 Genetic structure and function of the replicon RepFIIA. 26

Figure 1.7 Structure of type I and type II *par* loci and their models of mechanisms. 28

Figure 1.8 Post Segregational Killing (PSK) mediated by plasmid-encoded type II TA systems and regulation of TA operons. 33

Figure 1.9 Structure of IS and transposition 41

Figure 1.10 Regulation of virulence in *Shigella*. 49

Figure 1.11 Schematic of CRISPR-Cas9. 52

Figure 3.1 VapBC is essential for maintenance of the entire plasmid at 37°C. 76

Figure 3.2 VapBC exerts protection of nearby sequences. 79

Figure 3.3 Deletion of the T3SS PAI is mediated by least four different pairs of IS. 82

Figure 3.4 The RecBCD and RecFOR systems influence rates of T3SS PAI-related deletions. . 84

Figure 3.5 Distinct copies of IS*1294* mediate chromosomal integration of pINV, inducing a down-regulation of the expression of the T3SS PAI. 87

Figure 3.6 Excision of pINV is sufficient to restore T3SS activity. 88

Figure 3.7 The effect of ParAB and StbAB on the stability of the *S. flexneri* pINV in the presence or absence of the three TA systems. 91

Figure 3.8 Deletion of *parAB* stabilises pINV in *S. flexneri*, when grown either in M9 or in the absence of *virB*. 92

Figure 3.9 Plasmid copy number changes during growth in minimal medium. 94

Figure 4.1 CR$^-$ colonies of *S. sonnei* mostly result from loss of entire pINV. 100

Figure 4.2 VapC polymorphisms affect plasmid maintenance. 104

Figure 4.3 Loss of *S. sonnei* pINV is eliminated in a *S. flexneri* background. 105

Figure 4.4 Schematic of the sites of T3SS PAI-related deletions in *S. sonnei* 53G. 107

Figure 4.5 Distinct profile of IS on pINV from *S. flexneri* M90T and *S. sonnei* 53G. 108

Figure 5.1 Schematic of pCas9. 114

Figure 5.2 CRISPR-Cas9 system targeting *virB* leads to deletion of the T3SS PAI. 116

Figure 5.3 CRISPR-Cas9 system targeting the promoter of *vapBC* leads to plasmid loss while targeting *vapB* reduces bacterial viability. 120

Figure 5.4 Inducible CRISPR-Cas9 system targeting different components of *vapBC* reduces bacterial growth after 2 hours of induction. 122

Figure 6.1 Hypothetical model on the variation of the pINV copy number in *Shigella*. 139

2 List of Tables

Table 2.1 List of strains and their relative genotype. 56

Table 2.2 List of growth media 58

Table 2.3 List of plasmids 60

Table 2.4 List of oligos 62

3 List of Appendices

Appendix Figure 7.1 *vapBC*^{WT} is comparable to M90T. 143

Appendix Figure 7.2 Loss of CR binding is reduced at 21°C. 144

Appendix Figure 7.3 Emergence of CR^{-} colonies in *S. flexneri* results from spontaneous loss of the PAI. 145

Appendix Figure 7.4 CR^{-} isolates with wild-type multiplex PCR profiles show intact plasmid sequences in *S. flexneri*. 146

Appendix Figure 7.5 Schematic of the location of IS*1294* copies on pINV and on the chromosome of *S. flexneri* M90T. 147

Appendix Figure 7.6 Aligment of ParABS elements from *S. flexneri* pINV and P1 plasmid. .. 148

Appendix Figure 7.7 Aligment of StbABC elements from *S. flexneri* pINV with ParMRC elements from R1 plasmid. 149

Appendix Figure 7.8 Atomic structure of *S. sonnei* VapBC. 150

Appendix Figure 7.9 *S. sonnei* CS14 possesses a highly stable pINV. 151

Appendix Figure 7.10 Spacers and pCas9 plasmids designed to target *virB* and *vapBC*. 152

Appendix Figure 7.11 CRISPR-Cas9 systems targeting *vapBC* are inducible in *S. sonnei*. 153

Appendix Table 7.1 Frequency of T3SS PAI deletion variants, and alignment coverage, identity and GC content of the IS involved. 154

Appendix Table 7.2 Amino acid or nucleotide identity of ParABS and StbABC systems between *S. flexneri* and other *Shigella* spp.. 155

1 INTRODUCTION

Bacterial pathogens are the major cause of diarrheal disease both in children and the elderly (Collaborators, 2018). Among this group of organisms, *Shigella* is the most important genus, responsible for around 270 million cases and approximately 212,500 deaths globally every year, mostly in children under five and adults over 70 years of age (Collaborators, 2018). Shigellosis is endemic particularly in children living in low- and middle-income countries (LMICs), such as those in sub-Saharan Africa and South-East Asia, where lack of access to clean water and healthcare, inadequate sanitation and poor nutrition facilitate bacterial transmission and the development of disease (Kotloff *et al*., 2018; Collaborators, 2018).

This pathogen invades into and replicates within the intestinal epithelium, where it triggers inflammation in the sub-epithelial tissue and destruction of the intestinal mucosa (Schroeder and Hilbi, 2008). The symptoms of shigellosis can vary from mild watery diarrhoea to severe dysentery (Kotloff *et al*., 2018). WHO guidelines recommend that all individuals with shigellosis receive antibiotics (WHO, 2005); however, this therapy is becoming progressively more ineffective due to the increasing rise in antimicrobial resistance (AMR) (Baker and Chung The, 2018). Furthermore, there is still no vaccine candidate available against *Shigella*, so there is a pressing need to develop alternative strategies to prevent and treat shigellosis.

1.1 *Shigella*

The microbiologist Kyoshi Shiga is credited as the first to identify and characterise the bacterial cause of dysentery, by investigating an epidemic in Japan that caused more than 25,000 deaths during 1896 (Shiga, 1898). Initially called *Bacillus dysentericus*, the bacterium was renamed in his honour, Shiga bacillus or *Shigella,* and later as *Shigella dysenteriae*. In the following years, Simon Flexner isolated *Shigella flexneri* (Flexner, 1900) and Carl Sonn identified the causative agent of dysentery in Denmark, *Shigella sonnei* (Sonne, 1915). In the late 1930s, J.S.K. Boyd found another species in Bangalore, *Shigella boydii* (Boyd, 1938). Since its discovery, *Shigella* has been widely studied and several aspects have been characterised, such as its biology, epidemiology, evolution and pathogenesis.

1.1.1 Microbiology

Shigella is a Gram-negative bacterium of the Enterobacteriaceae family. It is facultative anaerobic, non-motile, $\sim$ 1 to 3 μm long bacillus (Schroeder and Hilibi, 2008). The genus *Shigella* is divided into four species, dependent upon biochemical, serological and clinical differences (Ewing *et al*., 1949): *S. flexneri*, *S. sonnei*, *S. dysenteriae* and *S. boydii*. *Shigella* spp. do not decarboxylate lysine (Maurelli *et al*., 1998) and are non-lactose fermenters, except for *S. sonnei*, which is a late lactose fermenter (Lan *et al*., 2002). According to the structure of their lipopolysaccharide (LPS) O antigen, each species can be divided into different serotypes (Hale and Keusch, 1996). *S. dysenteriae* (serogroup A) consists of 14 serotypes, while *S. flexneri* (serogroup B) has been classified into 15 serotypes, including several subtypes. *S. boydii* (serogroup C) is composed of 19 serotypes, while *S. sonnei* (serogroup D), is the most antigenically homogenous species, and consists of a single serotype (Kotloff *et al*., 2018).

1.1.2 Epidemiology

Besides the serotype classification, *Shigella* spp. also have particular geographic associations, which correlate with the level of economic development and poverty (Levine *et al*., 2007). *S. flexneri* and *S. sonnei* together account for ~90% of all shigellosis. *S. flexneri* is endemic in LMICs, causing over 60% of infections worldwide, while *S. sonnei* is the leading cause of endemic shigellosis in high-income countries, and is responsible for approximately 35% of cases (Livio *et al*., 2014). In contrast, *S. boydii* and *S. dysenteriae* infrequently cause dysentery, each representing only ~5% of all cases of shigellosis. *S. dysenteriae* is responsible for epidemic disease and causes the most severe disease and the majority of fatal cases (Kotloff *et al*., 1999). *S. dysenteriae* type 1 was the first *Shigella* strain found to produce the Shiga toxin (Stx), which leads to haemorrhagic colitis and serious complications (Lindberg *et al*., 1987; Gray *et al*., 2014), and has caused epidemics over the last 40 years in central America, south Asia, central and east Africa (Kotloff *et al*., 2018). *S. boydii* is largely restricted to the Indian subcontinent (Levine *et al*., 2007).

Shigella is a human specific pathogen and is transmitted by the fecal-oral route through ingestion of contaminated water or food (Kotloff *et al*., 2018). *Shigella* is able to survive in water (Ghosh and Sehgal, 2001) and its transmission is promoted by inadequate sanitation and poor hygiene. As ingestion of only 10 to 180 colony-forming units (CFU) of *S. dysenteriae* 1 and *S. flexneri*/*S. sonnei* can cause a symptomatic infection in healthy human volunteers (DuPont *et al*., 1989), *Shigella* is frequently spread by person-to-person contact (Kotloff, 2018).

Children under five years old are the population most at risk, resulting in around 64,000 deaths in 2016 (Collaborators, 2018). *Shigella* is the first and second most common aetiological cause of moderate-to-severe diarrhoeal disease among children aged 24–59 months and 12–23 months, respectively, living in sub-Saharan Africa and South East Asia (Kotloff *et al*., 2013). Another group of individuals that are particularly susceptible to *Shigella* are men who have sex with men (MSM). Behaviours such as unprotected sex, oro-anal contact, and multiple sexual partners are thought to favour transmission of *Shigella* in the MSM population (Gilbart *et al*., 2015). Furthermore, shigellosis is associated with HIV infection, which is thought to facilitate *Shigella* transmission, by impairing immunological responses to the bacterium, increasing the severity and duration of disease, as well as susceptibility to re-infection with the same serotype (Baker *et al*., 2015; Mohan *et al*., 2018). Finally, *Shigella* can also be acquired during travel to endemic countries (Gray *et al*., 2015). In fact, approximately 2–9% travellers' diarrhoea is estimated to be caused by *Shigella* (Saha *et al*., 2009). In the recent years, this has become a major issue, as most imported *Shigella* is multi-drug resistant, therefore, introducing AMR into new populations (Bowen *et al*., 2016).

1.1.2.1 The curious case of *S. sonnei*

During the 20th century, *S. sonnei* disseminated globally from Europe, leading to localised transmission (Holt *et al*., 2012). With rapid economic development, *S. sonnei* has become prevalent in industrialising regions in Asia, Latin America and the Middle East (Thompson *et al*., 2015). It is hypothesised that this phenomenon depends on three factors. First, as living standards improve, there is more access to clean drinking water, so individuals are not exposed to the bacterium *Plesiomonas shigelloides*, which possesses an identical O-antigen as *S. sonnei*, and do not develop cross-reactive natural immunity against *S. sonnei* (Sack *et al*., 1994). Secondly, *S. sonnei* can be phagocytosed and survive in the ubiquitous, free-living

amoeba *Acanthamoebae castellanii*, enabling the bacterium to survive sanitation processes, such as chlorination (Jeong *et al*., 2007). *S. flexneri* can also be phagocytosed by A. *castellanii*, but inhibits growth of the amoeba, reducing its ability to act as a vector (Saeed *et al*., 2012). Finally, *S. sonnei* has rapidly acquired antimicrobial resistance *via* mobile genetic elements, which allayed to the strong selective pressure from increasing antimicrobial use in industrialised areas and has contributed to the spread of antimicrobial resistant *S. sonnei* (Holt *et al*., 2012).

Another characteristic that differentiates *S. sonnei* from other *Shigella* spp. is its mode of transmission. In high-income settings, *S. sonnei* is mainly spread by direct person-to-person contact, commonly amongst groups of individuals living in close proximity, such as children in day care, MSM, nursing home patients and institutionalised people (Mahoney *et al*., 1993; Mohle-Boetani *et al*., 1995; McCall *et al*., 2000; Genobile *et al*., 2004; Baker *et al*., 2015; Simms *et al*., 2015).

1.1.3 Evolution

Phylogenetic relationships derived by comparison of whole genome sequences have shown that *Shigella* spp. evolved on four occasions from different *Escherichia coli* ancestors between 35,000 and 270,000 years ago (Lan and Reeves, 2002). The evolution of *Shigella* provides a clear example of convergent evolution, with *Shigella* founders independently acquired genes that conferred virulence on commensal bacteria (Zhang *et al*., 2012). Moreover, the similarities are evident when *Shigella* is compared with enteroinvasive *E. coli* (EIEC). EIEC possesses some biochemical characteristics of *E. coli*, while at the same time harbouring *Shigella*'s major virulence element, the large virulence plasmid, pINV. Recent sequence analysis revealed that EIEC is more closely related to *Shigella* than to non-pathogenic *E. coli*

(van den Beld *et al*., 2012). These studies, with other phylogenetic analyses (Ochman *et al*., 1983; Pupo *et al*., 1997; Rolland *et al*., 1998), suggest that *Shigella* and *E. coli* should be considered as members of the genus *Escherichia* (Fukushima *et al*., 2002). Despite this, *Shigella* has been kept as a separate genus for historical and medical reasons (Zuo *et al*., 2013).

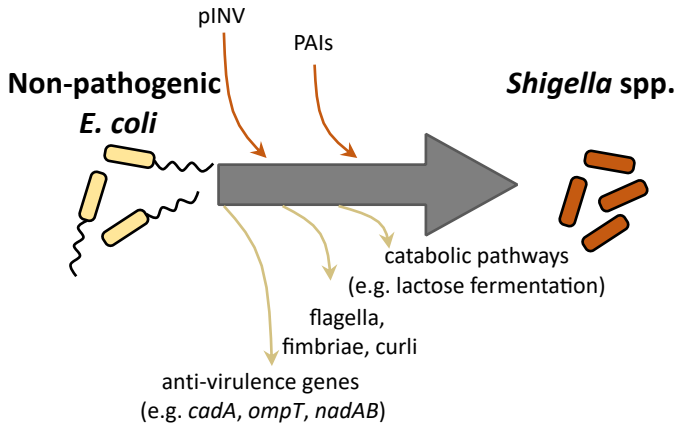


Figure 1.1 Summary of the evolution of *Shigella* spp. .

Shigella spp. evolved from non-pathogenic E. coli following acquisition of the virulence plasmid (pINV) and several pathogenicity islands (PAIs), and inactivation of anti-virulence genes, and genes involved in the biosynthesis of flagella, fimbriae, curli and catabolic pathways. Modified from Schroeder and Hilibi, 2008.

In terms of evolution *S. sonnei* can be considered an outlier from the other *Shigella* spp.. In fact, *S. sonnei* emerged more recently than other *Shigella* species (Shepherd *et al*., 2000). Also, unlike the other species of *Shigella*, *S. sonnei* uniquely possesses a cluster of genes that are essential for the synthesis, polymerisation and export of the O-Antigen (O-Ag) of its LPS and capsule, and are similar to the O-Ag encoding genes present in *Plesiomonas shigelloides* (Shepherd *et al*., 2000). Sequence comparison of this cluster from *S. sonnei* and *P. shigelloides* indicates that these genes diverged approximately 10,000 years ago, suggesting that *S. sonnei* did not arise before this time (Shepherd *et al*., 2000). A more recent phylogeographic analysis

suggests that current *S. sonnei* strains emerged from a common ancestor in the 17th century (Holt *et al*., 2012).

The critical event that led to *Shigella* becoming an important human pathogen is undoubtedly the acquisition of the virulence plasmid, which encodes a Type 3 Secretion System (T3SS) and most of its effectors (**Figure 1.1**) (Venkatesan *et al*., 2001; Baker and Chung The, 2016). However, further genetic changes have contributed to its transition from commensal to pathogen. In terms of acquisition, numerous pathogenesis-associated gene clusters, also known as pathogenicity islands (PAIs), have been incorporated into the chromosome of *Shigella,* contributing to its virulence by enhancing interactions with the host (**Figure 1.1**). For instance, SHI-1 encodes the ShET1 enterotoxin, the immunoglobulin A-like cytotoxic protease SigA and the mucinase Pic; SHI-2 encodes an aerobactin iron acquisition system and contains the genes encoding proteins involved in the suppression of T-cell signalling and the colicin I/V immunity (*shiA* and *shiD*) (Vokes *et al*., 1999; Ingersoll *et al*., 2002); SHI-3, which is found only in *S. boydii,* contains the genes for the aerobactin iron acquisition system (Purdy *et al*., 2001); SHI-O harbours genes for modifying the O-Ag, allowing serotype conversion and, hence, evasion of the immune responses (Huan *et al*., 1997; Ingersoll *et al*., 2002); and SRL (*Shigella* resistance locus) PAI contains genes that confer tetracycline, chloramphenicol, ampicillin and streptomycin resistance, and are involved in iron uptake (Luck *et al*., 2001; Turner *et al*., 2001). These PAIs are often associated with bacteriophage integrases, which mediated horizontal gene transfer. The Shiga toxin (Stx) in *S. dysenteriae* 1 is a well characterised example of bacteriophage-encoded element (McDonough and Butterton,2002). Insertion sequences (IS) are another category of mobile genetic elements which have shaped the genomic architecture of *Shigella*, by mediating gene acquisition and loss, and rearrangements (Buchrieser *et al*., 2000; Lan *et al*., 2001; Venkatesan *et al*., 2001; Yang *et al*., 2005; Siguier *et al*., 2014; Hawkey

et al., 2020). These elements are particularly abundant both in the chromosome and the plasmid; indeed, *Shigella* spp., have the highest number of IS in relation to genome size of any bacterium (Wagner and de la Chaux, 2008).

Shigella has also undergone substantial gene inactivation (**Figure 1.1**) (Escobar-Páramo *et al*., 2004; Schroeder and Hilibi, 2008). Nutrients are abundant inside the host, rendering, many catabolic pathways, such as lactose fermentation, unnecessary for an intracellular pathogen such as *Shigella* (Ito *et al*., 1991; Yang *et al*., 2005). As further evidence of convergent evolution in *Shigella* spp., gene inactivation in specific genomic regions and loci has involved different events, including deletion, point mutations, or inactivation by IS, in different strains (Yang *et al*., 2005; Feng *et al*., 2011).

Loss of flagellum biosynthesis is an example of functional gene loss (Tominaga *et al*., 2005). It is speculated this might have occurred because i) flagellar motility might not be compatible with intracellular movement mediated by actin polymerisation (Schroeder and Hilibi, 2008), or ii) flagellin, the main component of the flagellum, is a pathogen-associated molecular pattern (PAMP), so it might provoke immune responses (Ramos *et al*., 2004). A range of mutations have led to inactivation of Type 1 Fimbriae and curli production in *Shigella* spp. (**Figure 1.1**) (Sakellaris *et al*., 2000; Bravo *et al*., 2015). Loss of the outer membrane protease OmpT is also beneficial for *Shigella* spp., as it interferes with the polar localisation of the actin nucleator, IcsA (VirG), thus, inhibiting the intracellular motility (Nakata *et al*., 1993). Additionally, deletion of *cadA* and inactivation of *nadAB* also enhances virulence. The production of cadaverine, which is mediated by CadA, inhibits the activity of *Shigella* enterotoxins (Maurelli *et al*., 1998), while synthesis of the NAD precursor quinolinate by NadA and NadB strongly inhibits *Shigella* pathogenicity (Prunier *et al*., 2007). Finally, besides losing "anti-virulence" loci, *Shigella* has inactivated genes increasing adaptation and survival inside

the host. For instance, inactivation of *speG*, a gene encoding spermidine acetyltransferase, results in accumulation of the polyamine spermidine, which allows removal of free radicals, providing protection against oxidative stress in macrophages (Barbagallo *et al*., 2011).

1.1.4 Symptoms

Shigellosis is an acute intestinal infection which affects the mucosa of the rectum and large intestine (Schroeder *et al*., 2008). The incubation period is generally from 1 to 4 days but may be longer (7-8 days) in *S. dysenteriae* infection, which is also associated with long-term complications (Niyogi *et al*., 2005; Schroeder and Hilbi, 2008). Patients may present a varied range of symptoms from mild, watery diarrhoea to more severe dysentery, with abdominal pain, fever, with blood and mucus in the stool. *Shigella* can also cause extraintestinal infections, leading to meningitis, osteomyelitis, arthritis, splenic abscess, and rarely sepsis (Struelens *et al*., 1985; Keddy *et al*., 2012; Khan *et al*., 2013). Haematological changes associated with shigellosis include Haemolytic Uraemic Syndrome (HUS), which is caused by the Stx-expressing *S. dysenteriae* type 1.

Finally, several *Shigella* strains possess two other toxins that contribute to the symptoms of shigellosis. Chromosomally-encoded ShET1 and plasmid-encoded ShET2 (also known as OspD3) promote fluid secretion into the intestine, producing the characteristic watery diarrhoea in the early phases of shigellosis (Nataro *et al*., 1995; Fasano *et al*., 1997).

1.1.5 Pathogenesis

Knowledge about the *Shigella* virulence is largely based on studies carried out specifically on *S. flexneri*. This has led to the current view of *Shigella* infection as a multi-step process.

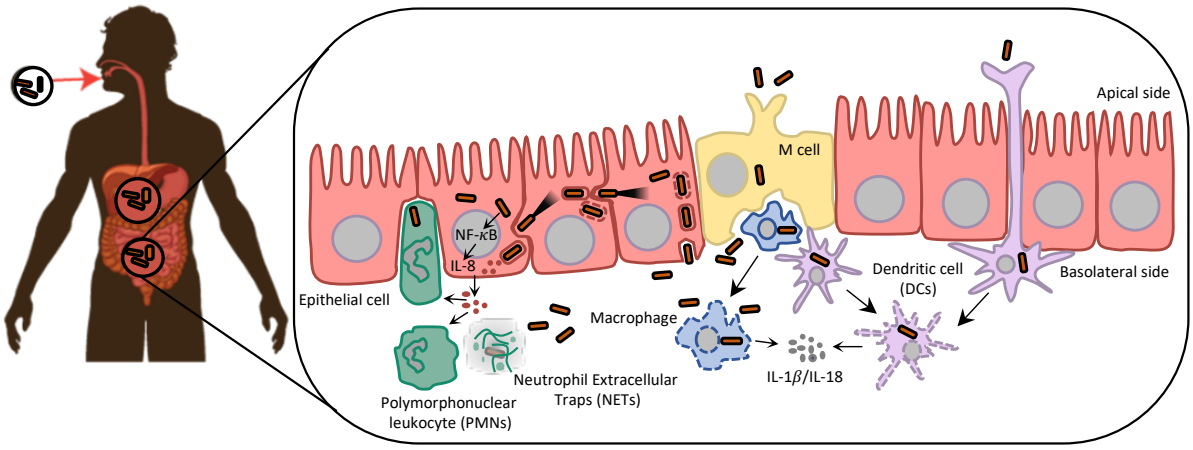


Figure 1.2 Shigella pathogenesis.

Following ingestion of contaminated water or food, *Shigella* reaches the gastro-intestinal tract. After surviving inside the stomach, Shigella arrives at the intestinal lumen, where it crosses the epithelial barrier by passing through M cells. After reaching the subepithelial M cell pocket region, the bacterium is then endocytosed by resident macrophages and DCs. Bacterial sampling can also occur by DCs residing between epithelial cells, through their extended dendrites. Phagocytosed virulent organisms induce an inflammatory cell death characterised by the release of pro-inflammatory cytokines, such as IL-1β and IL-18. Released live bacteria invade through the basolateral surface of epithelial cells, then spread from cell to cell. Infected epithelial cells produce cytokines and chemokines such as IL-8, which recruits PMNs. PMNs disrupt epithelial cell-cell junctions and eliminate bacteria by releasing NETs.

1.1.5.1 Gaining access to the intestinal epithelium

Once ingested, bacteria survive the low pH gastric environment thanks to sophisticated acid resistance mechanisms (Gorden and Small, 1993), such as Asr, a protein which acts as a proton sink in the periplasm, protecting essential proteins from denaturation (Tucker *et al*., 2002; Cheng *et al*., 2007), or the AR1 acid-resistance system that contributes to the survival of *Shigella* at a pH below 2.5 (Waterman and Small, 1996). Furthermore, studies on *S. flexneri* 2a have demonstrated that LPS also contributes to acid resistance, due to specific changes on the O-Ag and lipid A (Martinić *et al*., 2011).

After passing through the small intestine, bacteria reach the colonic epithelium. Here, *Shigella* competes with the microbiome, which represents the first barrier protecting the intestinal epithelium. The protective effect of the microbiome has been mainly demonstrated in animal models. *E. coli* colonisation of germ-free mice and Guinea pigs blocks the colonisation by *S. flexneri* (Formal *et al*., 1961; Maier and Hentges, 1972), while depletion of the murine flora by streptomycin treatment enables intestinal colonisation in mice (Pongpech *et al*., 1989; Martino *et al*., 2005). To compete against the microbiome, *Shigella* can secrete colicins, small molecules that have bactericidal activity against phylogenetically related bacteria (Anderson *et al*., 2016). Furthermore, *S. sonnei* possesses a Type VI Secretion system, which the bacterium uses to compete against other species, such as *E. coli*, to allow colonisation of new niches (Anderson *et al*., 2017).

Another barrier that *Shigella* encounters in the gastrointestinal tract is the mucus layer (Anderson *et al*., 2016). Mucus is mainly composed by oligomeric chains of glycoproteins, called mucins, which are secreted by goblet cells in the epithelial layer (McGuckin *et al*., 2011). All *Shigella* spp. can degrade mucus by secreting mucinases and neuraminidases (Haider *et al*., 1993); examples include PicA, SepA and EatA-like mucinases (Benjelloun-Touimi *et al*., 1995, 1998; Kumar *et al*., 2014).

1.1.5.2 Crossing the intestinal epithelium and survival inside the phagocytes

Upon reaching the colonic epithelium, *Shigella* gains access to the intestinal mucosa *via* M cells (microfold cells) as conduits for crossing the epithelial barrier (**Figure 1.2**). This particular type of intestinal epithelial cell (IEC) is distinguished from neighbouring enterocytes as their apical surface lacks microvilli, glycocalyx and mucus, and possesses broader microfolds that give the cell its name. The morphology of M cells is critical for their role, which is sampling

"self" and "non-self" antigens from the intestinal lumen *via* endocytosis or phagocytosis, and delivering the antigens to underlying mucosal lymphoid tissue (Neutra *et al*., 1999).

In vitro studies showed that *S. flexneri* is unable to invade monolayers of polarised epithelial cells, such as CaCo2 and HT29 cells, from the apical side (Mounier *et al*., 1992). These studies, in parallel with murine infection models (Yang *et al*., 2013), revealed that the preferential entry route of *Shigella* in IEC is *via* the basolateral aspect of cells (Köhler *et al*., 2002). Therefore, passage through M cells represents a fundamental step for epithelial invasion by *Shigella*. Other studies have shown that bacteria can also be sampled directly from the lumen by DCs residing between epithelial cells (Ng *et al*., 2008) (**Figure 1.2**).

After the bacterium reaches the subepithelial region, it encounters resident phagocytic cells, such as macrophages and DCs, which engulf the organism (**Figure 1.2**) (Ashida *et al*., 2011). However, once inside these cells, *Shigella* disrupts the phagocytic vacuole in a T3SS-dependent manner. Once the bacterium has escaped from the phagosome, it spreads through the cytosol. Here, it promotes the assembly of a specific inflammasome, known as NLRC4, which is triggered by bacterial molecules such as components of the T3SS, in the cytosol (Kanneganti *et al*., 2007; Koizumi *et al*., 2011). NLRC4 activates caspase-1, leading to subsequent pyroptosis. Pyroptosis is a well-defined mechanism of pro-inflammatory programmed cell death, accompanied by release of pro-inflammatory cytokines, such as IL-1β and IL-18, which mediate inflammation during infection with *S. flexneri* (Sansonetti *et al*., 2000; Bergsbaken *et al*., 2009).

1.1.5.3 Invasion, replication and spread across the epithelium

Live bacteria released by infected dead phagocytes invade through the basolateral surface of IEC (Ashida *et al*., 2011). Since epithelial cells are not phagocytic, *Shigella* uses the T3SS to

penetrate cells. Indeed, the bacterium injects several T3SS effectors into the host cell cytoplasm, where they interfere with cellular basal physiological processes. Among these numerous molecular effectors, a crucial role in the early steps of invasion is performed by IpaA, IpaB, IpaC, IpgD, IpgB1 and IpgB2. These proteins induce rearrangements in the host cell cytoskeleton and membranes, influencing actin polymerisation and filament reorganisation. This leads to the formation of large membrane protrusions, or "membrane ruffles", which allow the bacterium to be engulfed and internalised by micropinocytosis, a form of endocytosis (Tran Van Nhieu *et al*., 2003; Cossart and Sansonetti, 2004). Immediately following endocytosis, the bacterium is found a nascent phagosome, but in less than 15 minutes, it disrupts the vacuolar membrane and enters the cytoplasm (Clerc and Sansonetti, 1987; Shere *et al*., 1997).

Unlike phagocytes, IEC are not immune cells so they lack typical defence mechanisms that are necessary to limit infections. For this reason, they facilitate the replication and dissemination of *Shigella.* In addition to the typical processes of apoptotic and necrotic death (Carayol and Tran Van Nhieu, 2013), enterocytes deploy two other responses to bacterial infection: promotion of inflammation and autophagy.

In the first stages of infection, *Shigella* releases peptidoglycan (PGN) fragments during intracellular multiplication, which are sensed by the cytosolic receptor NOD1 (Nigro *et al*., 2008; Ogawa *et al*., 2009). Upon recognition, NOD1 activates its own oligomerisation and induces NF-κB transcription (Girardin *et al*., 2001; Marteyn *et al*., 2012). NF-κB is responsible for transcriptional activation of pro-inflammatory genes, principally cytokine and chemokine genes, including CXCL-8 (or IL-8 using the chemokine nomenclature) (Kasper *et al*., 2010). This molecule is a chemotactic factor that attracts PMNs, causing them to migrate from the

bloodstream toward the site of infection (Sansonetti, 1999). In addition to activation of NF-κB, NOD1 activates MAPKs (Mitogen-Activated Protein Kinases) which further activate pro-inflammatory gene expression (Strober *et al*., 2006). NF-κB and MAPKs signalling pathways are activated not only in infected IECs but also in nearby uninfected bystander cells (Ashida *et al*., 2011). The propagation of pro-inflammatory signals occurs through gap junctions, specialised intercellular connections which allow communication between adjacent cells. This leads to the rapid amplification of inflammation and recruitment of specialised immune cells (Stadnyk, 2002; Kasper *et al*., 2010).

In the later stages of infection, one of the leading responses against pathogens is autophagy (Carayol and Tran Van Nhieu, 2013). Autophagy is a highly regulated process that engages lysosomal degradation to eliminate pathogens (Levine *et al*., 2011). This cellular response can be activated by NOD1, which recruits Atg16L1 to the site of bacterial entry (Travassos *et al*., 2010).

Although *Shigella* is non-motile, it has developed a specialised mechanism to migrate through the host cytoplasm by exploiting actin polymerisation (**Figure 1.2**) (Bernardini *et al*., 1989). In particular, actin polymerisation is mediated by IcsA, a protein that localises at one pole of the bacterium and is encoded on the virulence plasmid. Upon binding to the bacterial pole, IcsA recruits and activates host neuronal Wiskott-Aldrich syndrome protein (N-WASP) and thence the Arp2/Arp3 complex, which promotes actin nucleation and elongation, propelling the bacterium through the cytoplasm (Egile *et al*., 1999). In addition, VirA, another pINV-encoded effector, localises at the opposite pole to IcsA and degrades α–tubulin to disrupt the compact intracellular microtubule network at the leading end of the bacterium (Yoshida *et al*., 2006). *S. flexneri* then reaches the host cell membrane and forms a protrusion that can be engulfed by neighbouring cells, *via* clathrin-dependent endocytosis (Fukumatsu *et al*., 2012). In this

way, the bacterium initiates a new replicative cycle in the next IEC, spreading across the intestinal epithelial layer "cell-by-cell" (**Figure 1.2**) (Bernardini *et al*., 1989; Phalipon *et al*., 2007; Ogawa and Sasagawa, 2009; Ashida *et al*., 2011). This process allows the organism to replicate and disseminate, while avoiding the exposure to extracellular components of the immune system.

1.1.5.4 Role of polymorphonuclear neutrophils

Another important cell type that interacts with *Shigella* is the polymorphonuclear neutrophil leukocyte (PMN), the most abundant circulating leukocyte. These cells act as a double-edged sword in terms of host defence (Ashida *et al*., 2011). Recruited during the early phases of infection, PMNs migrate from the blood into the intestinal lumen, passing between enterocytes, and disrupting epithelial cell-cell junctions (**Figure 1.2**). In this way, PMNs destroy the integrity of the epithelial cell barrier and grant luminal bacteria direct access to the submucosa (Perdomo *et al*., 1994). However, in the later stages of infection, IECs secrete IL-8, activating and recruiting PMNs to the site of infection, where they eliminate bacteria through release of NETs (Neutrophil Extracellular Traps) (**Figure 1.2**). NETS are composed of chromatin, associated with antimicrobial peptides and elastase, and arrest *Shigella*'s spread by cleaving virulence factors and killing the bacteria (Weinrauch *et al*., 2002; Brinkmann *et al*., 2004; Papayannopoulos *et al*., 2010). Therefore, in early stages PMNs can help *Shigella* by promoting epithelial invasion, while during later stages, the inflammatory response activates and recruits neutrophils to eliminate bacteria.

1.1.5.5 Manipulation of host immune responses

Several studies have demonstrated that *Shigella* is able to employ a variety of mechanisms to manipulate host innate immunity by delivering a subset of T3SS effectors, thus, promoting bacterial colonisation and survival (Ashida *et al*., 2015).

The ability of *Shigella* to inhibit autophagy has been widely described (Schroeder and Hilibi, 2008). IcsB is a bacterial protein involved in preventing autophagy (Ogawa *et al*., 2005). This effector is secreted into the host cytoplasm through the T3SS, and once inside, binds to bacterial IcsA. IcsB masks IcsA, preventing Atg5-mediated recognition and activation of autophagy (Ashida *et al*., 2011).

Preservation of the integrity of epithelial cells is fundamental for *Shigella* to guarantee its replication and dissemination. By delivering other effectors, the bacterium modulates and reduces inflammatory responses in epithelial cells (Carayol and Tran Van Nhieu, 2013). Among these effectors is VirA, which inhibits Rab1, a host GTPase implicated in vesicular trafficking, blocking CXCL-8 secretion (Dong *et al*., 2012). A drop in CXCL-8 release results in persistence of the bacterium inside the host and reduced recruitment of immunological cells.

Another group of effectors expressed by *Shigella* modulates MAPKs and NF-κB signalling pathways through post-translation modification (such as phosphorylation and ubiquitination). OspG is a secreted kinase that inhibits UbcH5, which ubiquitinates phosphorylated Iκ-Bα, an NF-κB inhibitor. In this way, OspG prevents Iκ-Bα degradation, suppressing NF-κB activity, and impairing production of pro-inflammatory cytokines (Kim *et al*., 2005).

Another effector that dampens host immune responses is OspF (Kim *et al*., 2008). This enzyme is a phosphothreonine lyase that irreversibly dephosphorylates MAP kinases (MAPKs) and ERKs (Extracellular signal Regulated Kinases), inhibiting transcription of a specific subset of pro-inflammatory genes, including NF-κB (Arbibe *et al*., 2007). Consequently, the bacterium

inhibits the expression of several NF-κB-dependent genes, such as those encoding cytokines and chemokines, which are fundamental for PMN and DC recruitment.

1.1.6 Treatment of shigellosis, antimicrobial resistance and *Shigella* vaccines

Maintenance of hydration and electrolyte balance is fundamental, albeit not sufficient, to treat shigellosis. Antibiotic therapy can reduce the duration of symptoms, clear the infection and interrupt further person-to-person transmission (Kotloff, 2018). In 2017, the WHO recommended the use of ciprofloxacin as first line therapy for children and adults affected by dysentery, and azithromycin, cefixime and ceftriaxone as second line therapy (WHO, 2017). Unfortunately, many *Shigella* strains are resistant to multiple antibiotics, including sulphonamides, tetracycline, chloramphenicol and fluoroquinolones, probably due to the overuse of these drugs (Puzari *et al*., 2018). Attention has been focused on antimicrobial-resistant *S. sonnei*, as this species has undergone rapid spread in LMICs, making AMR a serious problem in this species (Chung The *et al*., 2019; Thanh Duy *et al*., 2020). Appropriate use of antibiotics, regulation and surveillance are urgently needed to prevent further development and spread of antibiotic resistance in *Shigella* spp.. Furthermore, several alternative therapeutics against *Shigella*, have been suggested, such as use of probiotics, nanoparticles and phage therapy; however, none of these are available in the fight against this pathogen (Ranjbar and Farahani, 2019).

A valid solution to antibiotic-resistance is the development of a vaccine. To date there is no licensed vaccine available against *Shigella*, however, several vaccine candidates are under investigation which include live oral vaccines, killed whole-cell vaccines and subunit vaccines (Kotloff, 2018). The major factor that affects the development of a vaccine is the lack of an

effective animal model. Since humans are the primary reservoir of *Shigella* spp., it is still challenging to find suitable animal models that mimic human disease. In the past, mouse, rat, guinea pig, rabbit and non-human primates have been used but none of them faithfully reproduces the steps in pathogenesis observed in humans (Kim *et al*., 2013). This has been a major obstacle in the study of host-pathogen interactions, and particularly, the role of T3SS effectors (Liu *et al*., 2019). However, the rabbit ileal loop model has been employed for many years to study how *Shigella* responds to environments in the host and manipulates inflammatory signalling in the intestinal mucosa (Liu *et al*., 2019).

Another important factor that hinders vaccine development is the large number of different serotypes of *Shigella*. Antibody responses against *Shigella* LPS-associated O-Ag are generally robust and efficient in inducing a long-lasting protective immunity. However, responses are serotype-specific, with little evidence of cross-serotype immunity following natural infection (Livio *et al*., 2014; Mani *et al*., 2016). As including approximately 50 different serotypes in a vaccine would be technically challenging, one potential solution is to design a multivalent vaccine that protects against the most common serotypes, such as *S. flexneri* 2a, 3a and 6, and *S. sonnei*.

1.2 *Shigella*'s virulence

The clinical manifestations of shigellosis are the result of the complex interaction of numerous virulence factors, most of which are encoded on a large virulence or invasion plasmid, pINV. As mentioned in the previous section (Section 1.1.3), the major event that changed *Shigella* from a commensal into a pathogen was the acquisition of pINV. However, several virulence genes have also been acquired on the chromosome, demonstrating that the acquisition of virulence was a multi-step process.

1.2.1 The large virulence plasmid in *Shigella* spp.

Initial studies on virulence determinants in *Shigella* began by characterising chromosomal genes and secreted exotoxins (Flexner and Sweet, 1906; Falkow *et al*., 1963). In 1980, pINV was identified following characterisation of *S. sonnei* strains that displayed a change in their colony morphology. "Phase II" colonies appeared rough and irregular and were no longer able to invade into human intestinal epithelial cells (Kopecko *et al*., 1980). This was in contrast to the smooth, round "Phase I", wild-type colonies (Kopecko *et al*., 1980). Subsequently, these strains were shown to have lost a 120-megadalton plasmid, later identified as *Shigella* pINV (Sansonetti *et al*., 1981). By employing antibiotic resistance carried on transposons, Sansonetti and colleagues discovered that this plasmid was not self-transmissible, but could be mobilised using a conjugative plasmid into "Phase II" bacteria, re-establishing virulence and the "Phase I" phenotype. The presence of pINV was later confirmed in *S. flexneri* and transfer of the virulence plasmid into *E. coli* conferred an invasive phenotype (Sansonetti *et al*., 1982; 1983; Sasakawa *et al*., 1986).

Plasmids are extrachromosomal elements carrying accessory genes that can provide fitness benefits to the host bacterium, by enabling it to survive in new niches and resist to specific stresses. In the case of *Shigella*, pINV is essential for the invasion into, replication and survival within host cells, allowing the bacterium to spread in the mucosal layer.

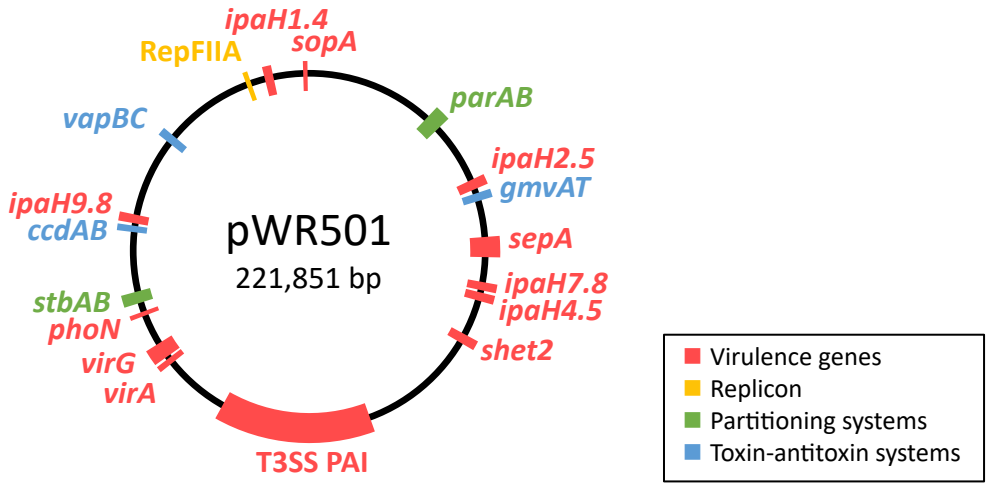


Figure 1.3 *S. flexneri* virulence plasmid pWR501.

S. flexneri pWR501 possesses a 32 kb pathogenicity island (T3SSPAI) encoding a type III secretion system (T3SS) and several virulence-related genes located outside of the T3SS PAI. The plasmid has a low copy number replicon (RepFIIA), and retention of the plasmid is promoted by three toxin–antitoxin systems (VapBC, GmvAT and CcdAB) and two partitioning systems (ParAB and StbAB). Genes important for virulence, plasmid replication, partitioning and post-segregational killing are highlighted and colour-coded (see key). Plasmid map constructed using previously published sequenced data (Venkatesan *et al*., 2001).

1.2.1.1 Evolution and phylogenesis

All *Shigella* spp. possess a pINV. Although all plasmids harbour a T3SS PAI, pINVs differ in their structure and the genetic organisation in the four *Shigella* spp., as a result of the independent recombination and acquisition events (Yang *et al*., 2005). In 2001, Lan and colleagues sequenced three pINV genes, *ipgD*, *mxiC*, and *mxiA*, from a wide range of isolates from the four *Shigella* spp., to characterise phylogenetic relationships between pINVs, and the

potential link between plasmid and chromosomal genes in each species (Lan *et al*., 2001). Two sequence forms of the plasmid were found, pINVA and pINVB, which, despite being relatively closely related, show distinct genetic features. Strains belonging to each *Shigella* species are not always associated with a specific plasmid form. Furthermore, the clustering of *Shigella* strains by plasmid largely correlates with the clustering based on chromosomal sequences, suggesting that the plasmid and the chromosome have evolved together in the majority of the *Shigella* strains analysed (Lan *et al*., 2001).

The virulence plasmid pWR100 and its Tn*501*-containing derivative pWR501 from *S. flexneri* 5a M90T (**Figure 1.3**) are the most studied pINV from *Shigella* spp. (Buchrieser *et al*., 2000; Venkatesan *et al*., 2001). Analysis of its GC content has revealed that the plasmid is a mosaic of genes from different origins (Buchrieser *et al*., 2000). In particular, the *osp* and T3SS PAI genes belong to a block with a similar GC content (average 34%) and were probably acquired from the same source. Similarly, the *icsA* and *virK* region (41% GC), and *sepA* (49% GC) may have been derived from a common source. Furthermore, the presence of two distinct partitioning systems (*parAB* and *stbAB,* 43% and 39% GC, respectively), a remnant of a third (*virB,* 34% GC) and a replicon with a distinct GC content (*repA*, 58% GC), suggests that the virulence plasmid is the result of a composite of elements from four plasmids (Buchrieser *et al*., 2000). DNA deletions, duplications and inversions have contributed to the organisation of these blocks in a distinct manner in each *Shigella* species (Buchrieser *et al*., 2000).

Additionally, the *Shigella* T3SS PAI shares homology with the PAI in *Yersinia* and *Pseudomonas syringae* plasmids, indicating they arose from a common ancestor. However, the T3SS PAI is likely to have been acquired much later in evolution than for *Yersinia* and *P. syringae*, or *Shigella* acquired the PAI from a different organism (Venkatesan *et al*., 2001).

1.2.1.2 Virulence determinants of pINV

The T3SS PAI on pINV is critical for virulence. Also known as the *ipa-mxi-spa* locus, this region is a 32 kb cluster of 37 ORFs, encoding the structural elements of the T3SS, secreted translocators and effectors, chaperones and transcriptional regulators (**Figure 1.4**). The T3SS PAI is necessary and sufficient for *Shigella* epithelial invasion, intracellular survival, and induction of pyroptosis in infected macrophages (Coburn *et al*., 2007). The PAI can be divided into four regions, depending on their function (Schroeder *et al*., 2008).

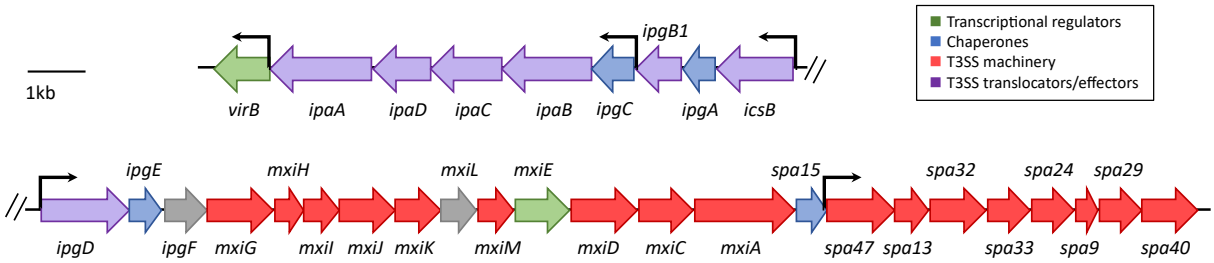


Figure 1.4 Genetic map of the T3SS PAI of the *S. flexneri* pWR501.

The genes form a ~32 kb cluster composed of four groups of genes: structural components of the T3SS, chaperones, molecular effectors and regulators (colour coded as in the key, modified from Schroeder and Hilibi, 2008).

The first region encodes secreted translocators and effectors which manipulate the host cell to allow bacterial entry, replication and survival inside the cellular cytoplasm. IpaA, IpaB, IpaC and IpaD are important effectors and protein antigens encoded in this region (Buysse *et al*., 1986). Specifically, IpaB and IpaC are components of the translocon, which associates with IpaD mediating membrane insertion of the translocon (Blocker *et al*., 1999). IpaA interacts with host cell vinculin, activating cytoskeleton rearrangements (Tran Van Nhieu *et al*., 1997). Important roles are performed by other effectors such as IpgB1 and IpgB2 during endocytosis; IpgD can trigger an anti-apoptotic pathway which partially impedes the death of infected host cells (Parsot, 2009).

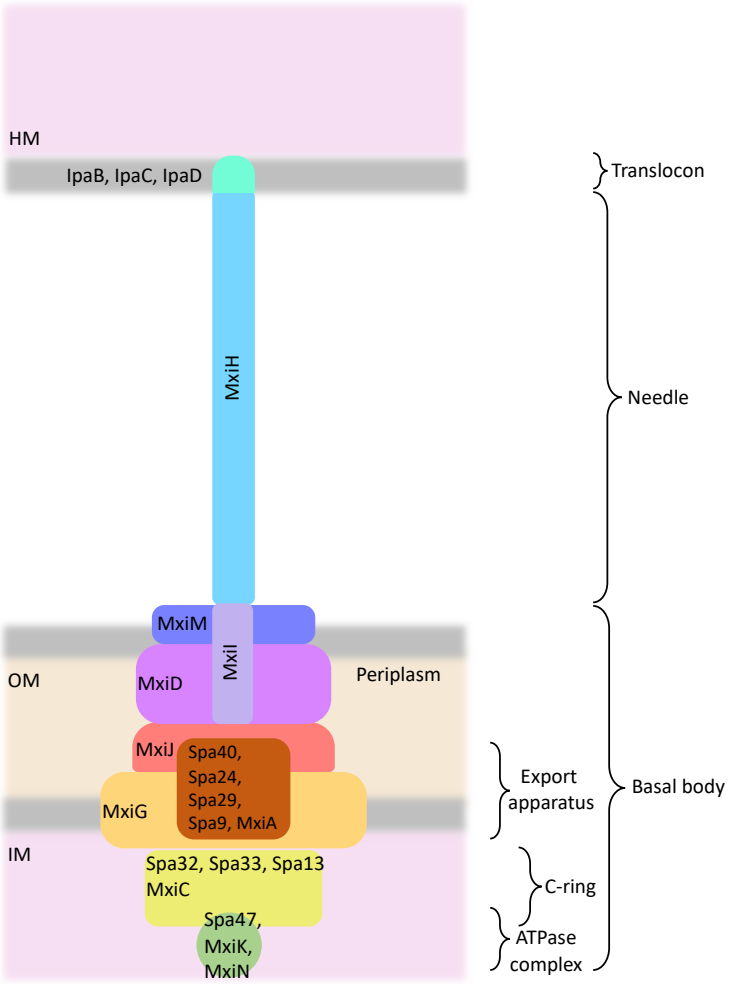


Figure 1.5 Architecture of T3SS in *Shigella*.

The T3SS is composed of a basal body that comprises an ATPase complex, a cytoplasmatic secretion machinery (C ring) and an export apparatus, a transmembrane basal body that spans inner (IM) and outer bacterial membranes (OM), and an extracellular needle equipped with a translocon on the tip, which mediates T3SS insertion in the host membrane (HM).

The genes of the second group occupy more than half the T3SS PAI and are required for T3SS assembly and function. In association with IpaB, IpaC and IpaD, these proteins build the T3SS apparatus (Blocker *et al.*, 1999). The secreton is organised into three parts (**Figure 1.5**) (Blocker *et al.*, 2001; Cordes *et al.*, 2003): i) a region within the cytoplasm, which serves as the motor of the secretion apparatus and interacts with the effector proteins; ii) a basal body, which passes through the Inner Membrane (IM), periplasm, peptidoglycan (PGN) and outer

Membrane (OM) and is composed of multiple copies of MxiG, MxiJ and MxiD; iii) a needle characterised by multiple MxiH and MxiI subunits, which form a helical polymer with an average length of 45 to 60 nm and an approximate diameter of 7-8 nm (Tamano *et al*., 2000; Sani *et al*., 2007).

The third region contains two transcriptional activators, VirB and MxiE (Adler *et al*., 1989), which regulate T3SS-associated genes on pINV (Kane *et al*., 2002; Kane and Dorman, 2012). This pair of activators regulates the expression of genes involved in the early and later stages of infection, respectively (Adler *et al*., 1989; Allaoui *et al*., 1993).

Finally, the fourth region is composed of genes encoding chaperones, including IpgA, IpgC, IpgE and Spa15. Their function is to stabilise the T3SS-secreted substrates in the bacterial cytoplasm (Parsot *et al*., 2003; van Erde *et al*., 2004). Notably, IpgC and Spa15 also participate in the transcriptional regulation of virulence genes outside the T3SS PAI (Mavris *et al*., 2002).

Several genes encoding T3SS effectors are located outside the PAI (**Figure 1.3**) (Venkatesan *et al*., 2001). These include *osp* family of genes *e.g. ospB*, *ospC1*, *ospD2*, *ospE2*, *ospF*, and *ospG*; *ipaH* genes (five in pWR100); genes involved in the intracellular motility such as *icsA*, *icsB*, *virA, virK* and *phoN2*; and other genes with different function such as *virF*, *msbB*, *sepA*, *ipgH*, *shET2*, *phoN* and *trcA*.

1.2.1.3 Maintenance of pINV

Maintenance of plasmids is fundamental to their propagation in bacterial populations. However, the presence of large plasmids such as pINV can be costly for bacteria, so their loss is prevented through highly regulated systems such as replicons, partitioning systems and toxin-antitoxin systems. These mechanisms are independent of the chromosome and there is

no evident regulatory cross-talk between the plasmid and chromosomal replication and segregation (Reyes-Lamothe *et al*., 2012).

1.2.1.3.1 Replication of pINV

pINV is a low copy number, non-conjugative plasmid, as the transfer genes on the plasmid are incomplete (**Figure 1.3**) (Venkatesan *et al*., 2001). Although the replicon of pINV in *Shigella* has not been extensively characterised, it is related to the RepFIIA replicon from the plasmid R100, sharing ~75% identity at the nucleotide and amino acid levels (**Figure 1.6**) (Venkatesan *et al*., 2001; Yang *et al*., 2005). The RepFIIA replicon belongs to the antisense RNA-regulated family of replicons (del Solar *et al*., 1998; Nordstrom *et al*., 2006), which generally contains a replication initiation protein (Rep) and an origin of replication (*ori*); replication proceeds unidirectionally according to the theta replication model (Pinto *et al*., 2012; Pilla and Tang, 2018). Binding of Rep to *ori* is essential to render the DNA in this region single-stranded and allow recruitment of the DNA replication machinery. However, replication is controlled by a short antisense transcript that binds to the 5' end of the longer of two Rep transcripts, preventing its translation (**Figure 1.6**). Because the antisense transcript is unstable, low concentrations of the antisense transcript are not sufficient to prevent replication. Therefore, inhibition of RepA translation only occurs when the plasmid copy number climbs, preventing further increase in the copy number (Pilla and Tang, 2018). In particular, in the RepFIIA replicon, translation of *repA* requires and is coupled with the translation of *tap*, a short leader sequence located upstream of the gene (Nordstrom *et al*., 2006). Translation of Tap prevents formation of a stable secondary RNA structure that blocks the *repA* ribosome binding site (RBS) (Blomberg *et al*., 1992, 1994). The 90 nt antisense RNA CopA is transcribed from the complementary strand of *repA* and binds the CopT target, which is upstream the leader sequence of *tap* with sequence complementary to *copA* (**Figure 1.6**) (Light *et al*., 1982;

Blomberg *et al*., 1992). This binding prevents translation of the Tap leader peptide and results in RNase III-dependent cleavage of both RNAs, thereby inhibiting *repA* translation (Malmgren *et al*., 1996; Nordstrom *et al*., 2006; Brantl *et al*., 2015). Furthermore, the sequence of the CopA antisense RNA determines the incompatibility group of the plasmid; for this reason, *copA* is also referred to as *incRNA18*, with pINV belonging to the IncFII incompatibility group (Funnell and Phillips, 2004). In fact, the presence of different plasmids with the same replicon in the cell is perceived as an increase in the copy number. This results in loss of one of the two plasmids, reflecting the incompatibility.

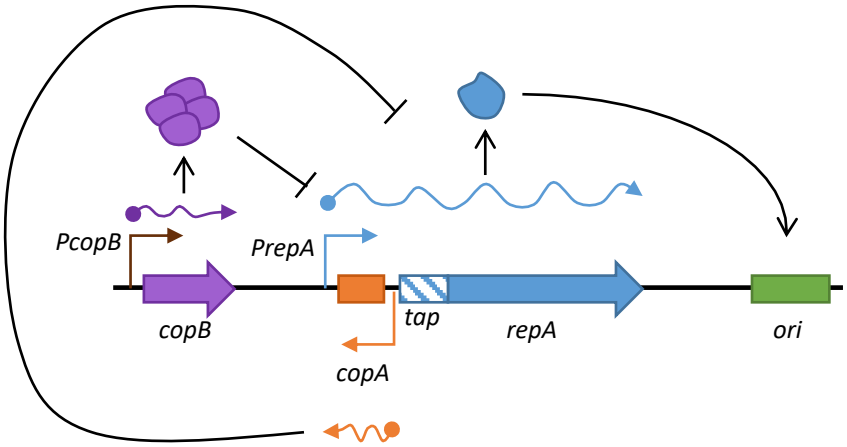


Figure 1.6 Genetic structure and function of the replicon RepFIIA.

The RepFIIA replicon is regulated by an unstable antisense RNA (CopA) which prevents RepA translation at high concentrations by binding to Tap, the untranslated leader of *repA* mRNA, essential for RepA translation. CopB also regulates the replicon by repressing *repA* transcription at its promoter *PrepA*. Modified from Pilla and Tang, 2018.

The second element that controls plasmid replication and copy number of RepFIIA is CopB (also named RepA2 in *Shigella*), a regulatory protein that represses *repA* transcription by binding to its promoter, *PrepA* (**Figure 1.6**). Transcription of *copB* is constitutive, so high CopB concentrations in the cell from a high plasmid copy number inhibit plasmid replication (Nordstrom *et al*., 2006; Pilla and Tang, 2018). Although the pINV replicon is highly homologous to the R100 replicon, the antisense RNA locus (*inc*) found in pINV differs from

other IncFII plasmids, suggesting that pINV falls into a different incompatibility group (Venkatesan *et al*., 2001).

1.2.1.3.2 Partitioning systems

Partitioning is another essential aspect for the maintenance of low copy plasmids. Partitioning systems directly segregate plasmids into daughter cells during cell division and are characterised by the possession of a nucleoside triphosphatase (NTPase) (either an ATPase or GTPase), which is necessary for the process of segregation, DNA *par* sites, which act as centromere-like elements and a centromere-binding protein (CBP), which binds the DNA *par* sites, forming a partitioning complex (Gerdes *et al*., 2010). Partitioning systems found on virulence plasmids can belong to the type I or type II family, which differ in their genetic organisation, the proteins involved and the mode of action (**Figure 1.7**) (Pilla and Tang, 2018).

ParABS systems

Type I loci (for example, *parABS* of P1 plasmid) are highly homologous to the chromosomal, ParAB partitioning systems. They contain Walker-type ATPases and the *par* sequence is generally located downstream *parB*. Type I systems can be further divided into type Ia, of which ParAB and SopAB are representatives, and type Ib, of which ParAB from pB171 is an exemplar (Ebersbach and Gerdes, 2005). Type Ib systems show several differences from type Ia systems. First of all, type Ib has *par* sites upstream of *parA* as well as downstream of *parB* (**Figure 1.7A**). Furthermore, in type Ib systems ParA does not possess a helix-turn-helix (HTH) motif at its N-terminus, which in type Ia ParA allows it to act as the main transcriptional regulator of the *par* operon. In contrast, in the type Ib system, ParB regulates *par* (Gerdes *et al*., 2000).

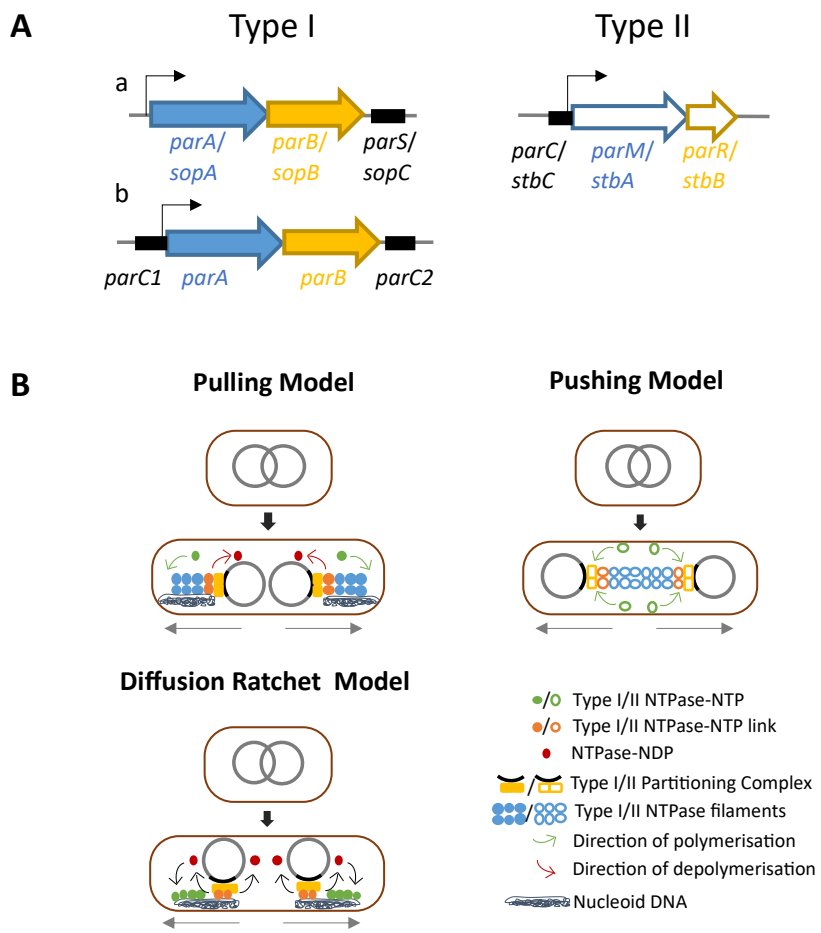


Figure 1.7 Structure of type I and type II *par* loci and their models of mechanisms.

(**A**) In type I and type II *par* loci, the gene encoding the nucleoside triphosphatase (NTPase, blue) is upstream of the centromere-binding protein (CBP, yellow), whereas the centromere-like element (*par*, grey) is either located downstream of the CBP (type Ia), upstream of the NTPase (type II) or at both positions (type Ib). (**B**) In the pulling model, NTPase-nucleoside 5'-triphosphates (NTPs) cooperatively bind to nucleoid DNA, forming filaments. Contact between the partitioning complex and the filaments induces NTP hydrolysis, resulting in disassembly of polymers. Each plasmid is pulled in the direction of NTPase depolymerisation. In the diffusion-ratchet model, the NTPase-NTP binds non-specifically to nucleoid DNA and interacts with the partitioning complex (NTPase–NTP link). This stimulates NTP hydrolysis, and the NTPase-nucleoside 5'-diphosphate (NDP) is then released from DNA, generating an area of NTPase-free nucleoid DNA around the site of interaction with the plasmid; movement of the plasmid is driven by the slow reassociation of the NTPase and its fast de-association from nucleoid DNA. In the pushing model, the partitioning complex increases the stability of NTPase-NTP, stimulating the formation of NTPase filaments that push the plasmid towards opposite poles. Eventually, NTP hydrolysis destabilises the filaments and the polymers are disassembled; in contrast to the pulling model, polymerisation and depolymerisation do not occur simultaneously (from Pilla and Tang, 2018).

In the case of P1 ParABS, the ParB binding sites in *parS* consist of short DNA sequences (defined as boxes A and B) that are arranged asymmetrically and located at either side of an Integration Host Factor (IHF) site, and bound by a HTH motif in ParB (Funnell, 1991; Funnell and Gagnier, 1993; Lobocka and Yarmolinsky, 1996; Surtees and Funnell, 1999). Furthermore, ParB interacts with *parS* as a dimer, with dimerisation mediated by the C-terminal domains (Surtees and Funnell, 2001). Following binding of one ParB dimer, more dimers are loaded on the *parS* sequence, forming a nucleoprotein structure composed by *parS* wrapped around a core of IHF and ParB molecules (Surtees and Funnell, 1999; 2001). Finally, ParB interacts with ParA. Similar to ParB, ParA binds nucleoid DNA, albeit non-specifically. According to the pulling model, when bound to ATP, ParA dimers interact with DNA through its HTH domain, leading to the formation of ParA filaments. Each filament then contacts the partitioning complex, and ParB stimulates the ParA dimer at the end of the filament to hydrolyse ATP. Conversion of ParA-ATP into ParA-ADP induces a conformational change in the protein that leads to its release from DNA and the filament, allowing further ParA-ATP dimers to interact with the partitioning complex. This leads to the progressive release of ParA dimers and moves the plasmid toward the direction of ParA depolymerisation (**Figure 1.7B**) (Gerdes *et al*., 2010). More recent *in vitro* studies have supported an alternative view of type I partitioning, termed the diffusion-ratchet model (**Figure 1.7B**) (Vecchiarelli *et al*., 2012; Brooks and Hwang, 2017).

In a cell-free system, the partitioning activity of the P1 and F plasmids were reconstituted, using a flowcell surface coated with non-specific DNA as an artificial nucleoid, and the system's dynamics visualised by total internal reflection fluorescence microscopy (TIRFM) (Vecchiarelli *et al*., 2013; 2014). It was found that there is no formation of ParA filaments, however, these proteins, when charged with ATP, associated with nucleoid DNA evenly. It is hypothesised that ParA binds dynamically and non-specifically to nucleoid DNA, and upon interaction with the

partitioning complex, formation of DNA bridges between ParA and ParB anchors the plasmid to the nucleoid DNA. However, ParB-stimulated ATP hydrolysis leads to accelerated dissociation of the ATPase from DNA which is significantly faster than the reassociation of the protein with DNA. This creates zones depleted of ParA surrounding and behind the moving partitioning complex, resulting in the movement of the plasmid towards areas of more concentrated ParA (Vecchiarelli *et al*., 2014). This asymmetric ParA distribution is proposed to force the plasmid to move in one direction, following so-called Brownian ratchet motion (**Figure 1.7B**) (Hu *et al*., 2017). Therefore, it is still not clear if asymmetric distribution of ParA causes or is the result of plasmid movement.

Finally, both ParA and ParB are involved in the autoregulation of the operon, which is an important aspect as over-expression of both proteins induces partitioning defects (Friedman and Austin, 1988) In particular, ADP-bound ParA stimulates binding to its operator sequences, repressing its transcription, thus high levels of ParB increase repression as ParB induces ATP hydrolysis when associated with *parS* sites, increasing the amount of ParA-ADP (Bouet and Funnell, 1999; Fung *et al*., 2001).

ParMRC systems

In contrast, type II systems are the most well-characterised plasmid partitioning systems, with the ParMRC of the *E. coli* R1 plasmid being a model system (Gerdes *et al*., 2010). Type II systems utilise an actin-like ATPase, ParM, and the *parMRC* promoter is located in the middle of *parC*, the centromere-like sequence that is upstream *parM* (**Figure 1.7A**) (Ebersbach and Gerdes, 2005). Similar to ParB, ParR, the CBP of type II systems, binds *parC*, forming the partitioning complex and autoregulating the *par* operon (Dam and Gerdes, 1994). Electron microscopy (EM) showed that ParR forms ring-shaped structures on *parC* and, in the specific

case of ParR from pB171, these structures are formed by dimers-of-dimers that produce ribbon-helix-helix (RHH) DNA-binding motifs (Møller-Jensen *et al*., 2002). In contrast, the atomic structure of ParM has shown that this protein is closely related to eukaryotic actin. ParM forms a two-stranded helix similar to actin (van den Ent *et al*., 2002; Orlova *et al*., 2007; Popp *et al*., 2008), and produces small bundles of three to five filaments (Salje *et al*., 2009). These bundles are generally located near the edge of the nucleoid, and the number of filaments can depend on the plasmid copy number, as it is suggested that one ParM filament carries one plasmid at each end (Choi *et al*., 2008). The dynamic instability of ParM is a feature reminiscent of eukaryotic microtubules (Gerdes *et al*., 2010). In particular, ParM filaments dynamically grow and shrink to explore the entire cell to bind a ParR-*parC* complex, then form a filament (Garner *et al*., 2004; 2007). This switch from long to short filaments is regulated by ATP hydrolysis, with ParM-ATP monomers forming more stable polymers than ParM-ADP. In fact, ParM-ATP monomers spontaneously polymerise (Gerdes *et al*., 2010). Further ParM-ATP monomers are added at both ends of the polymer, preserving the integrity of the filament. Once the filament end makes contact with the ParR-*parC* complex, stabilisation of the filament occurs (Gerdes *et al*., 2010). In fact, the ParR-*parC* complex forms a clamp that binds to the two terminal ParM-ATP subunits and allows addition of only one subunit to one protofilament at a time. Hydrolysis of ATP to ADP of the ParM subunit of one protofilament occurs, and dissociates it from the ParR-*parC* complex. A new subunit ParM-ATP is then added and the same process happens with another protofilament (Gerdes *et al*., 2010). This progressive addition of ParM subunits induces filament growth and “pushes” the plasmid towards the poles of the cell, until the spindles encounter resistance at the poles (Campbell and Mullins, 2007). For this reason, the mode of action of type II systems is considered to follow a pushing model (**Figure 1.7B**).

Partitioning systems in Shigella

There are two plasmid partitioning systems on pINV, ParAB and StbAB (Pilla and Tang, 2018). ParAB belongs to type I family and is highly homologous to ParAB from P1 plasmid (amino acid identity, ParA 75%; ParB 58%) (Buchreiser *et al*., 2000). StbAB belongs to type II family and is similar to the StbAB system from R100 plasmid (amino acid sequence identity, StbA 39%; StbB 29%) (Buchreiser *et al*., 2000). Interestingly, both systems are present on the pINV from *Shigella* spp. and are highly conserved, except *S. sonnei* which has lost *stbAB* (McVicker *et al*., 2019). ParAB from *S. flexneri* pWR100 is active and contributes to pINV partitioning (Sergueev *et al*., 2005). Furthermore, although the ParA, ParB and *parS* elements of this system show some sequence differences with the corresponding elements from the P1 plasmid, they are interchangeable (Sergueev *et al*., 2005). In addition, both ParAB and StbAB are functional at 37°C and 21°C when they are introduced into a test vector containing the origin of replication of pINV (McVicker *et al*., 2019).

1.2.1.3.3 Toxin-antitoxin systems

Toxin-antitoxin (TA) systems are another mechanism for plasmid maintenance that, in contrast to partitioning systems, ensures vertical transmission of plasmids to daughter cells by acting after cell division through a mechanism known as post-segregational killing (PSK). TA systems are usually composed of a protein toxin and a cognate antitoxin. The antitoxin is less stable than the toxin as it is specifically targeted by cellular proteases, belonging to Lon or Clp families (Brzozowska, and Zielenkiewicz, 2013). However, the antitoxin is produced at higher levels so that once degraded, it is quickly replenished (Gerdes and Maisonneuve, 2012). When the plasmid is inherited by a daughter cell, antitoxin expressed by the plasmid counteracts the activity of the stable toxin. If the plasmid is lost along with the genes encoding a TA system,

residual antitoxin in the daughter cell is degraded by proteases and is no longer replaced, so free toxin can kill the cell, leading to PSK (**Figure 1.8A**) (Pilla and Tang, 2018).

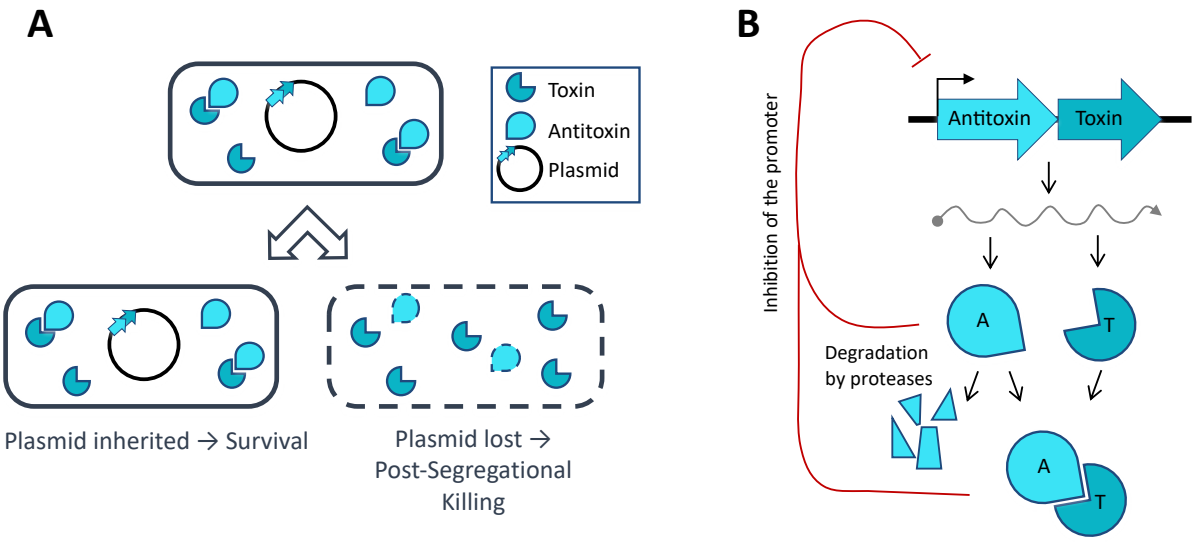


Figure 1.8 Post Segregational Killing (PSK) mediated by plasmid-encoded type II TA systems and regulation of TA operons.

(**A**) Type II TA systems are composed of a gene expressing a stable toxin protein and an unstable antitoxin protein, which is translated at higher levels. When the plasmid is present, the antitoxin is expressed and can bind the toxin, preventing it from associating with its cellular target. If the plasmid is lost, the unstable antitoxin is degraded by proteases and is not replaced, leaving the unbound toxin to kill the cell. (**B**) Both toxin and antitoxin genes are transcribed from the same operon with the antitoxin gene preceding the toxin gene. The TA complex regulates its own operon by a process known as conditional cooperativity.

TA systems are classified into six groups depending on the nature of the antitoxin and its mechanism of action (Lee and Lee, 2016). Type I systems are composed of a small RNA antitoxin, which has the ability to bind the mRNA of the toxin, inhibiting its translation. In type II systems, such as those found on *Shigella* pINV, both toxin and antitoxin are proteins, which form a protein-protein complex, resulting in toxin neutralisation. Type III systems are formed by an RNA antitoxin that inhibits the toxin by associating directly with it. In type IV systems, although both toxin and antitoxin are proteins, they do not directly interact but have opposing

activities on the same cellular target. In type V systems, the antitoxin is an endoribonuclease that cleaves the toxin mRNA, preventing its translation. Finally, type VI systems are composed by a protein antitoxin that promotes degradation of the toxin by binding to the toxin as an adaptor for proteolysis.

Biological functions of TA systems

When TA systems were first discovered, they were described as "addiction modules", as they were found to prevent loss of plasmids by PSK (Ogura and Hiraga, 1983; Gerdes *et al*., 1986). However, subsequent studies revealed two other important functions, abortive infection and persistence/tolerance (Gerdes *et al*., 2005; Gerdes and Maisonneuve, 2012; Harms *et al*., 2018). During abortive infection, TA systems inhibit bacteriophage propagation in a bacterial population by inducing altruistic suicide of infected cells. The TA systems Hok-Sok, MazEF and ToxIN are three examples of TA systems that significantly inhibit phage replication (Pecota and Wood, 1996; Fineran *et al*., 2009; Alawneh *et al*., 2016). In tolerance and persistence, bacteria survive a transient stress condition, such as exposure to antibiotics, by slowing down or stopping growth (thereby becoming dormant), allowing survival in hostile environments (Brauner *et al*., 2016). Tolerance generally involves the entire bacterial population while persistence affects a subpopulation (Brauner *et al*., 2016; Page and Peti, 2016; Harms *et al*., 2018). Furthermore, tolerance can be inherited, while persistence is a stochastic process (Brauner *et al*., 2016; Harms *et al*., 2018). Several pathways are implicated in the phenotypic transition into dormancy which can be mediated by TA systems. The RelE and MazF toxins can induce formation of persister cells in *E. coli* during exposure to against β-lactams, quinolones, and aminoglycosides (Tashiro *et al*., 2012; Tripathi *et al*., 2014), and the PhD-Doc and TacAT system have been linked to persistence in *Salmonella enterica* serovar Typhimurium (Helaine and Kugelberg, 2014).

Type II TA systems

Type II TA systems are the most widespread and characterised TA systems (Lee and Lee, 2016). Generally, type II toxin and antitoxin genes are part of the same operon with the antitoxin gene preceding the toxin gene. Due to this organisation, antitoxins are often more efficiently expressed than toxins (Harms *et al*., 2018). However, in stress conditions, the antitoxins are more quickly degraded by proteases, such as Lon or Clp (Chan *et al*. 2016).

Based on structure, type II toxins can be classified into nine families (ParE/RelE, MazF, HicA, VapC, HipA, FicT/Doc, AtaT/TacT, Zeta, and MbcT), while type II antitoxins are divided into four categories according to the N-terminal domain, which mediates binding to the TA promoter (helix-turn-helix (HTH), ribbon-helix helix (RHH), YefM/PhD-type, and VapB-type) (Zhang *et al*., 2020).

Transcriptional autoregulation in TA loci has been well described especially for type II TA systems, through a mechanism termed conditional cooperativity (Overgaard *et al*. 2008; Garcia-Pino *et al*. 2010). Binding of TA complexes to operator DNA occurs only when the toxin is in an optimal stoichiometric ratio with its cognate antitoxin (**Figure 1.8B**) (Hayes and Kędzierska, 2014; Cataudella *et al*., 2012), with the antitoxin in the complex binding to the operator in the TA locus promoter, partially inhibiting transcription. However, the DNA binding of the antitoxin is enhanced by the presence of the toxin, which behaves as a co-repressor. Therefore, intermediate ratios of toxin increase this affinity significantly whereas an excess of the toxin causes the release of the complex from the operator binding site. This mechanism fine tunes expression of TA systems, ensuring activation of the system when needed (Cataudella *et al*., 2012; Harms *et al*., 2018).

Plasmid TA systems in Shigella

Three plasmid-encoded TA systems have been characterised in *Shigella* spp. so far: VapBC, CcdAB and GmvAT. All TA systems are highly conserved in *Shigella*, with the exception of CcdAB and GmvAT, which are absent from *S. sonnei* (McVicker and Tang, 2016; Pilla and Tang, 2018;). In particular, *ccdAB* has been deleted from the *S. sonnei* pINV, with a remnant of the first 23 bp of *ccdA* remaining; in contrast, *gmvAT* has been entirely lost as part of a 4.9 kb deletion (McVicker and Tang, 2016). Loss of these systems has been detected across the entire phylogenetic tree of *S. sonnei*, suggesting that this occurred before or soon after this species emerged from *E. coli* (McVicker and Tang, 2016). Although all three TA systems are functional, only VapBC and GmvAT have been found to play a role in plasmid maintenance in *Shigella* (Sayeed *et al*., 2000; McVicker and Tang, 2016). VapBC is essential for pINV maintenance at 37°C, while GmvAT confers pINV maintenance at environmental temperatures (McVicker and Tang, 2016). However, aside from the influence of temperature, the need for multiple TA systems on pINV and other plasmids remains unclear. Chromosomal TA systems have been found to prevent large-scale deletion of dispensable genomic regions, such as super-integrons, integrative and conjugative element (ICE) and multidrug resistant genomic islands (Szekeres *et al*., 2007; Van Melderen and Saavedra De Bast, 2009; Wozniak and Waldor, 2009; Huguet *et al*., 2016). Several TA systems from super-integrons in various *Vibrionaceae* can stabilise plasmids or large genomic regions in *E. coli* (Christensen-Dalsgaard and Gerdes, 2006; Szekeres *et al*., 2007). This suggests that plasmid-encoded TA systems might increase genetic stability of neighbouring DNA, besides maintaining plasmids.

VapBC

VapBC (Virulence Associated Protein) is the best characterised TA system on pINV. *Shigella* VapC toxin is a site-specific endonuclease that cleaves $tRNA^{fMet}$ in the anticodon loop between nucleotides +38 and +39, *in vivo* and *in vitro* (Winther and Gerdes, 2011). It contains a Pil-T N-terminal (PIN) domain which contains a central β-sheet comprising five parallel β-strands, sandwiched between seven α-helices at both sides (Dienemann *et al*., 2011). Three conserved acidic residues (Asp7, Glu42 and Asp98) are clustered at the active site. The antitoxin, VapB, has an N-terminal four β-strand domain which mediates dimerisation with another VapB monomer, forming a DNA binding domain (Dienemann *et al*., 2011). Four copies of VapBC form a large symmetric complex of about 100 kDa. VapB binds to VapC through an extended C-terminus with the side chains of Arg64 and Gln66 forming interactions with conserved VapC residues. This interaction is thought to destabilise the VapC active site, displacing the divalent metal ions (Mn^{2+} or Mg^{2+}) that are coordinated by acidic residues at the active site, and required for their cleavage (Dienemann *et al*., 2011). The *vapBC* locus is located approximately 40 kb from the *ori* on *Shigella* pINV (**Figure 1.3**) and, similar to other TA loci, consists of an operon with the antitoxin gene preceding the toxin gene (Sayeed *et al*., 2000). *vapBC* expression is controlled by conditional cooperativity: the *vapBC* promoter region contains two VapBC binding sites, overlapping with the -10 and -35 sequences. Binding of the VapBC complex inhibits *vapBC* transcription (Robson *et al*., 2009; Dienemann *et al*., 2011). Furthermore, VapBC activity relies on the protease Lon (McVicker *et al*., 2019).

Interestingly, *vapBC* loci are widely distributed among pathogenic bacteria species (*e.g. Mycobacterium tuberculosis* possesses 47 *vapBC* loci) but its evolutionary benefit is still not clear. VapBC is also present on the *E. coli* F plasmid, with only a few nucleotide differences between *vapBC* in *Shigella* and *E. coli*, which lead to conservative substitutions, except for an

Arg to Leu change at position 89 in VapC (Radnedge *et al*., 1997). VapBC is also highly conserved in several *Salmonella* isolates. Originally found in the *Salmonella dublin* virulence plasmid and successively characterised in *S.* Typhimurium, VapBC contributes to the maintenance of pSLT, survival inside fibroblasts, and confers fitness to bacteria infecting epithelial cells (Lobato-Márquez *et al*., 2016). When compared with the locus in *S. flexneri*, VapBC is highly conserved in all the other *Shigella* spp., with two point mutations in *S. sonnei* and *S. boydii*, one in the antitoxin and one in the toxin gene (Thr58Ala and Lys32Arg, respectively); whilst, several mutations are present in *S. dysenteriae* (Thr58Ala and Gly68Asp for the antitoxin; Lys32Arg, Thr78Ile and Thr128Ile for the toxin) (Yang *et al*., 2005).

CcdAB

CcdAB, first discovered in *E. coli* F plasmid (Bernard and Couturier, 1992), is the TA system on pINV which is closest to the T3SS PAI, 42 kb away (**Figure 1.3**) (Venkatesan *et al*., 2001). CcdB targets the GyrA subunit of DNA gyrase, which converts relaxed DNA into a negatively supercoiled form (Menzel and Gellert *et al*., 1987; Bernard and Couturier, 1992). In particular, the topoisomerase acts by causing transient DNA breaks through which DNA can pass, increasing coiling. During this step, the enzyme binds to the termini of the cleaved DNA, forming a transient complex. The complex is then resolved and resealed by gyrase (Maxwell and Gellert, 1984; Menzel and Gellert *et al*., 1994). CcdB traps the gyrase in this transient form, preventing GyrA from re-joining double-stranded DNA breaks (Bernard *et al*., 1992). CcdA has a bi-functional role: it specifically antagonises the toxin, interacting with CcdB through its C-terminal region, and autoregulates *ccdAB* operon. Its anti-toxicity derives from its ability to form a high affinity $CcdA_2$-$(CcdB_2)_2$ complex: the antitoxin removes the toxin from the GyrA-CcdB complex, thus reactivating DNA gyrase (Bahassi *et al*., 1999; Dao-Thi *et al*., 2002). CcdA is metabolically unstable due to its sensitivity to Lon-dependent proteolysis (Van

Melderen *et al*., 1994). Transcriptional regulation results from the interaction of $CcdA_2CcdB_2$ complex with DNA, with CcdA specificity recognising three binding sites (Tam *et al*., 1989; Madl *et al*., 2006). The CcdAB system is well conserved in *Shigella* spp. and in other species like *Salmonella enterica*, *E. coli* and *Klebsiella oxytoca*. Although this system is functional in *E. coli*, its contribution to pINV maintenance is not clear, as it was shown that its deletion only has minor effect on the stability of the *S. flexneri* pINV at 37°C in the absence of VapBC (McVicker and Tang, 2016). Interestingly, in *S. sonnei*, only the first 23 bp of *ccdA* are present. Introduction of *ccdAB* into the *S. sonnei* pINV enhanced stability at 37°C but not at 21°C, suggesting that CcdAB might have temperature-dependent activity (McVicker and Tang, 2016).

GmvAT

Recent bioinformatic analysis revealed a third TA system on *Shigella* pINV: GmvAT (GNAT maintenance of virulence). This system is predicted to be composed of the antitoxin GmvA and the toxin GmvT, the latter hypothesised to be a Gcn5-related N-acetyltransferase (GNAT) (McVicker and Tang, 2016). This family of acetyltransferases is widespread in eukaryotes and prokaryotes, and acetylates a large number of substrates using Acetyl-Co-enzyme A (AcCoA) as the acetyl donor. Many roles have been described for bacterial GNATs including regulation of bacterial drug resistance, biofilm formation, regulation of transcription, stress response, virulence and metabolic flux (Xie *et al*., 2014). However, only two homologues of this TA system have been characterised, TacAT in *S. enterica* serovar Typhimurium and AtaRT in *E. coli* EHEC O157:H7, both of which are chromosomally encoded (Cheverton *et al*., 2016; Jurenas *et al*., 2017). Like GmvT, AtaT and TacT are GNAT enzymes that inhibit translation by acetylating the initiator $tRNA^{fMet}$ and elongator tRNAs, respectively. TacT has been shown to dimerise, forming a binding site for the substrate. Alternatively, dimerisation may allow acetylation of

two substrates. Furthermore, both GNAT TA systems auto-regulate their transcription (VanDrisse *et al*., 2017; Jurenas *et al*., 2017). At present, the specific target, the structure and role of GmvT in *Shigella* is unclear. Studies on the homologue TacT toxin showed that, following acetylation, the α amine group blocked by the acetyl cannot form a peptide bond with the carboxyl group of the amino acid loaded on the tRNA in the ribosomal P site, thereby blocking translation (Cheverton *et al*., 2016). Further analysis in *S. flexneri* showed that when bacteria are grown at 21°C, GmvAT is sufficient for plasmid retention. Furthermore, like VapB, GmvA is subject to Lon mediated degradation (McVicker *et al*., 2019). In addition, GmvAT is not found in all *Shigella* spp.: indeed, it is absent in *S. sonnei*, and the genes in *S. boydii* and *S. dysenteriae* have a few nucleotide changes compared with *S. flexneri* (for the *S. boydii* toxin, Ile57Val and Thr98Ile; for the *S. dysenteriae* toxin, Ile57Val; for the antitoxin, Val3Phe) (Yang *et al*., 2005).

1.2.1.4 Insertion Sequences

Insertion sequences (IS) represent an important component of bacterial genomes. Together with transposons, integrons and introns, IS are members of the “mobilome” of the genome, which consists of the mobile part of the genome that can be exchanged either intramolecularly, or between different DNA molecules within the same bacterial cell, or intercellularly, following the transfer of plasmid/bacteriophages into new hosts (Siguier *et al*., 2014).

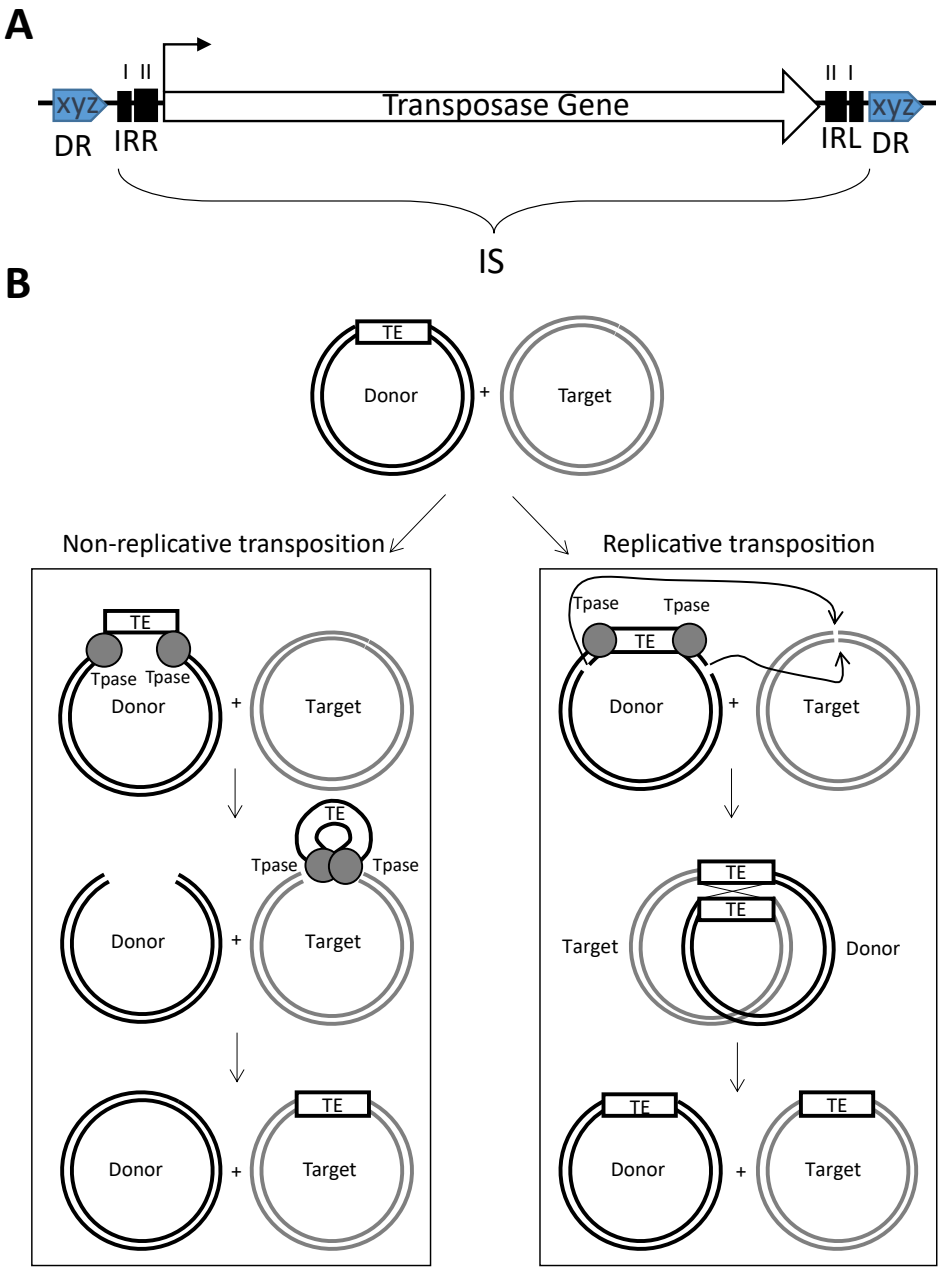


Figure 1.9 Structure of IS and transposition

(**A**) IS are composed of a transposase gene and two inverted repeats located at the right and left end of the IS (IRR and IRL), which are divided in two functional domains, I and II. Two direct repeats (DR) enclosing the IS are generated upon insertion (modified from Mahillon and Chandler, 1998). (**B**) Transposition of transposable elements (TE) can be non-replicative or replicative. In the non-replicative transposition the TE is excised by the transposase (Tpase) from the donor DNA and is integrated into the target DNA following nucleophilic attack by the TE. In the replicative transposition, the Tpase nicks the TE and ends attack the target sequence, producing a break. Strand transfer reaction occurs forming a co-integrate, which is eventually resolved into two molecules by recombination. As a result of this process, both the donor and the target sequence have a copy of the TE.

Structure

IS are defined as small transposable DNA elements, and only harbour genes involved in mobility, such as transposases (Tpase) and their regulatory elements (Mahillon and Chandler, 1998). In fact, in contrast with transposons, IS do not carry any accessory genes (for example, those encoding antibiotic resistance). Furthermore, the majority of IS have short terminal inverted-repeat sequences (IR) of 10-40 bp that are functional parts of the IS and are divided in two functional domains (**Figure 1.9A**). Domain I is and is involved in cleavage and strand transfer reactions that lead to transposition of the IS; domain II faces internally and is involved in Tpase recognition and binding (Mahillon and Chandler, 1998). IS*91*, IS*110*, and IS*200*/*605* families do not present this structure. Furthermore, the IR upstream the Tpase gene (IRL) can overlap with the promoter of the Tpase, suggesting that transcription of the Tpase gene might be autoregulated by Tpase binding (Mahillon and Chandler, 1998). Another general feature that applies to most of IS is the presence of short directly repeated sequences (DR) of the target DNA that flank the IS following insertion (**Figure 1.9A**). The length of the DR is generally between 2 and 14 bp and is IS-specific (Mahillon and Chandler, 1998). Homologous inter- or intra-molecular recombination between two IS elements, each with a different DR, will produce a hybrid element carrying one DR from each IS (Mahillon and Chandler, 1998). IS have been classified according to the (i) genetic organisation (*e.g.* number of ORFs they contain), (ii) structural domains in their Tpases, (iii) sequence of the IR, and (iv) the DR length. Based on these classifications, there are approximately 30 families of IS (Siguier *et al*., 2014).

Transposition

Transposition is a DNA recombination process that consists of the translocation of a transposable element from a donor site to a non-homologous target site. Transposition results

in DNA rearrangements, including inversions, deletions and replicon fusions, potentially causing large genome alterations. Transposition is catalysed by the Tpase, which are often multidomain proteins generally composed of a DNA-binding and a catalytic domain. The DNA binding domain is involved in the recognition of IR while the catalytic domain mediates DNA cleavage and strand transfer (Siguier *et al*., 2014). Tpases form multimeric nucleoprotein complexes, also known as transposomes, that can be composed by two or more Tpase monomers and can host accessory proteins, such as IHF and H-NS (Mizuuchi, 1992a, b; Haniford, 2006), and chaperones (Ton-Hoang *et al*., 2004) depending on the mode of transposition. As mentioned, IS families can be defined by the Tpase type. In fact, the nature of Tpase changes depending on the type of the chemical reaction they catalyse and employ for transposition. Tpases can belong to the DDE, DEDD, HUH and Serin groups, depending on the key amino acid residues at the catalytic site.

There are two form of transposition (**Figure 1.9B**): (i) non-replicative or "cut-and-paste" transposition, and (ii) replicative transposition. In non-replicative transposition, the transposable element is excised from the donor DNA molecule by the activity of the Tpase and is integrated into the target DNA following nucleophilic attack by the 3' end of the transposable element (Muñoz-López and García-Pérez, 2010). Host polymerases and ligases repair gaps between the transposable element and the target DNA. In replicative transposition, the transposase nicks the transposable element *via* a nucleophilic attack and the so ends attack the target sequence, producing a break. Strand transfer occurs (forming a strand transfer intermediate), and the host polymerase is then recruited to synthesise the missing DNA strand with ligase repairing any remaining nick, resulting in the formation of a co-integrate. Eventually, recombination resolves the two molecules, producing two copies of

the transposable element, one in the donor and one in the recipient site (Tavakoli and Derbyshire, 2001).

Type of transposition mechanism, number and type of DNA nicks generated in both the donor and the target sites (*e.g.* single / double stranded nick at one or both ends of the IS), vary according to the IS family (Mahillon and Chandler, 1998; Tavakoli and Derbyshire, 2001).

Impact on bacterial genome

IS are abundant in prokaryotic genomes and Tpases are the larger class of proteins in both prokaryotes and eukaryotes (Aziz *et al*., 2010). Genomic analysis has revealed that bacterial genomes often contain a high number of truncated IS that may represent the scar of ancestral transposition events (Siguier *et al*., 2014). Furthermore, the distribution of IS may be genera/species-specific, with some IS families limited to particular groups of bacteria (*e.g.* IS*1* in enterobacteria) (Mahillon and Chandler, 1998).

IS have profound changes on genome architecture. Given the abundance of IS in bacterial genomes, they are often substrates of several molecular rearrangements (Vandecraen *et al*., 2017). Homologous copies of the same IS type can flank specific sets of genes, mediating their horizontal transfer or deletion, therefore, acting as vectors. Furthermore, IS can insert inside a gene, inactivating it, or upstream of a gene regulating its expression. The latter can occur in two ways: i) the transcription of the IS can be extended into the downstream gene (Glansdorff *et al*., 1981); ii) the IS promoter can form a hybrid with the promoter of the nearby gene (Siguier *et al*., 2014). Many IS have -35 promoter elements located at the end of the IS, oriented outward (Prentki *et al*., 1986). If located at the right distance from a -10 box of a neighbouring gene, this can produce a strong promoter. In several cases, IS has affected antibiotic resistance by inserting inside genes encoding porins, thus reducing drug import (Vandecraen *et al*., 2017). Furthermore, IS insertions can also modulate indirectly antibiotic

resistance by affecting the expression of regulators, as well as virulence and metabolic pathways (Vandecraen *et al*., 2017).

Analysis of IS databases indicates that IS are more abundant in plasmids than chromosomes. In fact, it has been hypothesised that plasmids are an important reservoir of IS, and a major vector of IS transmission as well as transmission of accessory genes, such as virulence and AMR genes, which can be mediated also by IS (Siguier *et al*., 2014).

The abundance of IS can reflect the lifestyle of the bacterial host. For instance, IS expansion has been observed particularly in bacteria with host-restricted lifestyles or those which are adapting to an intracellular lifestyle (Moran and Plague, 2004; Siguier *et al*., 2014). In fact, in a rich nutritional environment, such that found intracellularly, many genes needed for free bacteria become dispensable and even detrimental (*e.g*. genes encoding high immunogenic elements, in the case of intracellular pathogens) (Casadevall, 2008; Siguier *et al*., 2014). By mediating transpositions and molecular rearrangements, IS can contribute to gene inactivation/loss. Following expansion, IS can mediate intrachromosomal recombinations and deletions that result in subsequent genome reduction, eventually eliminating IS themselves (Andersson and Andersson, 1999; Mahillon and Chandler, 1998; Lawrence *et al*., 2001; Mira *et al*., 2001). This will eventually produce an IS-free genome that can be "re-infected" by other IS.

IS in Shigella *spp.*

Shigella spp. are particularly enriched for IS (Yang *et al*., 2005). For instance, in *S. flexneri* strains Sf301 and 2457T, IS represent ~7% of their genomes (in contrast to <1 % in *E. coli*) and there are more than 300 pseudogenes disrupted by IS elements (Jin *et al*., 2002; Wei *et al*., 2003). Furthermore, IS are particularly abundant in pINV. For example, the *S. flexneri* pWR501 plasmid contains ~153 IS elements, representing approximately 53% of pINV ORFs

(Venkatesan *et al*., 2001). IS have significantly contributed to the evolution of *Shigella* spp. (Yang *et al*., 2005). In *Shigella*, independent IS expansion has been followed by genome reduction, resulting in the convergent evolution of different *Shigella* spp. which share common pathogenic traits (Hawkey *et al*., 2020). Interestingly, IS still contribute to continuing diversification of *S. sonnei* and *S. flexneri*. In all *Shigella* spp., almost all conserved pseudogenes have been disrupted by IS and analysis of the *Shigella* GEMS genomes has revealed that IS disrupted around 32-47% of metabolic genes (Hawkey *et al*., 2020). However, IS have also mediated the acquisition of important virulence elements. Particularly for pINV, the mosaic organisation of virulence genes separated by blocks of IS elements suggests that these regions have probably been acquired from other species and IS have contributed to the assembly of these blocks (Venkatesan *et al*., 2001). Furthermore, when pINVs from different species are compared, it is evident that some of these blocks have been rearranged, producing different gene organisations (Yang *et al*., 2005). The T3SS PAI is an example of an IS-mediated acquired element. This genetic region is highly homologous to the *yscM-yopD* gene cluster on the *Yersinia* low-Ca^{2+}-response (LCR) plasmid pCD1 and, interestingly, a partial copy of the *Yersinia* IS*100* is located also at one end of the *Shigella* PAI (Perry *et al*., 1998; Venkatesan *et al*., 2001). At the other end of the PAI there is a sequence of 139 bp which is highly homologous to the start sequence of IS*285* that borders the *yscM-yopD* cluster in pCD1. Another cluster of pINV-encoded genes whose acquisition is potentially IS-mediated is the O-Ag cluster on *S. sonnei* pINV, which contains genes essential for the synthesis (*wbg*), the transport (*wzx*) and polymerisation (*wzy*) of the O-Ag (Xu *et al*., 2002). This cluster is unique to *S. sonnei* and is identical to an O-Ag biosynthesis cluster in serotype 17 of *Plesiomonas shigelloides* (Jiang *et al*., 2005; Lai *et al*., 1998, Shepherd *et al*., 2000). However, in contrast to the O-Ag of *P*.

shigelloides, there is a IS*630* that divides the cluster in two, but without disrupting a gene (Jiang *et al*., 2005).

1.2.2 Chromosomal virulence genes

Several genes encoding IpaH T3SS effectors are located on the chromosome. Interestingly, expression of these genes is activated by the plasmid-encoded factor, MxiE. The IpaHs act to downregulate inflammatory responses during infection (Ashida *et al*., 2007). However, the number and types of chromosomal IpaHs vary between strains; it is hypothesised that some chromosome- and plasmid-encoded IpaHs are functionally redundant. Alternatively, the number of *ipaH* genes may affect the extent of host inflammatory responses to the bacterium, which can vary depending on the infecting strain (Ashida *et al*., 2007).

1.2.3 Regulation of virulence

The assembly of and secretion through the pINV-encoded T3SS are tightly regulated, as continual expression of the T3SS would be energetically costly for the bacterium. Therefore, *Shigella* employs a regulatory network which senses environmental changes and activates the expression of T3SS machinery when required.

Several factors influence T3SS expression, including temperature, pH, osmolarity, oxygen and iron concentrations (**Figure 1.10**) (Porter *et al*., 1994; Nakayama *et al*., 1995; Payne *et al*., 2006; Marteyn *et al*., 2012) with the main trigger being temperature. In particular, when the temperature reaches 37°C, the average temperature in the human body, a regulatory cascade is activated, principally involving two transcriptional activators, VirF and VirB (Durand *et al*., 2000). VirF is a member of the AraC family of transcriptional activators and is encoded outside the T3SS PAI. The expression of this gene is repressed by H-NS (Histone-like Nucleoid-

Structuring Protein) and activated by FIS (Factor for Inversion Stimulation protein). When the temperature is < 28°C (Prosseda *et al*., 2004), H-NS binds two sites within the *virF* promoter (Prosseda *et al*., 1998). These sites are separated by a DNA linker sequence that curves at lower temperatures, facilitating the formation of an H-NS temperature-sensitive nucleoprotein complex, which represses *virF* transcription (Prosseda *et al*., 2004, 2010). In fact, H-NS binding site I overlaps the -35 and -10 boxes of the *virF* promoter, preventing access of RNA polymerase. Binding site II is located far upstream and H-NS binding stabilises the nucleoprotein complex (Di Martino *et al*., 2016). H-NS access to the *virF* promoter relies on the temperature-dependent changes in topology of DNA in this region (Falconi *et al*., 1998). When the temperature increases, the DNA curvature at the *virF* promoter sequence progressively reduces. This disrupts the H-NS nucleoprotein complex, allowing FIS binding. FIS recognises four sites in the *virF* promoter region, two of which overlap with H-NS sites (Dorman, 2004). Another transcriptional factor, IHF (Integration Host Factor), also contributes to the activation of *virF* transcription by binding to the promoter (Porter and Dorman, 1994).

Expression of *virF* is also regulated at a translational level. VirF translation depends on the presence of two modified tRNA nucleosides (queuosine and 2-methylthio-N6-isopentenyladenosine). These tRNA modifications are necessary for VirF to reach levels that trigger activation of the regulatory cascade (Durand *et al*., 1994, 1997).

VirF binds directly upstream of the *virB* promoter at positions −17 to −105 (Tobe *et al*., 1995). VirB is a member of the ParB family of partitioning proteins and acts directly or indirectly on T3SS PAI-associated genes (Adler *et al*., 1989). *virB* is located just inside one end of the T3SS PAI. Like *virF*, *virB* is repressed by H-NS, which binds the *virB* sequence spanning positions −20 to +20 (Beloin and Dorman, 2003). However, VirF is essential for optimal *virB* expression (Tobe

et al., 1995), although the presence of VirF is not sufficient to activate *virB* transcription, with both a temperature shift and high-salt conditions required (Tobe *et al*., 1995). Thus, under specific conditions of temperature and osmolarity, VirF interacts with RNA polymerase and induces transcription when the DNA is negatively supercoiled and, thus, not bound by H-NS (Porter *et al*., 2002). Binding of IHF upstream of the VirF binding site further activates *virB* expression (Porter and Dorman, 1994). Expression of *virB* is also post-transcriptionally regulated, *via* the small regulatory RNA, RyhB (Murphy and Payne, 2007). Finally, VirB induces transcription of genes inside the T3SS PAI and some additional *osp* genes scattered across pINV, which encode the "first-set" of effector proteins (**Figure 1.10**) (Gall *et al*., 2005). VirB regulates transcription by interacting with *parS*-like sequences of the promoter of its target genes and polymerises along the DNA to counteract H-NS binding (Turner and Dorman, 2007).

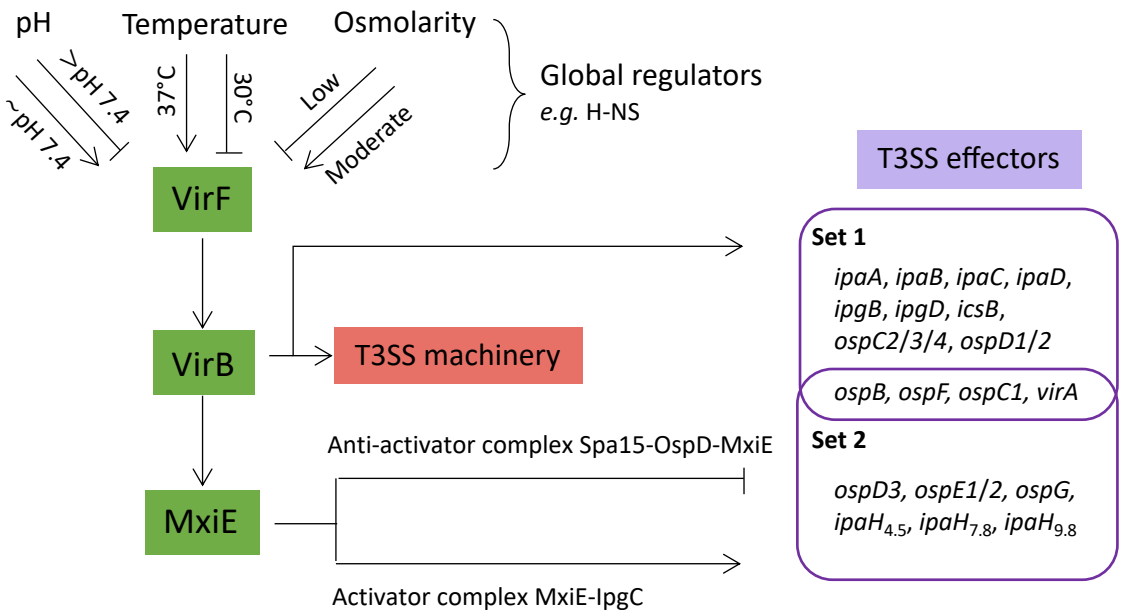


Figure 1.10 Regulation of virulence in *Shigella*.

Expression of *virF* is regulated by several signals such as pH, temperature and osmolarity through global regulators (*e.g.* H-NS). VirF, activates the expression of *virB*, which, in turn, promotes the expression of the genes within the PAI and additional genes located elsewhere in pINV (Set 1 genes). Set 1 genes include *mxiE*; MxiE activates expression of Set 2 genes once Set 1 effectors have been secreted (modified from Schroeder and Hilibi, 2008).

The presence of the VirF-VirB regulatory cascade is essential for virulence activation in *Shigella*. It is not clear why this two-step transcriptional activation system has emerged; however, it is plausible that this mechanism offers two regulatory checkpoints for T3SS expression (Kane and Dorman, 2012) and/or serves to amplify the activation of virulence.

Although the secretion machinery is assembled and the molecular effectors are expressed as soon as the temperature reaches 37°C, the T3SS apparatus remains inactive. In this inactivated form, substrates of the secreton are stored in the cytoplasm, in association with chaperones to prevent their degradation or inappropriate interactions (Menard *et al*., 1994; Page *et al*., 2001). Only upon direct contact with the host cell or exposure of the bacterium to Congo red (CR) dye is secretion induced and transcription of T3SS effectors increases (Menard *et al*., 1994; Parsot *et al*., 1995; Baharani *et al*., 1997).

At this step, another important regulator is MxiE, a transcriptional activator of the AraC-like family, controls the expression of the "second-set" effectors (**Figure 1.10**) (Allaoui *et al*., 1993). Secretion of IpaB and IpaC releases their chaperone IpgC, while secretion of OspD1 induces its dissociation from the chaperones Spa15 and MxiE, the latter then forming a complex with IpgC, triggering transcription of several virulence-associated genes located outside the PAI, such as *virA* and *ipaH* (Mavris *et al*., 2001; Parsot *et al*., 2005).

1.3 CRISPR-Cas systems and *Shigella*

The discovery of CRISPR-Cas has revolutionised genome engineering, and more recently CRISPR-Cas systems have been employed to target genes involved in antimicrobial resistance or virulence in bacteria (Bikard *et al*., 2014; Doudna and Charpentier, 2014; Kim *et al*., 2016; Rodrigues *et al*., 2019). CRISPR-Cas systems consist of RNA-guided nucleases that confer bacteria with adaptive immunity against infection with phages and other mobile genetic elements (Barrangou *et al*., 2007; Brouns *et al*., 2008; Marraffini and Sontheimer, 2008; Barrangou and Marraffini, 2014; Bondy-Denomy and Davidson, 2014).

1.3.1 Activity of CRISPR-Cas systems

CRISPR-Cas systems are composed of i) Cas proteins, which are important for acquisition of spacers, processing of CRISPR RNA (crRNA), and cleavage of target sequences, ii) the trans-activating crRNA (tracrRNA), which forms a duplex with the crRNA to direct DNA-cleavage, and iii) the crRNA, which is homologous to the target sequences (protospacer) (van der Oost *et al*., 2014). The action of CRISPR-Cas is usually divided into three stages: i) spacer acquisition, which is generally mediated by Cas1 and Cas2, and consists of the incorporation of small fragments of foreign DNA (protospacer), which are recognised, processed into sequences (spacers) and inserted into the CRISPR-array between repeated sequences (McGinn and Marraffini, 2018); ii) processing and maturation of the CRISPR locus transcript (pre-crRNA), which can be mediated by different factors (*e.g.* CRISPR-Cas associated proteins like Cas6, and/or housekeeping ribonucleases) (Deltcheva *et al*., 2011); iii) assembly of the crRNP complex, which consists of crRNA, tracrRNA and Cas proteins, and scans invading DNA for target sequences and mediates degradation (van der Oost *et al*., 2014).

CRISPR-Cas systems are classified into three main types and several subtypes, depending on the CRISPR repeat sequences, the *cas* gene, and proteins involved in crRNA processing, crRNP assembly and DNA targeting (Makarova *et al*., 2011).

In type II systems, Cas9 with the crRNA and the tracrRNA forms a complex which recognises the protospacer sequence in target DNA through base pairing with the crRNA; Cas9 introduces a double-strand break (DSB) in the protospacer three base pairs adjacent to the PAM (protospacer adjacent motif) sequence (**Figure 1.11**) (van der Oost *et al*., 2014; Knott and Doudna, 2018).

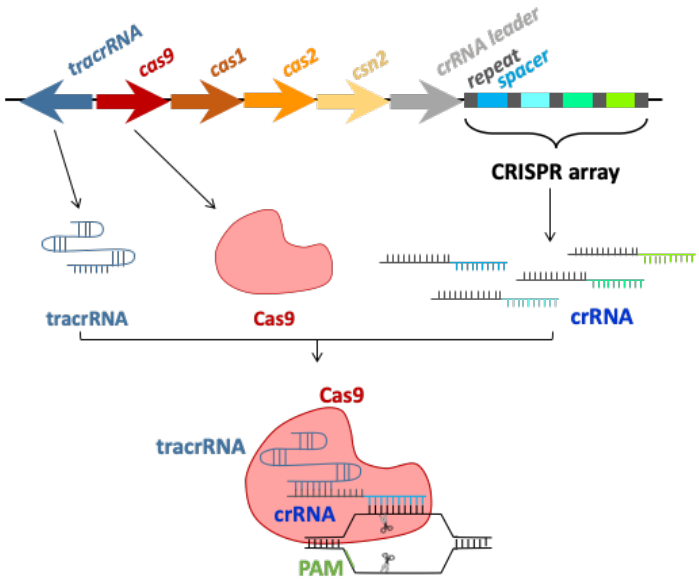


Figure 1.11 Schematic of CRISPR-Cas9.

The Type II CRISPR-Cas locus is composed of *tracrRNA*, *cas9* and other *cas* genes (*cas1*, *cas2* and *csn2*) and a CRISPR array that contains spacer sequences alternating with repeats. Cas9, tracrRNA and the crRNA form a complex that recognises the target DNA sequence through base pairing with the crRNA, and the Cas9 then introduces a DSB three base from the PAM sequence.

Type II CRISPR-Cas systems are the most characterised and extensively used for gene engineering, as they only require Cas9 to recognise and cleave DNA in an RNA-guided manner (**Figure 1.11**) (Gasiunas *et al*., 2012; Jinek *et al*., 2012). Additionally, the blunt DSB introduced by Cas9 offers the possibility to either introduce artificial point mutations at the target site or,

by using pairs of crRNA, generate large deletions or genomic rearrangements (Sander and Joung, 2014). In fact, DSB can be perfectly repaired *via* homologous recombination (HR), when there is a homologous sequence elsewhere in the genome or a double-stranded oligonucleotide is added which is homologous to the protospacer but carrying the desired mutation. If not, the non-homologous end joining (NHEJ) repair pathway can directly repair the DNA break without homologous sequences. This will result in either a faithful repair (with the DNA ends directly re-joined), or in mutagenic repair when the ends are remodelled by nucleases or polymerases before being sealed by ligation (Shuman and Glickman, 2007).

1.3.2 CRISPR-Cas systems in *Shigella*

Shigella spp. possess type I-E CRISPR sequences, and a variable number of *cas* genes associated have been found in *Shigella* strains (Guo *et al*., 2015; Wang *et al*., 2016). Although CRISPR-Cas sequences have been extensively analysed across the *Shigella* spp., their function and activity have not been investigated. Deletions, duplications and IS-mediated disruptions are commonly found in *Shigella cas* genes (Wang *et al*., 2016). Moreover, analysis of *Shigella* strains revealed presence of distinct spacer sequences, all of which matched plasmids and phages from a wide range of microorganisms (Yang *et al*., 2015), including the Stx-2 converting phage I, which mediated acquisition of the Stx2 in several clinical *Shigella* isolates (García-Aljaro *et al*., 2006). This suggests that CRISPR sequences may regulate virulence gene acquisition. Furthermore, some of the analysed spacers also share homology with *Shigella* chromosomal sequences (Yang *et al*., 2015). The presence of "self-target" spacers might drive loss of the CRISPR-Cas system, as it could be detrimental.

1.4 Aims

In enteropathogens including *Shigella* spp., plasmids make major contribution to virulence (Pilla and Tang, 2018). Besides carrying genes which are essential for invasion and survival inside the host, virulence plasmids possess mechanisms essential for their replication, segregation, and maintenance in a bacterial population. The virulence plasmid of *Shigella* spp. (pINV) is present in low copy number, and possesses multiple partitioning systems and toxin-antitoxin (TA) systems to guarantee its retention following replication. However, it is unclear why the plasmid needs multiple partitioning systems and TA systems. Furthermore, loss of virulence is often detected in *S. sonnei* and *S. flexneri* when they are grown in the laboratory (Kopecko *et al*., 1980; Maurelli *et al*., 1984; Sasakawa *et al*., 1986; Schuch and Maurelli, 1997; McVicker and Tang, 2016). The molecular events behind the spontaneous virulence loss has not been fully characterised.

The first aim of this work was to characterise the role of TA systems and partitioning systems on the maintenance of pINV, and the contribution of IS in the spontaneous loss of virulence in *S. flexneri*. The bacterium possesses three plasmid-encoded toxin-antitoxin (TA) systems, VapBC, CcdAB and GmvAT, which can mediate post-segregational killing if the plasmid is not inherited following cell division. I evaluated the contribution on plasmid maintenance of VapBC, as this TA system is critical for the retention of pINV. Analysis of chromosome-encoded TA systems in *E. coli*, *Vibrio cholerae* and *S. Typhimurium* (Szekeres *et al*., 2007; Wozniak and Waldor, 2009; Huguet *et al*., 2016) suggests that TA systems can prevent loss of discrete chromosomal regions. Therefore, I investigated if VapBC can act to maintain local sequences on pINV. Given their role in genome plasticity, I also examined the role of IS in the spontaneous

loss of T3SS-related genes in *S. flexneri*. Finally, I decided to assess the impact of the two partitioning systems on *S. flexneri* pINV, as they have not been investigated previously.

The second aim of this work was to evaluate the role of TA systems and IS on pINV maintenance in *S. sonnei*. Although the pINVs of *S. flexneri* and *S. sonnei* are closely related, they differ in their repertoire of TA systems and IS. Therefore, I assessed whether the absence of the two TA systems present in *S. flexneri*, CcdAB and GmvAT, as well as two amino acid substitutions in VapBC affect plasmid maintenance in *S. sonnei*. Additionally, I investigated the involvement of different IS in the spontaneous deletion of T3SS-related genes in *S. sonnei* and compared the number and types of IS on the pINV of *S. sonnei* with those on the pINV of *S. flexneri*.

Finally, I aimed to exploit my findings to design a CRISPR-Cas system to induce loss of virulence in *Shigella,* as the first step in devising a strategy to target *Shigella* spp.. Specifically, my aim was to exploit the CRISPR-Cas9 system to target *virB* and the TA system *vapBC*, two elements which are highly conserved among *Shigella* spp..

2 MATERIALS AND METHODS

2.1 Growth media and strains

Bacterial strains used in this study are shown in **Table 2.1** and were grown in media indicated in **Table 2.2**. Chloramphenicol was used at a final concentration of 5 μg/ml for *S. flexneri* and 20 μg/ml for *S. sonnei*; kanamycin was used at a final concentration of 50 μg/ml and carbenicillin was used at a final concentration of 100 μg/ml. When required, glucose or arabinose were added to media at a final concentration of 0.2% (w/v). Congo red (CR, Sigma Aldrich, St. Louis, USA) was added to tryptic soy agar (TSA) at a final concentration of 0.01 % w/v to make CR-TSA plates.

Table 2.1 List of strains and their relative genotype.

Strain	Genotype	References
M90T	*S. flexneri* 5a wild-type	Sansonetti *et al.*, 1982
BS176	pINV-cured *S. flexneri* M90T	Zychlinsky *et al.*, 1992
GMCT283	M90T *vapBC*^{WT}::*cat* in the wild-type locus	Pilla *et al.*, 2017
GMCT284	M90T *vapBC*^{D7A}::*cat* in the wild-type locus	Pilla *et al.*, 2017
GMCT285	M90T *vapBC*^{WT}::*cat* at position 100,792 bp of pWR100	Pilla *et al.*, 2017
GMCT286	M90T *vapBC*^{D7A}::*cat* at position 100,792 bp of pWR100	Pilla *et al.*, 2017
GPSf12	M90T *vapBC*^{WT}::*cat* in the wild-type locus Δ*recA*	This work
GPSf13	M90T *vapBC*^{WT}::*cat* in the wild-type locus Δ*recD*	This work
GPSf14	M90T *vapBC*^{WT}::*cat* in the wild-type locus Δ*recF*	This work
GMCT283_IVp	M90T *vapBC*^{WT}::*cat* in the wild-type locus, Isolate 4	Pilla *et al.*, 2017
GMCT283_VIp	M90T *vapBC*^{WT}::*cat* in the wild-type locus, Isolate 6	Pilla *et al.*, 2017
GMCT283_IVr1	M90T *vapBC*^{WT}::*cat* in the wild-type locus, Isolate 4 revertant 1	Pilla *et al.*, 2017
GMCT283_IVr2	M90T *vapBC*^{WT}::*cat* in the wild-type locus, Isolate 4 revertant 2	Pilla *et al.*, 2017

GMCT283_VIr1	M90T *vapBC*^{WT}::*cat* in the wild-type locus, Isolate 6 revertant 1	Pilla *et al*., 2017
GMCT283_VIr2	M90T *vapBC*^{WT}::*cat* in the wild-type locus, Isolate 6 revertant 2	Pilla *et al*., 2017
GMCT189	M90T *mxiH*::*sacB-neoR*	McVicker and Tang, 2016
GMCT356	M90T *mxiH*::*sacB-neoR* ∆*stbAB*	McVicker *et al*., 2019
GPSfA23	M90T *mxiH*::*sacB-neoR* ∆*parAB*	McVicker *et al*., 2019
GMCT213	M90T *mxiH*::*sacB-neoR ΔccdAB ΔgmvAT ΔvapBC*	McVicker and Tang, 2016
GMCT365	M90T *mxiH*::*sacB-neoR* ∆*stbAB ΔccdAB ΔgmvAT ΔvapBC*	McVicker *et al*., 2019
GPSfA23	M90T *mxiH*::*sacB-neoR* ∆*parAB ΔccdAB ΔgmvAT ΔvapBC*	McVicker *et al*., 2019
GPSfA31	M90T *mxiH*::*sacB-neoR* Δ*virB*::*cat*	McVicker *et al*., 2019
GPSfA32	M90T *mxiH*::*sacB-neoR ΔstbAB ΔvirB:cat*	McVicker *et al*., 2019
GPSfA35	M90T *mxiH*::*sacB-neoR ΔparAB ΔvirB:cat*	McVicker *et al*., 2019
53G	Wild-type *S. sonnei* 53G	Kopecko *et al.* 1980
GMCT323	53G *vapBC*^{WT}::*cat* in the wild-type locus	This work
GMCT364	53G *ccdAB*::*cat gmvAT*	This work
GPSs12	*S. sonnei* containing *vapBC*^{Sflex}::*cat* in the wild-type locus	Martyn *et al*., unpublished
GPSs15	*S. sonnei* containing *vapB*^{Sflex}*C*^{Ssonn}::*cat* in the wild-type locus	Martyn *et al*., unpublished
GPSs18	*S. sonnei* containing *vapB*^{Ssonn}*C*^{Sflex}::*cat* in the wild-type locus	Martyn *et al*., unpublished
GPSf1	BS176 containing pINV^{Ssonn} *vapBC*^{Ssonn}::*cat* in the wild-type locus	Martyn *et al*., unpublished
GPSf9	BS176 containing pINV^{Ssonn} *vapBC*^{Sflex}::*cat* in the wild-type locus	Martyn *et al*., unpublished
Helper plasmid	*E. coli* containing pRK2013	
CS14	*S. sonnei* CS14 (clinical isolate from Brasil)	Hollingshead *et al*., in preparation
CS14 control^{amp}	CS14 carrying pCas9^{amp}::control	This work
CS14 control^{kan}	CS14 carrying pCas9^{kan}::control	This work
CS14 control^{dual}	CS14 carrying pCas9^{amp}::control and pCas9^{kan}::control	This work
CS14 *virB*^{Pro}	CS14 carrying pCas9^{amp}::*virB*^{Pro}	This work
CS14 *vapBC*^{Pro}	CS14 carrying pCas9^{amp}::*vapBC*^{Pro}	This work
CS14 *virB*^{orf}	CS14 carrying pCas9^{kan}::*virB*^{orf}	This work

CS14 *vapB*^{orf}	CS14 carrying pCas9^{kan}::*vapB*^{orf}	This work
CS14 *virB*^{dual}	CS14 carrying pCas9^{amp}::*virB*^{Pro} and pCas9^{kan}::*virB*^{orf}	This work
CS14 *vapB*^{dual}	CS14 carrying pCas9^{amp}::*vapBC*^{Pro} and pCas9^{kan}::*vapB*^{orf}	This work
CS14 ind^{control}	CS14 carrying ind_pCas9::control	This work
CS14 ind^{vapP}	CS14 carrying ind_pCas9::*vapBC*^{Pro}	This work
CS14 ind^{vapB}	CS14 carrying ind_pCas9::*vapB*^{orf2}	This work
CS14 ind^{vapC}	CS14 carrying ind_pCas9::*vapC*^{orf}	This work

Table 2.2 List of growth media

Type	Brand
Tryptic Soy Broth (TSB)	Sigma Aldrich, St. Louis, USA
Tryptic Soy Agar (TSA): TSB[1] containing 1.5% (w/v) agar[2]	1) Sigma Aldrich, St. Louis, USA; 2) Oxoid, Basingstoke, UK
Lysogeny Broth (LB)	Invitrogen, Waltham, MA
LB solid medium: LB[1] containing 1.5% (w/v) agar[2]	1) Invitrogen, Waltham, MA; 2) Oxoid, Basingstoke, UK
Sucrose solid medium: 1% (w/v) tryptone, 1% (w/v) yeast extract and 10% (w/v) sucrose (added prior to pouring plates)	Sigma Aldrich, St. Louis, USA
Minimal medium (M9): 47.76 mM Na_2HPO_4 [1], 22.04 mM KH_2PO_4[1], 8.56 mM NaCl[1], 18.69 mM NH_4Cl[1] and 0.4 % (w/v) glucose[1], 12.5 μg/ml nicotinic acid[1], 2.0 mM $MgSO_4$ [1] which are added prior to pouring plates; contains 1.5% (w/v) agar[2]	1) Sigma Aldrich, St. Louis, USA; 2) Oxoid, Basingstoke, UK

2.2 Construction of strains

Lambda Red recombinease was employed for genetic manipulation of pINV (Datsenko and Wanner, 2000). Constructs included approximately 1 kb of sequence upstream and downstream of the gene interest, flanking by an antibiotic resistance cassette (see **Table 2.4** for primer sequences), and were assembled in pUC19 (Figurski *et al*., 1979) using NEBuilder HiFi DNA Assembly master mix (New England Biolabs) (**Table 2.3**). The resulting plasmids were then transformed into *Escherichia coli* DH5α and used as a template to amplify approximately 1 μg of the construct, which was then transformed by electroporation into *S. flexneri* or *S. sonnei* expressing the recombinase encoded by pKD46 (Datsenko and Wanner, 2000).

For manipulation of the chromosome, constructs consisting of approximately 600bp of sequence upstream and downstream of the gene interest (see **Table 2.4** for primer sequences) were assembled into pCONJ4S (Kaniga *et al*., 1991) using NEBuilder HiFi DNA Assembly master mix, and the resulting plasmids were transformed into *E. coli* SM10λ*pir.* For generating mutants, pCONJ4S derivatives were then transferred into *Shigella* by conjugation (Liu *et al*., 2020).

For the assembly of the CRISPR-Cas plasmids (pCas9 and its derivatives, **Table 2.3**), spacers were designed as single stranded 20 bp oligonucleotides homologous to chosen target sequences from *virB* or *vapBC* adjacent to the protospacer motif (PAM, 5'-NGG-3') (**Table 2.4**). Sense and antisense oligos were annealed to form duplex DNA, which was ligated into *Bsa*I-digested pCas9 or inducible pCas9 following the manufacturer's protocol (Jiang *et al*.,2013); plasmids were initially transformed into *E. coli* DH5α, then purified and finally transferred by electroporation into *S. sonnei* CS14.

During construction, bacteria were selected on CR-TSA plates containing appropriate antibiotics and incubated overnight at 37°C. Strains were checked by PCR and sequencing.

Table 2.3 List of plasmids

Plasmid	Genotype / Purpose	Antibiotic Resistance	Reference
pUC19	Cloning vector	Amp^r	Figurski *et al*., 1979
pUC19_nSf*vapBC*^{WT}::*cat*	Cloning vector to introduce *vapBC*^{WT}::*cat* into M90T to produce GMCT283	Amp^r, Cam^r	Pilla *et al*., 2017
pUC19_nSf*vapBC*^{D7A}::*cat*	Cloning vector to introduce *vapBC*^{D7A}::*cat* into M90T to produce GMCT284	Amp^r, Cam^r	Pilla *et al*., 2017
pUC19_eSf*vapBC*^{WT}::*cat*	Cloning vector to introduce *vapBC*^{WT}::*cat* ectopically into M90T to produce GMCT285	Amp^r, Cam^r	Pilla *et al*., 2017
pUC19_eSf*vapBC*^{D7A}::*cat*	Cloning vector to introduce *vapBC*^{WT}::*cat* ectopically into M90T to produce GMCT284	Amp^r, Cam^r	Pilla *et al*., 2017
pUC19_*mxiH*:: *sacB-neo*^{r}	Cloning vector to introduce *sacB-neo*^{r} into *mxiH* to produce strain GMCT189	Amp^r, Kan^r	McVicker and Tang, 2016
pGM173	Cloning vector to delete *stbAB* in M90T to produce GMCT356	Amp^r, Cam^r	McVicker *et al*., 2019
pUC19_Δ*parAB*::*cat*	Cloning vector to delete *parAB* in M90T to produce GPSfA23	Amp^r, Cam^r	McVicker *et al*., 2019
pUC19_Δ*vapBC*::*cat*	Cloning vector to delete *vapBC* in M90T to produce GMCT213	Amp^r, Cam^r	McVicker and Tang, 2016
pUC19_Δ*ccdAB*::*cat*	Cloning vector to delete *ccdAB* in M90T to produce GMCT213	Amp^r, Cam^r	McVicker and Tang, 2016
pUC19_Δ*gmvAT*::*cat*	Cloning vector to delete *ccdAB* in M90T to produce GMCT213	Amp^r, Cam^r	McVicker and Tang, 2016
pUC19__Δ*virB*::*cat*	Cloning vector to delete *virB* in M90T to produce GPSfA31	Amp^r, Cam^r	McVicker *et al*., 2019
pUC19_Ss*vapBC*::*cat*	Cloning vector to introduce *vapBC*::*cat* into 53G to produce GMCT323	Amp^r, Cam^r	Martyn *et al*., unpublished
pUC19_Ss*ccdAB*::*cat*	Cloning vector to introduce *ccdAB*::*cat* into 53G to produce GMCT364	Amp^r, Cam^r	Martyn *et al*., unpublished
pUC19_Ss*gmvAT*::*cat*	Cloning vector to introduce *gmvAT*::*cat* into 53G to produce GMCT364	Amp^r, Cam^r	Martyn *et al*., unpublished
pUC19_Ss*vapBC*^{Sflex}::*cat*	Cloning vector to introduce *vapBC*^{Sflex}::*cat* into 53G	Amp^r, Cam^r	Martyn *et al*., unpublished

pUC19_Ss*vapB*^{Sflex}*C*^{Ssonn}::*cat*	Cloning vector to introduce *vapB*^{Sflex}*C*^{Ssonn}::*cat* into 53G	Amp^r, Cam^r	Martyn *et al*., unpublished
pUC19_Ss*vapB*^{Ssonn}*C*^{Sflex}::*cat*	Cloning vector to introduce *vapB*^{Ssonn}*C*^{Sflex}::*cat* into 53G	Amp^r, Cam^r	Martyn *et al*., unpublished
pCP20	Temperature sensitive plasmid encoding FLP recombinase	Amp^r	Cherepanov and Wackernagel, 1995
pKD46	Plasmid expressing the λ red recombinase	Amp^r	Datsenko and Wanner, 2000
pCONJ4S	Cloning vector	Amp^r	Kaniga *et al*., 1991
pCONJ4S_Δ*recA*	Cloning vector to delete *recA*	Amp^r	This work
pCONJ4S_Δ*recD*	Cloning vector to delete *recD*	Amp^r	This work
pCONJ4S_Δ*recF*	Cloning vector to delete *recF*	Amp^r	This work
pGEM®-T Easy	Cloning vector	Amp^r	Promega Corporation
pGEM_*polA*	Model plasmid containing *polA* oligo for qPCR analysis	Amp^r	This work
pGEM_*repA*	Model plasmid containing *repA* oligo for qPCR analysis	Amp^r	This work
pRK2013	Helper plasmid for the tri-parental conjugation	Kan^r	Figurski and Helinski, 1979
pINV^{Ssonn} VapBC^{Ssonn}	*S. sonnei* pINV with *vapBC*^{Ssonn}::*cat*	Cam^r	This work
pINV^{Ssonn} VapBC^{Sflex}	*S. sonnei* pINV with *vapBC*^{Sflex}::*cat*	Cam^r	This work
pCas9	Plasmid encoding CRISPR-Cas9 system	Cam^r	Jiang *et al*., 2013
pCas9^{amp}::control	pCas9 with *amp*^{r} cassette and no target-specific oligo	Amp^r	This work
pCas9^{kan}::control	pCas9 with *kan*^{r} cassette and no target-specific oligo	Kan^r	This work
pCas9^{amp}::virB^{Pro}	pCas9^{amp} containing spacer targeting *virB* promoter	Amp^r	This work
pCas9^{amp}::*vapBC*^{Pro}	pCas9^{amp} containing spacer targeting *vapBC* promoter	Amp^r	This work
pCas9^{kan}::*virB*^{orf}	pCas9^{kan} containing spacer targeting *virB* ORF	Kan^r	This work
pCas9^{kan}::*vapB*^{orf}	pCas9^{kan} containing spacer targeting *vapB* ORF	Kan^r	This work
ind_pCas9::control	pCas9^{kan} with expression of *cas9* and *crRNA* under arabinose inducible promoter P_{BAD}	Kan^r	This work
ind_pCas9::*vapBC*^{Pro}	ind_pCas9 containing spacer targeting *vapBC* promoter	Kan^r	This work

ind_pCas9::*vapB*^{orf}	ind_pCas9 containing spacer targeting *vapB* ORF site 2	Kan^{r}	This work
ind_pCas9::*vapC*^{orf}	ind_pCas9 containing spacer targeting *vapC* ORF	Kan^{r}	This work

Table 2.4 List of oligos

Oligo	Sequence (5'- 3')	Purpose
	Construction of strains	
GP007	acggccagtgaattcgagctcgtgaagcgggtccgggtg	Construction of GMCT283, GMCT284, GMCT323, GPSs12, GPSs15 and GPSs18
GP008	tattgggccgttggtggaga	Construction of GMCT284 and GMCT286
GP009	tctccaccaacggcccaata	Construction of GMCT284 and GMCT286
GP010	ccatggctaattcccattcagctccagtcttcagttc	Construction of GMCT283, GMCT284, GMCT323, GPSs12, GPSs15 and GPSs18
GP011	agcctacacacctgttcatcagaaatcatctc	Construction of GMCT283, GMCT284, GMCT323, GPSs12, GPSs15 and GPSs18
GP012	ctatgaccatgattacgccaagcttggcgtagccgacaatttc	Construction of GMCT283, and GMCT284
GP013	acggccagtgaattcgagctctggatggcactcgtagaag	Construction of GMCT285 and GMCT286
GP015	aacgcttttaatggaaaccaccgtatttc	
GP016	agcctacacactactcaatttttttacatggtg	
GP017	ctatgaccatgattacgccaagcttatgcctgataagatagtcatg	
GP018	acaggcatcgtaaaagcgttgactgctaatc	
GP019	aacgcttttacgatgcctgttactgcca	
GP141	catgattacgccaagctgtaatcgtcgtgtgtgg	Construction of GPSs12, GPSs15 and GPSs18
GM174	tgtgtaggctggagctgctt	PCR FRT flanked *cat* cassette from pKD3
GM175	atgggaattagccatggtcc	
GP163	cggagtatagtaaaacgacggccagtgtttttttgactgacagaggttgtgatgactgac	Construction of GPSf12
GP164	gtattacccggcatgacaggagtaaaaatgaatctgtttatcagcaaaacctgggcgc	
EL127	catttttactcctgtcatgccggg	
EL128	actatactagcaggaaacagctatgacgttttatcttccatcagatagccacgatagagc	

EL145	cggagtatagtaaaacgacggccagtgttttatgatctgcaatatcagctttataccctg	Construction of GPSf15
GP167	caattacgtttattcccgcgaactaaacaacaaataaagacgatcgccacagaggatcat	
GP168	atgatcctctgtggcgatcgtctttatttgttgtttagttcgcgggaataaacgtaattg	
EL418	ctatactagcaggaaacagctatgacgtttttgtgatggaaagcgccgagcagattcgtc	
EL340	aaaataacggattaacccaagtataaatgagc	Construction of GPSf18
EL341	tcatttatacttgggttaatccgttattttgcgggtgagggacattacagtctcattggc	
EL342	cggagtatagtaaaacgacggccagtgttttcaatgccaattggtcaatctttgccaagc	
EL343	ctatactagcaggaaacagctatgacgtttttcatattcgaactcggtcacattggtgaa	
GM069	ctgaggcaaagaactcaaatcttg	Insertion of *sacB-neo*^{r} into *mxiH* of M90T
GM070	ccggatcctttatcctcacttatttttatc	
GM071	cccatatgtaatagggagcattcatg	
GM072	gtttctcttgaaccaccctgttttg	
GM073	ccggatccttttttaacccatcac	
GM074	cccatatgtgcaagcctcgtcgt	
GM124	cctgcgcgatactcatcataaacgtatatccctttgacatatcccggtatcaatcccacaatagatatacacaagacatatccacataaggaggcaaatatgtgtaggctggagctgcttc	Deletion of *vapBC* in M90T
GM125	cacctggcagcagatgcagcgtcgtgaagaggtgaacatcaatgtgcaccgggaacgcggggaggatgttgagccgggagatgatttctgatgaacaggatgggaattagccatggtcc	
GM176	ggtctttaccgcattgag	Deletion of *ccdAB* in M90T
GM177	aagcagctccagcctacacagataagccctcaaaaacatac	
GM178	aagcagctccagcctacacacatttcaccagtccctgttctc	
GM179	ggaccatggctaattcccatatgtcaggctagggtcaaaaatc	
GM180	agattaagaatgctttccggg	
GM183	ctccacaggttacagacac	Deletion of *gmvAT* in M90T
GM184	aagcagctccagcctacacataacagaacgctatacagaatc	
GM185	aagcagctccagcctacacacatttatttccattcaggctttac	
GM186	ggaccatggctaattcccatttgttcagtcagttgtgg	
GM187	ctcagtatgtatcgctgg	
GM241	acggccagtgaattcgagctcgctcatacaactgccagc	Deletion of *stbAB* in M90T
GM242	ccatggctaattcccattatttaaccaggccgtgc	
GM243	agcctacacagcaacagacaggaatgcaataattttg	
GM244	ctatgaccatgattacgccaagcttcgaactggctttcatgagc	
GM406	aaacgacggccagtgaattcgagctccgttataaaacgacgcttc	Deletion of *parAB* in M90T

GM407	catatggaccatggctaattcccatgaatattactgtaagatagtatgcc	
GM408	cttcgaagcagctccagcctacacacttctttgcttgaaaaagatagtttctc	
GM409	aacagctatgaccatgattacgccaagcttcaagagccgtatgacgcg	
GP130	ccagtgaattcgagctcaactatggaattaggtgcaagaagaag	Deletion of *virB* in M90T
GP131	ggaccatggctaattcccatcatcgaacactgatgtaaactgccc	
GP132	gtttacatcagtgttcgatgatgggaattagccatggtccatatg	
GP133	ctatggagctctcacatcagagtgtgtaggctggagctgctt	
GP134	gcagctccagcctacacactctgatgtgagagctccatagatg	
GP135	gctatgaccatgattacgccaagctccatactccagataattcgg	
GM208	acggccagtgaattcgagctcgtgaatgacctgtttgcc	Introduction of *ccdAB* in 53G
GM209	tacgctgcttcatgataagccctcaaaaacatac	
GM210	agggcttatcatgaagcagcgtattactg	
GM211	ccatggctaattcccatttatattccccagaacatcag	
GM212	agcctacacaatacggtgcaaacaggcc	
GM213	ctatgaccatgattacgccaagcttatgccatatctgtgatacatcag	
GM216	acggccagtgaattcgagctcggcttctccactgtgattccgtcgtg	Introduction of *gmvAT* in 53G
GM217	aatgaaatgccacggcagcctggcggct	
GM218	ggctgccgtggcatttcattatcaaaatcacattaaac	
GM219	ccatggctaattcccattcagactttataaaacaaggtattag	
GM220	agcctacacagttcaggtaaacgcctgttc	
GM221	ctatgaccatgattacgccaagcttgcgggtgtagtaagccct	
	Multiplex PCR	
*hns*_F	ttgcaaaggcgttgaatta	Detection of *hns*
*hns*_R	ttgcaaaggcgttgaatta	
*virB*_F	acatcagagctccacaagaa	Detection of *virB*
*virB*_R	agacgatagatggcgagaaa	
*virF*_F	cttagcttgttgcacagaga	Detection of *virF*
*virF*_R	aagatgggcttgatattccg	
*ori*_F	gtgacctcctcagaataatcc	Detection of *ori*
*ori*_R	aaaagatacattgcaccctgt	
*stbAB*_F	cttgaagatggcgaggtaag	Detection of *stbAB*
*stbAB*_R	gcatcccggttatctgaaag	
*ccdAB*_F	acttatagcgaattaatggtgtact	Detection of *ccdAB*
*ccdAB*_R	ccagtcctggatatgctca	
*vapBC*_F	ggaaaccaccgtatttctca	Detection of *vapBC*
*vapBC*_R	gggcaagttctgctcttatc	
*gmvAT*_F	cggtaaagcctgaatggaaa	Detection of *gmvAT*
*gmvAT*_R	taacagccttaatgccaacc	
	Identification of T3SS PAI deletions	

GPv1_fwd	tgaacaggccactcattg	Detection deletion variant 1 in M90T
GPv1_rev	ttacacatcccattgagg	
GPv234_fwd	ggaagaggggttgtgtgt	Detection deletion variants 2, 3 and 4 in M90T
GPv2_rev	acccgacgtgaaagtctg	
GPv3_rev	ccaatcccgtccagaagtcg	
GPv4_rev	taacagccttaatgccaacc	
GP99	cctatcgatggacatgaatgcagg	Detection IS*21*-mediated deletion in 53G
GP100	tcgcctgtttcaccgtatcttc	
GP145	ggagcggtatcaggcaaacg	Detection IS*1294*-mediated deletion in 53G
GP146	gctgacagtggtaatacgttg	
GP147	ccgaccatatccgccattg	Detection IS*1*-mediated deletion in 53G
GP148	cgtttgcctgataccgctcc	
qRT-PCR		
GP038	cagctgtttctgatgaggaagc	Quantification of *virF* expression
GP039	ccgtctagctctctctaatcgcagaaagacgccatctcttctcg	
GP042	gcaaaagagcatagcatccga	Quantification of *virB* expression
GP043	ccgtctagctctctctaatcgctttcgcgcgagacagattc	
GP048	gaagagctcaccttgcgtct	Quantification of *polA* expression
GP049	ccgtctagctctctctaatcgttccagtacctcttccgacg	
GP050	ggaacaagccctgaatccga	Quantification of *ipaB* expression
GP051	ccgtctagctctctctaatcggcgattgcccccagaataga	
qRT-Tag	ccgtctagctctctctaatcg	qRT-PCR tag
qPCR analysis for plasmid copy number quantification		
GP160	gcgcatgaaattgcaggtg	Amplification of *polA*
GP161	ttccagtacctcttccgacg	
GP154	acctaccagacggaatatgacc	Amplification of *repA*
GP155	cgctgtctgttttcccattcc	
CRISPR-Cas9 plasmids		
AH1	cacgcattgatttgagtcagc	Check of insertion of the oligo
AH2	cacgcattgatttgagtcagc	
AH23	gtatttagaaaaataaacaaataggggttcgcgacgtctcattttcgccaaaag	Construction of pCas9^{amp}
AH24	gatacttctattctactctgactgcaaaccaaaaaaacaagcgc	
AH25	cagagtagaatagaagtatcaaaaaaagcaccgactcg	
AH26	cctcctagctgactcaaatcaatgcgtgtttcataaagaccagtg	
AH27	cactggtctttatgaaacacgcattgatttgagtcagctaggagg	
AH28	gcctcactgattaagcattggtaatttttttaaggcagttattggtgc	
AH29	gcaccaataactgccttaaaaaaattaccaatgcttaatcagtgaggc	
AH30	gagacgtcgcggaacccctatttgtttatttttc	

AH33	cgtgttccgcttcctttaacgtctcattttcgccaaaag	Construction of pCas9^{kan}
AH34	cttcttgacgagttcttctgatttttttaaggcagttattggtgc	
AH35	taactgccttaaaaaaatcagaagaactcgtcaagaaggcgatag	
AH36	acttttggcgaaaatgagacgttaaaggaagcggaacacgtag	
AH37	gcaaaagaagagtagtgtgatcgtccattccgacagcatc	
AH38	gctgtcggaatggacgatcacactactcttcttttgcctattataac	
AH39	ccatgagcggcctcatttcttattctgagttacaacagtccg	
AH40	tgagcggcctcatttcttattctgagttacaacagtccgcacc	
AH60	tggacatttttgcctcctaaaataaaaagtttaaattaaatcc	Construction of ind_pCas9
AH61	acttttttattttaggaggcaaaaatgtccagacctgcag	
AH62	cctattgagtatttcttatccatatgtatatctccttcttaa	
AH63	agaaggagatatacatatggataagaaatactcaataggctt	
vapBC^{Pro}_F	aaacgatgccgggatatgtcaaag	Insertion of spacer in pCas9^{amp} and in ind_pCas9 to target the *vapBC* promoter
vapBC^{Pro}_R	aaaactttgacatatcccggcatc	
vapB^{orf}_F	aaacttacgccagccggagagacag	Insertion of spacer in pCas9^{kan} to target the *vapB* ORF
vapB^{orf}_R	aaaactgtctctccggctggcgtaa	
vapB^{orf2}_F	aaacgatggataacagggaacagccg	Insertion of spacer in ind_pCas9 to target the *vapB* ORF site 2
vapB^{orf2}_R	aaaacggctgttccctgttatccatc	
vapC^{orf}_F	aaacgcctgcgatcatttgatcaaag	Insertion of spacer in ind_pCas9 to target the *vapC* ORF
vapC^{orf}_R	aaaactttgatcaaatgatcgcaggc	
virB^{Pro}_F	aaacgtaactgttgatgatagtca	Insertion of spacer in pCas9^{amp} to target the *virB* promoter
virB^{Pro}_R	aaaattgactatcatcaacagtta	
virB^{orf}_F	aaacggaacgcttgctgcctgaa	Insertion of spacer in pCas9^{kan} to target the *virB* ORF
virB^{orf}_R	aaaattcaggcagcaagcgttcc	

2.3 Tri-parental mating

Tri-parental conjugation was employed to transfer pINV^{Ssonn} in *S. flexneri*. *S. sonnei* 53G possessing *vapBC*::*cat* (conferring chloramphenicol resistance) was used as the donor strain, *S. flexneri* BS176 was employed as the recipient strain (streptomycin resistant), and *E. coli* containing pRK2013 (conferring kanamycin resistance) constituted the helper strain. Strains were grown separately overnight at 37°C in 10 ml LB with appropriate antibiotics. Cultures

were washed in 10 ml PBS and re-suspended in 10 ml PBS. Bacteria were then mixed at a 1:1:5 ratio of donor, helper and recipient, respectively, in a 100 µl volume, spotted onto LB agar without antibiotics and incubated overnight at 37°C. Bacteria were harvested from spots, then re-suspended in 1 ml TSB, and finally plated onto CR-TSA containing chloramphenicol and streptomycin to select for transconjugants. Potential transconjugants were passaged twice on the same plates to remove remnants of pRK2013 and their sensitivity to kanamycin was confirmed with growth on CR-TSA containing this antibiotic. Strains were checked by PCR and sequencing.

2.4 CR-binding assay

For the assays performed in **Figure 3.1A**, **Figure 3.2B** and **Appendix Figure 7.2A**, *S. flexneri* was grown on CR-TSA plates overnight at 37°C to obtain single colonies. On at least two separate occasions, three CR^+ colonies from each strain were re-suspended in 5ml TSB and incubated at 30°C, with shaking at 180 r.p.m. for 16 hours; there is no detectable growth rate difference between CR^+ and CR^- bacteria at this temperature (McVicker and Tang, 2016), so CR^- bacteria do not outcompete CR^+ bacteria. After overnight growth, cultures were diluted in 5ml TSB to reach an OD_{600} = 0.001 and grown for approximately 50 generations at either 37°C or 21°C. Samples were diluted in PBS and plated onto CR-TSA, and incubated overnight at 37°C before CR^+ and CR^- colonies were counted.

For the remaining CR-binding assays performed in this work, to detect loss of CR-binding at earlier stages (*e.g*. after ~25 generations) and to prevent any plasmid loss before the start of the experiment, a *cat* cassette was inserted on pINV downstream of *vapBC* and *Shigella* spp. were grown on CR-TSA containing chloramphenicol overnight at 37°C to obtain single colonies. On three separate occasions, three independent CR^+ colonies were re-suspended in 5 ml TSB

liquid medium and incubated at 37°C with shaking at 180 r.p.m. for 16 hours (the equivalent of ~25 generations). Samples were diluted in PBS and plated onto CR-TSA, and incubated overnight at 37°C before counting CR^+ and CR^- colonies.

2.5 Virulence plasmid stability assay

S. flexneri strains possessing pINV *mxiH*::*sacB-neo*^{r} were grown from 15% glycerol stocks at 37°C for 16 hours or at 21°C for 96 hours when plated on solid LB medium or at 37°C for 30 hours when plated on solid M9 medium (**Table 2.2**.) to reach ~25 generations. On three separate occasions, three colonies were resuspended in 100 µl PBS, diluted and 5 µl aliquots were plated on plates with or without sucrose (**Table 2.2**.) to count the total number of CFU and the number of CFU that lost the *sacB-neo*^{r} cassette (on plates containing sucrose). The percentage of T3SS PAI loss was calculated by measuring the number of CFU on plates with sucrose and the number of CFU on plates without sucrose, and shown as a percentage.

2.6 Multiplex PCR

Multiplex PCR was used to detect *virF*, *virB*, and *ori*, the origin of pINV replication, as pINV-specific elements; *hns* was included as a positive chromosomal control. Bacteria were first lysed at 98°C for 5 min and 1 µl of each lysate was used as PCR template. Reactions included Q5® High-Fidelity polymerase (New England Biolabs) and were performed using an annealing temperature of 51.2°C and extension time of 1.5 min. For each biological replicate, eight emerging CR^- colonies were analysed.

2.7 Genome sequencing

For Illumina sequencing, genomic DNA was purified from colonies isolated after plating at 37°C overnight and grown in 5mL TSB at 37°C or 21°C for 16 hours and 24 hours, respectively. Charge Switch gDNA Mini Bacteria Kit (Thermo Fisher Scientific, MA, USA) was used to purify genomic DNA from these liquid cultures. DNA was sequenced at The Wellcome Trust Centre for Human Genetics at the University of Oxford. Sequence data were analysed using Snap Gene® for identifying single nucleotide polymorphisms and deletions, while BLAST Ring Image Generator (BRIG) was employed to align plasmid sequences (Alikhan *et al*., 2011). For both approaches, *S. flexneri* M90T was used as the reference (AL391753.1). The exact sites of T3SS PAI deletion were confirmed by amplification of the regions flanking the predicted deletions sites with following restriction enzyme digestion and sequencing of the amplification products.

For PacBio sequencing (Rhoads *et al*., 2015), DNA was recovered from a loop of colonies using Wizard Genomic DNA Purification kit (Promega, WI, USA) and sequencing was performed at The Earlham Institute, Norwich.

2.8 Purification and visualisation of pINV

Isolates were plated on CR-TSA from 15 % glycerol stocks and grown at 37°C overnight. One colony from each isolate was streaked again onto CR-TSA and grown as above. A loop of bacteria was then processed for pINV extraction as indicated in (Heringa and Herrick, 2007). DNA samples were run on 0.7 % agarose gels at 1 V/cm voltage for approximately 16 hours.

2.9 Protein Secretion

To activate secretion through the T3SS, bacteria were grown overnight at 37°C in TSB at 180 r.p.m., then sub-cultured to an OD_{600} = 0.05 into 10ml fresh TSB and grown in the same conditions until OD_{600} = 1 was reached. Cultures were centrifuged at 4000 r.p.m. for 15 minutes and pellets were resuspended in PBS to an OD_{600} = 5. T3SS secretion was induced by adding Congo red (final concentration, 0.02% w/v) followed by incubation at 37°C for 15 minutes. Bacteria were pelleted and supernatants containing secreted proteins were boiled in an equal volume of 2x SDS-PAGE loading buffer (0.1M Tris-HCl pH 6.8, 1.5 % SDS, 20 % glycerol, 25mM EDTA, 2 % β-mercaptoethanol, 150 µg/ml bromophenol blue) before loading on 10% SDS-PAGE gels. After electrophoresis, gels were silver stained using the SilverXpress Kit (Invitrogen LC6100) following manufacter's protocol.

2.10 RNA extraction and quantitative real-time PCR

Bacteria were grown at 37°C overnight in 5 ml TSB then sub-cultured to an OD A_{600} = 0.05 into 25ml fresh TSB and grown at 37°C until the OD_{600} ~ 1. 20ml of each culture medium were pelleted, resuspended in 460µl Resuspension Solution (200 µl 20% glucose, 200 µl Tris 25 mM pH 7.6 10 mM EDTA and 60 µl 0.5M EDTA) and lysed by using Lysing Matrix B Tubes (MP Biomedicals) in presence of 500 µl acid phenol. RNA was then purified by TRIzol-chloroform extraction, precipitated with isopropanol, and pellet was washed in 75 % ethanol. RNA samples were treated twice with DNase I (TURBO DNase; Ambion), each followed by further phenol-chloroform purification. RNA quality was checked by gel electrophoresis and NanoDrop analysis. First-strand cDNA was synthesised from 5µg total RNA using *polA*-, *ipaB*-, *virB*- and *virF*-specific primers with a unique 5' tag sequences not present in *S. flexneri* M90T

(**Table 2.4**) and reverse transcription was performed in presence of actinomycin, following RNase H (Life Technologies) treatment at 37°C for 20 minutes. After purification, cDNA was analysed by quantitative real-time PCR with Power SYBR green PCR master mix (Applied Biosystems) using *polA*-, *ipaB*-, *virB*- and *virF* -specific forward primer and the tag-specific reverse primer to ensure strand-specific cDNA amplification. StepOnePlus real-time PCR system was used to monitor the reaction. Results represent the average of four biological replicates and were normalised to *polA* cDNA levels by using the 2-ΔΔCt method (Livak and Schmittgen, 2001). Values are shown in relation to wild-type *S. flexneri* M90T 2-ΔCt levels, which are indicated as 1. Therefore, values less than 1 indicate decreased transcript levels of the target genes in the analysed samples. Statistical analysis was performed using one-way ANOVA with Tukey's multiple comparison test.

2.11 Quantitative PCR analysis to measure plasmid copy number

S. flexneri M90T possessing *vapBC*::*cat* was grown in the presence of chloramphenicol at 37°C for 16 hours or at 21°C for 96 hours when plated on CR-TSA or at 37°C for 30 hours when plated on solid M9 medium (**Table 2.2**) to reach ~ 25 generations. On three different occasions, three colonies were resuspended in 100 µl PBS. 50 µl aliquot of each suspension was diluted and plated in 5 µl spots on CR-TSA for CFU quantification. The remaining 50 µl were lysed at 98°C for 5 min and 1 µl of each lysate was analysed by quantitative PCR (qPCR) with Power SYBR green PCR master mix (Applied Biosystems) using *polA*-specific and *repA*-specific primers to measure chromosome and plasmid copy number, respectively.

For the determination of *polA* and *repA* copies, qPCR was run using known amounts of pGEM plasmids containing *polA* and *repA* (**Table 2.3**), serially diluted starting with 10^8 DNA copies/

µl. Standard curves were calculated by plotting the correlation of cycle threshold (Ct) to DNA template concentration and linear regression was performed using the equation y = Bx + A.

Once the number of chromosome and plasmid copies were measured in each lysate, these values were divided by the CFU counted for each sample and the ratio of plasmid to chromosome copy number (PCN/CCN, PCN = plasmid copy number, CCN = chromosome copy number) was used as the value to yield the plasmid copy number/cell.

2.12 Viability assay

S. sonnei harbouring the ind_pCas9 and derivatives were grown overnight at 37°C on CR-TSA containing kanamycin and glucose. On three separate occasions, three CR^+ colonies per strain were resuspended in 5 ml LB containing kanamycin and glucose. Samples were grown for 16 hours at 37°C, with shaking at 180 r.p.m.. Bacteria were then sub-cultured to an OD_{600} = 0.05 in 10 ml LB containing kanamycin and glucose, or kanamycin and arabinose, and grown for 4 hours at 37°C, with shaking at 180 r.p.m.. At time points 0, 1, 2, 3 and 4 hours cultures were serially diluted in PBS and plated in 5 µl spots on CR-TSA containing glucose, and grown at 37°C overnight.

2.13 Statistical analyses

All data were analysed using Graphpad Prism (version 8.4.1). The following statistical analyses were used as indicated in the legends of figures: *t*-test, one sample *t*-test/Wilcoxon signed-rank test and one-way ANOVA. For *t*-test and one-way ANOVA, data were first analysed for normal distribution using the Shapiro-Wilk test and depending on the result parametric or non-parametric *t*-test/one-way ANOVA were performed. For one sample *t*-test/Wilcoxon

signed-rank test analysis, data were log-transformed and divided by the respective control strain; then, the Shapiro-Wilk test was used to test for normal distribution. If data were normally distributed then a parametric one sample *t*-test was performed, comparing the mean of data of each strain to 1 (as 1 is the result of the ratio control/control) or to limit of the detection (l.o.d.= 0.00001, *i.e*. 0.001 % in graphs), which is equal to -5 (log l.o.d.) in the absence of the control strain. If the data were not normally distributed, Wilcoxon signed-rank test was performed, comparing the median of data of each strain to 1 or to l.o.d..

3 RESULTS: Analysis of the stability of pINV in *S. flexneri*

In this chapter, I describe how TA systems, IS, and partitioning systems contribute to the stability of pINV in *S. flexneri*.

3.1 The role of TA systems

Previous work has shown that TA systems play an important role in the maintenance of pINV in *Shigella*. In particular, of the three TA systems present on *S. flexneri* pINV, VapBC is essential and sufficient for pINV maintenance at 37°C, while GmvAT confers plasmid stability at environmental temperatures; the contribution of CcdAB on pINV maintenance is unclear (Sayeed *et al*., 2000; Sayeed *et al*., 2005; McVicker and Tang, 2016). Despite the presence of multiple TA systems, spontaneous avirulence still arises in this species (Sasakawa *et al*., 1986; Sasakawa *et al*., 1987; Schuch and Maurelli, 1997; McVicker and Tang, 2016). Furthermore, although these systems help to maintain pINV at different temperatures, it is not clear why there are multiple TA systems on pINV. Therefore, we decided to further investigate the impact of VapBC on the nature of the spontaneous loss of virulence in *S. flexneri* and the potential alternative roles of TA systems that might explain the presence of multiple functional TA systems on pINV.

3.1.1 VapBC efficiently maintains pINV

To understand the impact of VapBC on virulence loss in *S. flexneri*, a single point mutation was introduced into *vapC* to abolish VapC activity by alanine substitution of an aspartate residue at the active site (generating strain *vapBC*^{D7A}) (Winther and Gerdes, 2011); a *cat* cassette was introduced downstream of *vapBC* to facilitate genetic manipulation and to allow selection of bacteria with the plasmid at the start of experiments. To confirm that the selective marker had no effect on the expression of *vapBC* or on the stability of pINV, an isogenic strain was also constructed, *vapBC*^{WT}, with a wild-type copy of *vapBC* and the *cat* cassette in the same location as in *vapBC*^{D7A}. Avirulence in wild-type *S. flexneri* M90T, *vapBC*^{WT}, and *vapBC*^{D7A} was determined using a CR-binding assay, which employs the dye Congo red (CR) to detect avirulent colonies. Colonies of *Shigella* that are virulent and express the T3SS bind CR when grown on solid media containing this dye, giving rise to red (CR^{+}) colonies (Payne and Finkelstein, 1977). Because the T3SS is specifically expressed at 37°C (Porter and Dorman, 1994; Tobe *et al*., 1995), binding to CR is only evident for colonies grown at this temperature. The loss of T3SS expression results in white, avirulent colonies (CR^{-}) (Schuch and Maurelli, 1997). The proportion of CR^{-} colonies was measured after approximately 50 generations of growth at 37°C, the temperature of the human intestinal tract.

Consistent with previous work (McVicker and Tang, 2016; Schuch and Maurelli, 1997), inactivation of *vapC* led to a significant increase in the proportion of CR^{-} bacteria in *vapBC*^{D7A}, which reached ~90% of the population after 50 generations. In contrast, approximately 50% of *vapBC*^{WT} colonies were CR^{-} after 50 generations, which was not significantly different from the CR^{-} population arising from wild-type *S. flexneri* M90T (**Figure 3.1A**, *vapBC*^{WT} *vs*. *vapBC*^{D7A}, $p = 0.0016$; **Appendix Figure 7.1A**, *vapBC*^{WT} *vs*. M90T, $p = 0.1678$ by Welch's *t*-test), confirming

that introduction of the *cat* cassette downstream of *vapBC* did not affect the emergence of CR$^-$ bacteria.

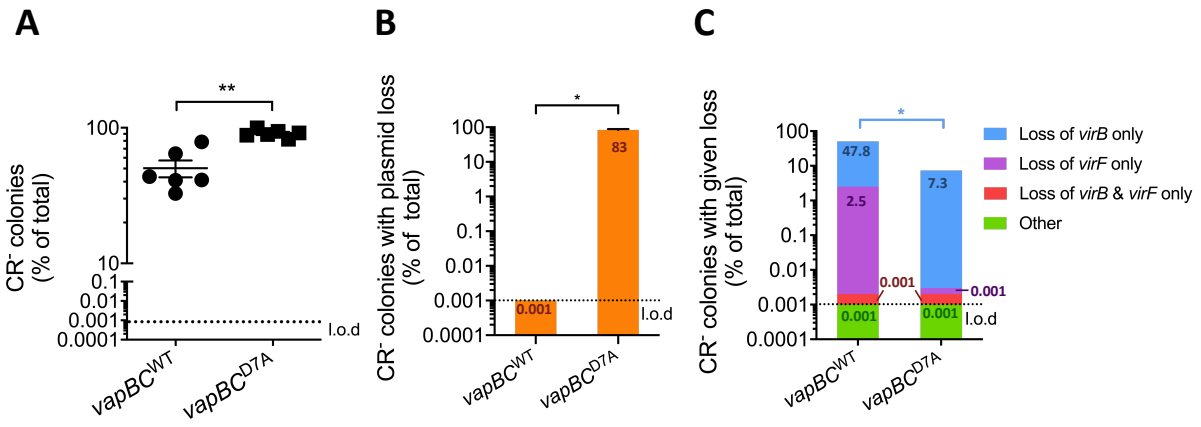


Figure 3.1 VapBC is essential for maintenance of the entire plasmid at 37°C.

(**A**) Number of CR$^-$ emerging from *S. flexneri vapBC*^{WT} and *vapBC*^{D7A} after approximately 50 generations of growth at 37°C, shown as a % of all colonies. Solid line: mean ± S.E.M. (n = 6 biological replicates). **, $p \leq 0.01$ by Welch's *t*-test. (**B**) and **C**) Number of CR$^-$ colonies lacking specified virulence-related genes (detected by multiplex PCR) and shown as a % of all colonies. Loss of the plasmid is inferred by loss of the origin of replication (**B**); "others" refers to CR$^-$ colonies that contain *virB*, *virF* and the origin of replication (**C**). The value for each category is indicated inside or next to the corresponding part of the bars. Eight independent CR$^-$ colonies were obtained from each of six biological replicates (*i.e.* 48 colonies were analysed per strain); data are shown as mean + S.E.M. in panel A and mean in panel B. *, $p \leq 0.05$; Wilcoxon signed-rank test (nonparametric comparison of median of strain/*vapBC*^{WT} to 1, for each event). l.o.d, limit of detection.

To identify the molecular events responsible for avirulence in *S. flexneri* in the presence of active or inactive VapBC, eight CR$^-$ colonies emerging from M90T, *vapBC*^{WT}, or *vapBC*^{D7A} on six independent occasions (*i.e.* 48 colonies in total for each strain) were analysed by multiplex PCR to detect the presence of *virB*, *virF*, the *ori* of pINV, and *hns* as a chromosomal control. As suggested previously (Schuch and Maurelli, 1997), since CR binding by colonies of *Shigella* is directly correlated with the expression of the T3SS PAI, we speculated that loss of genes encoding transcriptional activators of the T3SS PAI (*virB* and *virF*), or loss of pINV itself

(detected by amplifying the origin of replication of pINV, *ori*) are potential causes for the lack of T3SS expression, and hence loss of CR binding. After 50 generations, loss of pINV was the commonest event for *vapBC*^{D7A}, found in approximately 80% of all bacteria, but was below the limit of detection in both M90T and *vapBC*^{WT} (**Appendix Figure 7.1B**, **Figure 3.1B**; loss of pINV in *vapBC*^{WT} *vs*. *vapBC*^{D7A}, $p = 0.0312$, Wilcoxon singed-rank test). However, almost 50% of total bacteria in *vapBC*^{WT} lacked *virB*, with a similar result observed in M90T, confirming that presence of the *cat* cassette did not appear to affect the nature of emerging CR^{-} colonies (**Appendix Figure 7.1C**; loss of *virB* in *vapBC*^{WT} *vs*. M90T, $p = 0.4375$, Wilcoxon singed-rank test). In contrast, loss of *virB* was not as frequently observed in *vapBC*^{D7A} compared with *vapBC*^{WT} (**Figure 3.1C**, loss of *virB* in *vapBC*^{WT} *vs*. *vapBC*^{D7A}, $p = 0.312$, Wilcoxon singed-rank test).

Similar trends were observed when the same strains were grown at 21°C for 50 generations, with *vapBC*^{D7A} yielding a higher (but not significant) proportion of CR^{-} bacteria than *vapBC*^{WT} (**Appendix Figure 7.2A**, *vapBC*^{WT} *vs*. *vapBC*^{D7A}, $p = 0.1562$, Wilcoxon singed-rank test), mostly resulting from loss of pINV (**Appendix Figure 7.2B**, *vapBC*^{WT} *vs*. *vapBC*^{D7A}, $p = 0.0312$, Wilcoxon singed-rank test). However, the emergence of CR^{-} bacteria in all strains was much lower at this temperature than at 37°C, presumably due to the fitness cost imposed by the expression of the T3SS at 37°C and the corresponding growth advantage and proliferation of CR^{-} bacteria at the higher temperature.

These results confirm that VapBC is essential for the maintenance of pINV in *S. flexneri* at 37°C (McVicker and Tang, 2016; Schuch and Maurelli, 1997), and its inactivation through introduction of a non-functional allele of *vapC* leads a dramatic increase in the emergence of CR^{-} bacteria, mainly due to loss of pINV. Furthermore, in the presence of a functional copy of

VapBC, spontaneous loss of virulence in *S. flexneri* is mostly associated with loss of *virB*, probably from deletions involving theT3SS PAI.

3.1.2 VapBC maintains nearby sequences

Analysis of chromosome-encoded TA systems in *E. coli* showed that TA systems can prevent loss of discrete regions of the chromosome (Szekeres *et al*., 2007; Van Melderen, 2010). We, therefore, speculated that the presence of multiple TA systems on pINV could be due to their ability to maintain surrounding genetic sequences. To assess this, *vapBC* was deleted from its native location, and a wild-type or non-functional version was introduced into the T3SS PAI adjacent *virB*, at nt. 100,792 of pINV (Buchrieser *et al*., 2000), generating ectopic $vapBC^{WT}$ and ectopic $vapBC^{D7A}$, respectively (**Figure 3.2A**).

The emergence of CR^- colonies from these strains was quantified after 50 generations of growth at 37°C using the CR binding assay, and CR^- bacteria were analysed by multiplex PCR. There was a significant decrease in the proportion of CR^- colonies of ectopic $vapBC^{WT}$ after 50 generations compared with $vapBC^{WT}$, which has the TA system at its native site (**Figure 3.2B**, ectopic $vapBC^{WT}$ *vs*. $vapBC^{WT}$ $p = 0.0003$, one sample *t*-test). Furthermore, the nature of avirulence was strikingly different in ectopic $vapBC^{WT}$, as the loss of *virB* was very infrequent in CR^- colonies compared with $vapBC^{WT}$ (**Figure 3.2D**, loss of *virB* for ectopic $vapBC^{WT}$ *vs*. $vapBC^{WT}$ $p = 0.0312$, Wilcoxon signed-rank test), accounting for only 0.3% of all colonies. Instead, loss of *virF* represented the commonest event leading to loss of virulence, reaching almost 4% of all colonies, which was comparable with the rate of *virF* loss in $vapBC^{WT}$, suggesting that relocation of *vapBC* affects the loss of *virB* but not the loss of *virF* (**Figure 3.2D**, loss of *virF*, ectopic $vapBC^{WT}$ *vs*. $vapBC^{WT}$ $p = 0.5625$, Wilcoxon signed-rank test).

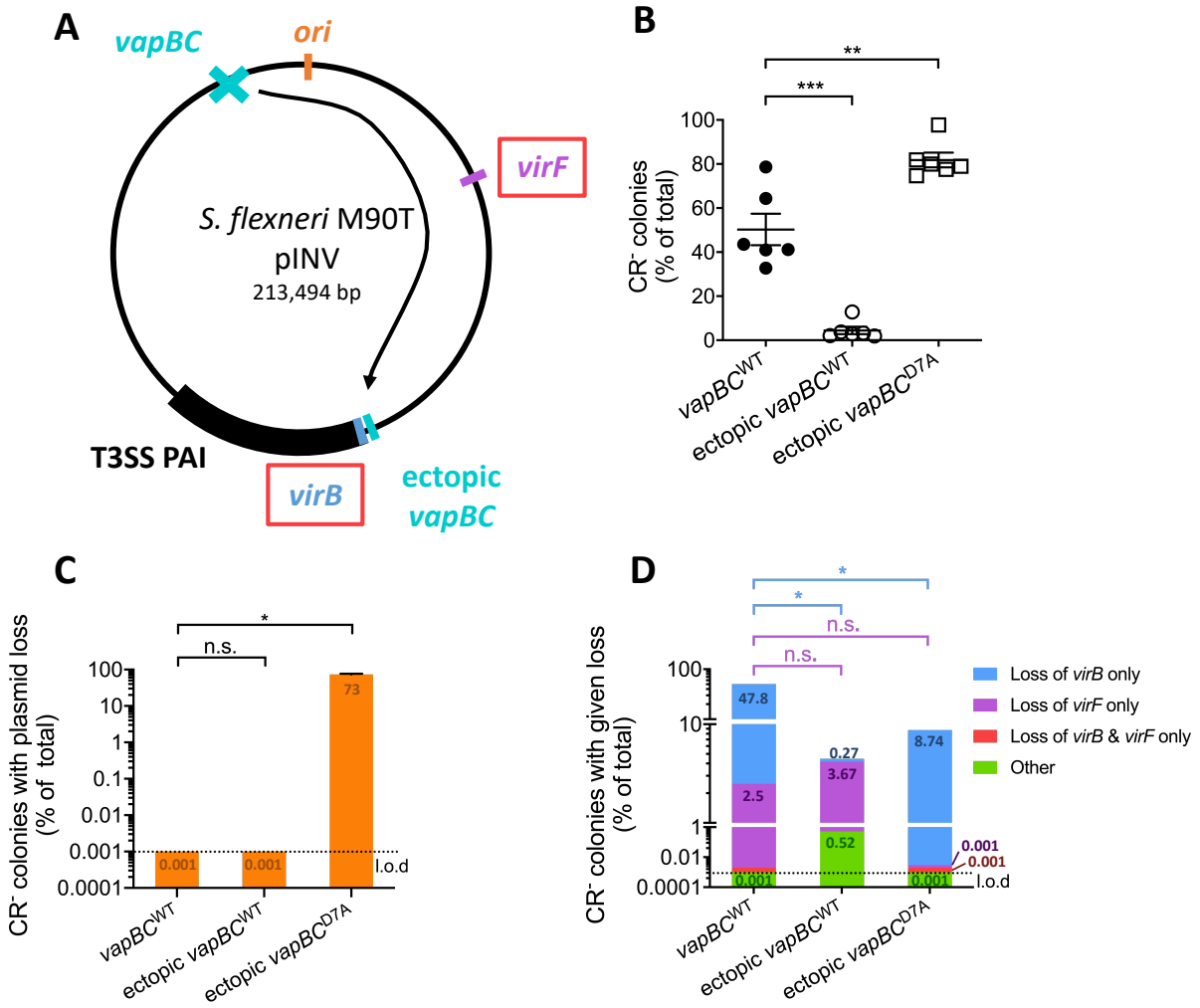


Figure 3.2 VapBC exerts protection of nearby sequences.

(**A**) Schematic representation of the repositioning of *vapBC* near the T3SS PAI in ectopic *vapBC* strains. (**B**) Number of CR$^-$ colonies emerging from *S. flexneri vapBC*^{WT}, ectopic *vapBC*^{WT} and ectopic *vapBC*^{D7A} after approximately 50 generations of growth at 37°C, shown as a % of all colonies. Solid line: mean ± S.E.M. (n = 6 biological replicates). **, $p \leq 0.01$ and ***, $p \leq 0.001$; one sample *t*-test (parametric comparison of mean test/control strain to 1). (**C** and **D**) Number of CR$^-$ colonies lacking specified virulence-related genes (detected by multiplex PCR) and shown as a % of all colonies. Loss of the plasmid is inferred by loss of the origin of replication (**C**); "others" refers to CR$^-$ colonies that contain *virB*, *virF* and the origin of replication (**D**). The value of each category is indicated inside/next to the corresponding part of the bar. Eight independent CR$^-$ colonies were obtained from each of six biological replicates (*i.e.* 48 colonies were analysed per strain); data are shown as mean + S.E.M. in panel C and mean in panel D. *, $p \leq 0.05$; n.s., not significant; Wilcoxon signed-rank test (nonparametric comparison of median of strain/vapBC^{WT} to 1, for each event). l.o.d, limit of detection.

Similar to *vapBC*^{D7A}, when the non-functional *vapBC* allele was introduced in the ectopic site (ectopic *vapBC*^{D7A}), almost 80% of bacteria were CR^{-} after 50 generations (**Figure 3.2B**, ectopic *vapBC*^{D7A} *vs*. *vapBC*^{WT} $p = 0.0078$, one sample *t*-test), most of which resulted from loss of the plasmid (approximately 73% of the total bacterial population) (**Figure 3.2C**, loss of plasmid, ectopic *vapBC*^{D7A} *vs*. *vapBC*^{WT} $p = 0.0312$, Wilcoxon signed-rank test). Again, similar trends were found when the same strains were grown at 21°C for 50 generations and, as observed previously, there was reduction in the emergence of CR^{-} colonies from all strains (**Appendix Figure 7.2**).

Taken together, these data show that VapBC acts to retain pINV and also influences the loss of surrounding genetic sequences. If the ectopic version of *vapBC* is inactive, loss of pINV significantly increases, demonstrating that VapBC contributes to the retention of pINV, irrespective of its position on the plasmid.

3.2 The role of Insertion Sequences

Insertion Sequences (IS) are an abundant component of pINV*,* often flanking virulence genes (Venkatesan *et al*., 2001). IS are known to play an important role in genetic rearrangements, including integrations or deletions (Tanaka *et al*., 2012; Lee *et al*., 2016; Vandecraen *et al*., 2016), conferring genetic plasticity and contributing to the evolution of *Shigella* species (Hawkey *et al*., 2020). IS*1*-mediated inactivation of *virF* in *S. flexneri* has been previously documented (Mills *et al*., 1992), confirming that the transposases in *Shigella* IS can be active, and can influence virulence. Therefore, we decided to investigate whether IS are involved in molecular rearrangements responsible for the spontaneous loss of virulence in *S. flexneri*, associated the loss of *virB* and/or *virF*, and to identify the mechanisms behind these events.

3.2.1 IS and homologous recombination pathways are involved in the deletion of the T3SS PAI

To further characterise the molecular events responsible for the emergence of CR^- in *S. flexneri*, genomic DNA was isolated from 20 CR^- colonies independently derived from *vapBC*^{WT} at 37°C, and analysed by whole genome sequencing using Illumina technology (Section 2.7). Alignment of plasmid sequences with BLAST Ring Image Generator using pINV from *S. flexneri* M90T as the reference (Buchrieser *et al*., 2000) showed that most of the CR^- colonies had lost the T3SS PAI as expected (**Appendix Figure 7.3**). In particular, 17 out of 20 CR^- colonies lacked sequences unique to the PAI, with the remaining three had no detectable loss of plasmid sequence. Furthermore, the regions of homology within the PAI in the 17 strains correspond to sequences found elsewhere on pINV or on the chromosome (**Appendix Figure 7.3**), such as IS elements (Venkatesan *et al*., 2001), so are likely to represent alignment artefact.

To define the sequences mediating deletion of the T3SS PAI, we performed PCR to amplify regions flanking the predicted deletion sites, and analysed the products by restriction enzyme digestion and sequencing. We identified four events leading to loss of the PAI, each involving a pair of homologous IS (**Figure 3.3**). The most frequent deletion (Variant 1) occurred in 11 out of 17 CR^- colonies and involved two copies of IS*Sfl4* (**Appendix Table 7.1** and **Figure 3.3**), spanning positions 87,663–90,277 bp and 164,461–167,075 bp of pINV (Buchrieser *et al*., 2000).

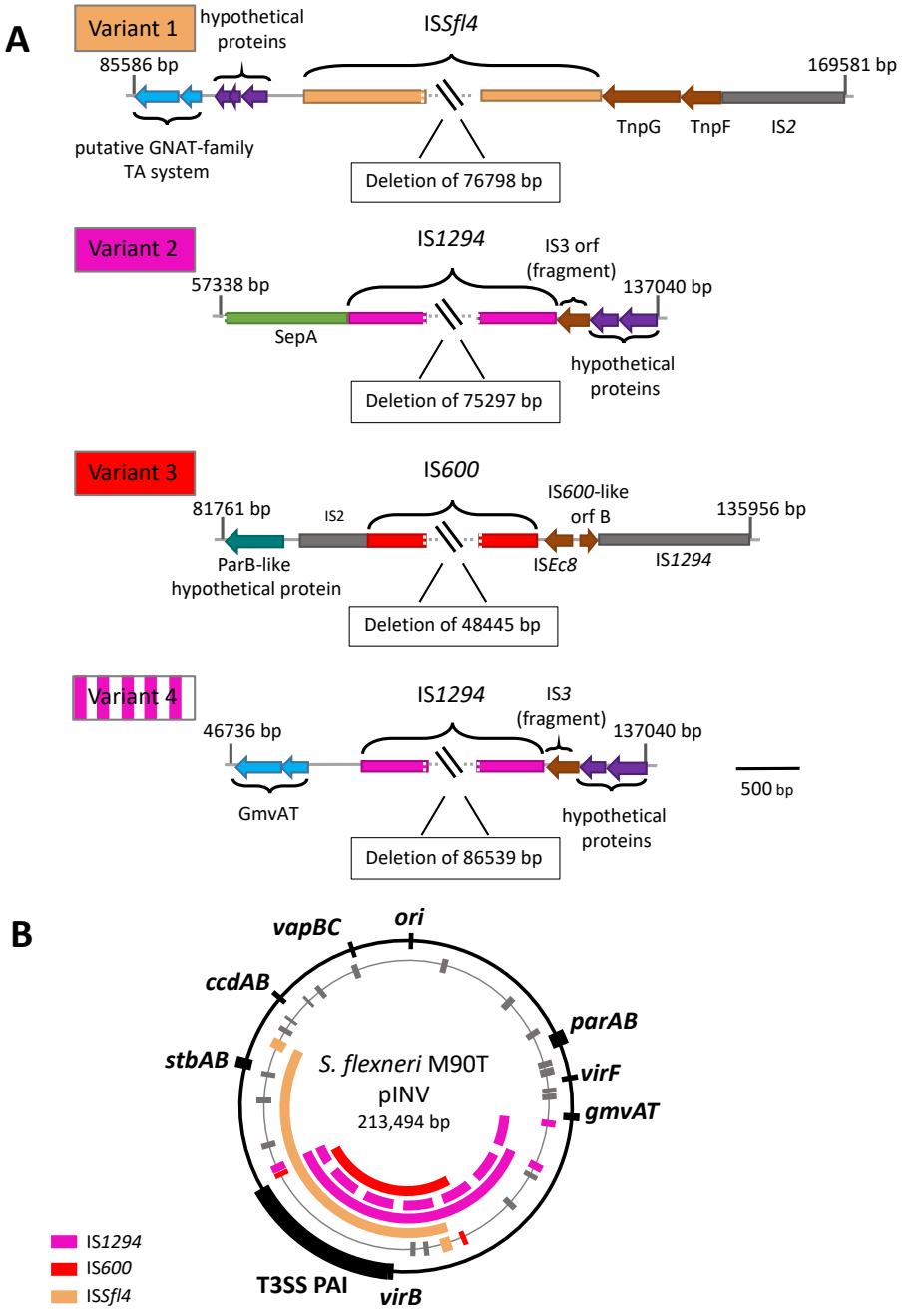


Figure 3.3 Deletion of the T3SS PAI is mediated by least four different pairs of IS.

(**A**) Deletions leading to loss of the PAI in CR$^-$ colonies are shown, with the sizes of the deleted fragments and ISs. ORFs are shown according to their predicted functions: TA systems (in cyan blue); hypothetical proteins (purple); virulence-associated proteins (green); homologues of ISs (grey); IS components (brown); homologues of proteins involved in plasmid partitioning (turquoise); ISs involved in the deletion (boxes colour-coded as in panel B). (**B**) Alignment of the four PAI deletions with pWR100 (AL391753.1) indicated as black ring; ISs are indicated as boxes coloured in grey or colour-coded for those involved in T3SS-PAI deletions. Deletion variants are shown with the same colour of the IS involved; to distinguish variant 2 from variant 4, the latter is shown with dashed line.

The second most frequent deletion event (Variant 2) occurred in four of CR^- colonies (**Appendix Table 7.1** and **Figure 3.3**) and included two copies of IS*1294*, located at positions 58,970–60,658 bp and 134,268–135,956 bp (Buchrieser *et al*., 2000). Variant 3 and 4 each accounted for a single CR^- colony (**Appendix Table 7.1** and **Figure 3.3**) and involved copies of IS*600* (83,866–85,129 bp and 132,311–133,574 bp) or IS*1294* (48,065–49,417 bp and 134,268–135,956 bp), respectively (Buchrieser *et al*., 2000).

Taken together, these results demonstrate that IS on pINV constitute "hot spots" for recombination, resulting in deletion of different regions of the plasmid that include the T3SS PAI (**Figure 3.3**).

We then investigated if the IS-mediated deletions occurred by homologous recombination, by examining the contribution of the RecBCD and RecFOR pathways. RecBCD is the main pathway involved in the repair of double-stranded DNA breaks (DSBs), while RecFOR generally mediates repair of single-stranded gaps (SSGs) but it may also substitute for RecBCD when this system is not functional (Spies and Kowalczykowski, 2005). Both pathways use RecA for the formation of recombinogenic nucleoprotein filaments on single-stranded DNA that is important in pairing of homologous DNA and strand exchange (Spies and Kowalczykowski, 2005). Therefore, to abolish both pathways together we constructed a Δ*recA* mutant, while, deletions of *recD* or *recF* were introduced to disrupt the RecBCD and RecFOR pathways individually, generating strains Δ*recD* and Δ*recF*. To better quantify the rates of the deletion events and minimise the proliferation of CR^- colonies, the emergence of CR^- colonies was measured after 25 generations of growth at 37°C using the CR binding assay, and CR^- bacteria were analysed by multiplex PCR.

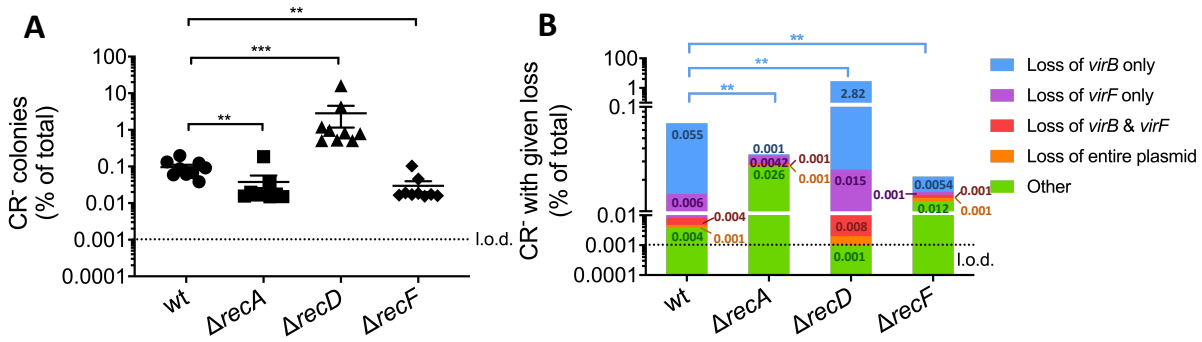


Figure 3.4 The RecBCD and RecFOR systems influence rates of T3SS PAI-related deletions.

(**A**) Number of CR$^-$ emerging from *S. flexneri* M90T wild-type (wt), and *S. flexneri* M90T lacking *recA* (Δ*recA*), *recD* (Δ*recD*) or *recF* (Δ*recF*) after approximately 25 generations at 37°C, shown as a % of all colonies. Solid line: mean ± S.E.M. (n = nine biological replicates); **, $p \leq 0.01$; ***, $p \leq 0.001$; one sample *t*-test (parametric comparison of mean test/wt strain to 1). (**B**) Number of CR$^-$ colonies lacking specified virulence-related genes were assayed by multiplex PCR and shown as a % of all colonies. Loss of the entire plasmid is inferred by loss of the origin of replication; "others" refers to CR$^-$ colonies that contain *virB*, *virF* and the origin of replication. The value of each category is indicated inside or next to the corresponding part of the bar. Eight independent CR$^-$ colonies were analysed from each of nine biological replicates (*i.e.* 72 colonies per strain); data are shown as mean. **, $p \leq 0.01$; ***, $p \leq 0.001$; Wilcoxon signed-rank test (nonparametric comparison of median of strain/wt to 1, for each event). l.o.d., limit of detection.

In the absence of RecA, there was a significant reduction in the number of CR$^-$ colonies emerging after 25 generations compared with wild-type bacteria (wt), with a similar effect observed in the *recF* mutant (**Figure 3.4A**, wt *vs*. Δ*recA*, p =0.0086; wt *vs*. Δ*recF*, p =0.0036, one sample *t*-test). Interestingly, loss of *virB* was significantly reduced in Δ*recA* and Δ*recF* compared with wild-type M90T (**Figure 3.4B**, wt *vs*. Δ*recA* or Δ*recF*, p =0.0039, Wilcoxon signed-rank test). Indeed in Δ*recA*, the loss of *virB* was not detected, while in the absence of RecF, the frequency of this event was just above the limit of detection (l.o.d). Similar to ectopic *vapBC*^{WT} (**Figure 3.2D**), the absence of loss of *virB* among CR$^-$ colonies of Δ*recA* rendered loss of *virF* the major event leading to the CR$^-$ phenotype, with levels similar to the wild-type strain (0.004% and 0.006% of all colonies for Δ*recA* and wt, respectively), suggesting that deletion

of *recA* affects deletion of the T3SS PAI, but not loss of *virF*. Surprisingly, an increase in CR^- colonies was observed in the Δ*recD* mutant, reaching 2.8% of the population after 25 generations, and resulted from an increase in *virB* loss (**Figure 3.4A**, wt *vs*. Δ*recD*, *p* =0.0039; **Figure 3.4**, wt *vs*. Δ*recD*, *p* =0.0002, one sample *t*-test and Wilcoxon test).

In conclusion, these results confirm that deletion of the T3SS PAI occurs by RecA-mediated homologous recombination and is abolished completely when both the RecBCD and RecFOR pathways are not functional, and is partially reduced when only RecFOR has been inactivated. However, it is not clear what is the contribution of RecBCD, as deletion of *recD* leads to an increase in T3SS PAI loss.

3.2.2 IS mediate reversible integration of pINV into the chromosome

As described in Section 3.1.1 and 3.1.2, the emergence of CR^- colonies was lower in *S. flexneri* following growth at 21°C than 37°C (**Appendix Figure 7.2**). Additionally, a considerable number of CR^- colonies emerging from wild-type *vapBC* at this temperature retain *virB*, *virF* and *ori* by multiplex PCR (indicated as "other" in **Appendix Figure 7.2C**), accounting for ~0.6% of all colonies in $vapBC^{WT}$. Illumina sequencing of nine CR^- colonies that emerged independently at 21°C from $vapBC^{WT}$ confirmed that these strains had not lost any plasmid sequence and had no detectable single nucleotide polymorphisms (**Appendix Figure 7.4**). We, therefore, hypothesised that these colonies possessed an intact plasmid, with alternative mechanisms responsible for the loss of virulence. Integration of pINV into chromosome has been previously described in *S. flexneri*, however, this event was detected only following exposure of bacteria to curing agents or selecting for acquisition of a second plasmid with an *ori* that is incompatible with pINV (Zagaglia *et al*., 1991). Therefore, we hypothesised that in

the nine CR$^-$ colonies isolated at 21°C with intact pINV sequence, plasmid integration could have occurred spontaneously.

To test this, plasmid DNA was extracted from each colony and analysed by agarose gel electrophoresis. Results demonstrate that a distinct band for pINV was missing in two out of nine CR$^-$ isolates (**Figure 3.5A**, isolate 4 and 6). As previous Illumina sequencing confirmed that neither isolate had lost any pINV sequence (**Appendix Figure 7.4**), we decided to analyse both these isolates by PacBio sequencing. An intact copy of pINV was found integrated into the chromosome *via* different copies of IS*1294* (**Figure 3.5B**) in both isolates. Integration in isolate 4 involved the chromosomal copy of IS*1294* at position 1,615,894–1,617,578 bp (Onodera *et al*., 2012) and pINV IS*1294* at 58,970–60,658 bp (Buchrieser *et al*., 2000) (*i.e.* IS1294_{pINV4} and IS$1294_{\text{Ch.1}}$, respectively, **Appendix Figure 7.5**); for isolate 6, chromosomal IS*1294* at position 1,872,287–1,874,073 bp (Onodera *et al*., 2012) and plasmid IS*1294* at position 205,186–206,599 bp are involved (*i.e.* IS1294_{pINV1} and IS$1294_{\text{Ch.3}}$, respectively, **Appendix Figure 7.5**) (Buchrieser *et al*., 2000).

To determine the mechanisms underlying the CR$^-$ phenotype in the pINV-integrated isolates, we examined mRNA levels of *ipaB*, *virB* and *virF* in isolate 4 and 6, and wild-type *S. flexneri* M90T. mRNA levels were measured by qRT-PCR in bacteria during exponential growth at 37°C, and results were normalised to the expression of the chromosomal gene, *polA* (**Figure 3.5C**). Levels of *ipaB* mRNA were significantly lower in the two strains with plasmid integration (**Figure 3.5C**, M90T *vs*. isolate 4 or isolate 6, $p < 0.0001$, one sample *t*-test). A similar reduction was detected for *virB* mRNA levels (**Figure 3.5C**, M90T *vs*. isolate 4, $p < 0.0001$ and M90T *vs*. isolate 6, $p = 0.0028$, one sample *t*-test) and *virF* (**Figure 3.5C**, M90T *vs*. isolate 4, $p = 0.0003$ and M90T *vs*. isolate 6, $p = 0.0015$, one sample *t*-test), suggesting that reduced T3SS activity in plasmid-integrated isolates is due to reduced expression of *virB* and *virF*.

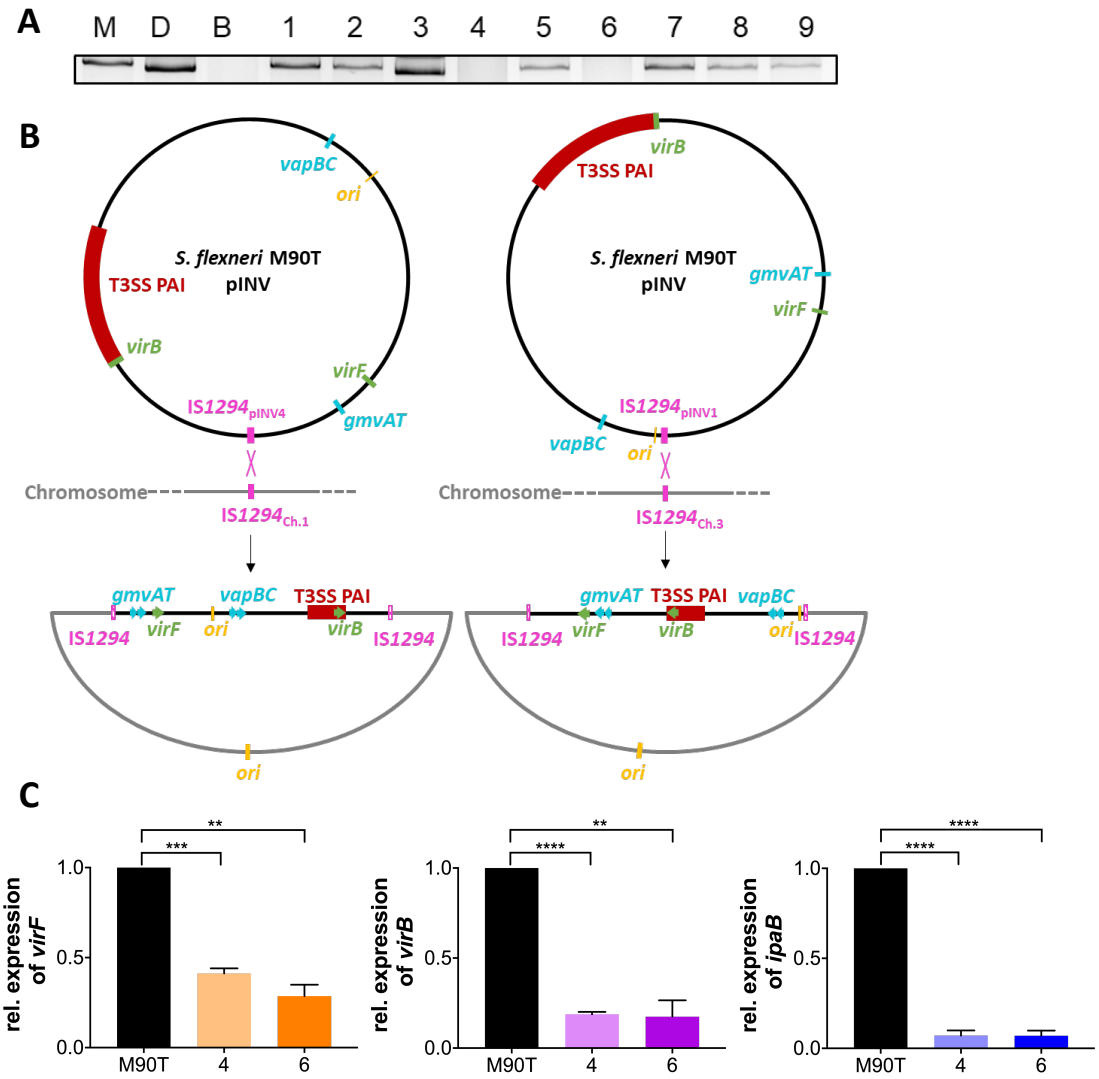


Figure 3.5 Distinct copies of IS*1294* mediate chromosomal integration of pINV, inducing a down-regulation of the expression of the T3SS PAI.

(**A**) Agarose gel electrophoresis of pINV DNA purified from nine independent CR^{-} colonies (lanes 1–9) emerging from *vapBC*^{WT} at 21°C that retained all the virulence-related genes tested by multiplex PCR. Lanes: *S. flexneri* (M) M90T, (B) BS176, (D) T3SS PAI-deleted M90T. (**B**) Schematic representation of chromosomal pINV integration in isolate 4 (left) and isolate 6 (right). The chromosomal and plasmid origins of replication are shown in yellow. ORFs are shown with the same colour coding as in Figure 3.2. (**C**) qRT-PCR analysis of *virF*, *virB* and *ipaB* in *S. flexneri* M90T, and isolates 4 and 6. Relative gene expression was determined by the 2-ΔΔ method. Data are shown as mean + S.E.M. of four biological replicates. **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$. Values analysed with one sample *t*-test (parametric comparison of mean test/control strain to 1).

Furthermore, we investigated whether integration of pINV is a reversible event. We examined isolate 4 and 6 for the emergence of CR^+ colonies, as phenotypic evidence for plasmid excision. The isolates were grown overnight in liquid media at 37°C and plated CR-TSA. CR^+ colonies emerged from both isolates (**Figure 3.6A**), and we analysed plasmid DNA from two independent CR^+ revertants (**Figure 3.6B**); results confirm that pINV had excised in these revertants.

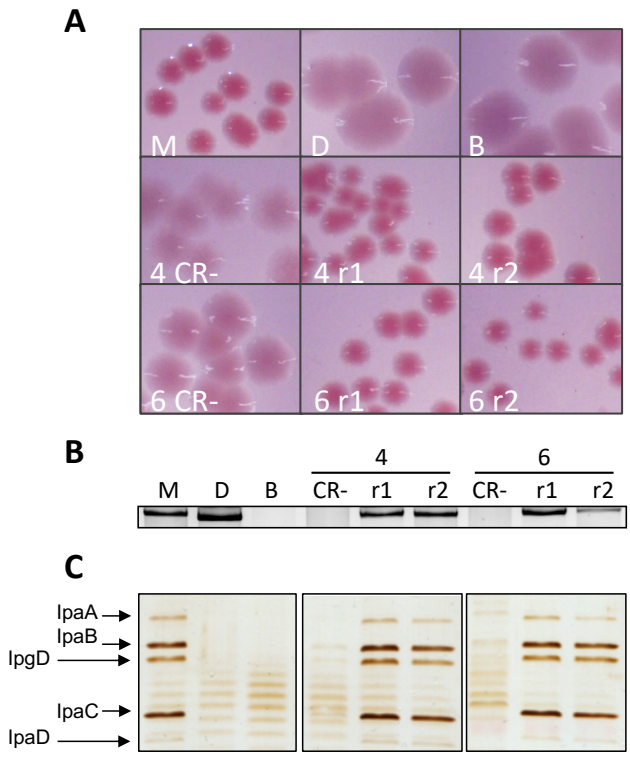


Figure 3.6 Excision of pINV is sufficient to restore T3SS activity.

(**A**) CR-TSA plates of strains grown at 37°C overnight: M90T (M), BS176 (B), T3SS PAI-deleted M90T (D), isolate 4 and two revertants (middle row, 4 r1 and 4 r2), and isolate 6 and its revertants (lower row, 6 r1 and 6 r2). (**B**) Agarose gel electrophoresis of purified plasmid DNA from control strains as for Panel A and plasmid-integrated isolates 4 and 6 (CR-), and two revertants (r1 and r2) derived from isolates 4 and 6. (**C**) Secretion through the T3SS induced by CR. Legend as for panel B. The bands corresponding to secreted effectors are indicated.

Restoration of T3SS activity in the revertants was confirmed by performing a CR-induced secretion assay. Bacteria were grown to exponential phase at 37°C in liquid media and

exposed to CR to induce secretion through the T3SS (Parsot *et al*., 1995). Silver staining to detect secreted proteins demonstrated that both isolates 4 and 6 failed to secrete the T3SS effectors, IpaA, IpaB, IpaC, IpaD and IpgD, confirming that down-regulation of the T3SS expression dramatically reduces the activity of the T3SS (**Figure 3.6C**). In contrast, the revertants secreted T3SS effectors at levels similar to *S. flexneri* M90T (**Figure 3.6C**), indicating that integration and excision of pINV provides a reversible mechanism that controls T3SS activity.

Taken together, these results demonstrate that in *S. flexneri* IS can be responsible for loss of virulence by mediating deletion of the T3SS PAI as well as integration of the virulence plasmid into chromosome, inducing a reduction in the levels of *virF* and *virB* mRNA, the two transcription factors involved in the activation cascade of T3SS PAI expression, and *ipaB*, which encodes a T3SS-secreted protein. However, in contrast with the T3SS PAI deletion, plasmid integration is a reversible event, as the plasmid can subsequently excise and therefore, restore the expression of the T3SS, and thence virulence.

3.3 The role of partitioning systems in pINV maintenance

Like *S. dysenteriae* and *S. boydii*, *S. flexneri* possesses two partitioning systems, ParAB and StbAB. Although these two systems have been well characterised in other bacterial species, their role in pINV maintenance has not been investigated previously.

ParAB belongs to the type Ia family of partitioning systems and is highly related to the well characterised ParAB system from the P1 plasmid (Abeles *et al.,* 1987), showing 75% and 58% of amino acid identity for ParA and ParB, respectively, with the *parS* nucleotide sequence 63% identical (**Appendix Figure 7.6**). In contrast, StbAB belongs to the type II family of partitioning

systems and shares low levels of identity with the well characterised ParMR system from the R1 plasmid (Gerdes and Molin, 1986; Jensen and Gerdes, 1997). In fact, StbA and StbB share 43% and 23% amino acid identity to ParM and ParR, respectively, while *stbC* has nucleotide identity of 55% with the *parC* site (**Appendix Figure 7.7**). Both partitioning systems are highly conserved across *Shigella* spp., with the only exception being *S. sonnei*, which only possesses ParAB (**Appendix Table 7.2**). Previous experiments preformed in the laboratory have shown that both systems are functional at 37°C and 21°C when they are introduced into a model vector containing the origin of replication from pINV (McVicker *et al.,* 2019). We therefore decided to examine how ParAB and StbAB contribute to pINV maintenance in *S. flexneri*, considering they belong to two distinct families of partitioning systems that act through distinct mechanisms (Section 1.2.1.3.2), and because it is unclear why they are both found on a single plasmid.

3.3.1 Contribution of single partitioning systems on pINV maintenance

To understand the contribution of each partitioning system to the maintenance of pINV, *parAB* and *stbAB* were deleted from *S. flexneri* pINV, generating Δ*parAB* and Δ*stbAB*, respectively. Loss of the T3SS PAI was measured after 25 generations of growth using a *sacB-neo*^{R} cassette inserted into *mxiH* of all strains, as described previously (McVicker and Tang 2016). As *mxiH* is part of the T3SS PAI, this double marker allows selection for bacteria that have the T3SS PAI and hence the plasmid (by resistance to kanamycin), as well as bacteria that have lost the T3SS PAI through pINV loss or PAI deletions (by resistance to sucrose) (McVicker and Tang 2016). Because temperature can influence the contribution of TA systems to pINV maintenance (McVicker and Tang 2016), we hypothesised that the two partitioning systems

may be involved in pINV segregation at different temperatures. Therefore, we tested the impact of *parAB* and *stbAB* at both 21°C and 37°C. There was no difference in the stability of pINV with or without *stbAB*, irrespective of temperature (**Figure 3.7A**, wt *vs*. Δ*stbAB, p* <0.9999 *;* **Figure 3.7B**, wt *vs*. Δ*stbAB, p* > 0.9999; one-way ANOVA with non-parametric Dunn's test), while deletion of *parAB* had no effect during growth at 37°C. However, at 21°C the absence of *parAB* significantly increased stability of pINV compared with the wild-type strain (**Figure 3.7A**, wt *vs*. Δ*parAB* at 37°C, *p* > 0.9999; **Figure 3.7B**, wt *vs*. Δ*parAB* at 21°C, *p* = 0.0157; one-way ANOVA with non-parametric Dunn's test).

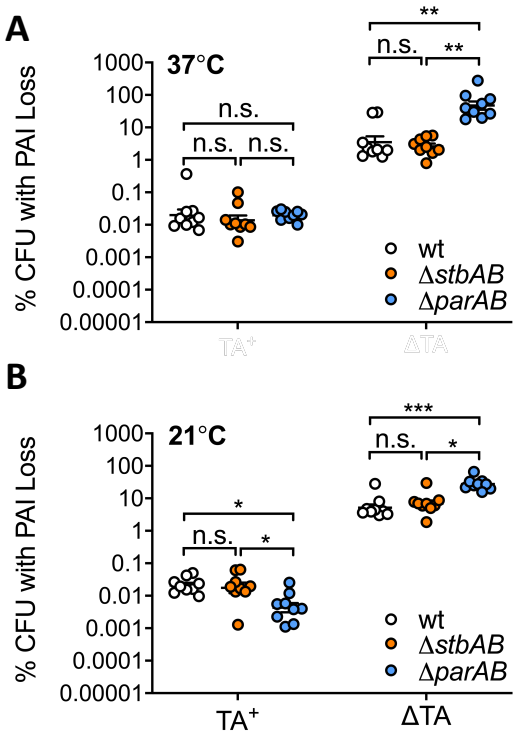


Figure 3.7 The effect of ParAB and StbAB on the stability of the *S. flexneri* pINV in the presence or absence of the three TA systems.

The percentage of colonies that lost the T3SS PAI was measured using *sacB-neo*^{R} in S. *flexneri* with both partitioning systems (wt), lacking *stbAB* (Δ*stbAB*), or *parAB* (Δ*parAB*), and in the presence or absence of MvpAT, GmvAT and CcdAB (TA^{+} and ΔTA, respectively) following growth at 37°C (**A**) and 21°C (**B**) for approximately 25 generations. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; n.s., not significant by one-way ANOVA with nonparametric Dunn's multiple comparison test (n = 9 colonies from three independent experiments; solid line, mean ± S.E.M).

The presence of TA systems on pINV could mask the effect of partitioning systems, as they could eliminate bacteria that have lost pINV by PSK. Therefore, we deleted *stbAB* and *parAB* in a strain lacking the TA systems, VapBC, CcdAB and GmvAT, and measured PAI loss using the *sacB-neo*^{R} assay following growth at 37°C or 21°C for 25 generations. Again, the absence of *stbAB* had no effect on pINV stability even without the three TA systems at both temperatures (**Figure 3.7A**, wt *vs*. Δ*stbAB, p* > 0.9999*;* **Figure 3.7B**, wt *vs*. Δ*stbAB, p* = 0.8156; one-way ANOVA with non-parametric Dunn's test). Interestingly, deletion of *parAB* in the triple TA system mutant resulted in a significant destabilisation of pINV at both temperatures (**Figure 3.7A**, wt *vs*. Δ*parAB, p* = 0.0026; **Figure 3.7B**, wt *vs*. Δ*parAB, p* = 0.0009; one-way ANOVA with non-parametric Dunn's test), showing the opposite effect to what was observed in the presence of TA systems at 21°C.

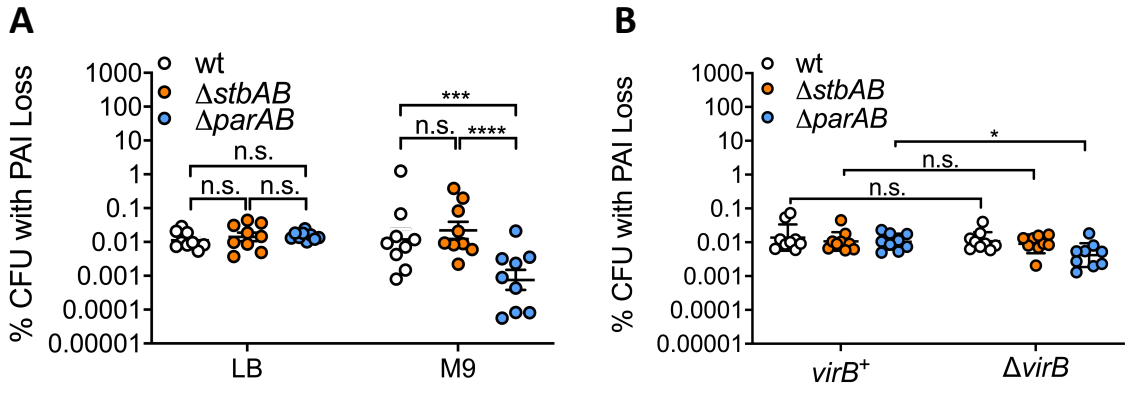


Figure 3.8 Deletion of *parAB* stabilises pINV in *S. flexneri*, when grown either in M9 or in the absence of *virB*.

The percentage of colonies with PAI loss was measured using *sacB-neo*^{R} in *S. flexneri* with both partitioning systems (wt), lacking *stbAB* (Δ*stbAB*) or *parAB* (Δ*parAB*), following growth in LB or M9 at 37°C (**A**), and in the presence or absence of *virB* (*virB*$^{+}$/Δ*virB*) at 37°C (**B**) for approximately 25 generations. **** $p < 0.0001$; *** $p < 0.001$; * $p < 0.05$; n.s., not significant by two-way ANOVA using parametric Tukey's multiple comparisons test in panel A, and one-way ANOVA with Dunn's multiple comparison test in panel B (n = 9 colonies from three independent experiments; solid line, mean ± S.E.M.).

Therefore, we decided to investigate further the variable role of ParAB on pINV stability at 21°C. As bacterial growth is slow at this temperature compared with 37°C, we decided to reduce the growth rate by another means, and tested the impact of Δ*parAB* and Δ*stbAB* in bacteria grown in minimal media (M9) at 37°C (**Figure 3.8A**). Removal of *parAB* significantly increased plasmid stability when bacteria were grown in M9 to a greater extent than observed during growth in LB at 21°C (**Figure 3.8A**, in M9 wt *vs*. Δ*parAB, p* = 0.0007; two-way ANOVA with parametric Tukey's test). Again, in M9 the absence of *stbAB* had no effect on pINV stability (**Figure 3.8A**, in M9 wt vs. Δ*parAB, p* = 0.6526; two-way ANOVA with parametric Tukey's test).

Finally, at 21°C expression of the T3SS is reduced, which we reproduced by deleting *virB*. Of note, VirB has homology with the partitioning protein, ParB (Watanabe *et al.,* 1990), and both recognise similar DNA sequences (Turner and Dorman, 2007), suggesting that there might be crosstalk between VirB and ParAB. Deletion of *virB* was introduced in both Δ*parAB* and Δ*stbAB*, and loss of the T3SS PAI was quantified following growth at 37°C as before (**Figure 3.8B**). In the absence of *virB*, there was no change in plasmid stability in both wt or Δ*stbAB* only (**Figure 3.8B**, *virB*$^+$ wt *vs*. Δ*virB* wt, $p > 0.9999$; *virB*$^+$ Δ*stbAB vs*. Δ*virB* Δ*stbAB*, $p > 0.9999$; one-way ANOVA with non-parametric Dunn's test); however, deletion of *parAB* showed a reduced albeit significant stabilisation of pINV compared with the *virB*$^+$ control (**Figure 3.8B**, *virB*$^+$ Δ*parAB vs*. Δ*virB* Δ*parAB*, $p = 0.0420$; one-way ANOVA with non-parametric Dunn's test).

In conclusion, these results demonstrate that ParAB is functional at both 37°C and 21°C, which is evident in the absence of TA systems. However, its contribution to pINV stability depends on the growth conditions. In the presence of TA systems at 21°C or in M9, ParAB is detrimental for pINV stability. A similar effect is also observed when T3SS expression is inhibited at 37°C

by deleting *virB*. However, in the absence of TA systems, ParAB contributes to pINV maintenance at both 37°C and 21°C. The role of StbAB is unclear, since the absence of this system had no detectable effect on plasmid stability under the condition tested.

3.3.2 Investigating pINV copy number as a potential factor influencing the effect of ParAB

As ParAB and other type I partitioning systems are generally found in large low-copy number plasmids (Reyes-Lamothe *et al.,* 2012), we hypothesised that the variable contribution of ParAB to plasmid stability could be explained by changes in pINV copy number that influence the efficacy of ParAB during segregation. For instance, if ParAB is more effective when pINV is present in fewer copies per cell, the detrimental effect of ParAB observed during growth at 21°C or in minimal media could reflect a higher copy number of pINV in these conditions.

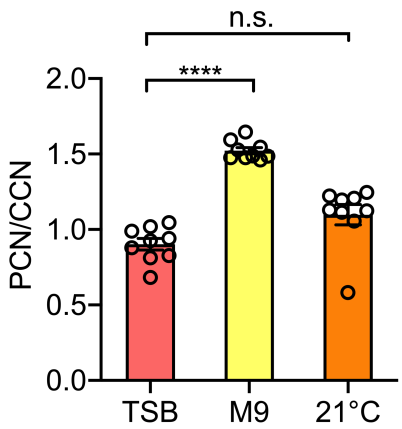


Figure 3.9 Plasmid copy number changes during growth in minimal medium.

Graph shows the ratio between plasmid copy number (PCN) and chromosome copy number (CCN) quantified by qPCR analysis in *S. flexneri* wild type grown in rich medium at 37°C (TSB) and at 21°C (21°C), and in minimal medium at 37°C (M9) for approximately 25 generations. Data are shown as mean ± S.E.M. (n = 9 biological repeats). ** $p < 0.01$; * $p < 0.05$; n.s., not significant; one sample *t*-test, nonparametric comparison of mean of condition/TSB to equal to 1 (n = 9 biological repeats).

To test this hypothesis, we measured the plasmid copy number (PCN) of pINV in *S. flexneri* M90T and expressed results as the ratio of the PCN to the chromosome copy number (PCN/CCN) (**Figure 3.9**). Quantitative PCR was used to measure the copies of the plasmid and the chromosome by detecting *repA* and *polA,* respectively. Known number of copies of pGEM plasmids containing *polA* and *repA* were used as a reference (Section 2.11). PCN was measured in bacteria grown in rich medium at 37°C (TSB) or at 21°C (21°C), or in minimal medium at 37°C (M9) for 25 generations (**Figure 3.9**).

At 37°C and in rich medium (TSB), the pINV was present in ~ 0.9 copies per cell, so presenting a PCN close to 1, which is the value we expected. However, the PCN of pINV increased significantly when bacteria were grown in minimal medium, reaching an average ~ 1.5 copies per cell (**Figure 3.9**,TSB *vs*. M9, $p < 0.0001$, one sample *t*-test). In contrast, pINV was found in ~ 1.1 copies per cell on average during growth at 21°C (**Figure 3.9**, TSB *vs*. 21°C, $p = 0.0510$, one sample *t*-test). However, the frequency distribution of the data in this condition shows that in more than half the biological replicates the PCN of pINV was closer to 1.2, suggesting that at 21°C pINV might be present at higher copy but not as marked as during growth in minimal media. Therefore, a distinct increase of the PCN of pINV was detected when bacteria were grown in minimal media, the same conditions in which removal of ParAB increased pINV stability (**Figure 3.8A**). A smaller, yet not significant, increase of PCN was also detected in bacteria grown at 21°C, which correlates with a reduced increase in the stability of pINV in the absence of ParAB (**Figure 3.8A**).

In conclusion, all these data are consistent with the hypothesis that variation of the pINV PCN influences the efficacy of ParAB in stabilising pINV. Specifically, the results suggest that when pINV is present in more than 1 copy per cell, ParAB is detrimental for plasmid retention.

3.4 Summary

In this chapter I describe how TA systems and partitioning systems contribute to the maintenance of pINV in *S. flexneri*, and how IS influence the genetic stability of this element.

In particular, I showed that, among the three TA systems present on the plasmid, VapBC is essential for pINV maintenance at 37°C and also stabilises the plasmid at 21°C (**Figure 3.1** and **Appendix Figure 7.2**). This system acts globally as well as locally, the latter by preventing loss of nearby genetic regions through recombination events, such as those mediated by IS (**Figure 3.2**). In fact, the major event leading to avirulence in *S. flexneri* is deletion of PAI-associated virulence genes mediated by recombination between by homologous IS (**Figure 3.3**). This is different from the result found in a previous study in which CR^- bacteria from a different strain (*S. flexneri* 2a) mostly arose following loss of pINV (Schuch and Maurelli, 1997). A pair of IS*Sfl4* is most commonly associated with T3SS PAI deletion in *S. flexneri* M90T, even though there are only two copies of this IS on pINV. There are other IS, such as IS*1294* and IS*600*, that occupy a far greater proportion of plasmid sequence, but are much less frequently involved in PAI deletions (**Appendix Table 7.1**). It is unclear why there is a difference in the frequency of the deletion variants. We hypothesise that this is influenced by the extent and length of homology, and the coverage of sequence of the IS pairs involved in the deletion. In fact, the percentage of coverage between the ISs involved in the less frequent deletions is lower than for the pair of IS*Sfl4* (**Appendix Table 7.1)**.

Furthermore, I demonstrated that these deletion events occur by RecA-mediated homologous recombination, and need both the RecBCD and RecFOR pathways to be functional (**Figure 3.4**). However, RecFOR is only partially involved, and the precise contribution of RecBCD is unclear, deletion of *recD* increased the frequency of PAI deletions, suggesting that its absence does

not disrupt the activity of this complex. This is consistent with the increased recombination in *recD* mutants observed previously (Lovett *et al.,* 1988) suggesting that, without RecD, RecBC retains some activity, and other proteins that are not required in the presence of RecD, such as RecJ, are recruited, possibly increasing the efficiency of recombination. Deletion of *recB* or *recC* (and complementation) would clarify this. Additionally, it would be interesting to identify the IS that cause T3SS PAI deletions in the absence of RecA and RecF. In particular, deletion of *virF* was still seen in the *recA* and *recD* mutants (**Figure 3.4**); characterising colonies lacking *virF* would identify the deletion sites and the potential IS/alternative elements involved.

Recombination events between homologous IS can also occur when one copy of the IS pair is present on the chromosome and the other one is on the plasmid. This results in the integration of the plasmid into the chromosome. Previous studies have shown chromosomal integration of *Shigella* pINV and the related plasmid from enteroinvasive *E. coli* but only after exposing bacteria to curing agents (such as rifampicin) or introducing other plasmids with pINV-incompatible replicons (Zagaglia *et al*., 1991). After these manipulations, pINV usually integrates into *metB*, leading to methionine auxotrophy. In contrast, I demonstrated that *S. flexneri* pINV integrates into the chromosome following intermolecular recombination between chromosomal and plasmid copies of IS*1294* (**Figure 3.5**). Plasmid integration induces a down-regulation of the expression of T3SS PAI genes, resulting in a dramatic reduction of T3SS activity (**Figure 3.5**); this is a reversible process as plasmid excision spontaneously occurs and restores the activity of the T3SS (**Figure 3.6**).

Finally, the reason why *S. flexneri* needs two partitioning systems remains an open question. Previous experiments have shown that both systems are active and capable of stabilising a model vector containing the origin of replication from pINV at both 37°C and 21°C (McVicker *et al*., 2019). However, when the contribution of each partitioning system is assessed on pINV,

the effect of their absence is masked by TA systems which eliminate the majority of bacteria that have lost the plasmid (**Figure 3.7**). In fact, removal of the TA systems from pINV demonstrated that ParAB is required to maintain the plasmid at both 37°C and 21°C. In contrast, in the presence of TA systems and during low growth rate (such as at 21°C or growth in minimal media), ParAB becomes detrimental for pINV stability (**Figure 3.8**). I hypothesised this is due to an increase in the PCN of pINV, which rises from 1 in higher growth rate conditions (rich medium, 37°C) to 1.2-1.5 in lower growth rate conditions (**Figure 3.9**) (at 21°C and in minimal medium, respectively). In fact, if ParAB is less effective at segregating pINV when the PCN > 1, then the increased PCN in low growth rate conditions might make ParAB detrimental for plasmid retention. Unfortunately, due to lack of time, it was not possible to quantify the PCN of pINV in Δ*virB,* to provide independent evidence to support this hypothesis. Furthermore, these results should be confirmed with microscopic visualisation of pINV and its segregation during cell division, in bacteria with or without functioning partitioning systems, and grown at the above tested conditions.

Finally, it is unclear how StbAB (a ParMR-homologue) contributes to pINV maintenance. Studies on the virulence plasmid of enteropathogenic *E. coli*, pB171, showed that this plasmid possesses two partitioning systems, ParAB and ParMR, which are each sufficient for plasmid segregation, although both systems are needed for optimal plasmid maintenance (Ebersbach and Gerdes, 2001). However, it has been demonstrated that there is regulatory crosstalk between the two systems, with ParB repressing the expression of ParMR (Ringgaard *et al*., 2006). Therefore, it would be interesting to verify if there is any regulatory crosstalk between ParAB and StbAB in *Shigella*.

4 RESULTS: Analysis of the stability of pINV in *S. sonnei*

In this chapter, I assessed the stability of pINV in *S. sonnei* by comparing it with *S. flexneri* pINV, and analysed how plasmid dynamics in *S. sonnei* are influenced by TA systems and IS.

4.1 The role of TA systems

Previous studies have shown that pINV^{Ssonn} is significantly less stable than pINV^{Sflex} (McVicker and Tang, 2016). Sequence analysis of pINV^{Ssonn} demonstrated that the plasmid possesses VapBC but lacks two other TA systems, CcdAB and GmvAT, that are found in all other *Shigella* species representatives (McVicker and Tang, 2016). More recently, a predicted RelBE TA system has been identified on pINV^{Ssonn}, located ~ 960 bp from the origin of replication of the plasmid (Martyn *et al*., unpublished). Although this system is functional, it does not contribute to plasmid maintenance at either 37°C or 21°C in laboratory conditions (Martyn *et al*., unpublished). Finally, comparison of the amino acid sequences of VapBC from *S. flexneri* and *S. sonnei*, revealed that both VapB and VapC from *S. sonnei* contain a single amino acid difference compared with the proteins in *S. flexneri*.

Therefore, the aim of work in this chapter was to further investigate how absence of CcdAB and GmvAT affects plasmid dynamics in *S. sonnei*, including loss of the plasmid and loss of discrete sequences, and to define the effect of the two amino acid substitutions found in VapBC on pINV^{Ssonn} maintenance.

4.1.1 The absence of *ccdAB* and *gmvAT* contributes to pINV loss

It was previously demonstrated, using an assay with *sacB-neo*^{R} positioned inside the T3SS PAI, that introduction of *ccdAB* and *gmvAT* into pINV^{Ssonn} at sites corresponding to their positions in *S. flexneri* reduced the emergence of avirulent bacteria in *S. sonnei* (McVicker and Tang, 2016). Although the *sacB-neo*^{R} assay is more sensitive than using the loss of CR binding to detect loss of the T3SS PAI, this technique does not distinguish loss of the plasmid from deletions affecting the T3SS PAI. Furthermore, it will not detect events (such as loss of *virF*) that lead to avirulence but do not involve the T3SS PAI.

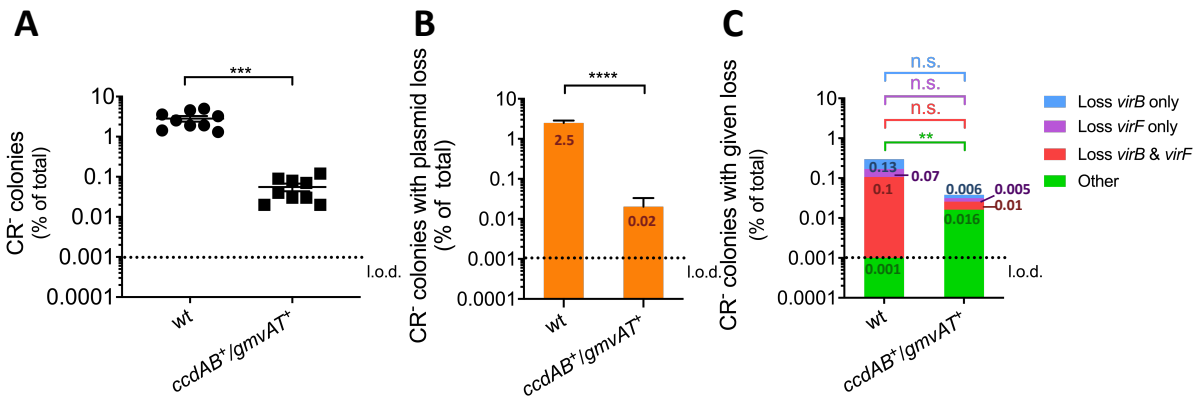


Figure 4.1 CR^{-} colonies of *S. sonnei* mostly result from loss of entire pINV.

(**A**) Number of CR^{-} colonies emerging from wild-type *S. sonnei* 53G (wt) and a strain with *ccdAB* and *gmvAT* introduced into pINV (*ccdAB*$^{+}$/*gmvAT*$^{+}$) after growth at 37°C for ~ 25 generations, shown as a % of all colonies. Solid line: mean ± S.E.M (n = 9 biological replicates). ***, $p \leq 0.001$ by parametric Welch's *t*-test. (**B** and **C**) Number of CR^{-} colonies lacking specified virulence-related genes were assayed by multiplex PCR and shown as a % of all colonies. Loss of the entire plasmid is inferred by loss of the origin of replication (B); "others" refers to CR^{-} colonies that contain *virB*, *virF* and the origin of replication (C). The value of each category is indicated inside the corresponding bar portion. Eight independent CR^{-} colonies were obtained from each of nine biological replicates (*i.e.* 72 colonies were analysed per strain); data are shown as mean + S.E.M. in panel B and mean in panel C. **, $p \leq 0.01$; ****; $p \leq 0.0001$; n.s., not significant; one sample *t*-test (parametric comparison of mean test/wt strain to 1) and Wilcoxon signed-rank test (nonparametric comparison of median of strain/wt to 1, for each event). l.o.d., limit of detection.

Therefore, to determine whether introduction of *ccdAB* and *gmvAT* into pINV^{Ssonn} has a global effect on the stability of the plasmid, the emergence of CR$^-$ colonies was examined in a *S. sonnei* strain, *ccdAB*$^+$/*gmvAT*$^+$, lacking the *sacB-neo*^{R} cassette but harbouring *ccdAB* and *gmvAT*, and the results were compared with the wild-type *S. sonnei* 53G (wt) (**Figure 4.1**). To prevent plasmid loss before the start of experiments, a chloramphenicol resistance cassette (*cat*) was inserted downstream of *vapBC* and *ccdAB* in wild-type *S. sonnei* and in *ccdAB*$^+$/*gmvAT*$^+$, respectively, and bacteria grown in the presence of the antibiotic prior to assays. After 25 generations of growth at 37°C, the presence of CcdAB and GmvAT resulted in a significant reduction in the emergence of CR$^-$ colonies, which dropped from ~ 2.8% of the population of wt bacteria to ~ 0.055% in *ccdAB*$^+$/*gmvAT*$^+$ (**Figure 4.1A**, $p = 0.0003$, Welch's *t*-test).

To identify the genetic events responsible for the emergence of CR$^-$ bacteria in these strains, CR$^-$ colonies arising from wild-type *S. sonnei* and *ccdAB*$^+$/*gmvAT*$^+$ were examined by multiplex PCR for the presence of *virB*, *virF*, and *ori*, as described previously (Section 3.1.1). Similar to *S. flexneri*, deletion events affecting T3SS-related genes are observed also in *S. sonnei*. However, in contrast to *S. flexneri*, loss of pINV is the leading cause of avirulence in wild-type *S. sonnei*, which was detected in ~ 2.5% of all colonies (**Figure 4.1B**). Interestingly, introduction of *ccdAB* and *gmvAT* resulted in a significant reduction, but not elimination, of plasmid loss which was found in only ~ 0.02% of all colonies (**Figure 4.1B**, for plasmid loss in wt *vs*. *ccdAB*$^+$/*gmvAT*$^+$, $p < 0.0001$, one sample *t*-test). Furthermore, no significant difference in the loss of *virB* and/or *virF* was detected in *ccdAB*$^+$/*gmvAT*$^+$ and wt (**Figure 4.1C**, loss of *virB* only, $p = 0.6250$; loss of *virF* only, $p = 0.8750$; loss of *virB* and *virF*, $p = 0.8125$; Wilcoxon signed-rank test). However, there was a significant increase in the "other" population of bacteria (*i.e.* CR$^-$ colonies that

retained *ori*, *virF* and *virB* by PCR) in *ccdAB*$^{+}$/*gmvAT*$^{+}$ compared with wt (**Figure 4.1C**, p = 0.0070, one sample *t*-test).

These results demonstrate that the genetic events responsible for emerging avirulent bacteria are distinct in *S. sonnei* and *S. flexneri*, with plasmid loss occuring at a significantly higher frequency in *S. sonnei* than *S. flexneri*. However, loss of T3SS-related genes is observed in *S. sonnei* as well as *S. flexneri*. Furthermore, introduction of CcdAB and GmvAT into pINV^{Ssonn} reduces but does not eliminate plasmid loss in *S. sonnei*, which still accounts for almost 50 % avirulent colonies in *ccdAB*$^{+}$/*gmvAT*$^{+}$.

4.1.2 Amino acid substitutions in VapBC contribute to pINV stability

As VapBC is the only TA system present in both *S. flexneri* and *S. sonnei*, and is essential for pINV^{Sflex} retention at 37°C (McVicker and Tang, 2016), we speculated that residual plasmid loss observed in *ccdAB*$^{+}$/*gmvAT*$^{+}$ could be due to differences in the *vapBC* sequence. Therefore, the nucleotide and predicted amino acid sequences of this TA system in *S. sonnei* were compared with those in *S. flexneri*. It was found that, while the promoter regions were identical, there is a single amino acid difference in VapB and in VapC from *S. sonnei* 53G and *S. flexneri* M90T (Martyn *et al*., unpublished). In the C-terminus of VapB, the Thr^{58} residue in *S. flexneri* is replaced by Ala in *S. sonnei*; this region is involved in neutralising VapC toxicity although this amino acid is not near the active site of VapC in the VapBC structure. In VapC, the Lys^{32} present in *S. flexneri* is replaced by Arg in *S. sonnei*; this substitution is located in a region of the protein that is neither part of the active site of the toxin nor the VapB-interacting domain (**Appendix Figure 7.8**) (Dienemann *et al*., 2011; Martyn *et al*., unpublished).

To investigate if these substitutions in VapBC are responsible for the increased loss of pINV^{Ssonn}, strains of *S. sonnei* containing pINV either with *S. sonnei vapBC* (generating strain VapBC^{Ssonn}), *vapBC* from *S. flexneri* (VapBC^{Sflex}), or a chimeric *vapBC* (*i.e.* VapB^{Sflex}C^{Ssonn} and VapB^{Ssonn}C^{Sflex}) were generated; a chloramphenicol resistance cassette (*cat*) was inserted downstream of *vapBC* in all strains to maintain pINV until the start of experiments by growing strains in the presence of the antibiotic. The emergence of CR$^-$ colonies in these strains was quantified after growth for ~25 generations at 37°C in the absence of chloramphenicol, and compared with wild-type *S. flexneri* M90T. The proportion of CR$^-$ colonies emerging at this stage was significantly higher in wild-type *S. sonnei* (VapBC^{Ssonn}) than in *S. flexneri* (**Figure 4.2A**, *S. sonnei* VapBC^{Ssonn} *vs*. *S. flexneri*, p = 0.0039, one sample *t*-test), accounting for 5.2% and 0.25% of the population of these strains, respectively.

Interestingly, a marked decrease in the number of CR$^-$ bacteria was observed in *S. sonnei* VapBC^{Sflex} and VapB^{Ssonn}C^{Sflex} compared with *S. sonnei* VapBC^{Ssonn}, both of which show similar levels of CR$^-$ bacteria compared with *S. flexneri* (**Figure 4.2A**, *S. flexneri vs*. *S. sonnei* VapBC^{Sflex}, p = 0.8203; *S. flexneri vs*. *S. sonnei* VapB^{Ssonn}C^{Sflex}; p = 0.4961, one sample *t*-test). In contrast, there was significant difference in the number of CR$^-$ bacteria emerging from VapB^{Sflex}C^{Ssonn} compared with *S. flexneri* (**Figure 4.2A**, *S. flexneri vs*. *S. sonnei* VapB^{Sflex}C^{Ssonn}, p > 0.0039, one sample *t*-test).

Next, the molecular events responsible for avirulence were assessed in the same strains by multiplex PCR as previously (Section 3.1.1). Of note, the presence of *vapC* from *S. flexneri* led to a significant reduction in plasmid loss compared with *vapC* from *S. sonnei* (**Figure 4.2B**, for plasmid loss, *S. sonnei* VapBC^{Ssonn} *vs*. VapBC^{Sflex}, p = 0.0039; *S. sonnei* VapBC^{Ssonn} *vs*. VapB^{Ssonn}C^{Sflex}, p = 0.0039; Wilcoxon signed-rank test).

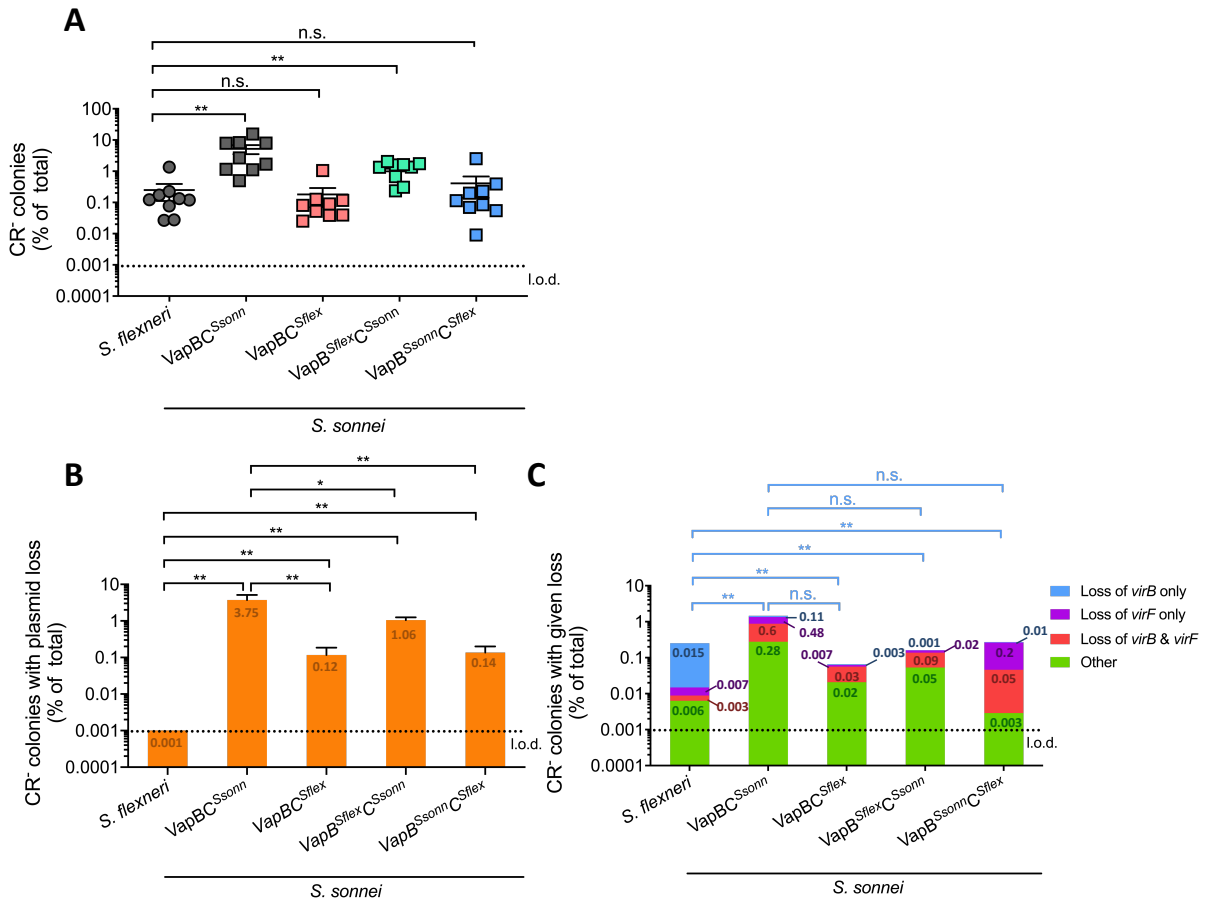


Figure 4.2 VapC polymorphisms affect plasmid maintenance.

(**A**) Number of CR$^-$ colonies emerging from wild-type *S. flexneri* or *S. sonnei* (VapBC^{Ssonn}), or *S. sonnei* with pINV harbouring *vapBC*^{Sflex} (VapBC^{Sflex}), *vapB*^{Sflex}*C*^{Ssonn} (VapB^{Sflex}C^{Ssonn}) or *vapB*^{Ssonn}*C*^{Sflex} (VapB^{Ssonn}C^{Sflex}) after growth at 37°C for ~ 25 generations, shown as a % of all colonies; solid line, mean ± S.E.M. (n = 9 biological replicates); **, $p \leq 0.01$; n.s., not significant; Wilcoxon signed-rank test (nonparametric comparison of median of strain/*S. flexneri* to 1). (**B** and **C**) Number of CR$^-$ colonies lacking specified virulence-related genes were assayed by multiplex PCR and shown as a % of all colonies. Loss of the entire plasmid is inferred by loss of the origin of replication (B); "others" refers to CR$^-$ colonies that contain *virB*, *virF* and the origin of replication (C). The value of each category is indicated inside the corresponding bar portion. Eight independent CR$^-$ colonies were obtained from each of three biological replicates (i.e. 72 colonies were analysed per strain); data are shown as mean + S.E.M. in panel B and mean in panel C. ****, $p \leq 0.0001$; ***, $p \leq 0.001$; **, $p \leq 0.01$; *, $p \leq 0.05$, n.s., not significant; Wilcoxon signed-rank test (nonparametric comparison of median of strain/ *S. flexneri* or *S. sonnei* to 1, for each event). l.o.d., limit of detection.

Furthermore, there was also a marked difference in the rate of plasmid loss in *S. flexneri* compared with *S. sonnei* VapBC^{Ssonn} or VapB^{Sflex}C^{Ssonn} (**Figure 4.2B** p = 0.0039 for both

comparisons, Wilcoxon signed-rank test). In contrast, replacement of *vapB* from *S. sonnei* with *vapB* from *S. flexneri* had less effect on plasmid loss (**Figure 4.2B**, for plasmid loss, *S. sonnei* VapBC^{Ssonn} *vs*. VapB^{Sflex}C^{Ssonn}, $p = 0.0391$; Wilcoxon signed-rank test).

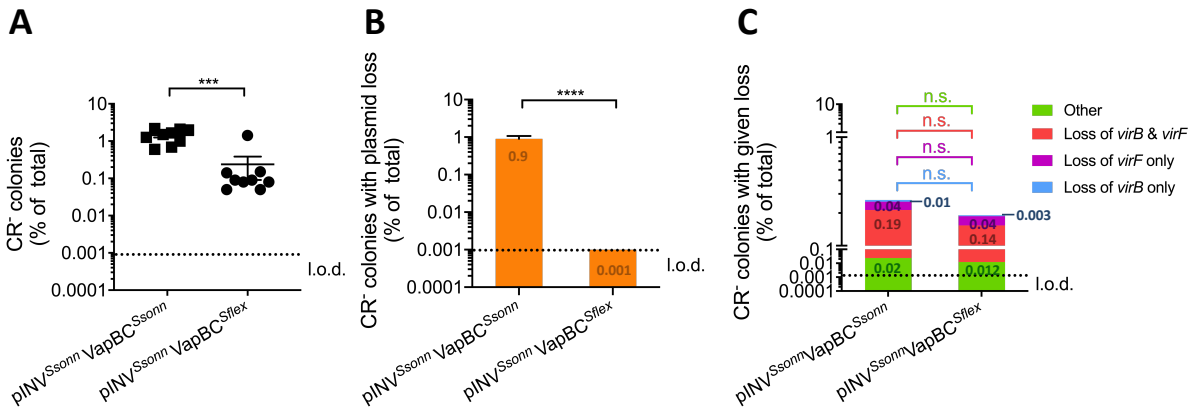


Figure 4.3 Loss of S. sonnei pINV is eliminated in a S. flexneri background.

(**A**) Number of CR$^-$ colonies emerging from *S. flexneri* carrying pINV^{Ssonn} vapBC^{Ssonn} or pINV^{Ssonn} *vapBC*^{Sflex} after growth at 37°C for ~ 25 generations, shown as a % of all colonies; solid line, mean ± S.E.M. (n = 9 biological replicates); ***, $p \leq 0.001$ by Mann-Whitney *t*-test. (**B**) and (**C**) Number of CR$^-$ colonies lacking specified virulence-related genes were assayed by multiplex PCR and shown as a % of all colonies. Loss of the entire plasmid is inferred by loss of the origin of replication (B); "others" refers to CR$^-$ colonies that contain *virB*, *virF* and the origin of replication (C). The value of each category is indicated inside the corresponding bar portion. Eight independent CR$^-$ colonies were obtained from each of three biological replicates (*i.e.* 72 colonies were analysed per strain); data are shown as mean + S.E.M. in panel B and mean in panel C. ****, $p \leq 0.0001$; n.s., not significant; one sample *t*-test (parametric comparison of mean test/control strain to 1) in panel B and Wilcoxon signed-rank test (nonparametric comparison of median of test/control strain to 1, for each event) in panel A. l.o.d., limit of detection.

Additionally, there were marked differences in the rates of *virB* and/or *virF* loss in *S. flexneri* and all *S. sonnei* strains. Loss of *virB* predominated in avirulent colonies of *S. flexneri* but was significantly less frequent in all *S. sonnei* strains (**Figure 4.2C**, loss of *virB*, *S. flexneri vs*. *S. sonnei* VapBC^{Ssonn} and *S. sonnei* VapB^{Ssonn}C^{Sflex} $p = 0.0078$; *S. flexneri vs*. *S. sonnei* VapBC^{Sflex} and *S. sonnei* VapB^{Sflex}C^{Ssonn}, $p = 0.0039$; Wilcoxon signed-rank test). Furthermore, loss of *virF* only

and loss of *virF* with *virB* was rarely detected in *S. flexneri*, but was more common than the loss of *virB* in all *S. sonnei* strains (**Figure 4.2C**).

These results demonstrate that the presence of *S. flexneri* VapC reduces the emergence of CR^- colonies, and more particularly, plasmid loss in *S. sonnei*. Furthermore, differences in the rates of loss of *virB* between *S. flexneri* and all *S. sonnei* strains provide for the first time evidence that distinct deletions of the T3SS PAI-related genes occur in these two species and are independent of any differences in VapBC.

To understand if the genetic background also affects the activity of VapBC, $pINV^{Ssonn}$ containing wild-type VapBC (generating strain $pINV^{Ssonn}$ $vapBC^{Ssonn}$) or VapBC from *S. flexneri* ($pINV^{Ssonn}$ $vapBC^{Sflex}$) was transferred into a *S. flexneri* strain lacking pINV (BS176) using tri-parental mating (**Figure 4.3**) (Section 2.3). Results of CR binding assays demonstrate that the presence of $VapBC^{Sflex}$ on $pINV^{Ssonn}$ significantly reduced the emergence of CR^- colonies compared with $VapBC^{Ssonn}$ even in *S. flexneri* (**Figure 4.3A**, *S. flexneri* $pINV^{Ssonn}$ $vapBC^{Ssonn}$ *vs.* $pINV^{Ssonn}$ $vapBC^{Sflex}$, $p = 0.0005$, Mann-Whitney *t*-test). Furthermore, multiplex PCR analysis of the CR^- colonies emerging from both strains showed that $pINV^{Ssonn}$ is lost less frequently in a *S. flexneri* genetic background than in *S. sonnei*, going from ~ 2.5% to 0.9% (**Figure 4.2B** and **Figure 4.3B**); plasmid loss is further reduced in the presence of $VapBC^{Sflex}$ (**Figure 4.3B**, *S. flexneri* $pINV^{Ssonn}$ $vapBC^{Ssonn}$ *vs.* $pINV^{Ssonn}$ $vapBC^{Sflex}$, $p < 0.0001$, one sample *t*-test). However, no change in the rates of *virB* and/or *virF* deletions was detected in the two genetic backgrounds (**Figure 4.3C**).

In conclusion, these findings confirm that the polymorphism in *vapC* is responsible for the different ability of $VapBC^{Ssonn}$ and $VapBC^{Sflex}$ to prevent pINV loss irrespective of genetic background.

4.2 The role of Insertion Sequences

As shown in the previous chapter, IS play an important role in the plasticity of pINV^{Sflex}, mediating deletions of the T3SS PAI and integration of pINV into the chromosome. IS also constitute a large proportion of pINV^{Ssonn}, and represent ~ 23% of all plasmid open reading frames. Therefore, we decided to investigate whether IS are similarly involved in molecular rearrangements that are responsible for the spontaneous loss of *virB* and/or *virF* in *S. sonnei* (**Figure 4.1C** and **Figure 4.2C**).

4.2.1 IS are involved in the deletion of the T3SS PAI in *S. sonnei*

To determine whether IS contribute to the loss of T3SS-related genes from pINV^{Ssonn}, we identified pairs of IS in the same orientation and flanking the T3SS PAI and/or *virF*.

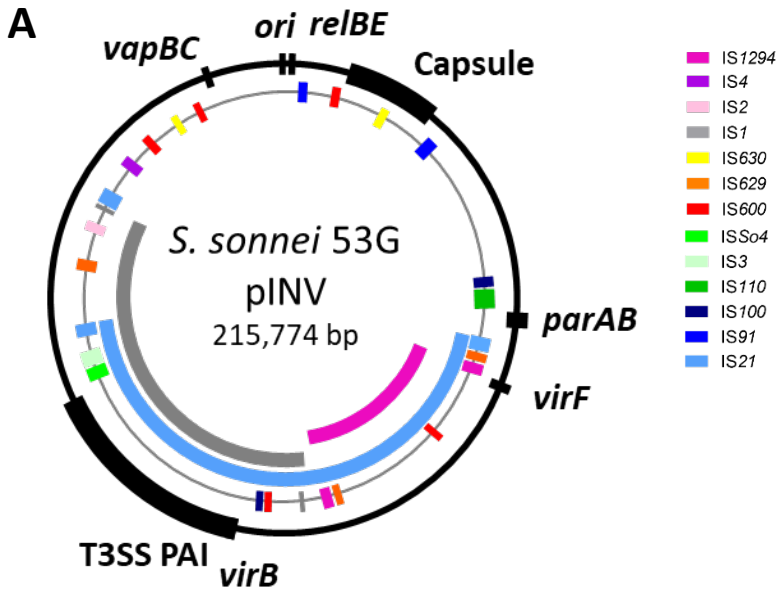


Figure 4.4 Schematic of the sites of T3SS PAI-related deletions in *S. sonnei* 53G.

Alignment of the three PAI deletions with *S. sonnei* 53G pINV (NC_016833) indicated as black ring; IS are indicated as in the colour-coded boxes.

CR^{-} colonies of wild-type *S. sonnei* isolated from previous CR binding assays were analysed by PCR, using primers binding to sequences flanking the predicted T3SS PAI deletions; PCR products of the size predicted for a deletion were then sequenced. This approach revealed the involvement of three pairs of IS in deletions on pINV^{Ssonn}: i) copies of IS*21* (**Figure 4.4**, blue arc, at positions 29,959–32,101 bp and 123,799–125,930 bp) mediated deletion of a 93.7 kb fragment, resulting in loss of the T3SS PAI and *virF*; ii) copies of IS*1* (**Figure 4.4**, grey arc, at 73,733–74500 and 146,228–146996 bp) led to loss of the T3SS PAI; while iii) copies of IS*1294* (**Figure 4.4**, purple arc, at 33,932–35,621 and 68,896–70,585 bp) were involved in deletion of *virF*.

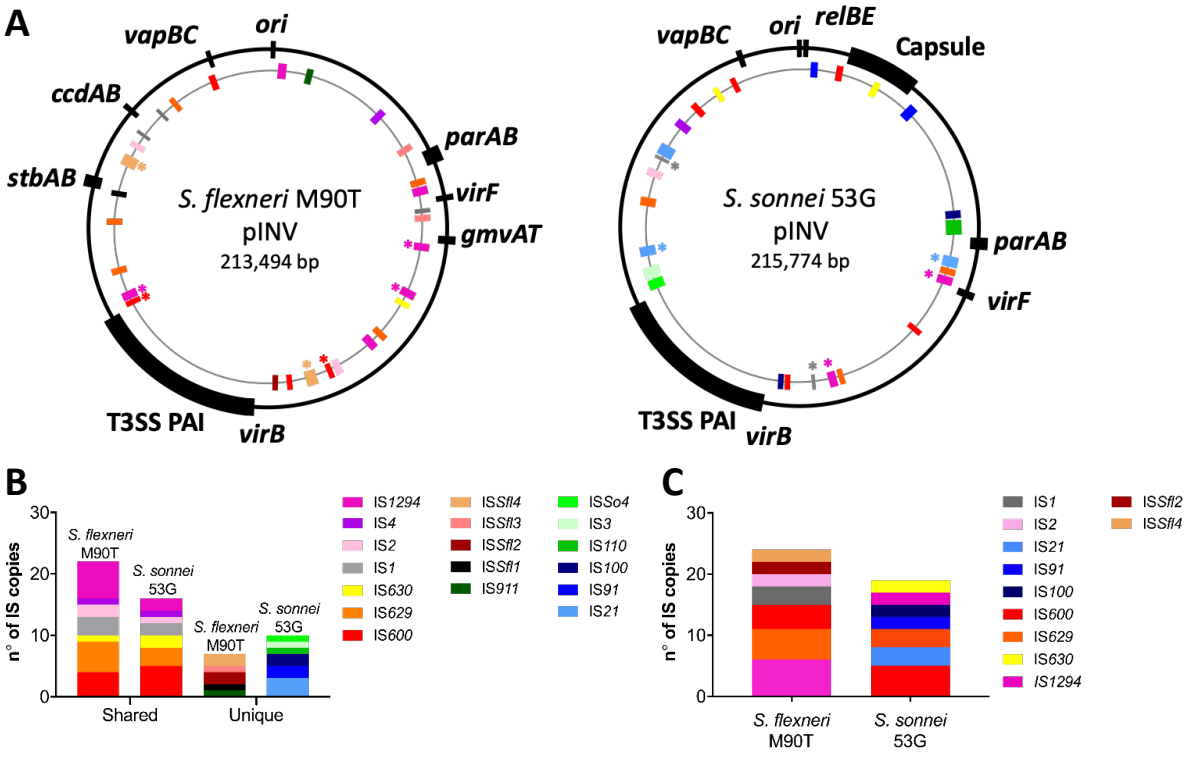


Figure 4.5 Distinct profile of IS on pINV from *S. flexneri* M90T and *S. sonnei* 53G.

(**A**) Schematic of the distribution of intact IS on pINV from *S. flexneri* M90T (AL391753.1) and *S. sonnei* 53G (NC_016833). IS are coloured as indicated; asterisks indicate the IS that lead to avirulent colonies in *S. flexneri* and *S. sonnei*. (**B**) Copies of IS that are found in both strains, or only found in one strain (unique). (**C**) IS that are present in more than one copy.

To establish whether the differences in the deletions leading to avirulence in *S. sonnei* compared to *S. flexneri* (Section 3.2.1) are due to the presence of distinct IS in the two species, the number of intact IS on pINV^{Ssonn} was examined and compared with those in pINV^{Sflex} (**Figure 4.5**). In *S. flexneri* M90T, pINV (pWR100) harbours a total of 29 different IS, consisting of 12 IS types; IS*Sfl4,* IS*1294*, and IS*600* are responsible for most T3SS PAI deletions (**Figure 4.5A**). IS*1294* is the most abundant IS on pINV^{Sflex}, present six times on the plasmid. There are five and four copies of IS*629* and IS*600* on the plasmid, respectively; only four IS are present in single copy (**Figure 4.5B** and **C**). In contrast, in pINV^{Ssonn} a total of 26 IS are found with 13 different types of intact IS. IS*600* is the most abundant IS in pINV^{Ssonn}, found five times on the plasmid; the remaining IS are present in up to three copies on the plasmid, with seven present only in single copy (**Figure 4.5B** and **C**). Therefore, there are substantial differences in the content of IS of the two plasmids, with pINV^{Sflex} harbouring more IS in multiple copies than pINV^{Ssonn}.

4.3 Summary

In this chapter, I show that *S. sonnei* loses virulence more frequently than *S. flexneri*, with plasmid loss constituting the major cause of avirulence in *S. sonnei*, consistent with previous reports (Kopecko *et al*., 1980; Sansonetti *et al*., 1980; 1981). I provide evidence that three factors contribute to differences between *S. sonnei* and *S. flexneri*. First, as shown previously (McVicker and Tang, 2016), absence of *ccdAB* and *gmvAT* increases the emergence of avirulence in *S. sonnei*. In fact, introduction of these two TA systems in pINV^{Ssonn} led to a significant increase in plasmid retention. Despite this, ~ 50 % of avirulent bacteria still resulted from plasmid loss when these TA systems were present on pINV^{Ssonn} (**Figure 4.1**), suggesting that the ancestral deletion of these systems from pINV^{Ssonn} only partially contributes to the high frequency of plasmid loss in *S. sonnei*. Secondly, a single amino acid substitution in position 32 of VapC from VapBC affects the efficacy of this system in retaining pINV^{Ssonn}. VapBC is the only functional TA system that contributes to plasmid maintenance in *S. sonnei* at 37°C in the laboratory (McVicker and Tang, 2016; Martyn *et al*., unpublished) and is also present in *S. flexneri*, where it plays a critical role in maintaining pINV^{Sflex} (Radnedge *et al*., 1997; Sayeed *et al*., 2000; Sayeed *et al*., 2005; McVicker and Tang, 2016). Substitution of VapC for the *S. flexneri* version resulted in a significant reduction in loss of pINV^{Ssonn} (**Figure 4.2**). Nonetheless, plasmid loss was still evident despite this change, suggesting that together with the deletion of *ccdAB* and *gmvAT*, the polymorphism in *vapC* is partially responsible for the higher plasmid instability observed in *S. sonnei* compared with *S. flexneri*. Substitution of *vapC* together with introduction of *ccdAB* and *gmvAT* in pINV^{Ssonn} could be performed to test this hypothesis.

Third, the ability of VapBC^{Ssonn} and VapBC^{Sflex} to prevent pINV loss seems to be influenced by the genetic background. In fact, when pINV^{Ssonn} was transferred into a *S. flexneri* strain lacking pINV, loss of this plasmid was reduced without any change to the plasmid. Indeed, plasmid

loss was under the l.o.d. among 72 CR^- colonies emerging from a strain in which $VapBC^{Ssonn}$ was replaced with $VapBC^{Sflex}$ (**Figure 4.3**). There might be differences between the species that impact the activity of Lon or sensitivity of $tRNA^{fMet}$ to VapC cleavage and thereby modulate the impact of VapBC during PSK.

It is unclear how the VapC substitution affects the ability of VapBC in maintaining $pINV^{Ssonn}$. Analysis of the atomic structure of the $VapBC^{Ssonn}$ complex did not show any change in the overall structure of VapBC (**Appendix Figure 7.8**) and there was no detectable change in the toxicity of VapC, its ability to cleave $tRNA^{fMet}$, or its susceptibility to neutralisation by VapB (Martyn *et al*., unpublished). For PSK, VapC must be liberated from VapB. Therefore, it is possible that the VapC polymorphism affects its affinity for VapB; reduced interaction between VapB and VapC might increase the ability of VapC to cause PSK. This could be tested by measuring the affinity of the two proteins by surface plasmon resonance. Alternatively, the change in VapC might affect degradation of VapB by the Lon protease, which could be examined in *in vitro* degradation assays. Furthermore, the different VapC alleles might affect its ability to cleave $tRNA^{fMet}$; this has not been observed in *in vivo* toxicity or cleavage assays performed so far (Martyn *et al*., unpublished).

Finally, similar to *S. flexneri*, avirulence also arises following deletions of T3SS-related genes in *S. sonnei*, but involves different IS than in *S. flexneri.* These intra-molecular rearrangements often go undetected in *S. sonnei* because of the high level of plasmid loss (**Figure 4.4**). The involvement of different pairs of IS in the deletion of T3SS-related genes correlates with the different repertoire and organisation of IS in $pINV^{Ssonn}$ and $pINV^{Sflex}$ (**Figure 4.5**). This is also likely to be influenced by the complement of TA systems, including GmvAT. This TA system is located only 7.4 kb from *virF* on $pINV^{Sflex}$ and within a region bounded by two homologous

copies of IS*1294* (**Figure 4.5A**). IS*1294*-mediated loss of *virF* is infrequently detected in *S. flexneri*, but is relatively common in *S. sonnei*, which lacks GmvAT (**Figure 4.2C**). Introduction of *gmvAT* into pINV^{Ssonn} at sites corresponding to its position in *S. flexneri* resulted in decreased loss of *virF* (**Figure 4.1C**). This is consistent with the results in the previous chapter that demonstrated that TA systems play a role in maintaining nearby sequences.

5 RESULTS: Evaluation of CRISPR-Cas gene editing system to induce loss of virulence in *Shigella*

In this chapter, I assessed the use of CRISPR-Cas gene editing to induce loss of virulence in *Shigella*. Recent studies have demonstrated that the CRISPR-Cas9 system, a simplified and engineered Type II CRISPR-Cas system from *Streptococcus pyogenes* (Jinek *et al*., 2012; Mali *et al*., 2013), can be exploited to eliminate specific bacterial strains by targeting distinct sequences in their genome, or to remove plasmids (Gomaa *et al*., 2014; Cao *et al*., 2017; Lauritsen *et al*., 2017).

Shigella spp. possess a type I-E CRISPR-Cas locus that often lacks *cas* genes and/or is disrupted by IS (Guo *et al*., 2015; Wang *et al*., 2016). Type I and type II CRISPR-Cas systems employ different mechanisms to recognise and cleave target DNA sequences (van der Oost *et al*., 2014). Therefore, even if the endogenous CRISPR-Cas system is functional in *Shigella*, it should not interfere with an introduced CRISPR-Cas9 system. Given the importance of the T3SS and thus pINV, for *Shigella* virulence, we speculated that we could induce loss of virulence in *Shigella* using CRISPR-Cas9

5.1 Testing *virB* and *vapBC* as potential CRISPR-Cas9 targets to induce loss of virulence in *Shigella*

To induce the loss of virulence in *Shigella*, we decided to use the CRISPR-Cas9 system encoded by pCas9, a plasmid that has been widely employed for gene editing in *E. coli* (Jiang *et al*., 2013). This plasmid contains a functional Type II CRISPR-Cas9 system from *S. pyogenes*

expressed under the control of its native promoter. *cas9* and *tracrRNA* are expressed constitutively with their promoters positioned close to each other but in opposite orientations (Jiang *et al*., 2013), while the *crRNA* is under the control of the *cas9* promoter and is composed of two direct repeats (DR) separated by a single spacer (**Figure 5.1**). The absence of *cas1*, *cas2* and *csn2* means that the system is unable to acquire new spacer sequences, allowing controlled targeting of protospacers identical to the inserted spacers (Jiang *et al*., 2013). Furthermore, an antibiotic cassette is present for selection (Jiang *et al*., 2013) (**Figure 5.1**).

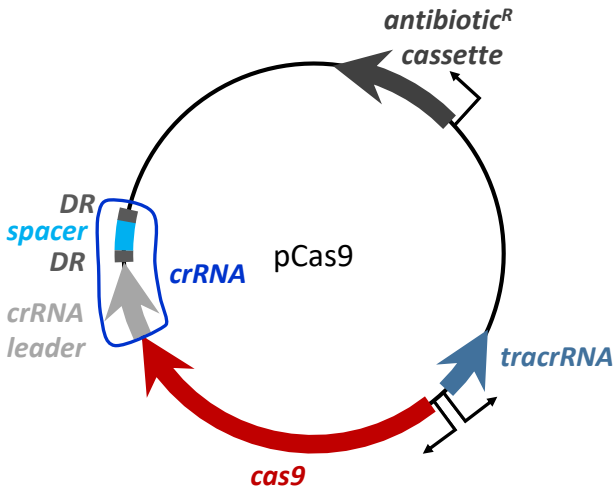


Figure 5.1 Schematic of pCas9.

Originally designed by Jiang *et al*., 2013. DR: direct repeat

Different versions of pCas9 were designed to target two distinct pINV loci, *virB* and *vapBC*. *virB* is located in the T3SS PAI, and encodes a transcriptional regulator essential for expression of the T3SS at 37°C (Dorman *et al*., 2001). Therefore, disruption of *virB* by CRISPR-Cas9 would result in the abrogation of the expression of the T3SS. The TA system VapBC is essential to maintain pINV at 37°C (McVicker and Tang, 2016; Schuch and Maurelli, 1997); consequently, its inactivation should result in the loss of the plasmid from *Shigella*.

I showed in the Chapter 3 and 4 that commonly used laboratory strains of *Shigella*, such as *S. sonnei* 53G and *S. flexneri* M90T, spontaneously lose virulence due to high loss of pINV or deletions of the T3SS PAI. Therefore, to minimise spontaneous loss of virulence and allow for accurate quantification of the efficacy of the engineered CRISPR-Cas9 systems in inducing loss of virulence, each system was tested in *S. sonnei* CS14. This strain is a clinical isolate from Brazil that harbours a highly stable pINV that is rarely lost by the host but still loses *virB via* T3SS PAI deletions although less frequently than *S. flexneri* M90T (**Appendix Figure 7.9,** Hollingshead *et al*., unpublished).

5.1.1 Targeting *virB* by CRISPR-Cas9 induces deletion of the T3SS PAI

To disrupt *virB*, two versions of pCas9 were designed, containing different spacers and antibiotic resistance cassettes: pCas9^{amp}, which possesses an *amp*^{R} cassette and contains a spacer targeting the -35 sequence of the *virB* promoter (pCas9^{amp}::*virB*^{Pro}); pCas9^{kan}, which possesses a *kan*^{R} cassette for selection, contains a spacer targeting the *virB* ORF encoding residue Gln^{170} (pCas9^{kan}::*virB*^{orf}) (**Appendix Figure 7.10**). The choice of spacers was influenced by the location of PAM sequences. pCas9^{amp}::control and pCas9^{kan}::control both containing a spacer with random sequence with no homology to any sequence in *S. sonnei* CS14, were also generated as controls for non-specific effects of pCas9 derivatives. All plasmids were introduced into *S. sonnei* CS14 by electroporation either alone (generating strains CS14 *virB*^{Pro}, CS14 *virB*^{orf}, CS14 control^{amp}, and CS14 control^{kan}), or in combination (producing CS14 *virB*^{dual} and CS14 control^{dual}). This resulted in bacteria harbouring a single plasmid containing one type of spacer, or two plasmids with different spacers. The rationale was to assess and compare the efficiency of CRISPR-Cas9 systems targeting different sites of the same locus, as I speculated that disruption of distinct parts of the locus might result in different events.

Furthermore, combining CRISPR-Cas9 systems targeting two sites in the same gene could induce deletions of the intervening region.

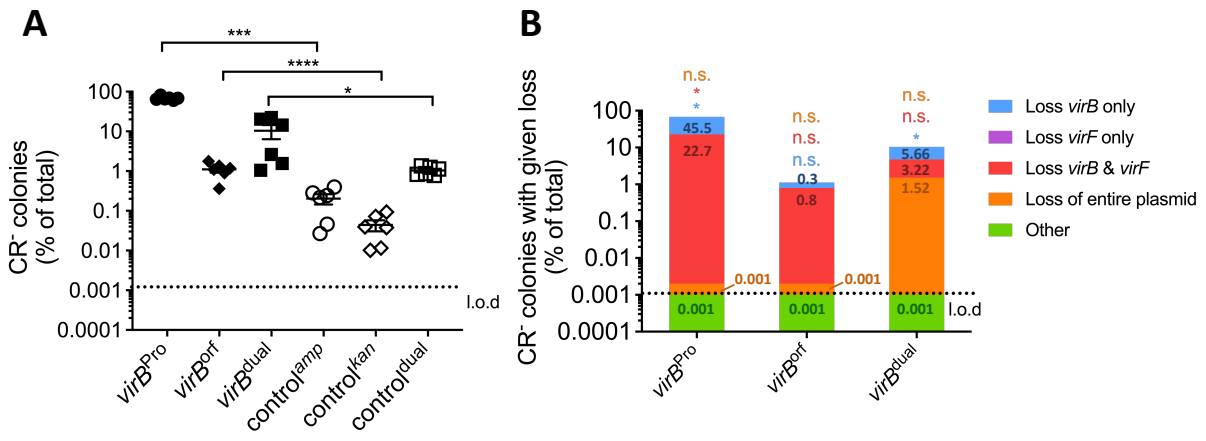


Figure 5.2 CRISPR-Cas9 system targeting *virB* leads to deletion of the T3SS PAI.

(**A**) Percentage of CR$^-$ colonies in the population of *S. sonnei* CS14 containing CRISPR-Cas9 systems targeting *virB* promoter (*virB*^{Pro}), *virB* ORF (*virB*^{orf}) and both (*virB*^{dual}) and systems with non-targeting control spacers (control^{amp}, control^{kan} and control^{dual}) after growth at 37°C for approximately 25 generations. Solid line: mean ± S.E.M. (n = 6 biological repeats). ****, p ≤ 0.0001; ***, *p* ≤ 0.001; *, *p* ≤ 0.05; one sample *t*-test (parametric comparison of mean test/control strain to 1). (**B**) Percentage of CR$^-$ colonies in the population of *S. sonnei* CS14 *virB*^{Pro}, *virB*^{orf}, or *virB*^{dual}, analysed by multiplex PCR with given genetic events leading to the emergence of CR$^-$ phenotype. Loss of the entire plasmid is inferred by loss of the origin of replication; "others" refers to CR- colonies that contain *virB*, *virF* and the origin of replication. Three CR$^-$ colonies were tested from each biological repeat (total 18 colonies for each strain); data are shown as mean. *, *p* ≤ 0.05; ns., not significant; Wilcoxon signed-rank test (nonparametric comparison of median of each event to the limit of detection, l.o.d.).

The efficiency of CRISPR-Cas9 targeting *virB* in inducing loss of virulence was analysed using CR binding assays, followed by multiplex PCR to characterise the genetic events in avirulent bacteria. On two separate occasions, three CR$^+$ colonies from each strain were grown at 37°C for approximately 25 generations in TSB containing antibiotics to select for the presence of pCas9 derivatives, then plated on CR-TSA with the same antibiotics (**Figure 5.2**). After 25 generations, approximately ~ 68% of colonies were CR$^-$ when the promoter of *virB* was

targeted (in strain CS14 *virB*^{Pro}), which was significantly higher than the corresponding control strain (**Figure 5.2A**, CS14 *virB*^{Pro} *vs*. CS14 control^{amp}, $p < 0.0001$; one sample *t*-test). There was a less marked but still significant increase in the percentage of CR^{-} colonies when the *virB* promoter and ORF were targeted together, reaching approximately 10% of the total population (**Figure 5.2A**, CS14 *virB*^{dual} *vs*. CS14 control^{dual}, $p = 0.0348$; one sample *t*-test). However, in CS14 *virB*^{orf} there was a more reduced, yet still significant, increase in the appearance of CR^{-} colonies, compared to the control strain (**Figure 5.2A**, CS14 *virB*^{orf} *vs*. CS14 control^{kan}, $p = 0.0004$; one sample *t*-test).

Multiplex PCR analysis revealed that in CS14 *virB*^{Pro} almost half of all colonies had lost *virB* and ~ 23% lost both *virB* and *virF* (**Figure 5.2B**, CS14 *virB*^{Pro}, loss of *virB vs*. l.o.d., $p = 0.0312$, Wilcoxon signed-rank test). In contrast, in CS14 *virB*^{orf} a reduction in the frequency of these events was observed (loss of *virB* or loss of *virB* and *virF*, < 1% of all colonies); in CS14 *virB*^{dual}, intermediate levels of loss of *virB* alone (3.22% of all colonies) or with *virF* (5.66% of all colonies) were detected (**Figure 5.2B**). Finally, plasmid loss was under l.o.d. in CS14 *virB*^{Pro} and CS14 *virB*^{orf}, but was detected in 1.5% of colonies in CS14 *virB*^{dual}; however, plasmid loss was only detected in one biological repeat (among six), indicating that it is a rare event.

To further examine whether the absence of *virB* by PCR resulted from deletion of this gene alone or larger deletions, 10 CR^{-} colonies lacking *virB* from each strain were analysed by PCR for the presence of *mxiG*, *mxiH*, *mxiI* and *mxiJ*, which are located in the PAI. Interestingly, these genes were absent from all colonies tested, consistent with loss of the T3SS PAI. Additionally, the sequence of *virB* was determined in 10 CR^{+} colonies emerging from strains harbouring *virB*-specific CRISPR-Cas9 to identify point mutations. Of note, no *virB* mutations

were identified, suggesting that DSBs either did not occur or were repaired by restoring the original sequence.

Taken together, these data show that targeting of *virB* by CRISPR-Cas9 induces an increase in the appearance of CR^- colonies in *S. sonnei*, with the majority of these colonies losing *virB* alone, with fewer losing *virB* with *virF*; both events likely result from deletion of the T3SS PAI. Furthermore, the CRISPR-Cas9 targeting the promoter of *virB* led to a higher rate of T3SS PAI deletion than when targeting the *virB* ORF; however, the system appears to be less efficient when both the promoter and the ORF of *virB* were targeted together.

5.1.2 Targeting of *vapBC* by CRISPR-Cas9 leads to distinct events

Next, two pCas9 derivatives were designed to target *vapBC* by CRISPR-Cas; pCas9^{amp} harbouring a spacer targeting the -35 sequence of the *vapBC* promoter (pCas9^{amp}::*vapBC*^{Pro}), and pCas9^{kan} containing the spacer targeting the *vapB* ORF (pCas9^{kan}::*vapB*^{orf}) encoding Thr^{46} (**Appendix Figure 7.10**), which is located in the region of VapB between its DNA binding domain and the C-terminal region that interacts with the active site of VapC (Dienemann *et al*., 2011). By targeting *vapB*, we aimed to eliminate the antitoxin of this important TA system and, therefore, induce killing by VapC. Again, spacers were chosen on the basis of the location of nearby PAM sequences. As before, plasmids were electroporated into *S. sonnei* CS14 individually (generating CS14 *vapBC*^{Pro}, CS14 *vapB*^{orf}) or together (producing CS14 *vapB*^{dual}). Bacteria containing pCas9^{amp}::control (CS14 control^{amp}) or pCas9^{kan}::control (CS14 control^{kan}), or both (CS14 control^{dual}) were included as controls.

As there were differences in the viability of strains, we first measured the number of surviving bacteria after 25 generations of growth at 37°C in TSB containing appropriate antibiotics

(**Figure 5.3A**). Reduced numbers of bacteria were recovered in the presence of CRISPR-Cas9 targeting *vapB*, with fewer than half the number of CFUs recovered of this strain compared with control bacteria harbouring pCas9^{kan}::control (**Figure 5.3A**, CS14 *vapB*^{orf} *vs.* CS14 control^{kan,} $p < 0.0001$; one sample *t*-test). In contrast, when the promoter of *vapBC* was targeted alone or together with the *vapB* ORF, bacteria grew as well as control strains (**Figure 5.3A**, CS14 *vapBC*^{Pro} *vs.* CS14 control^{amp}, $p = 0.1918$; CS14 *vapB*^{dual} *vs*. CS14 control^{dual}, $p = 0.5483$; one sample *t*-test).

Strains were also assessed using a CR binding assay, followed by multiplex PCR. Interestingly, in CS14 *vapBC*^{Pro} almost all colonies were CR^{-}, in contrast to the 0.2% of colonies of the control strain (**Figure 5.3B**, CS14 *vapBC*^{Pro} *vs.* CS14 control^{amp}, $p = 0.0001$; one sample *t*-test). However, in CS14 *vapB*^{dual} ~26% of all colonies were CR^{-}, which was significantly still higher than the control strain (**Figure 5.3B**, CS14 *vapB*^{dual} *vs*. CS14 control^{dual}, $p = 0.0162$; one sample *t*-test). Furthermore, in both CS14 *vapBC*^{Pro} and CS14 *vapB*^{dual}, all the emerging CR^{-} colonies lacked the plasmid (**Figure 5.3C**, CS14 *vapBC*^{Pro} and CS14 *vapB*^{dual} $p = 0.0312$; Wilcoxon signed-rank test). As before, the sequence of *vapBC* in CR^{+} colonies emerging from the three strains containing *vapB*-targeting CRISPR-Cas9 were examined to identify potential mutations following DSB introduced by Cas9. Similar to *virB*, sequencing of the *vapBC* locus revealed no colony tested had a point mutation in *vapB*, suggesting that either DSB were perfectly repaired or the CRISPR-Cas9 system failed to cleave at this site in these bacteria. No quantification and analysis of CR^{-} colonies was possible in the presence of CRISPR-Cas9 with *vapB*^{orf}, as viability was too low in this strain (**Figure 5.3A**).

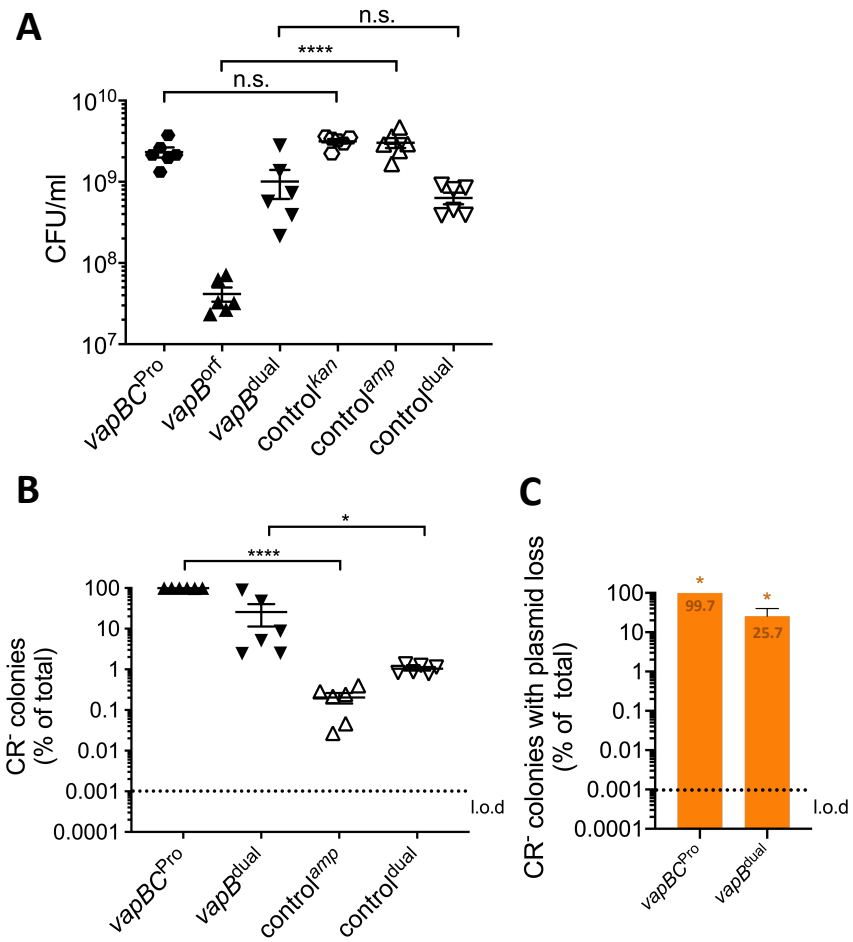


Figure 5.3 CRISPR-Cas9 system targeting the promoter of *vapBC* leads to plasmid loss while targeting *vapB* reduces bacterial viability.

(**A**) Number of viable *S. sonnei* CS14 containing CRISPR-Cas9 systems targeting the *vapBC* promoter (*vapBC*^{Pro}), the *vapB* ORF (*vapB*^{orf}), or both (*vapB*^{dual}), and the respective non-targeting spacer control strains (control^{amp}, control^{kan} and control^{dual}) following growth for approximately 25 generations at 37°C. Solid line: mean ± S.E.M. (n = 6 biological repeats). ****, $p < 0.0001$; n.s., not significant; one sample *t*-test (parametric comparison of mean test/control strain to 1). (**B**) Percentage of CR$^-$ colonies in the population of *S. sonnei* CS14 *vapBC*^{Pro} and *vapB*^{dual}, control^{amp} and control^{dual} after growth at 37°C for approximately 25 generations. Solid line: mean ± S.E.M. (n = 6 biological repeats). ****, $p \leq 0.0001$; *, $p \leq 0.05$; one sample *t*-test (parametric comparison of mean test/control strain to 1). (**C**) Percentage of CR$^-$ colonies in the population of *S. sonnei* CS14 *vapBC*^{Pro} and *vapB*^{dual} analysed by multiplex PCR with given genetic events leading to the CR$^-$ phenotype. Loss of the entire plasmid is inferred by loss of the origin of replication; "others" refers to CR$^-$ colonies that contain *virB*, *virF* and the origin of replication. Three CR$^-$ colonies were tested from each biological repeat (total 18 colonies for each strain); data are shown as mean + S.E.M.. *, $p \leq 0.05$; n.s., not significant; Wilcoxon signed-rank test (nonparametric comparison of median of each event tolimit of detection, l.o.d.).

Taken together, these data indicate that targeting of *vapBC* results in two distinct events, depending on the site targeted within the locus. Targeting of the *vapBC* promoter induces a marked loss of virulence resulting from loss of pINV, while targeting the *vapB* ORF dramatically impairs bacterial viability.

5.2 Inducible CRISPR-Cas9 targeting *vapBC*

These experiments provide proof-of-concept that CRISPR-Cas9 can be employed to induce loss of virulence in *Shigella* by targeting *virB* and *vapBC*. However, the constitutive expression of CRISPR-Cas9 in pCas9 makes it difficult to quantify the efficiency of this approach in editing the gene of interest, as the system will be active before the start of experiments. This is a particular problem when the activity of the system reduces bacterial viability, for instance with CRISPR-Cas9 targeting the *vapB* ORF. Furthermore, there was evident high variation in results from biological repeats of the same experiment, which might reflect differences in early events in bacteria before the start of the experiments. Therefore, an inducible pCas9 system was designed and constructed, in which the expression of *cas9* and the *crRNA* is arabinose-inducible, using the arabinose-inducible promoter P_{BAD} and its regulatory gene *araC* inserted upstream *cas9* (**Figure 5.4A**). Additionally, because targeting the promoter of *vapBC* induced a higher loss of virulence compared with *virB* (**Figure 5.2A** and **Figure 5.3B**), and targeting different sites of *vapBC* had different effects, inducible pCas9 was only engineered to edit *vapBC*. Furthermore, in addition to $vapBC^{Pro}$, two new spacers were designed, $vapB^{orf2}$ and $vapC^{orf}$. $vapB^{orf2}$ guides the CRISPR-Cas9 to cleave *vapB* at the codon for Gln^{66}, which is a residue essential for inhibiting VapC (**Figure 5.4B**) (Dienemann *et al*., 2011). In this way, disruption of the activity of VapB with $vapB^{orf2}$ might be more effective than targeting with $vapB^{orf}$, which is directed at a region of VapB between two functional domains. *vapC* is

designed to induce a DSB in *vapC* at the codon for Asp[98], an active site residue (**Figure 5.4B**) (Dienemann *et al*., 2011).

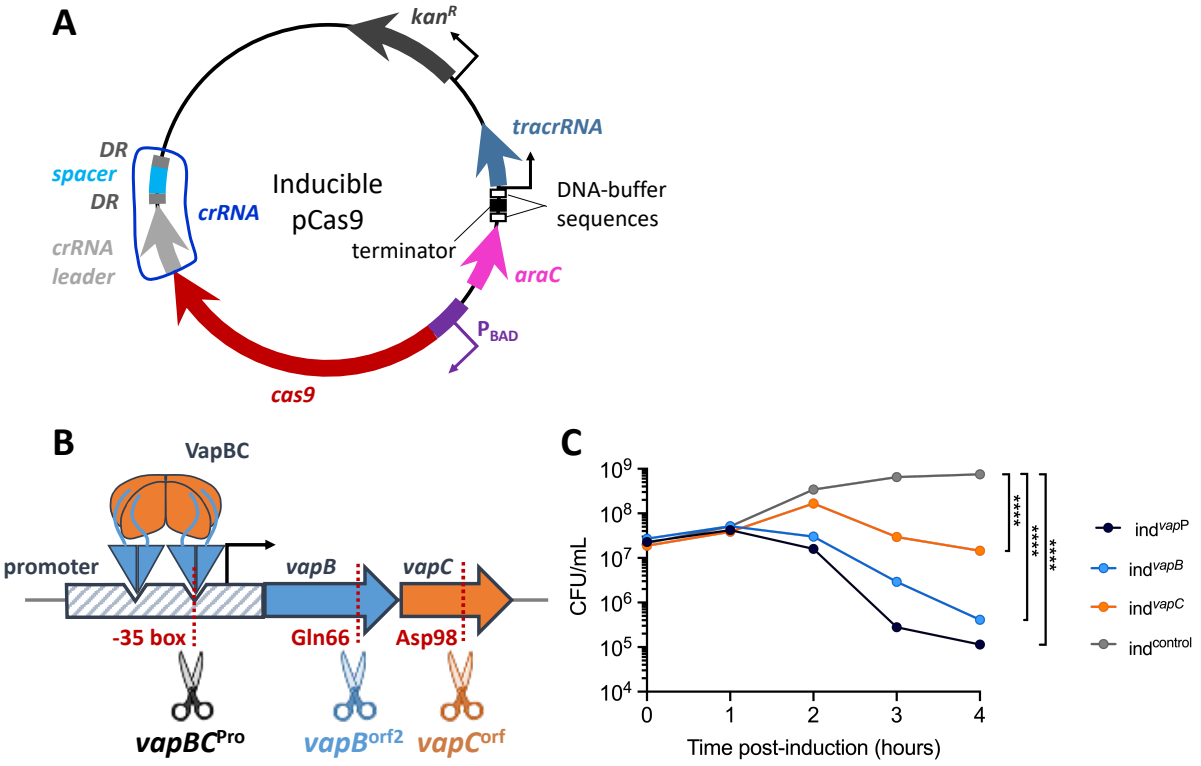


Figure 5.4 Inducible CRISPR-Cas9 system targeting different components of *vapBC* reduces bacterial growth after 2 hours of induction.

(**A**) Schematic of inducible pCas9. The arabinose inducible promoter pBAD with *araC* inserted upstream *cas9*. A terminator sequence enclosed by two DNA linker sequences was inserted downstream *araC* to prevent any disruption of *tracrRNA* transcription. (**B**) Schematic of the site of cleavage of the inducible CRISPR-Cas9 systems targeting the *vapBC* promoter (*vapBC*^{Pro}) at the -35 box, the VapB antitoxin at Gln66 (*vapB*^{orf2}), and the toxin at Asp98 (*vapC*^{orf}). (**C**) Viability of *S. sonnei* CS14 containing the inducible CRISPR-Cas9 systems targeting *vapBC* promoter (ind^{vapP}), *vapB* (ind^{vapB}), *vapC* (ind^{vapC}) and the non-targeting control (ind^{control}) at 0, 1, 2, 3 and 4 hours post-arabinose induction. Data are shown as mean + S.E.M. (n = 9 biological repeats). ****, $p \leq 0.0001$; one sample *t*-test (parametric comparison of mean test/control strain to 1).

Each inducible pCas9 harbouring either *vapBC*^{Pro}, *vapB*^{orf2}, *vapC*^{orf} or a non-targeting spacer was electroporated into *S. sonnei* CS14, generating CS14 ind^{vapP} , CS14 ind^{vapB} , CS14 ind^{vapC}

and CS14 ind^{control} (**Figure 5.4C**), and on two separate occasions the viability of three independent CR^{+} colonies emerging of these strains was assessed. Bacteria were first grown at 37°C for ~ 25 generations in LB containing glucose to repress expression of *cas9*, then subcultured at the same temperature for 4 hours in LB containing arabinose; the number of viable bacteria was quantified at 0, 1, 2, 3 and 4 hours following the addition of arabinose (**Figure 5.4C**). Interestingly, after 4 hours of arabinose induction, the presence of the *vapBC*-specific CRISPR-Cas9 systems caused a significant reduction in viability compared with the control system (**Figure 5.4C**, CS14 ind^{vapP}, CS14 ind^{vapB}, CS14 ind^{vapC} *vs*. CS14 ind^{control}, $p < 0.0001$, one sample *t*-test), with CS14 ind^{vapP} and ind^{vapC} showing the highest and the lowest decrease in survival, respectively. In contrast, no reduction of viability was observed when of *cas9* and the *crRNA* was repressed in the presence of glucose (**Appendix Figure 7.11**). Since no CR^{-} colonies were detected among the surviving bacteria of any strain, multiplex PCR was not performed.

These results demonstrate that introduction of a DSB in *vapBC via* an inducible CRISPR-Cas9 system reduces bacterial survival, with different efficiencies depending on the site of *vapBC* targeted. Furthermore, we did not detect plasmid loss when the promoter of *vapBC, vapB* or *vapC* were targeted. However, decrease in viability could be a result of activation of PSK caused by plasmid loss. This is in contrast with previous results whereby constitutive expression of CRISPR-Cas9 led to a decrease in bacterial viability only when *vapB*, and not its promoter, was targeted.

5.3 Summary

In this chapter, I validated the deployment of CRISPR-Cas gene editing to induce loss of virulence in *Shigella* by targeting the plasmid loci, *virB* and *vapBC*. Using a constitutively expressed CRISPR-Cas9 system, disruption of *virB* by targeting the promoter and/or the *virB* ORF induced the loss of virulence, which mainly resulted from deletion of the T3SS PAI (**Figure 5.2**). Two types of deletion were detected by PCR; the most abundant event involved the T3SS PAI only, while there was evidence of loss of the T3SS PAI and *virF* albeit less frequently. We speculate that the targeting of *virB* locus by the CRISPR-Cas9 selects for recombination events between homologous copies of IS, as this allows its repair of DSB, preventing loss of pINV and hence PSK. Amplification and sequencing of the predicted deletion sites should be performed to confirm the involvement of IS in the genetic events resulting in T3SS PAI deletions. Furthermore, the reason why deletion of the T3SS PAI only is more frequent than deletion of the T3SS PAI with *virF* probably depends on the frequency of recombination between the IS involved in each variant, a phenomenon that was also detected in the absence of CRISPR-Cas9 (**Appendix Figure 7.9**). In addition, the promoter sequence of *virB,* and in particular the -35 box, proved to be a more efficient target for CRISPR-Cas9, as a higher proportion of avirulent bacteria was obtained when this site was targeted. This may depend on the specificity of the spacer used, which might have reduced off-target activity, or the efficiency of targeting with different crRNA. Additionally, certain nucleotide changes could alter expression levels and be deleterious for bacteria through inappropriate T3SS expression.

Another approach used to induce loss of virulence in *Shigella* was to design a constitutively expressed CRISPR-Cas9 targeting *vapBC*, which is essential for pINV maintenance (McVicker and Tang, 2016; Schuch and Maurelli, 1997). As for *virB*, two CRISPR-Cas9 systems were generated, targeting the promoter of *vapBC* or its ORF (**Appendix Figure 7.10**). Interestingly,

depending on the site of the locus targeted, different events were induced; targeting of the *vapBC* promoter resulted in loss of pINV, while targeting of *vapB* led to a decrease in bacterial viability (**Figure 5.3**). Unfortunately, it is difficult to prove that the designed CRISPR-Cas9 systems actually introduced DSB in their target sequences as either the entire plasmid was lost or no point mutation was found in the target sequence. If the CRISPR-Cas9 cleaved its target, it is possible that cleavage of the promoter altered expression of both the antitoxin and the toxin; therefore, loss of pINV might occur due to low levels of the VapC toxin and impaired PSK. However, disruption of *vapB* could selectively deplete the bacterium of VapB, leaving VapC free to kill cells with consequent reduced viability.

Finally, for both approaches, the introduction of CRISPR-Cas9 targeting the promoter while at the same time targeting the ORF of the same locus was not as efficient at inducing avirulence compared with when the sites were targeted independently (*e.g.* CS14 *virB*^{Pro} *vs*. CS14 *virB*^{dual} in **Figure 5.2** and CS14 *vapBC*^{Pro} *vs*. CS14 *vapB*^{dual} in **Figure 5.3**). This might be due to an imbalance in number or ratio of pCas9 derivatives harbouring different spacers, leading to one system dominating over the other, and therefore, preventing both systems from acting simultaneously. Furthermore, the copy number of each pCas9 will be lower when both plasmids are together in bacteria than when there is only one pCas derivative, so the efficiency of each system might be reduced when two systems are present. To solve this problem, two different spacers could be inserted into pCas9, avoiding any issues caused by plasmid copy number. Furthermore, the *crRNAs* and the *cas9* could be expressed under the control of different promoters, with the protein expressed at higher levels than the RNA, to avoid any competition of the different crRNA for the protein.

The CRISPR-Cas9 system is constitutively expressed in pCas9, so may be active as soon as the plasmid is introduced into bacteria. Therefore, bacterial populations at the start of experiments are likely to contain sub-populations that have been subjected to CRISPR-Cas9 activity. Therefore, to control the activity of CRISPR-Cas9 and, thus, ensure that strains have not been edited before the start of the experiments, a new pCas9 plasmid was designed, containing *cas9* and the *crRNA* under the arabinose-inducible promoter, P_{BAD} (**Figure 5.4A**). In this way, expression of CRISPR-Cas9 targeting pINV was more controlled and results more consistent. Furthermore, as *vapBC* was a more promising target than *virB*, the inducible CRISPR-Cas9 system was engineered to target different sites only of *vapBC* including key residues in the antitoxin and toxin, in addition to the -35 sequence of the promoter (**Figure 5.4B**). Surprisingly, upon activation of the CRISPR-Cas9, a marked decrease in bacterial survival was detected irrespective of the site of *vapBC* targeted; furthermore, even when the promoter and the toxin were targeted, no plasmid loss was detected (**Figure 5.4C**).

The difference observed in the effect of the constitutively expressed and the arabinose-induced *vapBC*-targeting CRISPR-Cas9 systems is likely due to the different expression of the two systems. In *E. coli*, DSB are generally repaired through HR as the bacterium lacks NHEJ (Wilson *et al*., 2003). However, an alternative end-joining mechanism (A-EJ) has been characterised in *E. coli*, which allows repair of DSBs through microhomology (Romain *et al*., 2010) and this mechanism is sufficient to repair CRISPR-Cas9-induced DSB (Huang *et al*., 2019). *Shigella* is likely to not possess NHEJ and instead employs A-EJ to repair DSB, given its relatedness to *E. coli* (Lan and Reeves, 2002). As the efficiency of A-EJ decreases with increasing of numbers of DSB, high level of expression of the CRISPR-Cas9 could be cytotoxic because of the accumulation of un-repaired DSB. Indeed, un-repaired DSB could lead to loss of pINV, and induce PSK. To confirm this hypothesis, a gene encoding GFP, which is not

involved in plasmid maintenance and cannot be repaired by HR, could be introduced in pINV and targeted by the inducible CRISPR-Cas9 system; the viability of this strain could be compared with a strain that harbours the same system but has *vapC* inactivated. In this way, if the un-repaired DSB induces loss of pINV, this would be detected in the strain with inactive *vapC* which would be also GFP^-, while reduction of viability would be observed in the strain with active *vapC*.

Furthermore, it would be interesting to perform microfluidic live cell microscopy to visualise the loss of pINV *in vivo* after inducing the CRISPR-Cas9 system. For instance, plasmid loss could be detected in cells by introducing a constitutively expressed transcriptional repressor on pINV, (*e.g*. TetR or AraC), with a fluorophore under its control on the chromosome (fP). A second different fluorophore under constitutive promoter would be expressed from the chromosome (fC). If the plasmid is present, only fC would be expressed as fP expression would be repressed by the plasmid-encoded regulator. When the plasmid is lost, both fP and fC would be expressed. Similarly, the presence of the T3SS PAI could be monitored by introducing the transcriptional repressor next to *virB*. Alternatively, the presence of the plasmid could be followed by fluorescently-tagged ParB, which binds to pINV *parS* sequences and would be detected as fluorescent foci (Deghelt *et al*., 2014); plasmid loss would then be shown by the disappearance of these foci. Furthermore, PSK could be assessed by measuring the dilution of the fC, which would occur if translation is inhibited while there is continuing expansion of the cytosol; in fact, inhibition of translation by VapC is likely to occur before bacterial growth stops.

Finally, as observed previously, different levels of viability were observed depending on the spacer used. It is possible this is merely due to the specificity of the spacer and the resulting off-target effects. Several strategies have been proposed to ameliorate target specificity by

altering the crRNA sequence. For instance, addition of two guanine nucleotides to the 5' end of the *crRNA* significantly decreases off-target editing while too high or too low GC content of the 5-6 nt sequence downstream the PAM sequence reduce target specificity (Zhang *et al*., 2015). Therefore, *crRNA* designing could be optimised and different sites within the same locus could be tested.

6 Discussion

Plasmids play a pivotal role in the virulence of enteropathogens such as *Shigella* spp. (Pilla and Tang, 2018). Although these genetic elements vary highly in their structure and contribution to virulence in different species and even strains, they share many features. For instance, virulence plasmids possess mechanisms that promote their maintenance by regulating replication, partitioning, and post-segregational killing. Virulence plasmids are generally large (> 40 kb) and present in low copy number in bacteria. Therefore, their maintenance is a fundamental aspect of their biology. Furthermore, another important feature of virulence plasmids is genetic plasticity, which facilitates acquisition or loss of genes, and structural rearrangements through translocations and inversions.

The *Shigella* plasmid pINV is a model virulence plasmid for several reasons. First of all, besides encoding the T3SS, which is essential for the invasion and survival of the bacterium inside the host, it possesses a highly regulated low-copy number replicon and multiple maintenance systems that ensure its retention following replication. *Shigella* pINV is also exemplar of genetic plasticity. The abundance of IS in pINV has enabled acquisition of several genes, including the T3SS PAI, and shaped the structure of the plasmid, which varies between the species (Yang *et al*., 2005).

The study of plasmid biology is of increasing relevance. For instance, understanding maintenance of virulence and/or resistance plasmids by an important pathogen such as *Shigella* provides strategies that can be used to prevent and/or treat shigellosis. In fact, insights into pINV maintenance could be employed to eliminate the plasmid from the bacterium, thus, inducing loss of virulence. As related maintenance systems are found on

many plasmids (Pilla and Tang, 2018), these strategies could be also applied to virulence plasmids of other enteropathogens, as well as antibiotic resistance plasmids, which pose a major public health threat (Pilla and Tang, 2018; Chandler, 2019). Furthermore, characterisation of the molecular rearrangements that affect plasmids should shed light on the processes that lead to the evolution of virulence.

6.1 TA systems as plasmid guardians and potential clues of plasmid evolution

One intriguing aspect of *Shigella* pINV is the presence of multiple TA systems, as one TA system should be sufficient to prevent the loss of a plasmid. Nonetheless, this is a common feature among virulence plasmids in enteropathogenic bacteria. For example, the virulence plasmid of *Salmonella enterica* serovar Typhimurium pSLT harbours *ccdAB* and *vapBC*, while pB171, the virulence plasmid of enteropathogenic *E. coli* (EPEC), possesses *ccdAB* and *relBE* (Tobe *et al*., 1995; Lobato-Marquez *et al*., 2015; Pilla and Tang, 2018). The reason why there are multiple TA systems on a single element is still unknown, however, characterisation of the activity of TA systems on pINV might provide some insights.

In this work I showed that the pINV-encoded TA system VapBC acts to maintain the entire plasmid as well as discrete regions. When the TA locus was moved next to the T3SS PAI, it prevented the emergence of avirulent colonies with T3SS PAI deletions (**Figure 3.2**). Local stabilisation has been observed before with chromosomally-encoded TA systems. In *E. coli*, *Vibrio cholerae* and *S. enterica* serovar Typhimurium, TA systems stabilise adjacent genomic regions and can maintain superintegrons (SI), integrative and conjugative elements (ICE), and multidrug resistance islands (Szekeres *et al*., 2007; Wozniak and Waldor, 2009; Huguet *et al*., 2016). This function could also apply to other TA systems on pINV, such as GmvAT. Loss of *virF*

with or without *virB* is more common in *S. sonnei* pINV (pINV^{Ssonn}) than pINV of *S. flexneri* (pINV^{Sflex}) (**Figure 4.2C**). GmvAT is located only 7.4 kb from *virF* on pINV^{Sflex} within a region enclosed by two copies of IS*1294* (Buchrieser *et al*., 2000). Interestingly, introduction of *gmvAT* (together with *ccdAB*) into pINV^{Ssonn} led to a reduction in loss of *virF* alone or together with *virB* although this result was not statistically significant (**Figure 4.2C**). Therefore, the absence of *gmvAT* from pINV^{Ssonn} might explain why deletion of *virF* is more common in *S. sonnei*. To confirm this hypothesis, *gmvAT* should be introduced alone into pINV^{Ssonn} and loss of *virF* should be assessed in this strain at 21°C, the temperature at which GmvAT functions (McVicker and Tang, 2016).

Previous studies suggested that different TA systems might provide efficient plasmid maintenance in distinct environments. For instance, TA systems on pINV are known to act at different temperatures (McVicker and Tang, 2016). VapBC is essential and sufficient for plasmid maintenance at 37°C, temperature in the human body, while GmvAT, as mentioned above, confers plasmid stability at 21°C, a temperature *Shigella* encounters in the environment. Interestingly, GmvAT is not present in *S. sonnei*, which shows higher plasmid loss at 21°C than at 37°C (McVicker and Tang, 2016). Loss of GmvAT in *S. sonnei* has been proposed to reflect an ongoing transition of the bacterium into a more host-associated pathogen that is mainly transmitted by person-to-person contact (McVicker and Tang, 2016). This suggests that comparison of TA loci between species can provide important hints about the evolution of the bacterium. Nonetheless, it is not clear what role the third pINV-encoded TA system CcdAB plays and in which conditions it is active (McVicker and Tang, 2016).

Furthermore, growth conditions may affect availability of the toxin substrates or cofactors, which, in turn, could influence the activity of toxins. For instance, downregulation of the

expression of $tRNA^{fMet}$ might increase the toxicity of VapC. Alternatively, low abundance of AcCoA could make GmvT less able to inhibit translation.

Given the local effect of TA systems, it is interesting that both CcdAB and GmvAT have been deleted from $pINV^{Ssonn}$ (McVicker and Tang, 2016). Although introduction of CcdAB and GmvAT into $pINV^{Ssonn}$ increased plasmid maintenance, plasmid loss was still responsible for a high proportion of avirulent colonies (**Figure 4.1B**). In this work I demonstrate that besides lacking CcdAB and GmvAT, $pINV^{Ssonn}$ possesses two amino acid substitutions in VapBC, one in VapB and one in VapC. The latter renders the plasmid less stable in *S. sonnei*. Substitution of VapC for the *S. flexneri* version significantly reduced, but did not eliminate, plasmid loss (**Figure 4.1B**). These results suggest that the presence of CcdAB and GmvAT, and the VapC substitution are responsible for high frequency plasmid loss in *S. sonnei*. Interestingly, plasmid loss was reduced to below the limit of detection when the $pINV^{Ssonn}$ carrying $VapBC^{Sflex}$ was present in a *S. flexneri* strain (**Figure 4.3C**), suggesting that the genetic background influences the ability of VapBC to prevent plasmid loss. This might be due to differences in the target of VapC ($tRNA^{fMet}$), either in its sequence or expression, and is currently under investigation by the group.

These insights could be exploited to construct strains of *S. sonnei* with more stable plasmids that could be used to examine host-pathogen interactions as well as to develop and test candidate vaccines in human/animal models.

It is unclear why $pINV^{Ssonn}$ has undergone changes that have reduced its maintenance. Loss of pINV might be beneficial in certain circumstances. In fact, the presence of a fast replicating subpopulation of bacteria lacking pINV could “distract” the host immune system and/or cause hyperinflammation by increasing the bacterial load and enhancing damage of the intestinal epithelium, allowing more virulent bacteria to penetrate the tissue. Additionally, the plasmid

could be re-acquired with the assistance of a helper plasmid. *S. sonnei* has acquired more AMR plasmids than *S. flexneri* (Holt *et al*., 2012). As many AMR plasmids encode conjugation systems, it is plausible that AMR plasmids confer an additional advantage to *S. sonnei* by facilitating the spread and re-acquisition of pINV.

6.2 The double-edged sword of IS on pINV

Spontaneous loss of virulence in *S. flexneri* and *S. sonnei* has been documented since the first molecular studies on these pathogens (Kopecko *et al*., 1980; Maurelli *et al*., 1984; Sasakawa *et al*., 1986; Chosa *et al*., 1989; Schuch and Maurelli, 1997). However, the genetic events responsible for this have not been characterised previously. In this work, I defined the molecular mechanisms that lead to loss of virulence in both *S. flexneri* and *S. sonnei*. Although plasmid loss is prevalent in *S. sonnei*, both species undergo deletion of T3SS-related genes. In particular, I demonstrate that *S. flexneri* mainly loses the entire T3SS PAI through four distinct IS-mediated deletions (**Figure 3.3**). Similarly, deletion of the entire T3SS PAI is observed in *S. sonnei* but is mediated by different IS and often includes *virF* (**Figure 4.4**).

IS are known to have a crucial contribution to the biology of plasmids and, indeed, in the evolution of pathogens. By mediating different types of molecular rearrangements, IS contribute to HGT and gene inactivation (Vandecraen *et al*., 2017). Undoubtedly, IS have played a major role in the evolution of *Shigella* spp., which have undergone IS expansion and subsequent genome degradation (Hawkey *et al*., 2020). IS have mediated loss of core *E. coli* metabolic genes as well as the acquisition of virulence traits, such as the T3SS PAI and other plasmid-encoded virulence genes (Venkatesan *et al*., 2001; Hawkey *et al*., 2020). Interestingly, although the repertoire of IS varies, IS expansion with gene reduction has been observed in all species of *Shigella* (Hawkey *et al*., 2020). The pINVs of *S. flexneri* and *S. sonnei* display

distinct IS profiles (**Figure 4.5B** and **C**), which dictate the nature of T3SS-related deletions (**Figure 4.5A**). While these events have been observed mainly in the laboratory, it is unclear whether they occur during the natural lifecycle of *Shigella* spp.. It is interesting to note that no TA system is found next to or within the T3SS PAI which could prevent its loss. It is likely that loss of T3SS-related genes is detrimental during infection, as bacteria that are T3SS-negative cannot invade cells and, therefore, could be eliminated by the host immune system. Additionally, IS-mediated deletion of T3SS-related genes may represent a trade-off for genetic plasticity, which enables the acquisition of virulence genes such as the T3SS PAI. Alternatively, similar to plasmid loss, deletion of T3SS-related genes could represent an advantage during infection as T3SS-negative bacteria would replicate faster and act as an immune decoy in the intestinal tract.

Interestingly, deletion of T3SS-related genes is not the only IS-mediated event detected in *S. flexneri* that leads to avirulence. Homologous IS on the chromosome and pINV recombine, leading to integration of the plasmid into the chromosome (**Figure 3.5B**). This results in downregulation of T3SS PAI gene expression and consequently impaired activity of the T3SS (**Figure 3.5C** and **Figure 3.6C**). However, excision of pINV occurs, restoring T3SS secretion (**Figure 3.6**), and demonstrating that pINV integration is a reversible mechanism for *Shigella* to circumvent the metabolic cost imposed by the T3SS. Experiments are underway to test if plasmid integration also occurs in *S. sonnei* and to identify potential IS that mediate this event. Plasmid integration has been observed previously in *S. flexneri*, but this was not IS-mediated and was not spontaneous; the plasmid integrated into *metB*, leading to methionine auxotrophy (Zagaglia *et al*., 1991). Other IS-mediated plasmid integration events have been documented in the late 70s, which involved the F plasmid in *E. coli* (Davidson *et al*., 1975; Ohtsubo and Ohtsubo 1977; Holloway, 1979). Furthermore, Vandenbosch *et al*. observed

chromosomal integration of the virulence plasmid of *S*. Typhimurium, which, also decreased virulence (Vandenbosch *et al*., 1987). However, the events mediating integration in *Salmonella* were not characterised. Co-integration by recombination between plasmids belonging to different incompatibility groups can also occur, and generate chimeric plasmids (Aviv *et al*., 2014; Chen *et al*., 2017), such as pCoo from enterotoxigenic *E. coli* (ETEC) (Froehlich *et al*., 2005).

The molecular process by which T3SS activity is reduced following pINV integration is unknown. As described in Section 1.2.3, the transcriptional repressor H-NS binds to the promoters of *virF* and *virB* in a temperature-dependent manner due to alterations in DNA topology (Falconi *et al*., 1998). Increase in H-NS repression of the *virB* promoter has been previously documented following forced integration of pINV in *S. flexneri* and enteroinvasive *E. coli* (Colonna *et al*., 1995). However, our qRT-PCR results show that spontaneous pINV integration also affects *virF* expression (**Figure 3.5**), suggesting that H-NS could similarly repress *virF* promoter. The *E. coli* and *Salmonella* chromosomes are organised into topologically distinct regions which are defined as structured DNA macrodomains (Postow *et al*., 2004; Valens *et al*., 2004; Espeli *et al*., 2008). We speculate that *Shigella* has a similar heterogenous chromosomal architecture and, as other IS could also mediate integration, the exact site and orientation of pINV integration may influence gene expression.

The mechanisms by which IS recombine are not fully understood. Here I demonstrate that IS-mediated deletion of T3SS PAI is abolished when both pathways of HR, RecBCD and RecFOR are non-functional (**Figure 3.4**); however, deletion of *virF* still occurs. A similar result was also observed by Schuch and Maurelli (1997). In *S.sonnei,* deletion of *virF* is IS*1294*-mediated (**Figure 4.4**) and it is likely that homologous pairs of this IS similarly mediate loss of *virF* in *S.*

flexneri, as multiple copies of IS*1294* surround this gene in pINV^{Sflex} (**Figure 4.5**). This should be confirmed by further sequence analyses. Involvement of IS*1294* in the deletion of *virF* would suggest that not all IS-mediated recombination events depend on HR.

Involvement of HR in pINV integration was not investigated. However, deletion of *recA* did not reduce the emergence of avirulent bacteria with intact plasmid sequence (**Figure 3.4B**), indicating that there might be other HR-independent mechanisms mediating pINV integration. Additionally, Zagaglia *et al*., demonstrated that excision of pINV from the chromosome is RecA-dependent (Zagaglia *et al*., 1991). Finally, it would be interesting to assess whether Tpases actively mediate T3SS-related deletions/integration or if IS contribute solely as substrates for HR. Examining this aspect is technically challenging, as the *Shigella* genome is particularly enriched in IS (especially in those involved in T3SS-related deletions/integration), and introducing point mutations into single Tpases/IR is difficult. However, systematic mutation of all copies of a specific IS type (*e.g*. the nine copies of IS*1294* in the *S. flexneri* genome) using CRISPR-Cas9 gene editing, followed by quantification of virulence loss, could be an approach to assess the role of Tpases in T3SS-related deletions. Additionally, it would be interesting to understand if the mode of transposition influences the contribution of the IS to these events.

In conclusion, my findings indicate that IS are important components of pINV, as they have not only contributed to the evolution of this pathogen and shaped the architecture of the plasmid, but are also active elements that still influence the biology of this plasmid by regulating the maintenance of virulence.

6.3 Partitioning systems: two better than one?

Another interesting aspect of pINV is the presence of two partitioning systems, ParAB and StbAB, which belong to two distinct *par* families with different mechanisms of action (**Figure 1.7**). Both systems are functional when tested on a model vector containing the pINV replicon, however, their contribution to pINV maintenance is not fully understood (McVicker *et al*., 2019). The presence of TA systems masks the potential effect of the absence of each partitioning system, as TA-mediated PSK eliminates the majority of bacteria that have lost the plasmid (**Figure 3.7**). However, in the absence of TA systems, the presence of ParAB helps maintain the plasmid at both 37°C and 21°C (**Figure 3.7**). Interestingly, the role of ParAB in pINV maintenance alters when bacteria grow slowly. Indeed, during growth at 21°C or in minimal media, pINV is maintained more efficiently in bacteria without ParAB maintain pINV more efficiently (**Figure 3.7B** and **Figure 3.8A**), indicating that ParAB is detrimental to partitioning in these conditions. I hypothesise that ParAB activity is influenced by plasmid copy number (PCN), which I found varies between 1 and 1.5 depending on the growth rate (**Figure 3.9**). This concept is explained in the following model (**Figure 6.1**).

Following pINV replication and cell division, segregation should produce two identical bacterial cells, each containing a single copy of the plasmid (**Figure 6.1**, step **a**). However, if this process is defective, it would generate two different bacteria, one with two copies of the plasmid, and the other with none (**Figure 6.1**, step **b**). Hypothetically, in a non-selective environment, these different bacteria will be present in equal amounts. In contrast, under conditions of rapid growth (*e.g.* at 37°C, in rich medium), bacteria with a single copy of pINV will dominate the population compared with those containing two copies of the plasmid, because of the fitness cost imposed by carrying the extra copy of pINV, and because fast bacterial growth gives less time for plasmids to replicate and then segregate, leading to

progressive lowering of the plasmid copy number. Additionally, the activity of TA systems will eliminate bacteria without pINV. These factors will steer the PCN towards an average of 1 (**Figure 3.9**). If ParAB is more efficient at maintaining pINV when it is present in ~1 copy, under these conditions ParAB will contribute to plasmid retention (**Figure 3.7**). In contrast, during slow growth, such as at 21°C or in minimal media, maintenance of two pINV copies might not be so costly for the bacterium, and the cell would have more time to segregate multiple copies of pINV. Therefore, bacteria with one or more copies of pINV would increase in the population. Again, in the presence of active TA systems, the majority of the cells that lack pINV will be eliminated. Therefore, the average PCN will tend to increase towards ~1.5 (**Figure 3.9**). If ParAB is less effective at segregation when the PCN is > 1, then under these circumstances, ParAB might be detrimental for plasmid retention (**Figure 3.7B** and **Figure 3.8A**). Similarly, in the absence of *virB*, the fitness cost of pINV is reduced (as the T3SS PAI is not expressed) so there would be a concomitant increase in the PCN as the proportion of bacterial cells with two pINV copies rises, rendering ParAB detrimental for pINV maintenance (**Figure 3.8B**). This hypothesis could be tested by measuring PCN in a Δ*virB* strain.

Finally, in the absence of TA systems, in fast or slow-growth conditions, bacterial cells lacking a plasmid would not be eliminated by PSK and will outcompete those which possess pINV. Furthermore, this outgrowth is likely to be more evident in conditions of fast growth. In these conditions, the average PCN would be expected to be < 1, and ParAB will be important for plasmid maintenance (**Figure 3.7**). To further confirm this hypothesis PCN should be quantified in the absence of TA systems, in both fast- and slow-growth conditions.

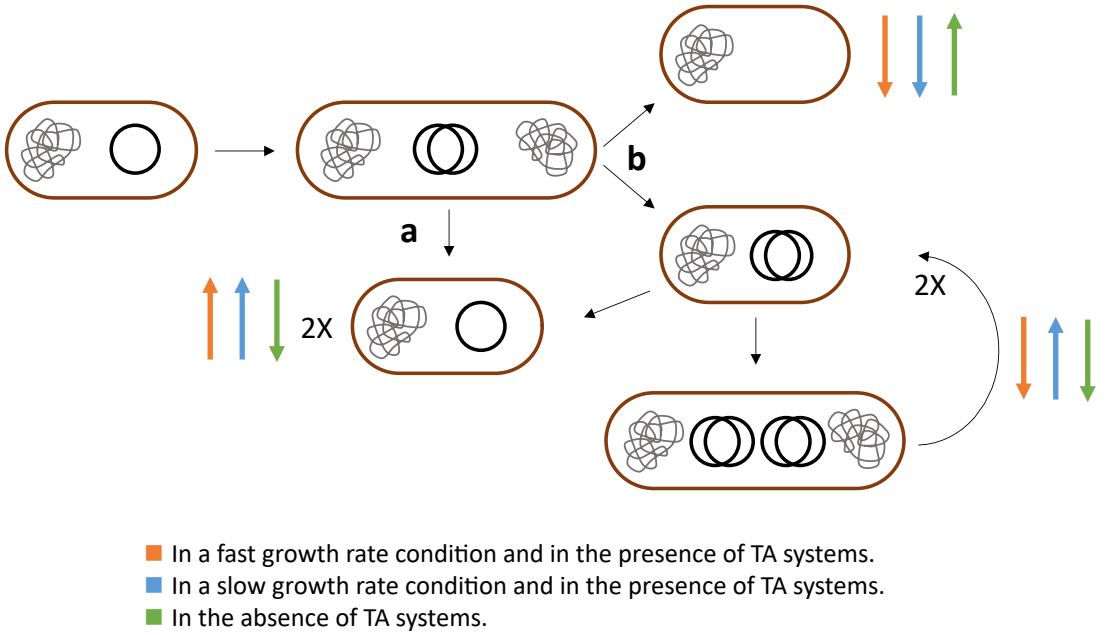


Figure 6.1 Hypothetical model on the variation of the pINV copy number in *Shigella*.

Following genomic replication, bacterial cells containing a different number of pINV copies (none, one or two) can originate due to correct (a) or incorrect (b) partitioning. Colour-coded arrows show the increase or decrease of the specific bacterial cell type in the population, according to the growth condition and presence/absence of TA systems. pINV is indicated in black, the chromosome in grey.

The role of the conditions in which StbAB contributes to pINV maintenance is not known, as its deletion did not affect plasmid stability in any of the conditions tested (**Figure 3.7** and **Figure 3.8**). It is possible that it may contribute *in vivo* or its function might be redundant given the presence of other pINV-encoded elements, including ParAB. The presence of multiple partitioning systems is common in virulence plasmids in enteropathogenic bacteria. For instance, pB171 contains two *par* loci, *par1* and *par2*, which are homologous to *parMRC* (*stbAB*) and *parABS,* respectively. In contrast to the *S. flexneri* partitioning systems, these loci are adjacent to each other on pB171 and share a common *par* site but are oriented in opposite directions (Ebersbach and Gerdes, 2000). Although each partitioning system is sufficient for plasmid segregation, both systems are needed for optimal plasmid maintenance. By regulatory crosstalk, ParB from *par2* represses the transcription of *par1* in addition to

regulating its own expression (Ringaard *et al*., 2007); this crosstalk might indicate the ongoing replacement of the *par2* system with the *par1* system, which is consistent with the higher prevalence of type I loci in virulence plasmids (Ringaard *et al*., 2007). StbAB is conserved in all *Shigella* spp. except for *S. sonnei*. As *S. sonnei* is considered to be the most recent *Shigella* species to have emerged from *E. coli* (Holt *et al*., 2012), it is possible that replacement of *stbAB* with *parAB* has been completed in this species. Alternatively, multiple partitioning systems might contribute to plasmid incompatibility, as explained before (Chapter 1, Section 1.2.1.3.2).

6.4 CRISPR-Cas9 as a novel tool to treat shigellosis and investigate PSK

Identification of elements that confer pINV maintenance in *Shigella* provides potential targets to induce the loss of virulence in this important pathogen. CRISPR-Cas9 has proven to be an efficient method for gene editing and for targeting antimicrobial resistance and virulence genes in bacteria (Bikard *et al*., 2014; Citorik *et al*., 2014; Doudna and Charpentier, 2014; Kim *et al*., 2016; Rodrigues *et al*., 2019).

Here, I tested CRISPR-Cas9 as a tool to render *Shigella* avirulent. For this, two approaches were assessed. First, I employed constitutively expressed CRISPR-Cas9 to target the T3SS transcriptional regulator *virB*, with the aim of inducing loss of the expression of the T3SS PAI. In the second approach, using the same system I targeted *vapBC*, to induce either loss of the plasmid or PSK mediated by VapC, according to the component of the system targeted. Targeting *virB* induced a significant increase in the loss of virulence, which mainly resulted from deletion of the T3SS PAI (**Figure 5.2**). In contrast, when *vapBC* was targeted, different events occurred; targeting of the *vapBC* promoter resulted in loss of pINV, while targeting

vapB led to a decrease in bacterial viability (**Figure 5.3**). Second, to control the activity of the CRISPR-Cas9 system before the start of experiments, an arabinose-inducible CRISPR-Cas9 system was used to target the same locus (**Figure 5.4A**). Of note, a decrease in bacterial survival was detected irrespective of the site of *vapBC* targeted by CRISPR-Cas9; no plasmid loss was detected, even when the promoter or *vapC* were targeted (**Figure 5.4C**). As discussed at the end of Chapter 5, I speculate that the difference observed in the effect of the constitutively expressed and the arabinose-induced *vapBC*-targeting CRISPR-Cas9 systems depends on the different expression of the two systems and thus, the ability of *Shigella* to repair DSB.

These experiments provide-proof-of-principle that CRISPR-Cas9 can be employed to induce loss of virulence in *Shigella* by targeting *virB* and *vapBC*. However, several questions still need to be addressed. Initially, it will be important to establish whether the CRISPR-Cas9 systems actually introduce a DSB as no point mutations were found in the target sequence; instead CRISPR-Cas9 led to plasmid loss, T3SS PAI deletions, and/or a reduction in viability. Potentially, an antibiotic resistance cassette could be introduced in pINV, next to *virB*, to prevent loss of both the PAI and pINV, following CRISPR-Cas9 targeting. It is not clear how *Shigella* would repair DSB introduced by the CRISPR-Cas9 systems. Even though hypotheses have been proposed about potential A-EJ mechanisms, further experiments should be conducted to characterise the effect of the CRISPR-Cas9 system on *Shigella* (Chayot *et al*., 2010).

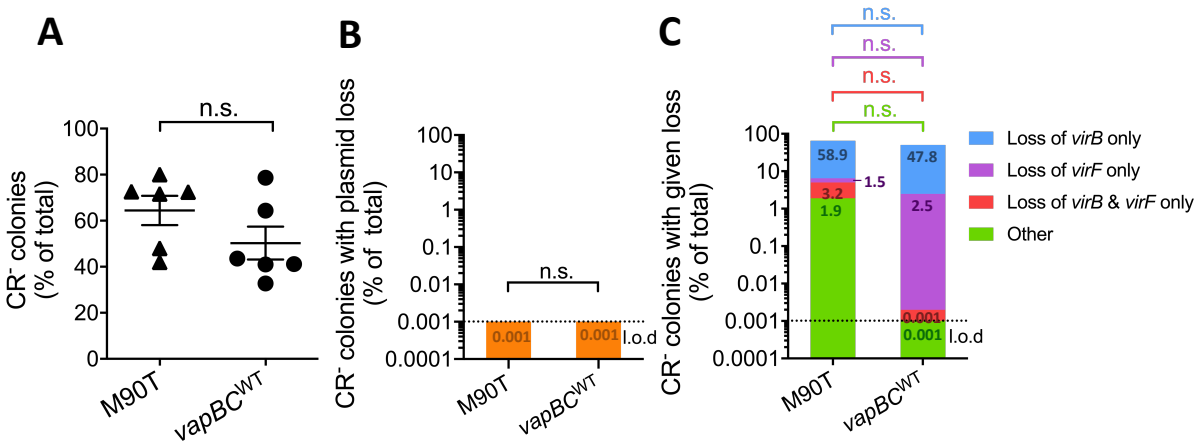
More broadly, the CRISPR-Cas9 system could be employed to target other plasmid-encoded elements essential for the spread of plasmids in bacteria. For instance, bacterial conjugation is the main mechanism of transfer of AMR plasmids among enteropathogenic bacteria and inhibition of this process has been proven to be a promising strategy to prevent dissemination of AMR (de la Cruz and Davies, 2000; Getino *et al*., 2015; Ripoll-Rozada *et al*., 2016).

Additionally, CRISPR-Cas9 could be as a tool to characterise and visualise PSK *in vivo*. The dynamics of PSK (*e.g.* stalling of translation/replication, growth arrest, death) and consequent changes in bacteria (*e.g.* growth, morphology, inhibition of translation) have not been investigated. Using the approach previously discussed (Section 5.3), microfluidic live cell microscopy could be employed for this purpose.

Given the potential of this approach to eliminate virulence in *Shigella*, a long-term goal would be to translate this finding into a clinical application. However, besides optimisation of the system itself, equally important will be the delivery of CRISPR-Cas for therapeutic purposes. HGT of the plasmid encoding the CRISPR-Cas9 system could be achieved by transduction or, alternatively, by conjugation through a probiotic carrier strain (Citorik *et al*., 2014; Pursey *et al*., 2018). Both approaches have been successfully tested in previous studies, however, several are the challenges encountered, such as transfer efficacy and the complexity of microbial communities in the gastrointestinal tract (Beisel *et al*., 2014; Pursey *et al*., 2018).

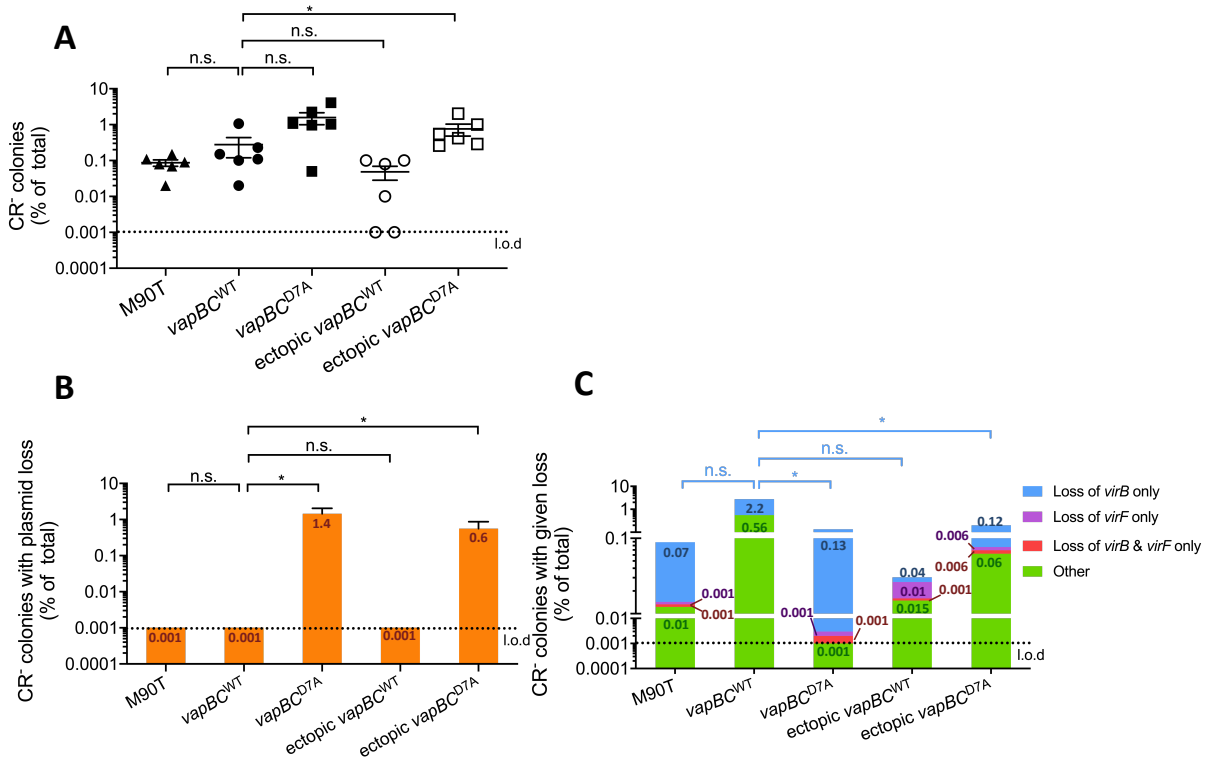
7 Appendix

Appendix Figure 7.1 *vapBC*^{WT} is comparable to M90T.



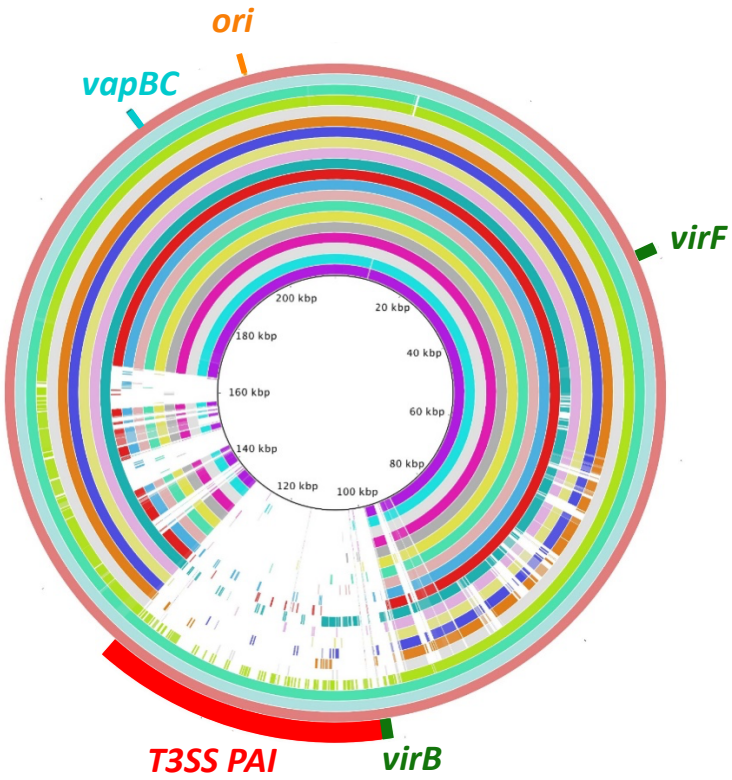
(**A**) Number of CR$^-$ colonies emerging from *S. flexneri* M90T and *vapBC*^{WT}, after approximately 50 generations of growth at 37°C, shown as a % of all colonies. Solid line: mean ± S.E.M. (n = 6 biological replicates). n.s., not significant; Welch's *t*-test. (**B**) and (**C**) Number of CR$^-$ colonies lacking specified virulence-related genes were assayed by multiplex PCR and shown as a % of all colonies. Loss of the plasmid is inferred by loss of the origin of replication (B); "others" refers to CR$^-$ colonies that contain *virB*, *virF* and the origin of replication (C). The value of each category is indicated inside or next to the corresponding bar portion. Eight independent CR$^-$ colonies obtained from the indicated strains grown in six biological repeats (total 48 colonies) were analysed; data are shown as mean + S.E.M. in panel B and mean in panel C. n.s., not significant; Wilcoxon signed-rank test (nonparametric comparison of median of strain/vapBC^{WT} to 1, for each event). l.o.d, limit of detection.

Appendix Figure 7.2 Loss of CR binding is reduced at 21°C.



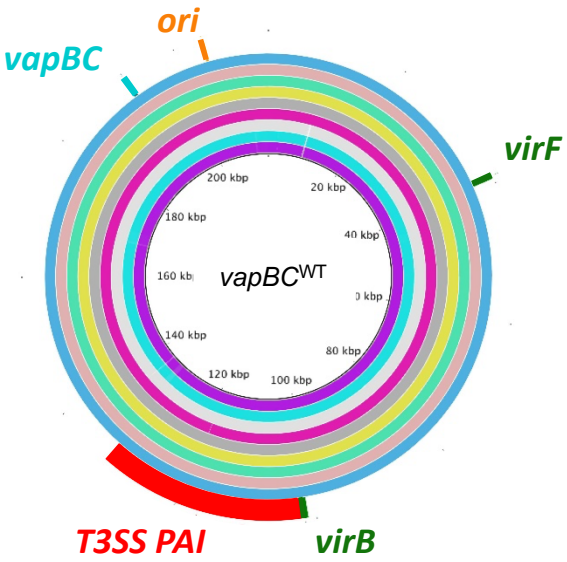
(**A**) Number of CR$^-$ colonies emerging from *S. flexneri* M90T, *vapBC*^{WT}, *vapBC*^{D7A}, ectopic *vapBC*^{WT} and ectopic *vapBC*^{D7A} after approximately 50 generations of growth at 37°C, shown as % of all colonies. Solid line: mean ± S.E.M (n = 6 biological replicates). (**B**) and (**C**) Number of CR$^-$ colonies lacking specified virulence-related genes were assayed by multiplex PCR and shown as a % of all colonies. Loss of the plasmid is inferred by loss of the origin of replication (B); "others" refers to CR$^-$ colonies that contain *virB*, *virF* and the origin of replication (C). The value of each category is indicated inside or next to the corresponding bar portion. Eight independent CR$^-$ colonies obtained from the indicated strains grown in six biological repeats (total 48 colonies) were analysed; data are shown as mean + S.E.M. in panel B and mean in panel C. *, $p \leq 0.05$ and n.s., not significant; Wilcoxon signed-rank test (nonparametric comparison of median of strain/ *vapBC*^{WT} to 1, for each event). l.o.d, limit of detection.

Appendix Figure 7.3 Emergence of CR$^-$ colonies in *S. flexneri* results from spontaneous loss of the PAI.



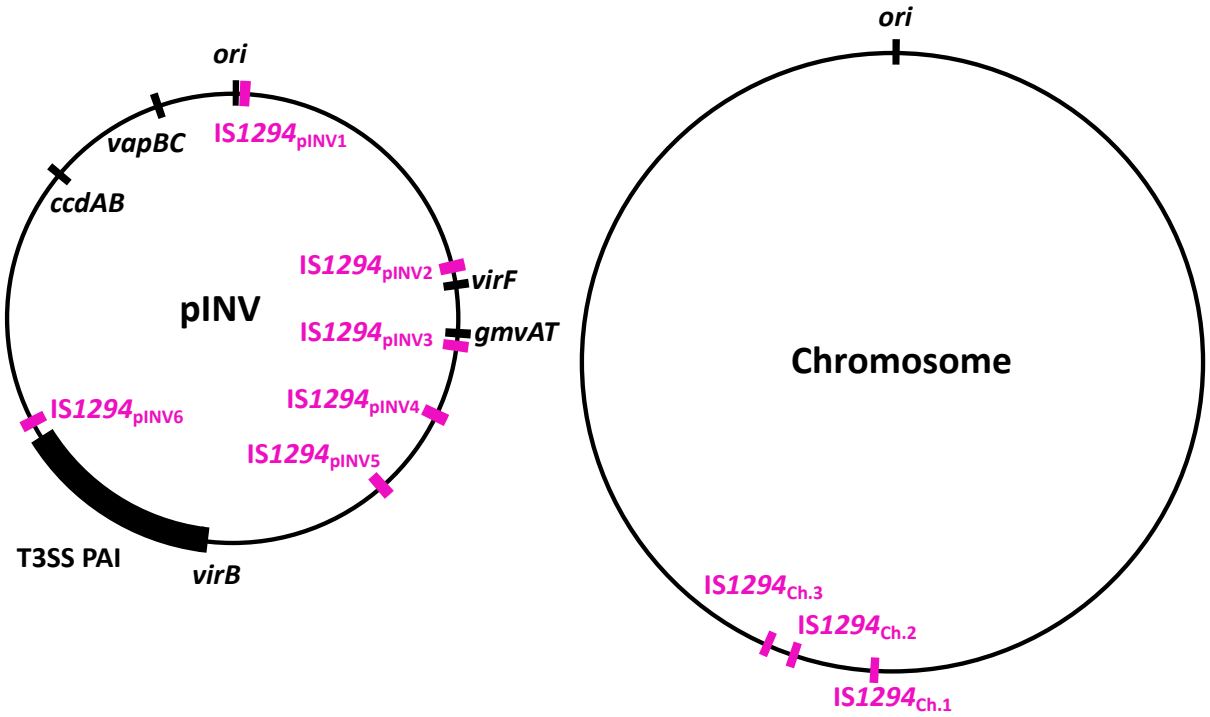
BLAST Ring Image Generator (BRIG 0.95 and BLASTN v2.2.29) alignment of plasmid sequences from twenty CR$^-$ colonies that had emerged from *vapBC*^{WT} at 37°C; each ring represents the plasmid from an independent CR$^-$ colony. *S. flexneri* M90T pWR100 (AL391753.1, inner black ring) shown as the reference.

Appendix Figure 7.4 CR$^-$ isolates with wild-type multiplex PCR profiles show intact plasmid sequences in *S. flexneri*.



BLAST Ring Image Generator (BRIG 0.95 and BLASTN v2.2.29) alignment of plasmid sequences from nine CR$^-$ colonies that had emerged from *vapBC*^{WT} at 21°C and presented all multiplex PCR products ("other" category); each ring represents the plasmid from an independent CR$^-$ colony. S. *flexneri* M90T pWR100 (AL391753.1, inner black ring) shown as the reference.

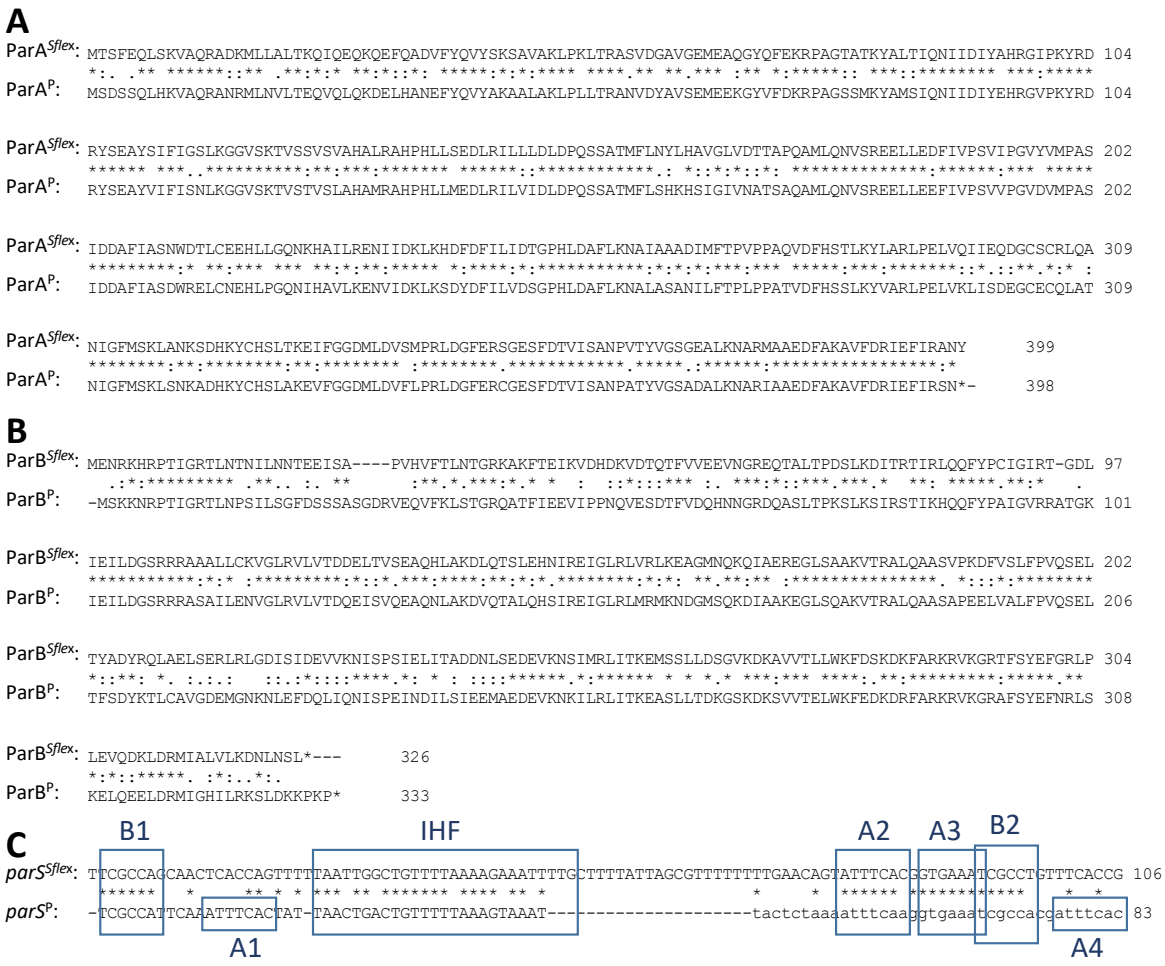
Appendix Figure 7.5 Schematic of the location of IS*1294* copies on pINV and on the chromosome of *S. flexneri* M90T.



Appendix Figure 7.6 Schematic of the location of IS*1294* copies on pINV and on the chromosome of *S. flexneri* M90T.

IS*1294* copies are indicated in purple and the sequences deposited with the accession numbers AL391753.1 and CM001474.1, respectively, were used as reference.

Appendix Figure 7.6 Alignment of ParABS elements from *S. flexneri* pINV and P1 plasmid.



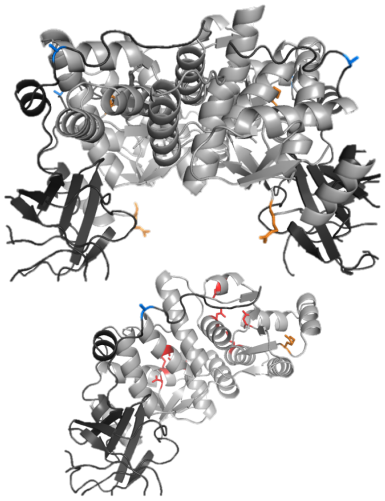
Amino acid alignment of ParA (**A**) and ParB (**B**), and nucleotide alignment of *parS* (**C**) from *S. flexneri* pINV (AL391753) and P1 plasmid (NC_005856) by CLUSTAL O (1.2.4). Asterisks indicate identical amino acids/nucleotides, colons show amino acids with similar properties, and periods indicate amino acids sharing weakly similar properties. In panel C ParB binding boxes and the IHF binding site are indicated.

Appendix Figure 7.7 Alignment of StbABC elements from *S. flexneri* pINV with ParMRC elements from R1 plasmid.



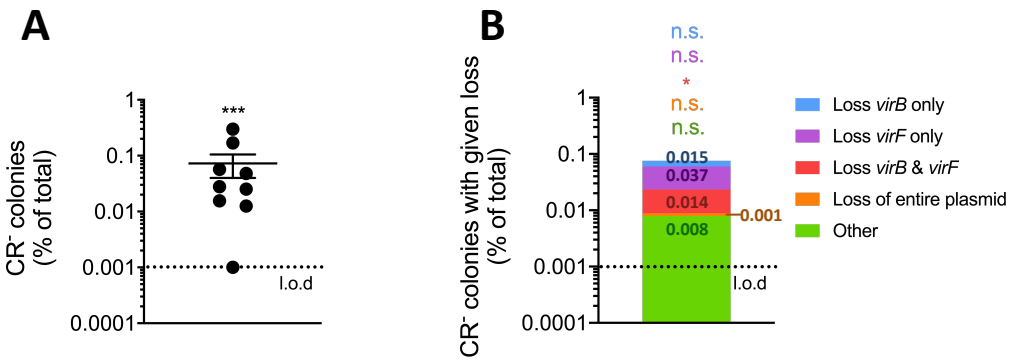
Amino acid alignment of StbA/ParM (**A**) and StbB/ParR (**B**), and nucleotide alignment of *stbC*/*parC* (**C**) from *S. flexneri* pINV (AL391753) and R1 plasmid (KY749247) by CLUSTAL O (1.2.4). Asterisks indicate identical amino acids/nucleotides, colons show amino acids with similar properties, and periods indicate amino acids sharing weakly similar properties.

Appendix Figure 7.8 Atomic structure of *S. sonnei* VapBC.



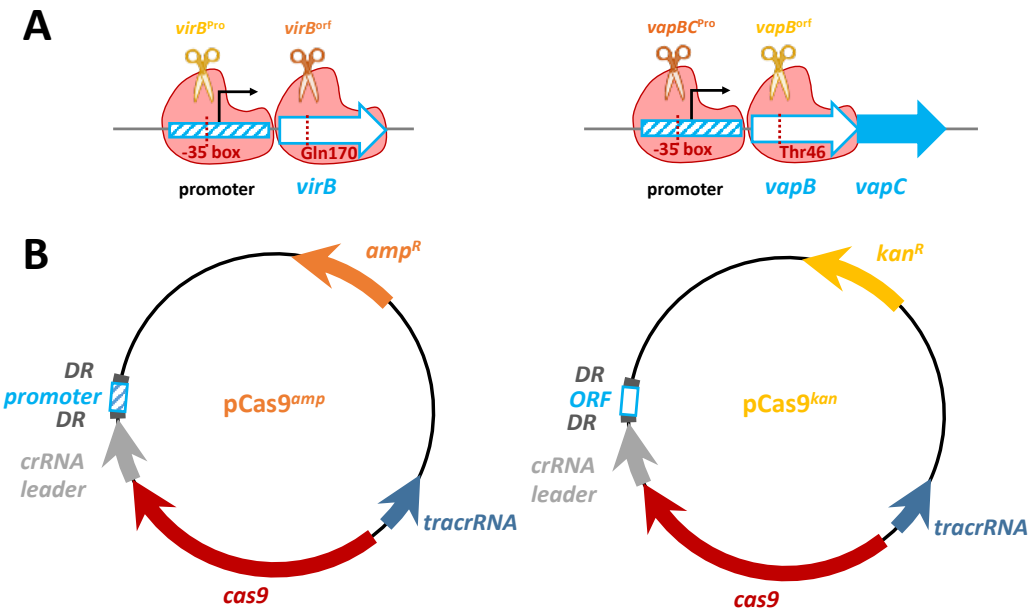
Atomic structure of the *S. sonnei* 53G VapBC hetero-octomer (top) and the $VapB^{Ssonn}$ dimer in complex with a $VapC^{Ssonn}$ dimer (bottom); $VapC^{Ssonn}$ is in light grey, and $VapB^{Ssonn}$ is dark grey. The four residues comprising the VapC active site (D7, E42, D98 and E119) are highlighted in red. The location of the polymorphisms in VapB (T58A) and VapC (K32R) are highlighted in blue and orange, respectively. From Martyn *et al*., unpublished.

Appendix Figure 7.9 *S. sonnei* CS14 possesses a highly stable pINV.



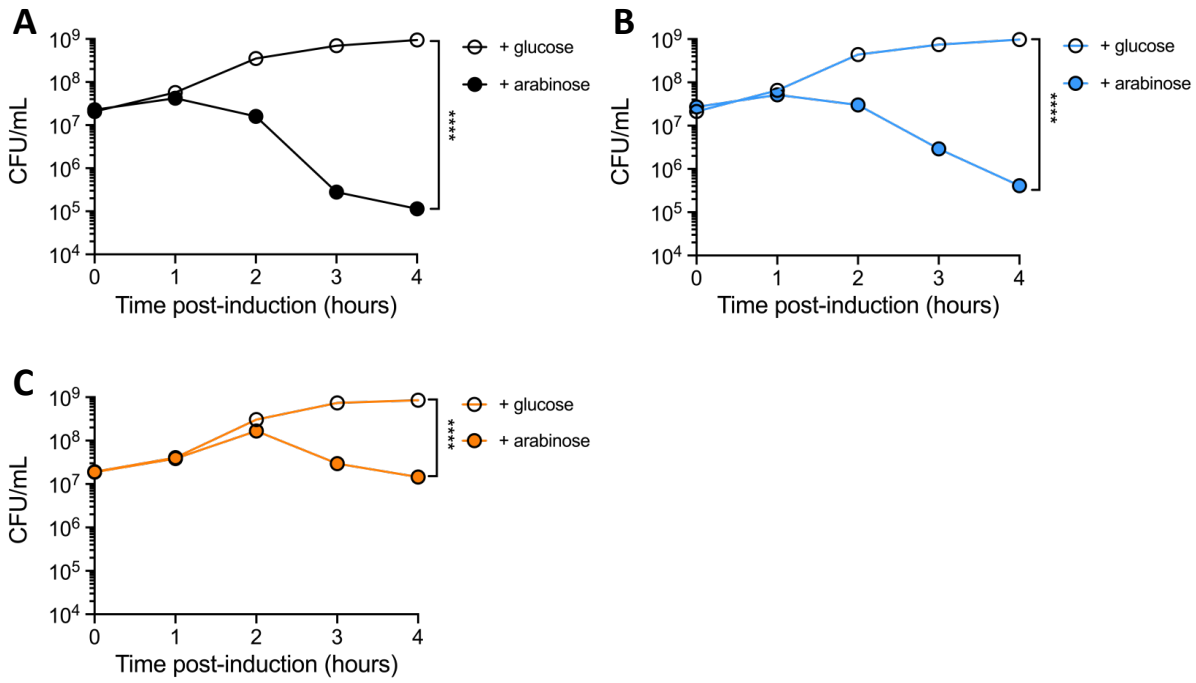
(**A**) Number of CR$^-$ colonies emerging from *S. sonnei* CS14 after growth at 37°C for approximately 25 generations, shown as a % of all colonies. Solid line: mean ± S.E.M (n = 9 biological repeats). ***, $p < 0.001$; one sample *t*-test (parametric comparison of mean strain to l.o.d.). (**B**) Number of CR$^-$ colonies analysed by multiplex PCR with given genetic events leading to the emergence of CR$^-$ phenotype, shown as a % of all colonies. Loss of the entire plasmid is inferred by loss of the origin of replication; "others" refers to CR- colonies that contain *virB*, *virF* and the origin of replication. The value of each category is indicated inside the corresponding bar portion. Three CR$^-$ colonies were tested from each biological repeat; data are shown as mean. *, $p \leq 0.05$; ns., not significant; Wilcoxon signed-rank test (nonparametric comparison of median of each event to limit of detection, l.o.d.).

Appendix Figure 7.10 Spacers and pCas9 plasmids designed to target *virB* and *vapBC*.



(**A**) For both *virB* and *vapBC*, two spacers were designed to target each a distinct site of the locus: one spacer targeted the -35 box of the promoter (spacers *virB*^{Pro} and *vapBC*^{Pro}), while the second spacer targeted the ORF of *virB* and *vapB* at Gln170 and Thr46, respectively (spacers *virB*^{orf} and *vapB*^{orf}). (**B**) Two versions of pCas9 were designed: pCas9^{amp} (orange), which contains an *amp*^{R} cassette and the spacer targeting the promoter sequence of *virB* or *vapBC*; pCas9^{kan} (yellow), which contains a *kan*^{R} cassette and the spacer targeting the ORF of *virB* or *vapB*. Each single spacer (blue dashed or blue outlined box) was inserted in pCas9 between the two direct repeats (DR, dark grey).

Appendix Figure 7.11 CRISPR-Cas9 systems targeting *vapBC* are inducible in *S. sonnei*.



Viability of *S. sonnei* CS14 containing the CRISPR-Cas9 systems *vapP* (**A**), *vapB* (**B**) and *vapC* (**C**), following growth with 0.2% glucose or 0.2% arabinose for 0, 1, 2, 3 and 4 hours. Data are shown as mean (n = 6 biological repeats). ****, $p < 0.0001$ by one sample *t*-test, parametric comparison of mean test/control strain to 1.

Appendix Table 7.1 Frequency of T3SS PAI deletion variants, and alignment coverage, identity and GC content of the IS involved.

IS pair	Frequency	Coverage (%)	Identity (%)	GC Content (%)
Variant 1 (IS*Sfl4*)	11/17	100	99	52.3/52.2
Variant 2 (IS*1294*)	4/17	99	99	56.1/56.2
Variant 3 (IS*600*)	1/17	98	99	49.1/49.2
Variant 4 (IS*1294*)	1/17	80	99	56.6/56.2

The frequency shows the ratio of the number of colonies presenting the specific deletion variant to the total number of colonies presenting T3SS PAI deletions detected by sequencing (Appendix Figure 7.1). The DNA sequences of the IS in each pair were aligned using Nucleotide BLAST®.

Appendix Table 7.2 Amino acid or nucleotide identity of ParABS and StbABC systems between *S. flexneri* and other *Shigella* spp..

	S. flexneri					
	ParA	ParB	*parS*	StbA	StbB	*stbC*
S. sonnei	100	99	100	/	/	/
S. boydii	100	100	100	99	99	99
S. dysenteriae	99	100	99	99	99	99

Values are shown as percentages of amino acid identity (for ParA and ParB, and StbA and StbB) or nucleotide identity (for *parS* and *stbC*). Amino acid/ nucleotide sequences were obtained from *S. flexneri* pWR100 (AL391753.1), *S. sonnei* 53G plasmid A (NC_016833), *S. boydii* pBS512_211 (CP001062) and *S. dysenteriae* pSD1 (CP000035) and aligned using CLUSTAL O (1.2.4).

8 Bibliography

Abeles, A. L. and S. J. Austin (1987). "P1 plasmid replication requires methylated DNA." EMBO J 6(10): 3185-3189.

Adler, B., C. Sasakawa, T. Tobe, S. Makino, K. Komatsu and M. Yoshikawa (1989). "A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of *Shigella flexneri*." Mol Microbiol 3(5): 627-635.

Alawneh, A. M., D. Qi, T. Yonesaki and Y. Otsuka (2016). "An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the *Escherichia coli* MazF toxin of a toxin-antitoxin module." Mol Microbiol 99(1): 188-198.

Alikhan, N. F., N. K. Petty, N. L. Ben Zakour and S. A. Beatson (2011). "BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons." BMC Genomics 12: 402.

Allaoui, A., R. Menard, P. J. Sansonetti and C. Parsot (1993). "Characterisation of the *Shigella flexneri ipgD* and *ipgF* genes, which are located in the proximal part of the mxi locus." Infect Immun 61(5): 1707-1714.

Anderson, M., P. J. Sansonetti and B. S. Marteyn (2016). "*Shigella* Diversity and Changing Landscape: Insights for the Twenty-First Century." Front Cell Infect Microbiol 6: 45.

Andersson, J. O. and S. G. Andersson (1999). "Insights into the evolutionary process of genome degradation." Curr Opin Genet Dev 9(6): 664-671.

Arbibe, L. and P. Sansonetti (2007). "*Shigella flexneri* modulates host cell epigenetic information as a strategy to shape the transcriptional response." Med Sci (Paris) 23(3): 238-240.

Ashida, H., H. Mimuro and C. Sasakawa (2015). "*Shigella* manipulates host immune responses by delivering effector proteins with specific roles." Front Immunol 6: 219.

Ashida, H., M. Ogawa, M. Kim, S. Suzuki, T. Sanada, C. Punginelli, H. Mimuro and C. Sasakawa (2011). "*Shigella* deploy multiple countermeasures against host innate immune responses." Curr Opin Microbiol 14(1): 16-23.

Ashida, H., T. Toyotome, T. Nagai and C. Sasakawa (2007). "*Shigella* chromosomal IpaH proteins are secreted via the type III secretion system and act as effectors." Mol Microbiol 63(3): 680-693.

Aziz, R. K., M. Breitbart and R. A. Edwards (2010). "Transposases are the most abundant, most ubiquitous genes in nature." Nucleic Acids Res 38(13): 4207-4217.

Bahassi, E. M., M. A. Salmon, L. Van Melderen, P. Bernard and M. Couturier (1995). "F plasmid CcdB killer protein: *ccdB* gene mutants coding for non-cytotoxic proteins which retain their regulatory functions." Mol Microbiol 15(6): 1031-1037.

Bahrani, F. K., P. J. Sansonetti and C. Parsot (1997). "Secretion of Ipa proteins by *Shigella flexneri*: inducer molecules and kinetics of activation." Infect Immun 65(10): 4005-4010.

Baker, K. S., T. J. Dallman, P. M. Ashton, M. Day, G. Hughes, P. D. Crook, V. L. Gilbart, S. Zittermann, V. G. Allen, B. P. Howden, T. Tomita, M. Valcanis, S. R. Harris, T. R. Connor, V. Sintchenko, P. Howard, J. D. Brown, N. K. Petty, M. Gouali, D. P. Thanh, K. H. Keddy, A. M. Smith, K. A. Talukder, S. M. Faruque, J. Parkhill, S. Baker, F. X. Weill, C. Jenkins and N. R. Thomson (2015). "Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study." Lancet Infect Dis 15(8): 913-921.

Baker, S. and H. C. The (2018). "Recent insights into Shigella." Curr Opin Infect Dis 31(5): 449-454.

Barbagallo, M., M. L. Di Martino, L. Marcocci, P. Pietrangeli, E. De Carolis, M. Casalino, B. Colonna and G. Prosseda (2011). "A new piece of the *Shigella* Pathogenicity puzzle: spermidine accumulation by silencing of the *speG* gene [corrected]." PLoS One 6(11): e27226.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." Science 315(5819): 1709-1712.

Barrangou, R. and L. A. Marraffini (2014). "CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity." Mol Cell 54(2): 234-244.

Beisel, C. L., A. A. Gomaa and R. Barrangou (2014). "A CRISPR design for next-generation antimicrobials." Genome Biol 15(11): 516.

Benjelloun-Touimi, Z., P. J. Sansonetti and C. Parsot (1995). "SepA, the major extracellular protein of *Shigella flexneri*: autonomous secretion and involvement in tissue invasion." Mol Microbiol 17(1): 123-135.

Bergsbaken, T., S. L. Fink and B. T. Cookson (2009). "Pyroptosis: host cell death and inflammation." Nat Rev Microbiol 7(2): 99-109.

Bernard, P. and M. Couturier (1992). "Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes." J Mol Biol 226(3): 735-745.

Bernardini, M. L., J. Mounier, H. d'Hauteville, M. Coquis-Rondon and P. J. Sansonetti (1989). "Identification of *icsA*, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin." Proc Natl Acad Sci U S A 86(10): 3867-3871.

Bikard, D., C. W. Euler, W. Jiang, P. M. Nussenzweig, G. W. Goldberg, X. Duportet, V. A. Fischetti and L. A. Marraffini (2014). "Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials." Nat Biotechnol 32(11): 1146-1150.

Blocker, A., P. Gounon, E. Larquet, K. Niebuhr, V. Cabiaux, C. Parsot and P. Sansonetti (1999). "The tripartite type III secreton of *Shigella flexneri* inserts IpaB and IpaC into host membranes." J Cell Biol 147(3): 683-693.

Blocker, A., N. Jouihri, E. Larquet, P. Gounon, F. Ebel, C. Parsot, P. Sansonetti and A. Allaoui (2001). "Structure and composition of the *Shigella flexneri* "needle complex", a part of its type III secreton." Mol Microbiol 39(3): 652-663.

Blomberg, P., H. M. Engdahl, C. Malmgren, P. Romby and E. G. Wagner (1994). "Replication control of plasmid R1: disruption of an inhibitory RNA structure that sequesters the *repA* ribosome-binding site permits tap-independent RepA synthesis." Mol Microbiol 12(1): 49-60.

Blomberg, P., K. Nordstrom and E. G. Wagner (1992). "Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation." EMBO J 11(7): 2675-2683.

Bondy-Denomy, J. and A. R. Davidson (2014). "When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness." J Microbiol 52(3): 235-242.

Bouet, J. Y. and B. E. Funnell (1999). "P1 ParA interacts with the P1 partition complex at *parS* and an ATP-ADP switch controls ParA activities." EMBO J 18(5): 1415-1424.

Bowen, A., J. Grass, A. Bicknese, D. Campbell, J. Hurd and R. D. Kirkcaldy (2016). "Elevated Risk for Antimicrobial Drug-Resistant *Shigella* Infection among Men Who Have Sex with Men, United States, 2011-2015." Emerg Infect Dis 22(9): 1613-1616.

Boyd, JS., (1938). "The antigenic structure of the mannitol-fermenting group of dysentery bacilli." The Journal of Hygene(38): 477-499.

Brantl, S. (2015). "Antisense-RNA mediated control of plasmid replication - pIP501 revisited." Plasmid 78: 4-16.

Brauner, A., O. Fridman, O. Gefen and N. Q. Balaban (2016). "Distinguishing between resistance, tolerance and persistence to antibiotic treatment." Nat Rev Microbiol 14(5): 320-330.

Bravo, V., A. Puhar, P. Sansonetti, C. Parsot and C. S. Toro (2015). "Distinct mutations led to inactivation of type 1 fimbriae expression in Shigella spp." PLoS One 10(3): e0121785.

Brinkmann, V., U. Reichard, C. Goosmann, B. Fauler, Y. Uhlemann, D. S. Weiss, Y. Weinrauch and A. Zychlinsky (2004). "Neutrophil extracellular traps kill bacteria." Science 303(5663): 1532-1535.

Brooks, A. C. and L. C. Hwang (2017). "Reconstitutions of plasmid partition systems and their mechanisms." Plasmid 91: 37-41.

Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin and J. van der Oost (2008). "Small CRISPR RNAs guide antiviral defense in prokaryotes." Science 321(5891): 960-964.

Brzozowska, I. and U. Zielenkiewicz (2013). "Regulation of toxin-antitoxin systems by proteolysis." Plasmid 70(1): 33-41.

Buchrieser, C., P. Glaser, C. Rusniok, H. Nedjari, H. D'Hauteville, F. Kunst, P. Sansonetti and C. Parsot (2000). "The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri." Mol Microbiol 38(4): 760-771.

Buysse, J. M., C. K. Stover, E. V. Oaks, M. Venkatesan and D. J. Kopecko (1987). "Molecular cloning of invasion plasmid antigen (ipa) genes from *Shigella flexneri*: analysis of *ipa* gene products and genetic mapping." J Bacteriol 169(6): 2561-2569.

Campbell, C. S. and R. D. Mullins (2007). "In vivo visualisation of type II plasmid segregation: bacterial actin filaments pushing plasmids." J Cell Biol 179(5): 1059-1066.

Cao, Q. H., H. H. Shao, H. Qiu, T. Li, Y. Z. Zhang and X. M. Tan (2017). "Using the CRISPR/Cas9 system to eliminate native plasmids of *Zymomonas mobilis* ZM4." Biosci Biotechnol Biochem 81(3): 453-459.

Carayol, N. and G. Tran Van Nhieu (2013). "Tips and tricks about *Shigella* invasion of epithelial cells." Curr Opin Microbiol 16(1): 32-37.

Casadevall, A. (2008). "Evolution of intracellular pathogens." Annu Rev Microbiol 62: 19-33.

Cataudella, I., A. Trusina, K. Sneppen, K. Gerdes and N. Mitarai (2012). "Conditional cooperativity in toxin-antitoxin regulation prevents random toxin activation and promotes fast translational recovery." Nucleic Acids Res 40(14): 6424-6434.

Chan, W. T., M. Espinosa and C. C. Yeo (2016). "Keeping the Wolves at Bay: Antitoxins of Prokaryotic Type II Toxin-Antitoxin Systems." Front Mol Biosci 3: 9.

Chayot, R., B. Montagne, D. Mazel and M. Ricchetti (2010). "An end-joining repair mechanism in Escherichia coli." Proc Natl Acad Sci U S A 107(5): 2141-2146.

Cheng, F., J. Wang, J. Peng, J. Yang, H. Fu, X. Zhang, Y. Xue, W. Li, Y. Chu and Q. Jin (2007). "Gene expression profiling of the pH response in *Shigella flexneri* 2a." FEMS Microbiol Lett 270(1): 12-20.

Cherepanov, P. P. and W. Wackernagel (1995). "Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant." Gene 158(1): 9-14.

Cheverton, A. M., B. Gollan, M. Przydacz, C. T. Wong, A. Mylona, S. A. Hare and S. Helaine (2016). "A *Salmonella* Toxin Promotes Persister Formation through Acetylation of tRNA." Mol Cell 63(1): 86-96.

Choi, C. L., S. A. Claridge, E. C. Garner, A. P. Alivisatos and R. D. Mullins (2008). "Protein-nanocrystal conjugates support a single filament polymerisation model in R1 plasmid segregation." J Biol Chem 283(42): 28081-28086.

Christensen-Dalsgaard, M. and K. Gerdes (2006). "Two *higBA* loci in the *Vibrio cholerae* superintegron encode mRNA cleaving enzymes and can stabilise plasmids." Mol Microbiol 62(2): 397-411.

Chung The, H., C. Boinett, D. Pham Thanh, C. Jenkins, F. X. Weill, B. P. Howden, M. Valcanis, N. De Lappe, M. Cormican, S. Wangchuk, L. Bodhidatta, C. J. Mason, T. N. T. Nguyen, T. Ha Thanh, V. P. Voong, V. T. Duong, P. H. L. Nguyen, P. Turner, R. Wick, P. J. Ceyssens, G. Thwaites, K. E. Holt, N. R. Thomson, M. A. Rabaa and S. Baker (2019). "Dissecting the molecular evolution of fluoroquinolone-resistant Shigella sonnei." Nat Commun 10(1): 4828.

Citorik, R. J., M. Mimee and T. K. Lu (2014). "Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases." Nat Biotechnol 32(11): 1141-1145.

Clerc, P. and P. J. Sansonetti (1987). "Entry of *Shigella flexneri* into HeLa cells: evidence for directed phagocytosis involving actin polymerisation and myosin accumulation." Infect Immun 55(11): 2681-2688.

Coburn, B., I. Sekirov and B. B. Finlay (2007). "Type III secretion systems and disease." Clin Microbiol Rev 20(4): 535-549.

Collaborators, G. B. D. D. D. (2018). "Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016." Lancet Infect Dis 18(11): 1211-1228.

Colonna, B., M. Casalino, P. A. Fradiani, C. Zagaglia, S. Naitza, L. Leoni, G. Prosseda, A. Coppo, P. Ghelardini and M. Nicoletti (1995). "H-NS regulation of virulence gene expression in

enteroinvasive *Escherichia coli* harboring the virulence plasmid integrated into the host chromosome." J Bacteriol 177(16): 4703-4712.

Cordes, F. S., K. Komoriya, E. Larquet, S. Yang, E. H. Egelman, A. Blocker and S. M. Lea (2003). "Helical structure of the needle of the type III secretion system of *Shigella flexneri*." J Biol Chem 278(19): 17103-17107.

Cossart, P. and P. J. Sansonetti (2004). "Bacterial invasion: the paradigms of enteroinvasive pathogens." Science 304(5668): 242-248.

Dam, M. and K. Gerdes (1994). "Partitioning of plasmid R1. Ten direct repeats flanking the *parA* promoter constitute a centromere-like partition site *parC*, that expresses incompatibility." J Mol Biol 236(5): 1289-1298.

Dao-Thi, M. H., D. Charlier, R. Loris, D. Maes, J. Messens, L. Wyns and J. Backmann (2002). "Intricate interactions within the ccd plasmid addiction system." J Biol Chem 277(5): 3733-3742.

Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products." Proc Natl Acad Sci U S A 97(12): 6640-6645.

de la Cruz, F. and J. Davies (2000). "Horizontal gene transfer and the origin of species: lessons from bacteria." Trends Microbiol 8(3): 128-133.

Deghelt, M., C. Mullier, J. F. Sternon, N. Francis, G. Laloux, D. Dotreppe, C. Van der Henst, C. Jacobs-Wagner, J. J. Letesson and X. De Bolle (2014). "G1-arrested newborn cells are the predominant infectious form of the pathogen *Brucella abortus*." Nat Commun 5: 4366.

del Solar, G., R. Giraldo, M. J. Ruiz-Echevarria, M. Espinosa and R. Diaz-Orejas (1998). "Replication and control of circular bacterial plasmids." Microbiol Mol Biol Rev 62(2): 434-464.

Deltcheva, E., K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel and E. Charpentier (2011). "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Nature 471(7340): 602-607.

Di Martino, M. L., M. Falconi, G. Micheli, B. Colonna and G. Prosseda (2016). "The Multifaceted Activity of the VirF Regulatory Protein in the Shigella Lifestyle." Front Mol Biosci 3: 61.

Dienemann, C., A. Bøggild, K. S. Winther, K. Gerdes and D. E. Brodersen (2011). "Crystal structure of the VapBC toxin-antitoxin complex from *Shigella flexneri* reveals a hetero-octameric DNA-binding assembly." J Mol Biol 414(5): 713-722.

Dong, N., Y. Zhu, Q. Lu, L. Hu, Y. Zheng and F. Shao (2012). "Structurally distinct bacterial TBC-like GAPs link Arf GTPase to Rab1 inactivation to counteract host defenses." Cell 150(5): 1029-1041.

Dorman, C. J. (2004). "Virulence Gene Regulation in *Shigella*." Ecosal Plus 1(1).

Doudna, J. A. and E. Charpentier (2014). "Genome editing. The new frontier of genome engineering with CRISPR-Cas9." Science 346(6213): 1258096.

DuPont, H. L., M. M. Levine, R. B. Hornick and S. B. Formal (1989). "Inoculum size in shigellosis and implications for expected mode of transmission." J Infect Dis 159(6): 1126-1128.

Durand, J. M., G. R. Bjork, A. Kuwae, M. Yoshikawa and C. Sasakawa (1997). "The modified nucleoside 2-methylthio-N6-isopentenyladenosine in tRNA of *Shigella flexneri* is required for expression of virulence genes." J Bacteriol 179(18): 5777-5782.

Durand, J. M., B. Dagberg, B. E. Uhlin and G. R. Björk (2000). "Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the virF gene." Mol Microbiol 35(4): 924-935.

Durand, J. M., N. Okada, T. Tobe, M. Watarai, I. Fukuda, T. Suzuki, N. Nakata, K. Komatsu, M. Yoshikawa and C. Sasakawa (1994). "*vacC*, a virulence-associated chromosomal locus of *Shigella flexneri*, is homologous to *tgt*, a gene encoding tRNA-guanine transglycosylase (Tgt) of *Escherichia coli* K-12." J Bacteriol 176(15): 4627-4634.

Ebersbach Gitte and K. Gerdes (2005). "Plasmid Segregation Mechanisms." Annual Review of Genetics.

Egile, C., T. P. Loisel, V. Laurent, R. Li, D. Pantaloni, P. J. Sansonetti and M. F. Carlier (1999). "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 146(6): 1319-1332.

Escobar-Paramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec and E. Denamur (2004). "A specific genetic background is required for acquisition and expression of virulence factors in Escherichia coli." Mol Biol Evol 21(6): 1085-1094.

Espeli, O., R. Mercier and F. Boccard (2008). "DNA dynamics vary according to macrodomain topography in the E. coli chromosome." Mol Microbiol 68(6): 1418-1427.

Ewing, W. H. (1949). "*SHIGELLA* NOMENCLATURE." J Bacteriol 57(6): 633-638.

Falconi, M., B. Colonna, G. Prosseda, G. Micheli and C. O. Gualerzi (1998). "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 17(23): 7033-7043.

Falkow, S., H. Schneider, L. S. Baron and S. B. Formal (1963). "Virulence of *Escherichia-Shigella* Genetic Hybrids for the Guinea Pig." J Bacteriol 86: 1251-1258.

Fasano, A., F. R. Noriega, F. M. Liao, W. Wang and M. M. Levine (1997). "Effect of Shigella enterotoxin 1 (ShET1) on rabbit intestine in vitro and in vivo." Gut 40(4): 505-511.

Feng, Y., Z. Chen and S. L. Liu (2011). "Gene decay in *Shigella* as an incipient stage of host-adaptation." PLoS One 6(11): e27754.

Figurski, D. H. and D. R. Helinski (1979). "Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans." Proc Natl Acad Sci U S A 76(4): 1648-1652.

Fineran, P. C., T. R. Blower, I. J. Foulds, D. P. Humphreys, K. S. Lilley and G. P. Salmond (2009). "The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair." Proc Natl Acad Sci U S A 106(3): 894-899.

Flexner, S. and J. E. Sweet (1906). "The Pathogenesis of Experimental Colitis, and the Relation of Colitis in Animals and Man." J Exp Med 8(4): 514-535.

Formal, S. B., G. Dammin, H. Sprinz, D. Kundel, H. Schneider, R. E. Horowitz and M. Forbes (1961). "Experimental *Shigella* infections. V. Studies in germ-free guinea pigs." J Bacteriol 82: 284-287.

Friedman, S. A. and S. J. Austin (1988). "The P1 plasmid-partition system synthesizes two essential proteins from an autoregulated operon." Plasmid 19(2): 103-112.

Froehlich, B., J. Parkhill, M. Sanders, M. A. Quail and J. R. Scott (2005). "The pCoo plasmid of enterotoxigenic Escherichia coli is a mosaic cointegrate." J Bacteriol 187(18): 6509-6516.

Fukumatsu, M., M. Ogawa, S. Arakawa, M. Suzuki, K. Nakayama, S. Shimizu, M. Kim, H. Mimuro and C. Sasakawa (2012). "*Shigella* targets epithelial tricellular junctions and uses a noncanonical clathrin-dependent endocytic pathway to spread between cells." Cell Host Microbe 11(4): 325-336.

Fukushima, M., K. Kakinuma and R. Kawaguchi (2002). "Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the gyrB gene sequence." J Clin Microbiol 40(8): 2779-2785.

Fung, E., J. Y. Bouet and B. E. Funnell (2001). "Probing the ATP-binding site of P1 ParA: partition and repression have different requirements for ATP binding and hydrolysis." EMBO J 20(17): 4901-4911.

Funnell, B. E. (1991). "The P1 plasmid partition complex at *parS*. The influence of *Escherichia coli* integration host factor and of substrate topology." J Biol Chem 266(22): 14328-14337.

Funnell, B. E. and L. Gagnier (1993). "The P1 plasmid partition complex at *parS*. II. Analysis of ParB protein binding activity and specificity." J Biol Chem 268(5): 3616-3624.

Funnell, E. B. and J. G. Phillips (2004). Plasmid Biology, American Society for Microbiology Press.

Gall, T. L., M. Mavris, M. C. Martino, M. L. Bernardini, E. Denamur and C. Parsot (2005). "Analysis of virulence plasmid gene expression defines three classes of effectors in the type III secretion system of *Shigella flexneri*." Microbiology 151(Pt 3): 951-962.

Garcia-Aljaro, C., M. Muniesa, J. Jofre and A. R. Blanch (2006). "Newly identified bacteriophages carrying the *stx2g* Shiga toxin gene isolated from *Escherichia coli* strains in polluted waters." FEMS Microbiol Lett 258(1): 127-135.

Garcia-Pino, A., Y. Sterckx, G. Vandenbussche and R. Loris (2010). "Purification and crystallisation of Phd, the antitoxin of the *phd*/*doc* operon." Acta Crystallogr Sect F Struct Biol Cryst Commun 66(Pt 2): 167-171.

Garner, E. C., C. S. Campbell and R. D. Mullins (2004). "Dynamic instability in a DNA-segregating prokaryotic actin homolog." Science 306(5698): 1021-1025.

Garner, E. C., C. S. Campbell, D. B. Weibel and R. D. Mullins (2007). "Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog." Science 315(5816): 1270-1274.

Gasiunas, G., R. Barrangou, P. Horvath and V. Siksnys (2012). "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria." Proc Natl Acad Sci U S A 109(39): E2579-2586.

Genobile, D., J. Gaston, G. F. Tallis, J. E. Gregory, J. M. Griffith, M. Valcanis, D. Lightfoot and J. A. Marshall (2004). "An outbreak of shigellosis in a child care centre." Commun Dis Intell Q Rep 28(2): 225-229.

Gerdes, K., S. K. Christensen and A. Lobner-Olesen (2005). "Prokaryotic toxin-antitoxin stress response loci." Nat Rev Microbiol 3(5): 371-382.

Gerdes, K., M. Howard and F. Szardenings (2010). "Pushing and pulling in prokaryotic DNA segregation." Cell 141(6): 927-942.

Gerdes, K. and E. Maisonneuve (2012). "Bacterial persistence and toxin-antitoxin loci." Annu Rev Microbiol 66: 103-123.

Gerdes, K., J. Moller-Jensen and R. Bugge Jensen (2000). "Plasmid and chromosome partitioning: surprises from phylogeny." Mol Microbiol 37(3): 455-466.

Gerdes, K. and S. Molin (1986). "Partitioning of plasmid R1. Structural and functional analysis of the parA locus." J Mol Biol 190(3): 269-279.

Getino, M., D. J. Sanabria-Rios, R. Fernandez-Lopez, J. Campos-Gomez, J. M. Sanchez-Lopez, A. Fernandez, N. M. Carballeira and F. de la Cruz (2015). "Synthetic Fatty Acids Prevent Plasmid-Mediated Horizontal Gene Transfer." mBio 6(5): e01032-01015.

Ghosh, A. R. and S. C. Sehgal (2001). "Survivability of *Shigella dysenteriae* type 1 and *S. flexneri* 2a in laboratory conditions simulating aquatic environment." Indian J Med Res 114: 199-206.

Gilbart, V. L., I. Simms, C. Jenkins, M. Furegato, M. Gobin, I. Oliver, G. Hart, O. N. Gill and G. Hughes (2015). "Sex, drugs and smart phone applications: findings from semistructured interviews with men who have sex with men diagnosed with *Shigella flexneri* 3a in England and Wales." Sex Transm Infect 91(8): 598-602.

Girardin, S. E., R. Tournebize, M. Mavris, A. L. Page, X. Li, G. R. Stark, J. Bertin, P. S. DiStefano, M. Yaniv, P. J. Sansonetti and D. J. Philpott (2001). "CARD4/Nod1 mediates NF-kappaB and JNK activation by invasive *Shigella flexneri*." EMBO Rep 2(8): 736-742.

Glansdorff, N., D. Charlier and M. Zafarullah (1981). "Activation of gene expression by IS2 and IS3." Cold Spring Harb Symp Quant Biol 45 Pt 1: 153-156.

Gomaa, A. A., H. E. Klumpe, M. L. Luo, K. Selle, R. Barrangou and C. L. Beisel (2014). "Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems." mBio 5(1): e00928-00913.

Gorden, J. and P. L. Small (1993). "Acid resistance in enteric bacteria." Infect Immun 61(1): 364-367.

Gray, M. D., K. A. Lampel, N. A. Strockbine, R. E. Fernandez, A. R. Melton-Celsa and A. T. Maurelli (2014). "Clinical isolates of Shiga toxin 1a-producing *Shigella flexneri* with an epidemiological link to recent travel to Hispañiola." Emerg Infect Dis 20(10): 1669-1677.

Guo, X., Y. Wang, G. Duan, Z. Xue, L. Wang, P. Wang, S. Qiu, Y. Xi and H. Yang (2015). "Detection and analysis of CRISPRs of Shigella." Curr Microbiol 70(1): 85-90.

Haider, K., A. Hossain, C. Wanke, F. Qadri, S. Ali and S. Nahar (1993). "Production of mucinase and neuraminidase and binding of *Shigella* to intestinal mucin." J Diarrhoeal Dis Res 11(2): 88-92.

Hale, T. L. and G. T. Keusch (1996). *Shigella*. University of Texas Medical Branch at Galveston.

Haniford, D. B. (2006). "Transpososome dynamics and regulation in Tn10 transposition." Crit Rev Biochem Mol Biol 41(6): 407-424.

Harms, A., D. E. Brodersen, N. Mitarai and K. Gerdes (2018). "Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology." Mol Cell 70(5): 768-784.

Hawkey, J., J. M. Monk, H. Billman-Jacobe, B. Palsson and K. E. Holt (2020). "Impact of insertion sequences on convergent evolution of Shigella species." PLoS Genet 16(7): e1008931.

Hayes, F. and B. Kedzierska (2014). "Regulating toxin-antitoxin expression: controlled detonation of intracellular molecular timebombs." Toxins (Basel) 6(1): 337-358.

Helaine, S. and E. Kugelberg (2014). "Bacterial persisters: formation, eradication, and experimental systems." Trends Microbiol 22(7): 417-424.

Heringa S. D., M. J. D., Herrick J. B. (2007). "A simple, rapid method for extracting large plasmid DNA from bacteria." Nature Procedings.

Holt, K. E., S. Baker, F. X. Weill, E. C. Holmes, A. Kitchen, J. Yu, V. Sangal, D. J. Brown, J. E. Coia, D. W. Kim, S. Y. Choi, S. H. Kim, W. D. da Silveira, D. J. Pickard, J. J. Farrar, J. Parkhill, G. Dougan and N. R. Thomson (2012). "*Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe." Nat Genet 44(9): 1056-1059.

Hu, L., A. G. Vecchiarelli, K. Mizuuchi, K. C. Neuman and J. Liu (2017). "Brownian Ratchet Mechanism for Faithful Segregation of Low-Copy-Number Plasmids." Biophys J 112(7): 1489-1502.

Huan, P. T., D. A. Bastin, B. L. Whittle, A. A. Lindberg and N. K. Verma (1997). "Molecular characterisation of the genes involved in O-antigen modification, attachment, integration and excision in *Shigella flexneri* bacteriophage SfV." Gene 195(2): 217-227.

Huguet, K. T., M. Gonnet, B. Doublet and A. Cloeckaert (2016). "A toxin antitoxin system promotes the maintenance of the IncA/C-mobilizable *Salmonella* Genomic Island 1." Sci Rep 6: 32285.

Ingersoll, M., E. A. Groisman and A. Zychlinsky (2002). "Pathogenicity islands of *Shigella*." Curr Top Microbiol Immunol 264(1): 49-65.

Ito, H., N. Kido, Y. Arakawa, M. Ohta, T. Sugiyama and N. Kato (1991). "Possible mechanisms underlying the slow lactose fermentation phenotype in *Shigella* spp." Appl Environ Microbiol 57(10): 2912-2917.

Jensen, R. B. and K. Gerdes (1997). "Partitioning of plasmid R1. The ParM protein exhibits ATPase activity and interacts with the centromere-like ParR-parC complex." J Mol Biol 269(4): 505-513.

Jeong, H. J., E. S. Jang, B. I. Han, K. H. Lee, M. S. Ock, H. H. Kong, D. I. Chung, S. Y. Seol, D. T. Cho and H. S. Yu (2007). "*Acanthamoeba*: could it be an environmental host of *Shigella*?" Exp Parasitol 115(2): 181-186.

Jiang, W., D. Bikard, D. Cox, F. Zhang and L. A. Marraffini (2013). "RNA-guided editing of bacterial genomes using CRISPR-Cas systems." Nat Biotechnol 31(3): 233-239.

Jiang, Y., F. Yang, X. Zhang, J. Yang, L. Chen, Y. Yan, H. Nie, Z. Xiong, J. Wang, J. Dong, Y. Xue, X. Xu, Y. Zhu, S. Chen and Q. Jin (2005). "The complete sequence and analysis of the large virulence plasmid pSS of *Shigella sonnei*." Plasmid 54(2): 149-159.

Jin, Q., Z. Yuan, J. Xu, Y. Wang, Y. Shen, W. Lu, J. Wang, H. Liu, J. Yang, F. Yang, X. Zhang, J. Zhang, G. Yang, H. Wu, D. Qu, J. Dong, L. Sun, Y. Xue, A. Zhao, Y. Gao, J. Zhu, B. Kan, K. Ding, S. Chen, H. Cheng, Z. Yao, B. He, R. Chen, D. Ma, B. Qiang, Y. Wen, Y. Hou and J. Yu (2002). "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157." Nucleic Acids Res 30(20): 4432-4441.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science 337(6096): 816-821.

Jurenas, D., A. Garcia-Pino and L. Van Melderen (2017). "Novel toxins from type II toxin-antitoxin systems with acetyltransferase activity." Plasmid 93: 30-35.

Kane, C. D., R. Schuch, W. A. Day, Jr. and A. T. Maurelli (2002). "MxiE regulates intracellular expression of factors secreted by the *Shigella flexneri* 2a type III secretion system." J Bacteriol 184(16): 4409-4419.

Kane, K. A. and C. J. Dorman (2012). "VirB-mediated positive feedback control of the virulence gene regulatory cascade of *Shigella flexneri*." J Bacteriol 194(19): 5264-5273.

Kaniga, K., I. Delor and G. R. Cornelis (1991). "A wide-host-range suicide vector for improving reverse genetics in gram-negative bacteria: inactivation of the *blaA* gene of *Yersinia enterocolitica*." Gene 109(1): 137-141.

Kanneganti, T. D., M. Lamkanfi and G. Núñez (2007). "Intracellular NOD-like receptors in host defense and disease." Immunity 27(4): 549-559.

Kasper, C. A., I. Sorg, C. Schmutz, T. Tschon, H. Wischnewski, M. L. Kim and C. Arrieumerlou (2010). "Cell-cell propagation of NF-κB transcription factor and MAP kinase activation amplifies innate immunity against bacterial infection." Immunity 33(5): 804-816.

Keddy, K. H., A. Sooka, P. Crowther-Gibson, V. Quan, S. Meiring, C. Cohen, T. Nana, C. Sriruttan, S. Seetharam, A. Hoosen, P. Naicker, E. Elliott, S. Haffejee, A. Whitelaw, K. P. Klugman, R. Group for Enteric and A. Meningeal Disease Surveillance in South (2012). "Systemic shigellosis in South Africa." Clin Infect Dis 54(10): 1448-1454.

Khan, W. A., J. K. Griffiths and M. L. Bennish (2013). "Gastrointestinal and extra-intestinal manifestations of childhood shigellosis in a region where all four species of Shigella are endemic." PLoS One 8(5): e64097.

Kim, D. W., H. Chu, D. H. Joo, M. S. Jang, J. H. Choi, S. M. Park, Y. J. Choi, S. H. Han and C. H. Yun (2008). "OspF directly attenuates the activity of extracellular signal-regulated kinase during invasion by *Shigella flexneri* in human dendritic cells." Mol Immunol 45(11): 3295-3301.

Kim, D. W., G. Lenzen, A. L. Page, P. Legrain, P. J. Sansonetti and C. Parsot (2005). "The *Shigella flexneri* effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes." Proc Natl Acad Sci U S A 102(39): 14046-14051.

Kim, J. S., D. H. Cho, M. Park, W. J. Chung, D. Shin, K. S. Ko and D. H. Kweon (2016). "CRISPR/Cas9-Mediated Re-Sensitisation of Antibiotic-Resistant *Escherichia coli* Harboring Extended-Spectrum beta-Lactamases." J Microbiol Biotechnol 26(2): 394-401.

Kim, Y. J., S. G. Yeo, J. H. Park and H. J. Ko (2013). "*Shigella* vaccine development: prospective animal models and current status." Curr Pharm Biotechnol 14(10): 903-912.

Knott, G. J. and J. A. Doudna (2018). "CRISPR-Cas guides the future of genetic engineering." Science 361(6405): 866-869.

Köhler, H., S. P. Rodrigues and B. A. McCormick (2002). "*Shigella flexneri* Interactions with the Basolateral Membrane Domain of Polarised Model Intestinal Epithelium: Role of

Lipopolysaccharide in Cell Invasion and in Activation of the Mitogen-Activated Protein Kinase ERK." Infect Immun 70(3): 1150-1158.

Koizumi, Y., C. Toma, N. Higa, T. Nohara, N. Nakasone and T. Suzuki (2012). "Inflammasome activation via intracellular NLRs triggered by bacterial infection." Cell Microbiol 14(2): 149-154.

Kopecko, D. J., O. Washington and S. B. Formal (1980). "Genetic and physical evidence for plasmid control of *Shigella sonnei* form I cell surface antigen." Infect Immun 29(1): 207-214.

Kotloff, K. L., J. P. Nataro, W. C. Blackwelder, D. Nasrin, T. H. Farag, S. Panchalingam, Y. Wu, S. O. Sow, D. Sur, R. F. Breiman, A. S. Faruque, A. K. Zaidi, D. Saha, P. L. Alonso, B. Tamboura, D. Sanogo, U. Onwuchekwa, B. Manna, T. Ramamurthy, S. Kanungo, J. B. Ochieng, R. Omore, J. O. Oundo, A. Hossain, S. K. Das, S. Ahmed, S. Qureshi, F. Quadri, R. A. Adegbola, M. Antonio, M. J. Hossain, A. Akinsola, I. Mandomando, T. Nhampossa, S. Acacio, K. Biswas, C. E. O'Reilly, E. D. Mintz, L. Y. Berkeley, K. Muhsen, H. Sommerfelt, R. M. Robins-Browne and M. M. Levine (2013). "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 382(9888): 209-222.

Kotloff, K. L., M. S. Riddle, J. A. Platts-Mills, P. Pavlinac and A. K. M. Zaidi (2018). "Shigellosis." Lancet 391(10122): 801-812.

Kotloff, K. L., J. P. Winickoff, B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak and M. M. Levine (1999). "Global burden of Shigella infections: implications for vaccine development and implementation of control strategies." Bull World Health Organ 77(8): 651-666.

Kumar, P., Q. Luo, T. J. Vickers, A. Sheikh, W. G. Lewis and J. M. Fleckenstein (2014). "EatA, an immunogenic protective antigen of enterotoxigenic Escherichia coli, degrades intestinal mucin." Infect Immun 82(2): 500-508.

Lai, V., L. Wang and P. R. Reeves (1998). "*Escherichia coli* clone *Sonnei* (*Shigella sonnei*) had a chromosomal O-antigen gene cluster prior to gaining its current plasmid-borne O-antigen genes." J Bacteriol 180(11): 2983-2986.

Lan, R., B. Lumb, D. Ryan and P. R. Reeves (2001). "Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli*." Infect Immun 69(10): 6303-6309.

Lan, R. and P. R. Reeves (2002). "*Escherichia coli* in disguise: molecular origins of *Shigella*." Microbes Infect 4(11): 1125-1132.

Lauritsen, I., A. Porse, M. O. A. Sommer and M. H. H. Norholm (2017). "A versatile one-step CRISPR-Cas9 based approach to plasmid-curing." Microb Cell Fact 16(1): 135.

Lawrence, J. G., R. W. Hendrix and S. Casjens (2001). "Where are the pseudogenes in bacterial genomes?" Trends Microbiol 9(11): 535-540.

Lee, H., T. G. Doak, E. Popodi, P. L. Foster and H. Tang (2016). "Insertion sequence-caused large-scale rearrangements in the genome of *Escherichia coli*." Nucleic Acids Res 44(15): 7109-7119.

Lee, K. Y. and B. J. Lee (2016). "Structure, Biology, and Therapeutic Application of Toxin-Antitoxin Systems in Pathogenic Bacteria." Toxins (Basel) 8(10).

Levine, B., N. Mizushima and H. W. Virgin (2011). "Autophagy in immunity and inflammation." Nature 469(7330): 323-335.

Levine, M. M., K. L. Kotloff, E. M. Barry, M. F. Pasetti and M. B. Sztein (2007). "Clinical trials of *Shigella* vaccines: two steps forward and one step back on a long, hard road." Nat Rev Microbiol 5(7): 540-553.

Light, J., E. Riise and S. Molin (1985). "Transcription and its regulation in the basic replicon region of plasmid R1." Mol Gen Genet 198(3): 503-508.

Lindberg, A. A., J. E. Brown, N. Strömberg, M. Westling-Ryd, J. E. Schultz and K. A. Karlsson (1987). "Identification of the carbohydrate receptor for Shiga toxin produced by *Shigella dysenteriae* type 1." J Biol Chem 262(4): 1779-1785.

Liu, G., G. Pilla and C. M. Tang (2019). "*Shigella* host: Pathogen interactions: Keeping bacteria in the loop." Cell Microbiol 21(11): e13062.

Liu, G., S. E. Beaton, A. G. Grieve, R. Evans, M. Rogers, K. Strisovsky, F. A. Armstrong, M. Freeman, R. M. Exley and C. M. Tang (2020). "Bacterial rhomboid proteases mediate quality control of orphan membrane proteins." EMBO J 39(10): e102922.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods 25(4): 402-408.

Livio, S., N. A. Strockbine, S. Panchalingam, S. M. Tennant, E. M. Barry, M. E. Marohn, M. Antonio, A. Hossain, I. Mandomando, J. B. Ochieng, J. O. Oundo, S. Qureshi, T. Ramamurthy, B. Tamboura, R. A. Adegbola, M. J. Hossain, D. Saha, S. Sen, A. S. Faruque, P. L. Alonso, R. F. Breiman, A. K. Zaidi, D. Sur, S. O. Sow, L. Y. Berkeley, C. E. O'Reilly, E. D. Mintz, K. Biswas, D. Cohen, T. H. Farag, D. Nasrin, Y. Wu, W. C. Blackwelder, K. L. Kotloff, J. P. Nataro and M. M. Levine (2014). "Shigella isolates from the global enteric multicenter study inform vaccine development." Clin Infect Dis 59(7): 933-941.

Lobato-Márquez, D., L. Molina-García, I. Moreno-Córdoba, F. García-Del Portillo and R. Díaz-Orejas (2016). "Stabilisation of the Virulence Plasmid pSLT of *Salmonella Typhimurium* by Three Maintenance Systems and Its Evaluation by Using a New Stability Test." Front Mol Biosci 3: 66.

Lobocka, M. and M. Yarmolinsky (1996). "P1 plasmid partition: a mutational analysis of ParB." J Mol Biol 259(3): 366-382.

Luck, S. N., S. A. Turner, K. Rajakumar, H. Sakellaris and B. Adler (2001). "Ferric dicitrate transport system (Fec) of *Shigella flexneri* 2a YSH6000 is encoded on a novel pathogenicity island carrying multiple antibiotic resistance genes." Infect Immun 69(10): 6012-6021.

Madl, T., L. Van Melderen, N. Mine, M. Respondek, M. Oberer, W. Keller, L. Khatai and K. Zangger (2006). "Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA." J Mol Biol 364(2): 170-185.

Mahillon, J. and M. Chandler (1998). "Insertion sequences." Microbiol Mol Biol Rev 62(3): 725-774.

Mahoney, F. J., T. A. Farley, D. F. Burbank, N. H. Leslie and L. M. McFarland (1993). "Evaluation of an intervention program for the control of an outbreak of shigellosis among institutionalised persons." J Infect Dis 168(5): 1177-1180.

Maier, B. R. and D. J. Hentges (1972). "Experimental Shigella infections in laboratory animals. I. Antagonism by human normal flora components in gnotobiotic mice." Infect Immun 6(2): 168-173.

Makarova, K. S., D. H. Haft, R. Barrangou, S. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost and E. V. Koonin (2011). "Evolution and classification of the CRISPR-Cas systems." Nat Rev Microbiol 9(6): 467-477.

Mali, P., K. M. Esvelt and G. M. Church (2013). "Cas9 as a versatile tool for engineering biology." Nat Methods 10(10): 957-963.

Malmgren, C., H. M. Engdahl, P. Romby and E. G. Wagner (1996). "An antisense/target RNA duplex or a strong intramolecular RNA structure 5' of a translation initiation signal blocks ribosome binding: the case of plasmid R1." RNA 2(10): 1022-1032.

Mani, S., T. Wierzba and R. I. Walker (2016). "Status of vaccine research and development for Shigella." Vaccine 34(26): 2887-2894.

Marraffini, L. A. and E. J. Sontheimer (2008). "CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA." Science 322(5909): 1843-1845.

Marteyn, B., A. Gazi and P. Sansonetti (2012). "*Shigella*: a model of virulence regulation in vivo." Gut Microbes 3(2): 104-120.

Martinic, M., A. Hoare, I. Contreras and S. A. Alvarez (2011). "Contribution of the lipopolysaccharide to resistance of *Shigella flexneri* 2a to extreme acidity." PLoS One 6(10): e25557.

Martino, M. C., G. Rossi, I. Martini, I. Tattoli, D. Chiavolini, A. Phalipon, P. J. Sansonetti and M. L. Bernardini (2005). "Mucosal lymphoid infiltrate dominates colonic pathological changes in murine experimental shigellosis." J Infect Dis 192(1): 136-148.

Maurelli, A. T., B. Blackmon and R. Curtiss, 3rd (1984). "Temperature-dependent expression of virulence genes in Shigella species." Infect Immun 43(1): 195-201.

Maurelli, A. T., R. E. Fernandez, C. A. Bloch, C. K. Rode and A. Fasano (1998). ""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 95(7): 3943-3948.

Mavris, M., P. J. Sansonetti and C. Parsot (2002). "Identification of the cis-acting site involved in activation of promoters regulated by activity of the type III secretion apparatus in *Shigella flexneri*." J Bacteriol 184(24): 6751-6759.

Maxwell, A. and M. Gellert (1984). "The DNA dependence of the ATPase activity of DNA gyrase." J Biol Chem 259(23): 14472-14480.

McCall, B., R. Stafford, S. Cherian, K. Heel, H. Smith, N. Corones and S. Gilmore (2000). "An outbreak of multi-resistant *Shigella sonnei* in a long-stay geriatric nursing centre." Commun Dis Intell 24(9): 272-275.

McDonough, M. A. and J. R. Butterton (1999). "Spontaneous tandem amplification and deletion of the *shiga* toxin operon in *Shigella dysenteriae* 1." Mol Microbiol 34(5): 1058-1069.

McGinn, J. and L. A. Marraffini (2019). "Molecular mechanisms of CRISPR-Cas spacer acquisition." Nat Rev Microbiol 17(1): 7-12.

McGuckin, M. A., S. K. Linden, P. Sutton and T. H. Florin (2011). "Mucin dynamics and enteric pathogens." Nat Rev Microbiol 9(4): 265-278.

McVicker, G., S. Hollingshead, G. Pilla and C. M. Tang (2019). "Maintenance of the virulence plasmid in *Shigella flexneri* is influenced by Lon and two functional partitioning systems." Mol Microbiol 111(5): 1355-1366.

McVicker, G. and C. M. Tang (2016). "Deletion of toxin-antitoxin systems in the evolution of *Shigella sonnei* as a host-adapted pathogen." Nat Microbiol 2: 16204.

Menard, R., P. Sansonetti and C. Parsot (1994). "The secretion of the *Shigella flexneri* Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD." EMBO J 13(22): 5293-5302.

Menzel, R. and M. Gellert (1987). "Modulation of transcription by DNA supercoiling: a deletion analysis of the *Escherichia coli gyrA* and *gyrB* promoters." Proc Natl Acad Sci U S A 84(12): 4185-4189.

Mills, J. A., M. M. Venkatesan, L. S. Baron and J. M. Buysse (1992). "Spontaneous insertion of an IS1-like element into the *virF* gene is responsible for avirulence in opaque colonial variants of *Shigella flexneri* 2a." Infect Immun 60(1): 175-182.

Mira, A., H. Ochman and N. A. Moran (2001). "Deletional bias and the evolution of bacterial genomes." Trends Genet 17(10): 589-596.

Mizuuchi, K. (1992). "Transpositional recombination: mechanistic insights from studies of mu and other elements." Annu Rev Biochem 61: 1011-1051.

Mizuuchi, K. (1992). "Polynucleotidyl transfer reactions in transpositional DNA recombination." J Biol Chem 267(30): 21273-21276.

Mohan, K., M. Hibbert, G. Rooney, M. Canvin, T. Childs, C. Jenkins, I. Simms, P. Kirwan, V. Delpech, Z. Yin, G. Hughes and N. Field (2018). "What is the overlap between HIV and shigellosis epidemics in England: further evidence of MSM transmission?" Sex Transm Infect 94(1): 67-71.

Mohle-Boetani, J. C., M. Stapleton, R. Finger, N. H. Bean, J. Poundstone, P. A. Blake and P. M. Griffin (1995). "Communitywide shigellosis: control of an outbreak and risk factors in child day-care centers." Am J Public Health 85(6): 812-816.

Moller-Jensen, J., R. B. Jensen, J. Lowe and K. Gerdes (2002). "Prokaryotic DNA segregation by an actin-like filament." EMBO J 21(12): 3119-3127.

Moran, N. A. and G. R. Plague (2004). "Genomic changes following host restriction in bacteria." Curr Opin Genet Dev 14(6): 627-633.

Mounier, J., T. Vasselon, R. Hellio, M. Lesourd and P. J. Sansonetti (1992). "*Shigella flexneri* enters human colonic Caco-2 epithelial cells through the basolateral pole." Infect Immun 60(1): 237-248.

Munoz-Lopez, M. and J. L. Garcia-Perez (2010). "DNA transposons: nature and applications in genomics." Curr Genomics 11(2): 115-128.

Murphy, E. R. and S. M. Payne (2007). "RyhB, an iron-responsive small RNA molecule, regulates *Shigella dysenteriae* virulence." Infect Immun 75(7): 3470-3477.

Nakata, N., T. Tobe, I. Fukuda, T. Suzuki, K. Komatsu, M. Yoshikawa and C. Sasakawa (1993). "The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the *ompT* and *kcpA* loci." Mol Microbiol 9(3): 459-468.

Nakayama, S. and H. Watanabe (1995). "Involvement of cpxA, a sensor of a two-component regulatory system, in the pH-dependent regulation of expression of Shigella sonnei virF gene." J Bacteriol 177(17): 5062-5069.

Nataro, J. P., J. Seriwatana, A. Fasano, D. R. Maneval, L. D. Guers, F. Noriega, F. Dubovsky, M. M. Levine and J. G. Morris, Jr. (1995). "Identification and cloning of a novel plasmid-encoded enterotoxin of enteroinvasive *Escherichia coli* and *Shigella* strains." Infect Immun 63(12): 4721-4728.

Neutra, M. R., N. J. Mantis, A. Frey and P. J. Giannasca (1999). "The composition and function of M cell apical membranes: implications for microbial pathogenesis." Semin Immunol 11(3): 171-181.

Ng, L. G., A. Hsu, M. A. Mandell, B. Roediger, C. Hoeller, P. Mrass, A. Iparraguirre, L. L. Cavanagh, J. A. Triccas, S. M. Beverley, P. Scott and W. Weninger (2008). "Migratory dermal dendritic cells act as rapid sensors of protozoan parasites." PLoS Pathog 4(11): e1000222.

Nigro, G., L. L. Fazio, M. C. Martino, G. Rossi, I. Tattoli, V. Liparoti, C. De Castro, A. Molinaro, D. J. Philpott and M. L. Bernardini (2008). "Muramylpeptide shedding modulates cell sensing of *Shigella flexneri*." Cell Microbiol 10(3): 682-695.

Niyogi, S. K. (2005). "Shigellosis." J Microbiol 43(2): 133-143.

Nordstrom, K. (2006). "Plasmid R1--replication and its control." Plasmid 55(1): 1-26.

Ochman, H., T. S. Whittam, D. A. Caugant and R. K. Selander (1983). "Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*." J Gen Microbiol 129(9): 2715-2726.

Ogawa, M. and C. Sasakawa (2009). "[*Shigella* strategy for host defense and mucosal infection]." Tanpakushitsu Kakusan Koso 54(8 Suppl): 988-995.

Ogawa, M., T. Yoshimori, T. Suzuki, H. Sagara, N. Mizushima and C. Sasakawa (2005). "Escape of intracellular *Shigella* from autophagy." Science 307(5710): 727-731.

Ogura, T. and S. Hiraga (1983). "Mini-F plasmid genes that couple host cell division to plasmid proliferation." Proc Natl Acad Sci U S A 80(15): 4784-4788.

Onodera, N. T., J. Ryu, T. Durbic, C. Nislow, J. M. Archibald and J. R. Rohde (2012). "Genome sequence of *Shigella flexneri* serotype 5a strain M90T Sm." J Bacteriol 194(11): 3022.

Orlova, A., E. C. Garner, V. E. Galkin, J. Heuser, R. D. Mullins and E. H. Egelman (2007). "The structure of bacterial ParM filaments." Nat Struct Mol Biol 14(10): 921-926.

Overgaard, M., J. Borch, M. G. Jorgensen and K. Gerdes (2008). "Messenger RNA interferase RelE controls *relBE* transcription by conditional cooperativity." Mol Microbiol 69(4): 841-857.

Page, A. L., M. Fromont-Racine, P. Sansonetti, P. Legrain and C. Parsot (2001). "Characterisation of the interaction partners of secreted proteins and chaperones of *Shigella flexneri*." Mol Microbiol 42(4): 1133-1145.

Page, R. and W. Peti (2016). "Toxin-antitoxin systems in bacterial growth arrest and persistence." Nat Chem Biol 12(4): 208-214.

Papayannopoulos, V., K. D. Metzler, A. Hakkim and A. Zychlinsky (2010). "Neutrophil elastase and myeloperoxidase regulate the formation of neutrophil extracellular traps." J Cell Biol 191(3): 677-691.

Parsot, C. (2009). "Shigella type III secretion effectors: how, where, when, for what purposes?" Curr Opin Microbiol 12(1): 110-116.

Parsot, C., E. Ageron, C. Penno, M. Mavris, K. Jamoussi, H. d'Hauteville, P. Sansonetti and B. Demers (2005). "A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in *Shigella flexneri*." Mol Microbiol 56(6): 1627-1635.

Parsot, C., C. Hamiaux and A. L. Page (2003). "The various and varying roles of specific chaperones in type III secretion systems." Curr Opin Microbiol 6(1): 7-14.

Parsot, C., R. Menard, P. Gounon and P. J. Sansonetti (1995). "Enhanced secretion through the *Shigella flexneri* Mxi-Spa translocon leads to assembly of extracellular proteins into macromolecular structures." Mol Microbiol 16(2): 291-300.

Payne, S. M., E. E. Wyckoff, E. R. Murphy, A. G. Oglesby, M. L. Boulette and N. M. Davies (2006). "Iron and pathogenesis of *Shigella*: iron acquisition in the intracellular environment." Biometals 19(2): 173-180.

Payne, S. M. and R. A. Finkelstein (1977). "Detection and differentiation of iron-responsive avirulent mutants on Congo red agar." Infect Immun 18(1): 94-98.

Pecota, D. C. and T. K. Wood (1996). "Exclusion of T4 phage by the *hok*/*sok* killer locus from plasmid R1." J Bacteriol 178(7): 2044-2050.

Perdomo, O. J., J. M. Cavaillon, M. Huerre, H. Ohayon, P. Gounon and P. J. Sansonetti (1994). "Acute inflammation causes epithelial invasion and mucosal destruction in experimental shigellosis." J Exp Med 180(4): 1307-1319.

Perry, R. D., S. C. Straley, J. D. Fetherston, D. J. Rose, J. Gregor and F. R. Blattner (1998). "DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1 of *Yersinia pestis* KIM5." Infect Immun 66(10): 4611-4623.

Phalipon, A. and P. J. Sansonetti (2007). "*Shigella*'s ways of manipulating the host intestinal innate and adaptive immune system: a tool box for survival?" Immunol Cell Biol 85(2): 119-129.

Pilla, G., G. McVicker and C. M. Tang (2017). "Genetic plasticity of the *Shigella* virulence plasmid is mediated by intra- and inter-molecular events between insertion sequences." PLoS Genet 13(9): e1007014.

Pilla, G. and C. M. Tang (2018). "Going around in circles: virulence plasmids in enteric pathogens." Nat Rev Microbiol 16(8): 484-495.

Pinto, U. M., K. M. Pappas and S. C. Winans (2012). "The ABCs of plasmid replication and segregation." Nat Rev Microbiol 10(11): 755-765.

Pongpech, P., D. J. Hentges, W. W. Marsh and M. E. Eberle (1989). "Effect of streptomycin administration on association of enteric pathogens with cecal tissue of mice." Infect Immun 57(7): 2092-2097.

Popp, D., A. Narita, T. Oda, T. Fujisawa, H. Matsuo, Y. Nitanai, M. Iwasa, K. Maeda, H. Onishi and Y. Maeda (2008). "Molecular structure of the ParM polymer and the mechanism leading to its nucleotide-driven dynamic instability." EMBO J 27(3): 570-579.

Porter, M. E. and C. J. Dorman (1994). "A role for H-NS in the thermo-osmotic regulation of virulence gene expression in *Shigella flexneri*." J Bacteriol 176(13): 4187-4191.

Porter, M. E. and C. J. Dorman (2002). "In vivo DNA-binding and oligomerisation properties of the Shigella flexneri AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis." J Bacteriol 184(2): 531-539.

Postow, L., C. D. Hardy, J. Arsuaga and N. R. Cozzarelli (2004). "Topological domain structure of the Escherichia coli chromosome." Genes Dev 18(14): 1766-1779.

Prentki, P., B. Teter, M. Chandler and D. J. Galas (1986). "Functional promoters created by the insertion of transposable element IS1." J Mol Biol 191(3): 383-393.

Prosseda, G., M. Falconi, M. Giangrossi, C. O. Gualerzi, G. Micheli and B. Colonna (2004). "The virF promoter in Shigella: more than just a curved DNA stretch." Mol Microbiol 51(2): 523-537.

Prosseda, G., P. A. Fradiani, M. Di Lorenzo, M. Falconi, G. Micheli, M. Casalino, M. Nicoletti and B. Colonna (1998). "A role for H-NS in the regulation of the virF gene of Shigella and enteroinvasive *Escherichia coli*." Res Microbiol 149(1): 15-25.

Prosseda, G., A. Mazzola, M. L. Di Martino, D. Tielker, G. Micheli and B. Colonna (2010). "A temperature-induced narrow DNA curvature range sustains the maximum activity of a bacterial promoter in vitro." Biochemistry 49(13): 2778-2785.

Prunier, A. L., R. Schuch, R. E. Fernandez, K. L. Mumy, H. Kohler, B. A. McCormick and A. T. Maurelli (2007). "*nadA* and *nadB* of *Shigella flexneri* 5a are antivirulence loci responsible for the synthesis of quinolinate, a small molecule inhibitor of Shigella pathogenicity." Microbiology 153(Pt 7): 2363-2372.

Pupo, G. M., D. K. Karaolis, R. Lan and P. R. Reeves (1997). "Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies." Infect Immun 65(7): 2685-2692.

Purdy, G. E. and S. M. Payne (2001). "The SHI-3 iron transport island of *Shigella boydii* 0-1392 carries the genes for aerobactin synthesis and transport." J Bacteriol 183(14): 4176-4182.

Pursey, E., D. Sunderhauf, W. H. Gaze, E. R. Westra and S. van Houte (2018). "CRISPR-Cas antimicrobials: Challenges and future prospects." PLoS Pathog 14(6): e1006990.

Puzari, M., M. Sharma and P. Chetia (2018). "Emergence of antibiotic resistant *Shigella* species: A matter of concern." J Infect Public Health 11(4): 451-454.

Radnedge, L., M. A. Davis, B. Youngren and S. J. Austin (1997). "Plasmid maintenance functions of the large virulence plasmid of *Shigella flexneri*." J Bacteriol 179(11): 3670-3675.

Ramos, H. C., M. Rumbo and J. C. Sirard (2004). "Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa." Trends Microbiol 12(11): 509-517.

Ranjbar, R. and A. Farahani (2019). "*Shigella*: Antibiotic-Resistance Mechanisms And New Horizons For Treatment." Infect Drug Resist 12: 3137-3167.

Reyes-Lamothe, R., E. Nicolas and D. J. Sherratt (2012). "Chromosome replication and segregation in bacteria." Annu Rev Genet 46: 121-143.

Rhoads, A. and K. F. Au (2015). "PacBio Sequencing and Its Applications." Genomics Proteomics Bioinformatics 13(5): 278-289.

Ringgaard, S., G. Ebersbach, J. Borch and K. Gerdes (2007). "Regulatory cross-talk in the double par locus of plasmid pB171." J Biol Chem 282(5): 3134-3145.

Ripoll-Rozada, J., Y. Garcia-Cazorla, M. Getino, C. Machon, D. Sanabria-Rios, F. de la Cruz, E. Cabezon and I. Arechaga (2016). "Type IV traffic ATPase TrwD as molecular target to inhibit bacterial conjugation." Mol Microbiol 100(5): 912-921.

Robson, J., J. L. McKenzie, R. Cursons, G. M. Cook and V. L. Arcus (2009). "The *vapBC* operon from *Mycobacterium smegmatis* is an autoregulated toxin-antitoxin module that controls growth via inhibition of translation." J Mol Biol 390(3): 353-367.

Rodrigues, M., S. W. McBride, K. Hullahalli, K. L. Palmer and B. A. Duerkop (2019). "Conjugative Delivery of CRISPR-Cas9 for the Selective Depletion of Antibiotic-Resistant Enterococci." Antimicrob Agents Chemother 63(11).

Rolland, K., N. Lambert-Zechovsky, B. Picard and E. Denamur (1998). "*Shigella* and enteroinvasive *Escherichia coli* strains are derived from distinct ancestral strains of E. coli." Microbiology 144 (Pt 9): 2667-2672.

Sack, D. A., A. T. Hoque, A. Huq and M. Etheridge (1994). "Is protection against shigellosis induced by natural infection with Plesiomonas shigelloides?" Lancet 343(8910): 1413-1415.

Saeed, A., D. Johansson, G. Sandstrom and H. Abd (2012). "Temperature Depended Role of *Shigella flexneri* Invasion Plasmid on the Interaction with *Acanthamoeba castellanii*." Int J Microbiol 2012: 917031.

Saha, T., M. Murhekar, Y. J. Hutin and T. Ramamurthy (2009). "An urban, water-borne outbreak of diarrhoea and shigellosis in a district town in eastern India." Natl Med J India 22(5): 237-239.

Sakellaris, H., N. K. Hannink, K. Rajakumar, D. Bulach, M. Hunt, C. Sasakawa and B. Adler (2000). "Curli loci of *Shigella* spp." Infect Immun 68(6): 3780-3783.

Salje, J., B. Zuber and J. Lowe (2009). "Electron cryomicroscopy of *E. coli* reveals filament bundles involved in plasmid DNA segregation." Science 323(5913): 509-512.

Sander, J. D. and J. K. Joung (2014). "CRISPR-Cas systems for editing, regulating and targeting genomes." Nat Biotechnol 32(4): 347-355.

Sani, M., A. Allaoui, F. Fusetti, G. T. Oostergetel, W. Keegstra and E. J. Boekema (2007). "Structural organisation of the needle complex of the type III secretion apparatus of *Shigella flexneri*." Micron 38(3): 291-301.

Sansonetti, P. J., J. Arondel, M. Huerre, A. Harada and K. Matsushima (1999). "Interleukin-8 controls bacterial transepithelial translocation at the cost of epithelial destruction in experimental shigellosis." Infect Immun 67(3): 1471-1480.

Sansonetti, P. J., T. L. Hale, G. J. Dammin, C. Kapfer, H. H. Collins, Jr. and S. B. Formal (1983). "Alterations in the pathogenicity of *Escherichia coli* K-12 after transfer of plasmid and chromosomal genes from *Shigella flexneri*." Infect Immun 39(3): 1392-1402.

Sansonetti, P. J., D. J. Kopecko and S. B. Formal (1981). "*Shigella sonnei* plasmids: evidence that a large plasmid is necessary for virulence." Infect Immun 34(1): 75-83.

Sansonetti, P. J., D. J. Kopecko and S. B. Formal (1982). "Involvement of a plasmid in the invasive ability of *Shigella flexneri*." Infect Immun 35(3): 852-860.

Sansonetti, P. J., A. Phalipon, J. Arondel, K. Thirumalai, S. Banerjee, S. Akira, K. Takeda and A. Zychlinsky (2000). "Caspase-1 activation of IL-1beta and IL-18 are essential for *Shigella flexneri*-induced inflammation." Immunity 12(5): 581-590.

Sasakawa, C., K. Kamata, T. Sakai, S. Y. Murayama, S. Makino and M. Yoshikawa (1986). "Molecular alteration of the 140-megadalton plasmid associated with loss of virulence and Congo red binding activity in *Shigella flexneri*." Infect Immun 51(2): 470-475.

Sayeed, S., T. Brendler, M. Davis, L. Reaves and S. Austin (2005). "Surprising dependence on postsegregational killing of host cells for maintenance of the large virulence plasmid of *Shigella flexneri*." J Bacteriol 187(8): 2768-2773.

Sayeed, S., L. Reaves, L. Radnedge and S. Austin (2000). "The stability region of the large virulence plasmid of *Shigella flexneri* encodes an efficient postsegregational killing system." J Bacteriol 182(9): 2416-2421.

Schroeder, G. N. and H. Hilbi (2008). "Molecular pathogenesis of Shigella spp.: controlling host cell signaling, invasion, and death by type III secretion." Clin Microbiol Rev 21(1): 134-156.

Schuch, R. and A. T. Maurelli (1997). "Virulence plasmid instability in *Shigella flexneri* 2a is induced by virulence gene expression." Infect Immun 65(9): 3686-3692.

Sergueev, K., A. Dabrazhynetskaya and S. Austin (2005). "Plasmid partition system of the P1par family from the pWR100 virulence plasmid of *Shigella flexneri*." J Bacteriol 187(10): 3369-3373.

Shepherd, J. G., L. Wang and P. R. Reeves (2000). "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 68(10): 6056-6061.

Shere, K. D., S. Sallustio, A. Manessis, T. G. D'Aversa and M. B. Goldberg (1997). "Disruption of IcsP, the major *Shigella* protease that cleaves IcsA, accelerates actin-based motility." Mol Microbiol 25(3): 451-462.

Shiga, K., (1898). "Ueber den Erreger der Dysenterie in Japan." Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene 23: 599-600.

Shuman, S. and M. S. Glickman (2007). "Bacterial DNA repair by non-homologous end joining." Nat Rev Microbiol 5(11): 852-861.

Siguier, P., E. Gourbeyre and M. Chandler (2014). "Bacterial insertion sequences: their genomic impact and diversity." FEMS Microbiol Rev 38(5): 865-891.

Simms, I., N. Field, C. Jenkins, T. Childs, V. L. Gilbart, T. J. Dallman, P. Mook, P. D. Crook and G. Hughes (2015). "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 20(15).

Sonne, C., (1938). "The bacteriology of atoxic dysentery bacilli (paradysentery bacilli)." Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene(75): 408-456.

Spies, M. K. C. (2005). Homologous Recombination by the RecBCD and RecF Pathways. The Bacterial Chromosome. H. N. P. Washington, D.C., ASM Press.

Stadnyk, A. W. (2002). "Intestinal epithelial cells as a source of inflammatory cytokines and chemokines." Can J Gastroenterol 16(4): 241-246.

Strober, W., P. J. Murray, A. Kitani and T. Watanabe (2006). "Signalling pathways and molecular interactions of NOD1 and NOD2." Nat Rev Immunol 6(1): 9-20.

Struelens, M. J., D. Patte, I. Kabir, A. Salam, S. K. Nath and T. Butler (1985). "*Shigella* septicemia: prevalence, presentation, risk factors, and outcome." J Infect Dis 152(4): 784-790.

Surtees, J. A. and B. E. Funnell (1999). "P1 ParB domain structure includes two independent multimerisation domains." J Bacteriol 181(19): 5898-5908.

Surtees, J. A. and B. E. Funnell (2001). "The DNA binding domains of P1 ParB and the architecture of the P1 plasmid partition complex." J Biol Chem 276(15): 12385-12394.

Szekeres, S., M. Dauti, C. Wilde, D. Mazel and D. A. Rowe-Magnus (2007). "Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection." Mol Microbiol 63(6): 1588-1605.

Tam, J. E. and B. C. Kline (1989). "The F plasmid *ccd* autorepressor is a complex of CcdA and CcdB proteins." Mol Gen Genet 219(1-2): 26-32.

Tamano, K., S. Aizawa, E. Katayama, T. Nonaka, S. Imajoh-Ohmi, A. Kuwae, S. Nagai and C. Sasakawa (2000). "Supramolecular structure of the *Shigella* type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors." EMBO J 19(15): 3876-3887.

Tanaka, K. H., S. Dallaire-Dufresne, R. K. Daher, M. Frenette and S. J. Charette (2012). "An insertion sequence-dependent plasmid rearrangement in *Aeromonas salmonicida* causes the loss of the type three secretion system." PLoS One 7(3): e33725.

Tashiro, Y., K. Kawata, A. Taniuchi, K. Kakinuma, T. May and S. Okabe (2012). "RelE-mediated dormancy is enhanced at high cell density in *Escherichia coli*." J Bacteriol 194(5): 1169-1176.

Tavakoli, N. P. and K. M. Derbyshire (2001). "Tipping the balance between replicative and simple transposition." EMBO J 20(11): 2923-2930.

The, H. C., D. P. Thanh, K. E. Holt, N. R. Thomson and S. Baker (2016). "The genomic signatures of *Shigella* evolution, adaptation and geographical spread." Nat Rev Microbiol 14(4): 235-250.

Thanh Duy, P., T. N. Thi Nguyen, D. Vu Thuy, H. Chung The, F. Alcock, C. Boinett, H. N. Dan Thanh, H. Thanh Tuyen, G. E. Thwaites, M. A. Rabaa and S. Baker (2020). "Commensal Escherichia coli are a reservoir for the transfer of XDR plasmids into epidemic fluoroquinolone-resistant Shigella sonnei." Nat Microbiol 5(2): 256-264.

Thompson, C. N., P. T. Duy and S. Baker (2015). "The Rising Dominance of *Shigella sonnei*: An Intercontinental Shift in the Etiology of Bacillary Dysentery." PLoS Negl Trop Dis 9(6): e0003708.

Tobe, T., M. Yoshikawa and C. Sasakawa (1995). "Thermoregulation of *virB* transcription in *Shigella flexneri* by sensing of changes in local DNA superhelicity." J Bacteriol 177(4): 1094-1097.

Tominaga, A., R. Lan and P. R. Reeves (2005). "Evolutionary changes of the *flhDC* flagellar master operon in *Shigella strains*." J Bacteriol 187(12): 4295-4302.

Ton-Hoang, B., C. Turlan and M. Chandler (2004). "Functional domains of the IS1 transposase: analysis in vivo and in vitro." Mol Microbiol 53(5): 1529-1543.

Tran Van Nhieu, G., A. Ben-Ze'ev and P. J. Sansonetti (1997). "Modulation of bacterial entry into epithelial cells by association between vinculin and the *Shigella* IpaA invasin." EMBO J 16(10): 2717-2729.

Tran Van Nhieu, G., C. Clair, R. Bruzzone, M. Mesnil, P. Sansonetti and L. Combettes (2003). "Connexin-dependent inter-cellular communication increases invasion and dissemination of *Shigella* in epithelial cells." Nat Cell Biol 5(8): 720-726.

Travassos, L. H., L. A. Carneiro, M. Ramjeet, S. Hussey, Y. G. Kim, J. G. Magalhães, L. Yuan, F. Soares, E. Chea, L. Le Bourhis, I. G. Boneca, A. Allaoui, N. L. Jones, G. Nuñez, S. E. Girardin and D. J. Philpott (2010). "Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry." Nat Immunol 11(1): 55-62.

Tripathi, A., P. C. Dewan, S. A. Siddique and R. Varadarajan (2014). "MazF-induced growth inhibition and persister generation in *Escherichia coli*." J Biol Chem 289(7): 4191-4205.

Tucker, D. L., N. Tucker and T. Conway (2002). "Gene expression profiling of the pH response in *Escherichia coli*." J Bacteriol 184(23): 6551-6558.

Turner, E. C. and C. J. Dorman (2007). "H-NS antagonism in *Shigella flexneri* by VirB, a virulence gene transcription regulator that is closely related to plasmid partition factors." J Bacteriol 189(9): 3403-3413.

Turner, S. A., S. N. Luck, H. Sakellaris, K. Rajakumar and B. Adler (2001). "Nested deletions of the SRL pathogenicity island of *Shigella flexneri* 2a." J Bacteriol 183(19): 5535-5543.

Valens, M., S. Penaud, M. Rossignol, F. Cornet and F. Boccard (2004). "Macrodomain organization of the Escherichia coli chromosome." EMBO J 23(21): 4330-4341.

van den Beld, M. J. and F. A. Reubsaet (2012). "Differentiation between *Shigella*, enteroinvasive *Escherichia coli* (EIEC) and noninvasive *Escherichia coli*." Eur J Clin Microbiol Infect Dis 31(6): 899-904.

van den Ent, F., J. Moller-Jensen, L. A. Amos, K. Gerdes and J. Lowe (2002). "F-actin-like filaments formed by plasmid segregation protein ParM." EMBO J 21(24): 6935-6943.

van der Oost, J., E. R. Westra, R. N. Jackson and B. Wiedenheft (2014). "Unravelling the structural and mechanistic basis of CRISPR-Cas systems." Nat Rev Microbiol 12(7): 479-492.

van Eerde, A., C. Hamiaux, J. Pérez, C. Parsot and B. W. Dijkstra (2004). "Structure of Spa15, a type III secretion chaperone from Shigella flexneri with broad specificity." EMBO Rep 5(5): 477-483.

Van Melderen, L., P. Bernard and M. Couturier (1994). "Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria." Mol Microbiol 11(6): 1151-1157.

Van Melderen, L. and M. Saavedra De Bast (2009). "Bacterial toxin-antitoxin systems: more than selfish entities?" PLoS Genet 5(3): e1000437.

Van Melderen, L. (2010). "Toxin-antitoxin systems: why so many, what for?" Curr Opin Microbiol 13(6): 781-785.

Vandecraen, J., M. Chandler, A. Aertsen and R. Van Houdt (2017). "The impact of insertion sequences on bacterial genome plasticity and adaptability." Crit Rev Microbiol 43(6): 709-730.

VanDrisse, C. M., A. R. Parks and J. C. Escalante-Semerena (2017). "A Toxin Involved in *Salmonella* Persistence Regulates Its Activity by Acetylating Its Cognate Antitoxin, a Modification Reversed by CobB Sirtuin Deacetylase." mBio 8(3).

Vecchiarelli, A. G., L. C. Hwang and K. Mizuuchi (2013). "Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism." Proc Natl Acad Sci U S A 110(15): E1390-1397.

Vecchiarelli, A. G., K. Mizuuchi and B. E. Funnell (2012). "Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria." Mol Microbiol 86(3): 513-523.

Vecchiarelli, A. G., Y. Seol, K. C. Neuman and K. Mizuuchi (2014). "A moving ParA gradient on the nucleoid directs subcellular cargo transport via a chemophoresis force." Bioarchitecture 4(4-5): 154-159.

Venkatesan, M. M., M. B. Goldberg, D. J. Rose, E. J. Grotbeck, V. Burland and F. R. Blattner (2001). "Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*." Infect Immun 69(5): 3271-3285.

Vokes, S. A., S. A. Reeves, A. G. Torres and S. M. Payne (1999). "The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island." Mol Microbiol 33(1): 63-73.

Wagner, A. and N. de la Chaux (2008). "Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences." Mol Genet Genomics 280(5): 397-408.

Wang, P., B. Zhang, G. Duan, Y. Wang, L. Hong, L. Wang, X. Guo, Y. Xi and H. Yang (2016). "Bioinformatics analyses of Shigella CRISPR structure and spacer classification." World J Microbiol Biotechnol 32(3): 38.

Watanabe, H., E. Arakawa, K. Ito, J. Kato and A. Nakamura (1990). "Genetic analysis of an invasion region by use of a Tn3-lac transposon and identification of a second positive regulator gene, *invE*, for cell invasion of *Shigella sonnei*: significant homology of *invE* with ParB of plasmid P1." J Bacteriol 172(2): 619-629.

Waterman, S. R. and P. L. Small (1996). "Identification of sigma S-dependent genes associated with the stationary-phase acid-resistance phenotype of Shigella flexneri." Mol Microbiol 21(5): 925-940.

Wei, J., M. B. Goldberg, V. Burland, M. M. Venkatesan, W. Deng, G. Fournier, G. F. Mayhew, G. Plunkett, D. J. Rose, A. Darling, B. Mau, N. T. Perna, S. M. Payne, L. J. Runyen-Janecky, S. Zhou, D. C. Schwartz and F. R. Blattner (2003). "Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T." Infect Immun 71(5): 2775-2786.

Weinrauch, Y., D. Drujan, S. D. Shapiro, J. Weiss and A. Zychlinsky (2002). "Neutrophil elastase targets virulence factors of enterobacteria." Nature 417(6884): 91-94.

WHO, (2005). "Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* 1.", from https://apps.who.int/iris/bitstream/handle/10665/43252/924159330X.pdf.

WHO, (2017). "The selection and use of essential medicines.", from http://ezproxyprd.bodleian.ox.ac.uk:2635/medicines/publications/essentialmedicines/en/.

Winther, K. S. and K. Gerdes (2011). "Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA." Proc Natl Acad Sci U S A 108(18): 7403-7407.

Wozniak, R. A. and M. K. Waldor (2009). "A toxin-antitoxin system promotes the maintenance of an integrative conjugative element." PLoS Genet 5(3): e1000439.

Xie, L., J. Zeng, H. Luo, W. Pan and J. Xie (2014). "The roles of bacterial GCN5-related N-acetyltransferases." Crit Rev Eukaryot Gene Expr 24(1): 77-87.

Xu, D. Q., J. O. Cisar, N. Ambulos, Jr., D. H. Burr and D. J. Kopecko (2002). "Molecular cloning and characterisation of genes for *Shigella sonnei* form I O polysaccharide: proposed biosynthetic pathway and stable expression in a live salmonella vaccine vector." Infect Immun 70(8): 4414-4423.

Yang, F., J. Yang, X. Zhang, L. Chen, Y. Jiang, Y. Yan, X. Tang, J. Wang, Z. Xiong, J. Dong, Y. Xue, Y. Zhu, X. Xu, L. Sun, S. Chen, H. Nie, J. Peng, J. Xu, Y. Wang, Z. Yuan, Y. Wen, Z. Yao, Y. Shen, B. Qiang, Y. Hou, J. Yu and Q. Jin (2005). "Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery." Nucleic Acids Res 33(19): 6445-6458.

Yang, J. Y., S. N. Lee, S. Y. Chang, H. J. Ko, S. Ryu and M. N. Kweon (2013). "A Mouse Model of Shigellosis by Intraperitoneal Infection." J Infect Dis.

Yang, C., P. Li, W. Su, H. Li, H. Liu, G. Yang, J. Xie, S. Yi, J. Wang, X. Cui, Z. Wu, L. Wang, R. Hao, L. Jia, S. Qiu and H. Song (2015). "Polymorphism of CRISPR shows separated natural groupings of Shigella subtypes and evidence of horizontal transfer of CRISPR." RNA Biol 12(10): 1109-1120.

Yoshida, S., Y. Handa, T. Suzuki, M. Ogawa, M. Suzuki, A. Tamai, A. Abe, E. Katayama and C. Sasakawa (2006). "Microtubule-severing activity of Shigella is pivotal for intercellular spreading." Science 314(5801): 985-989.

Zagaglia, C., M. Casalino, B. Colonna, C. Conti, A. Calconi and M. Nicoletti (1991). "Virulence plasmids of enteroinvasive *Escherichia coli* and *Shigella flexneri* integrate into a specific site on the host chromosome: integration greatly reduces expression of plasmid-carried virulence genes." Infect Immun 59(3): 792-799.

Zhang, X. H., L. Y. Tee, X. G. Wang, Q. S. Huang and S. H. Yang (2015). "Off-target Effects in CRISPR/Cas9-mediated Genome Engineering." Mol Ther Nucleic Acids 4: e264.

Zhang, Y. and K. Lin (2012). "A phylogenomic analysis of *Escherichia coli* / *Shigella* group: implications of genomic features associated with pathogenicity and ecological adaptation." BMC Evol Biol 12: 174.

Zhang, S.P., Q. Wang, S. W. Quan, Yu X. Q., Wang Y., Guo D. D., Peng L., Feng H. Y. and He Y. X. (2020). "Type II toxin–antitoxin system in bacteria: activation, function, and mode of action." Biophysics Reports 6: 68-79.

Zuo, G., Z. Xu and B. Hao (2013). "*Shigella* strains are not clones of *Escherichia coli* but sister species in the genus Escherichia." Genomics Proteomics Bioinformatics 11(1): 61-65.

Zychlinsky, A., M. C. Prevost and P. J. Sansonetti (1992). "*Shigella flexneri* induces apoptosis in infected macrophages." Nature 358(6382): 167-169.