1 Introduction

The emergence of antibiotic resistance in bacterial pathogens is one of the major challenges facing public health policies in the future. This thesis describes the outcome of studies into multiple drug resistance in one of humankind's major diseases, enteric fever.

1.1 Enteric fever

1.1.1 Definition

Enteric fever is a systemic illness, characterised by prolonged fever, caused by *Salmonella enterica* subspecies *enterica* serovar Typhi (*S.* Typhi) or *S*. Paratyphi A, B and C.

1.1.1.1 Typhoid fever

Typhoid fever is a systemic, febrile illness caused by the bacterium *Salmonella enterica* serovar Typhi (*S.* Typhi), a human-restricted *Salmonella*. Although the disease incidence has greatly declined since the early 20th century, due principally to the provision of good sanitation in Europe and the USA, it remains a major public health problem in many parts of the world, with an estimate of 21.6 million cases resulting in 216,500 deaths during the year 2000 (Crump, Luby & Mintz 2004). Typhoid is a disease of the poorer developing countries and, to a smaller extent, travellers visiting these endemic regions. Hence, the people who get typhoid are normally located in regions with poor health monitoring and support. As a bacterial disease, typhoid can be treated by the use of antibiotics. However, without appropriate treatment, 12% to 16% of patients may develop prolonged clinical disease and some can even die of severe

disease or complications such as gastrointestinal perforation, toxaemia, meningitis or endocarditis. Unlike many other acute bacterial diseases, between 1 to 5% of typhoid patients become chronic carriers (defined as excretion of *S.* Typhi in urine or stools for more than one year) (Bhan, Bahl & Bhatnagar 2005). The presence of a pool of asymptomatic carriers can complicate control of the disease.

1.1.1.2 Paratyphoid fever

A very similar but possibly less severe disease, paratyphoid fever, is caused predominantly by *S.* Paratyphi A. Although previously estimated to cause approximately a quarter of the incidence of typhoid fevers (Crump, Luby & Mintz 2004), studies from India and Nepal suggest that paratyphoid fever can contribute up to half of all cases of enteric fever (Bhan, Bahl & Bhatnagar 2005) in some regions. Surveillance has revealed an incidence of *S.* Paratyphi A ranging from 14% of enteric fever episodes in Indonesia to 15% in Pakistan, 24% in India and in China *S.* Paratyphi A is more common (64%) than *S.* Typhi (Ochiai *et al.* 2005).

1.1.2 The pathogens

Salmonella is a genus of Gram negative, facultatively anaerobic, rod-shaped bacteria that can usually utilise citrate as a carbon source and do not ferment lactose. Based on genomic DNA comparisons, the genus is divided into two species, *Salmonella enterica* and *Salmonella bongori* (Brenner *et al.* 2000, Tindall *et al.* 2005); and further subdivided into subspecies and serovars. *S. enterica* is currently subdivided into 6 subspecies in which subspecies I (*Salmonella enterica* subspecies *enterica*) accounts for 99% of all human and animal infections. There are more than 2500 known serovars defined by serology (the Kauffman-White classification scheme), illustrating the broad antigenic diversity of the species. The Kauffman-White scheme is based on the

determination of the "O" antigen type (polysaccharides associated with lipopolysaccharide of the bacterial outer membrane), the presence or absence of the "Vi" (**Vi**rulence) capsule, and the "H" antigens (proteins associated with bacterial flagella) of phase 1 (motile phase) and phase 2 (non-motile phase). Individual bacterial isolates are typed using sets of specific antisera raised to different typing *Salmonella* strains.

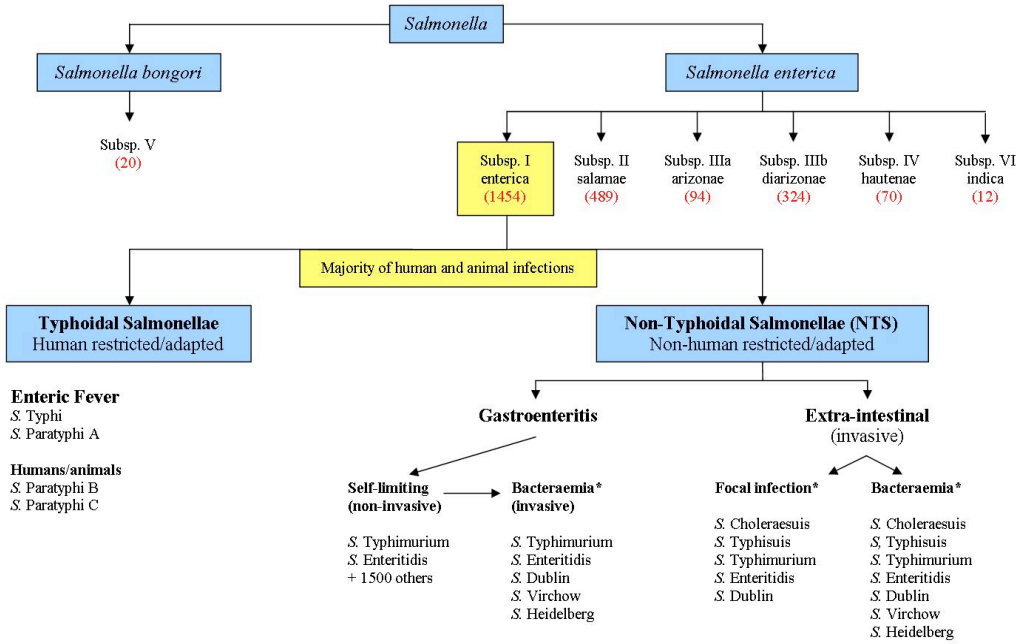


Figure 1-1 Diagram showing the classification of the *Salmonella* genus and its diseases.
Figure adapted from Langridge, Wain & Nair, 2009

1.1.2.1 Salmonella enterica serovar Typhi

Colonies of *S.* Typhi are by definition non-lactose fermenters that produce a characteristic biochemical pattern in Kligler iron agar (acid without gas, an alkaline slant and a moderate amount of H_2S production). Serotyping defines *S.* Typhi isolates by the presence of the lipopolysaccharide antigen O9, 12 (group D), protein flagella antigen H_d and Vi polysaccharide capsular antigen. Other flagella antigen variants, such as H_j, have been reported.

1.1.2.2 Salmonella enterica serovar Paratyphi A

S. Paratyphi A can be distinguished from *S.* Typhi by the inability to produce H_2S. Serologically, *S.* Paratyphi A is positive for O1, 2, 12 lipopolysaccharide antigens and H_a flagella antigen. *S.* Paratyphi A do not produce Vi capsule, although most *S.* Paratyphi C isolates do.

1.1.3 Treatments of enteric fever

Antibiotics have been used to treat typhoid for over 50 years. Chloramphenicol was first introduced to treat typhoid fever in 1948 (Woodward, Smadel 1948). However, resistance to chloramphenicol was reported in 1972 and subsequently further resistance emerged to all first line drugs including ampicillin, trimethoprim and sulphamethoxazole. Multiple drug resistance (MDR) developed, appearing sporadically during the 1970s but becoming significantly established by the end of the 1990s (Rowe, Ward & Threlfall 1997). As a direct consequence, fluoroquinolones (ciprofloxacin and ofloxacin) then become the treatment of choice, along with extended spectrum cephalosporins (ceftriaxone and cefixime) and azithromycin as alternative treatments for resistant organisms. In Central, South and South East Asia, where fluoroquinolones are widely used, isolates with reduced susceptibility to ciprofloxacin have become relatively common (Parry, Threlfall 2008) and isolates that are fully resistant to ciprofloxacin have also been reported (Capoor *et al.* 2009). For enteric fever, in many parts of Asia where isolates with reduced susceptibility to ciprofloxacin and MDR are becoming common, azithromycin, gatifloxacin and ceftriaxone can be used. Fluoroquinolones remain the best option in areas where resistance is uncommon (Parry, Beeching 2009).

1.1.4 Antibiotic resistance

Resistance to all of the drugs described above that are used for treatment has been reported. The genetic basis of this resistance has been defined in many studies. Resistance to the older first line drugs is mainly plasmid encoded. In contrast, resistance to fluoroquinolones is normally chromosomally mediated via mutations in topoisomerases (Figure 1-2). The plasmid borne resistance is usually mediated by IncHI1 plasmids. Thus there is a close association between *S.* Typhi and resistance mediated by the IncHI1 incompatibility type plasmids.

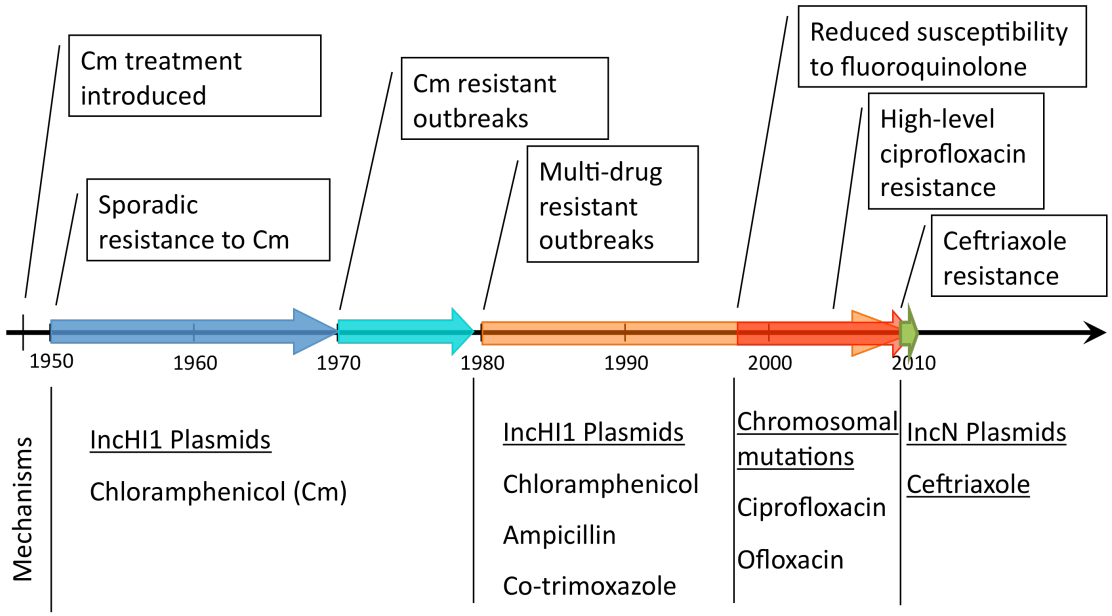


Figure 1-2 Development of antibiotic resistance in *S.* Typhi.

1.1.4.1 Plasmid borne resistance

1.1.4.1.1 Chloramphenicol

Chloramphenicol was introduced for the treatment of typhoid fever in 1948 (Woodward and Smadel, 1948). Although sporadic cases of chloramphenicol resistant typhoid fever were reported two years later (Colquhoun, Weetch 1950), resistance to chloramphenicol

took a long time to become established in the *S.* Typhi population. The first reported antibiotic resistant typhoid fever outbreak occurred in May 1972 in Kerala, India (Paniker, Vimala 1972). Chloramphenicol resistance in this outbeak was demonstrated to be plasmid borne in *S.* Typhi. In the same year, two other chloramphenicol resistant outbreaks were also documented in Mexico and Vietnam, both caused by *S.* Typhi harbouring IncH plasmids (Gangarosa *et al.* 1972, Butler *et al.* 1973).

1.1.4.1.2 Ampicillin

Ampicillin was introduced for the treatment of typhoid fever in 1962 (Maddock 1962). During the large epidemic in Mexico in 1972, isolates resistant to both chloramphenicol and ampicillin was reported. However, resistance to these two drugs was transferred independently by two separate plasmids (Olarte, Galindo 1973). The plasmids conferring resistance to chloramphenicol were later identified as incompatibility group H whilst the ampicillin resistance plasmids were of the incompatibility group I or A/C (Datta, Olarte 1974).

1.1.4.1.3 Co-trimoxazole

Co-trimoxazole was first described for the treatment of typhoid fever in 1972 (Jaques, Van der Heyde 1972). In 1980, single resistance to chloramphenicol, ampicillin and co-trimoxazole was described in Bangkok (Vongsthongsri, Tharavanij 1980). Until the 1980s, there was no report of single isolates harbouring resistance to all three first line drugs.

1.1.4.1.4 Multi-drug resistance (MDR)

MDR, defined as resistance to all first-line antimicrobials (ampicillin, co-trimoxazole, and chloramphenicol), emerged gradually in *S.* Typhi. The first reported MDR outbreak occurred in Kashmir, India in 1988 (Kamili *et al.* 1993). By 1990, there were reports of

MDR *S.* Typhi strains from India, Pakistan and the Arabian Gulf (Bhan *et al.*, 2005). Through the course of the emergence of MDR in *S.* Typhi, plasmids of incompatibility groups IncA/C and IncI were observed in the short-term (Datta, Olarte 1974, Datta, Richards & Datta 1981) and failed to disseminate whereas plasmids of IncH group were detected repeatedly in outbreak cases and such plasmids have conferred long-term resistance. Indeed, MDR *S.* Typhi strains harbouring IncHI1 plasmids are now globally distributed (Hampton *et al.*, 1998).

1.1.4.1.5 Cephalosporins

Cephalosporins (ceftriaxone and cefixime) were one of the few choices left for the treatment of enteric fever after ciprofloxacin resistance developed in *S.* Typhi and *S.* Paratyphi A. The first reported trial for the use of ceftriaxone to treat typhoid fever was conducted in Bangladesh in 1988 (Islam *et al.* 1988). Although resistance to third generation cephalosporins in non-typhoidal salmonellae had been reported as early as 1989 (Garbarg-Chenon *et al.* 1989), resistance in *S.* Typhi remains rare. The first cases of reduced susceptibility or resistance to ceftriaxone were documented last year in Bangladesh and Kuwait (Pontali *et al.* 2008, Rotimi *et al.* 2008). A most recent case of ceftriaxone resistant *S.* Typhi was reported from an Iraqi woman who returned to Germany after a month's long visit in Iraq. Interestingly, this resistance was encoded on an IncN plasmid of ~50 kp carrying the $bla_{\text{CTX-M-15}}$ and *qnrB2* genes (Pfeifer, Matten & Rabsch 2009).

1.1.4.2 Chromosomally encoded resistance

1.1.4.2.1 Nalidixic acid

Fluoroquinolones have become the treatment of choice for typhoid fever after the emergence of MDR *S.* Typhi. However, there have been many reports of nalidixic acid

resistant (Na^R) *S.* Typhi which exhibit decreased susceptibility to ciprofloxacin. Such isolates have been implicated in a poorer clinical response to fluoroquinolones (Parry *et al.* 1998, Wain *et al.* 1997). There was a major outbreak of MDR *S.* Typhi in Tajikistan in 1997 spread by contaminated drinking water affected nearly 9,000 individuals leading to 95 deaths. This epidemic MDR *S.* Typhi strain developed resistance to nalidixic acid and reduced susceptibility to ciprofloxacin (Mermin *et al.* 1999) during the outbreak.

1.1.4.2.2 Fluoroquinolones

Under fluoroquinolone treatment, a simultaneous increase in resistance levels to fluoroquinolones and a decline in the percentage of MDR in the *S.* Typhi population has been reported in some regions (Mandal, Mandal & Pal 2004). However, the emergence of high-level ciprofloxacin resistance in *S.* Typhi (Hasan *et al.* 2005) and *S.* Paratyphi A (Adachi *et al.* 2005, Joshi, Amarnath 2007) has not been generally sustained. Because reduced susceptibility to fluoroquinolones is now very common, causing patients to respond poorly to treatment (Wain *et al.* 1997), the treatment for resistant typhoid fever now depends on third generation cephalosporins and azithromycin (Parry 2004). However, resistance to both these antibiotics can be plasmid mediated.

1.1.4.3 Resistance in *S.* Paratyphi A

While MDR *S.* Typhi has been isolated globally since the 1980s, *S.* Paratyphi A remained predominantly susceptible to antibiotics (Kapil *et al.* 1997, Woods *et al.* 2006). In Pakistan, however, there has been a reported increasing incidence of MDR *S.* Paratyphi A (Hasan *et al.* 2008). In addition, in Nepal, the MDR rate in *S.* Paratyphi A was higher (7%) than the rate in *S.* Typhi (5%) (Pokharel *et al.* 2006). MDR *S.* Paratyphi A isolates were also recently reported to be on the increase in North India

(Mohanty *et al.* 2006). In China, however, although *S.* Paratyphi A is largely resistant to nalidixic acid, this is chromosomally mediated and plasmid borne MDR remains very low (Ochiai *et al.* 2005). In Europe, the reported levels of MDR S. Paratyphi A rose from 9% in 1999 to 25% in 2001 in travellers, most of whom were returning from the Indian subcontinent where resistant strains were endemic (Threlfall *et al.* 2003). Recent data from Pakistan shows a decline in MDR in *S.* Paratyphi A but not in *S.* Typhi (Hasan *et al.* 2008).

1.1.5 Epidemiology – disease burden

The global incidence of typhoid fever has been reported to be 21.6 million per year (Crump, Luby & Mintz 2004). However, this is a rough estimation since most typhoid endemic areas lack facilities to confirm diagnosis and the incidents of typhoid fever fluctuate from region to region, especially in developing countries (Figure 1-3).

Enteric fevers are predominantly transmitted by the faecal-oral route via contaminated food and water. Poor sanitation and hygiene remain the most important risk factors for enteric fevers. Typhoid fever is endemic in many parts of the developing world, particularly in the Indian subcontinent and South East Asia, but many cases are also reported from elsewhere in Asia and in Africa (Parry 2004). In the US and most of Europe, typhoid fever is mainly a disease of returning travellers.

The incidence of paratyphoid fever is less well-defined than typhoid fever. It is estimated that a quarter of enteric fevers may be caused by *S.* Paratyphi A (Crump, Luby & Mintz 2004). However, *S.* Paratyphi A is thought to be increasing in some areas of Asia, especially in India, Nepal and China (Ochiai *et al.* 2005). Clinical manifestation of *S.* Paratyphi A infection may be as severe as *S.* Typhi (Maskey *et al.* 2006).

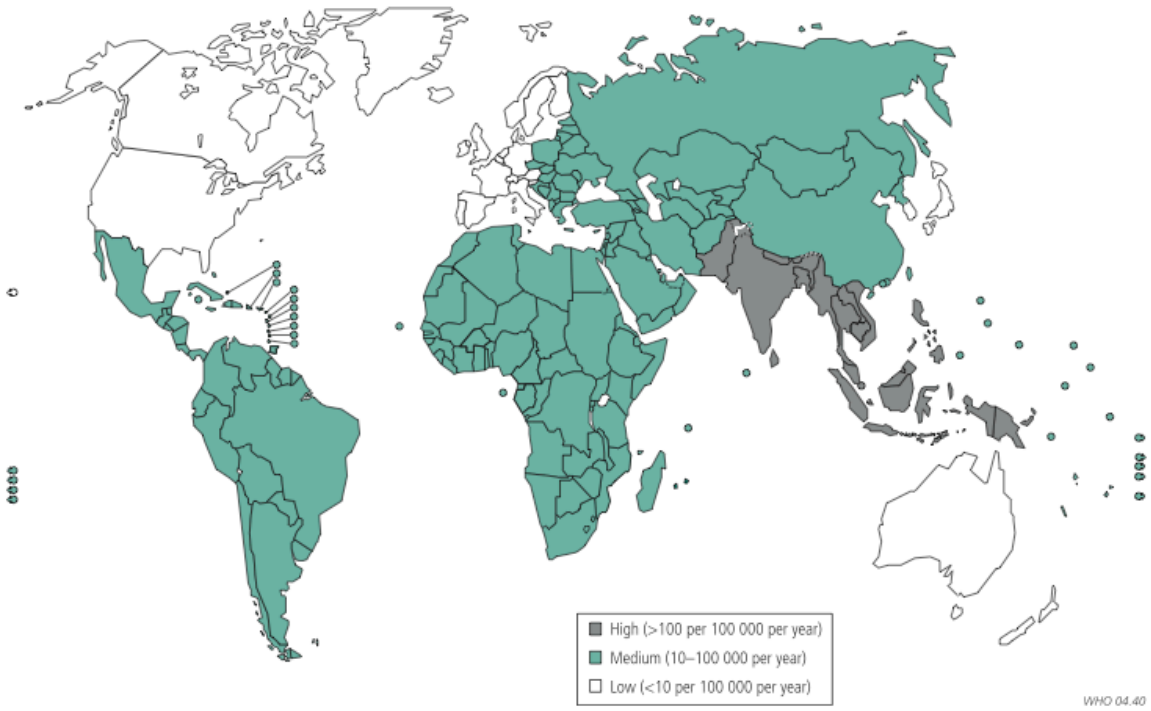


Figure 1-3 Geographical distribution of typhoid fever.
Figure is adapted from Crump, Luby & Mintz (2004).

1.1.6 Salmonella typing methods

Table 1-1 Common typing methods

Technique	Brief description	Reference
Phenotypic		
Serotyping	Also a classification method to define serovars. Agglutination of somatic "O" and flagella "H" antigens.	(Kauffmann 1950, Popoff, Bockemuhl & Gheesling 2004)
Phage type	Susceptibility of a bacterial isolate to a panel of bacteriophage	(Craigie, Yen 1938, Threlfall, Frost 1990)
R-type (Antibiotic resistance)	Susceptibility to different antibiotics	
Genotypic		
RAPD	Random amplification of polymorphic DNA: PCR amplification from random primers	(Shangkuan, Lin 1998)
PCR for specific genes or islands	PCR for resistance genes, pathogenicity factors or metabolic markers	(Mortimer *et al.* 2004, Kidgell *et al.* 2002)
Plasmid profile	Analysis of plasmids harboured by isolate	(Connerton *et al.* 2000)

RFLP of plasmids	Restriction fragment length polymorphism: Digestion of plasmid DNA with a restriction enzyme	(Wain *et al.* 2003)
PFGE	Pulsed field gel electrophoresis: Restriction digestion of genomic DNA and separation on agarose gel using a pulsed electrical field	(Thong *et al.* 1994)
(F)ALFP	Fluorescent amplified fragment length polymorphism: PCR-based modification of PFGE. Fluorescent markers included to improve discrimination of fragments	(Nair *et al.* 2000)
VNTR/MLVA	Variable number of tandem repeats/Multiple loci VNTR analysis: Size of PCR products represents copy number of short repetitive sequences. Increase the number of VNTRs for more discrimination power.	(Liu *et al.* 2003)
Ribotyping	Analysis of rRNA genes using either restriction digests or Southern blot.	(Altwegg, Hickman-Brenner & Farmer 1989)
MLEE	Multilocus enzyme electrophoresis: Separate and detect activity and isoelectric point of enzymes	(Selander *et al.* 1990)
MLST	Multilocus sequence typing: Compare sequences of house keeping genes.	(Kidgell *et al.* 2002)
SNP-typing	Detection of defined SNP variations on the genome.	(Octavia, Lan 2007, Le *et al.* 2007, Roumagnac *et al.* 2006)

Although serotyping was developed more than fifty years ago (Kauffmann 1950), it is still the primary method for typing and classification of *Salmonella* in clinical diagnostic laboratories today. Serotyping of *Salmonella* uses the agglutination of the somatic “O” and flagella “H” antigen based on the Kauffmann-White scheme (Popoff, Bockemuhl & Gheesling 2004). Full serotyping to define *Salmolnella* serovars is not always easy to perform and it is predominantly performed in reference laboratories, due to the limitation of acquiring antisera and the difficulty in detection of phase 2 “H” antigen. For epidemiology and evolutionary studies, different methods to subtype *Salmonella* serovars, especially serovar Typhi, have been developed. These methods can be considered to be either as phenotypically or genotypically based. Different

techniques can be more useful in particular settings and circumstances and usually a combination of several methods is employed. The summary of common typing techniques is presented in Table 1-1.

1.1.6.1 Phenotypic techniques

Phage typing is a traditional method that has been employed over many years to differentiate further within the serotypes of *Salmonella*. Phage typing for serovar Typhi was developed in the late 1930s (Craigie, Yen 1938) in which the lysis patterns of different phages are compared to a set of typing strains. For *S.* Typhi, it is based on the detection of the Vi capsule by specific phages that target this antigen as receptor. There are approximately 140 Vi phage types for serovar Typhi. While this is a fairly robust and discriminating typing approach, phage typing is only limited to reference laboratories due to difficulty in maintaining all the phages. Also, as a phenotypic approach biological drift in phage, typing strains and laboratory procedure is inevitable. Further, a high percentage of degraded Vi-positive strains or untypeable Vi-positive strains in certain areas also reduces the usefulness of this technique (Trung *et al.* 2007, Le *et al.* 2004).

Another phenotypic method is R-typing, which groups *Salmonella* according to antibiotic resistance profiles. Since the majority of antibiotic resistant genes are located on plasmids, this method provides limited information about the background genetic identity of the isolate. R-typing is used in combination with other typing methods to report resistance profiles of clinical strains.

1.1.6.2 Genotypic techniques

Genotypic methods use molecular biological tools for subtyping of *Salmonella* serovars. They can be broadly divided into extrachromosomal typing and chromosomal typing methods.

Extrachromosomal typing includes plasmid profiling and plasmid RFLP (Restriction Fragment Length Polymorphism). Plasmid profiling is used specifically in outbreaks of antibiotic resistant *Salmonella*. This method is of low discriminatory power and can be misleading as different strain types can contain plasmids of similar size (Connerton *et al.* 2000). Plasmid RFLP is used to increase the resolution of plasmid profiling by restriction digestion of plasmid DNA (Wain *et al.* 2003).

Pulse field gel electrophoresis (PFGE) is a technique of higher discriminatory power and is useful for the investigation of local outbreaks (Thong *et al.* 1994). The use of this technique has become the gold standard for the international comparison of isolates in the public health system. Considerable attempts have been made to improve the standardisation of the method and networks such as Pulsenet and Enternet have been formed to promote international collaboration to ensure reproducibility and comparability of PFGE data. While it is not technically difficult, PFGE is quite labour and skill demanding to produce standardised results. This is a good method for subtyping, especially in local epidemiology but it provides little phylogenetic information. It can also give a false impression of the true phylogenetic distance of isolates when looking in the context of global epidemiology.

Random amplification of polymorphic DNA (RAPD) (Shangkuan, Lin 1998) and (Fluorescent) amplified fragment length polymorphism ((F)AFLP) (Nair *et al.* 2000) are less commonly used methods exploiting PCR-based techniques to increase the discriminatory power of typing. Different types are visualised by separating digested

DNA using electrophoresis, which has the inherent disadvantages for standardisation and comparing results between laboratories. These two techniques do not give information about phylogenetic relationship of isolates. Other PCR-based methods include analysis of gene profile coding for antibiotic resistance (Kidgell *et al.* 2002) and pathogenicity markers (Mortimer *et al.* 2004). These are techniques that provide additional information to describe clinical isolates and add supporting information for other subtyping schemes.

Variable number of tandem repeat (VNTR) produces data on the copy number of short repetitive sequence of individual isolates by determining the size of PCR products generated from across specific repeats (Liu *et al.* 2003). Multiple loci VNTR analysis (MLVA) looks at many VNTR loci to increase the discriminatory power. MLVA has been shown to be useful in distinguishing closely related *S.* Typhi isolates but gives conflicting phylogenetic relationships compare to those inferred from SNP typing (Octavia, Lan 2009).

The methods mentioned above mainly focus on capturing the variations between isolates in order to increase their discriminatory power. While such approaches are useful for distinguishing isolates locally over a short period of time, they can be misleading for global epidemiology. A more appropriate approach for long term, global epidemiological studies with many bacteria is multilocus enzyme electrophoresis (MLEE) and multilocus sequence typing (MLST). MLEE looks at the variation in electrophoretic mobility of enzymes and correctly identifies lineages of isolates (Selander *et al.* 1990). This approach has contributed to much of our understanding of the global epidemiology and population structure of infectious agents (Selander *et al.* 1987). MLST was developed to overcome a major problem associated with MLEE, that results are difficult to compare between laboratories while essentially use the same

experimental approaches (Maiden *et al.* 1998). MLST identifies alleles and sequence types directly from the actual sequences of (normally) seven housekeeping genes. The overwhelming advantage of this method is that the same results should be achieved wherever the test is performed and any comparison of results is simple and absolute in nature. Since this method monitors the accumulation of'neutral' mutations in housekeeping genes, it provides some phylogenetic inference on the relationship between isolates. It, therefore, can be used as a classification scheme and has the potential of replacing classical serotyping (Sukhnanand *et al.* 2005). MLST is useful for identifying serovar Typhi because this is a monomorphic group (Achtman 2008). However, it does not have the resolution to discriminate within *S.* Typhi (Kidgell *et al.* 2002). There is a requirement for a typing method to interrogate the evolutionary and population structure of monomorphic/monophyletic groups such as *S.* Typhi.

1.1.6.3 Future typing methods

The typing of a monomorphic group of bacteria presents a technical challenge in that more genetic variation needs to be investigated simultaneously for the identification and classification of subgroups. SNP typing uses single base pair variation at a large number of loci for the typing of bacteria. The recent rapid advances in DNA sequencing technologies have been exploited to re-sequence multiple genomes of *S.* Typhi and these projects have been used as a basis for SNP discovery (Roumagnac *et al.* 2006, Holt *et al.* 2008). SNP were identified by exploiting globally representative collections of *S.* Typhi. The paper by Holt *et al.* 2008 provides information on the development of SNP identification methods. It is essential to build SNP typing approaches based on globally representative bacterial isolates. Octavia and Lan (Octavia, Lan 2007) used PCR-restriction enzyme digestion to genotype S. Typhi using 38 genome-wide SNPs. However, these SNPs were identified using only two genomes, *S.* Typhi Ty2 and CT18,

which introduced phylogenetic bias to the results. A better strategy for SNP discovery, to avoid phylogenetic discovery bias, is crucial for the development of SNP genotyping methods (Roumagnac *et al.* 2006). Increasing the number of SNPs used for typing can increase the discriminatory power of the system but this could also increase the cost and technological challenge. There are a number of commercially available platforms for large scale SNP typing, such as iFLEX by Sequenome (Gabriel, Ziaugra & Tabbaa 2009) or GoldenGate assay by Illumina (Butler, Ragoussis 2008) that can be adapted to simultaneously monitor a high number of specific bacterial SNPs in single DNA samples. The development of SNP typing methods will provide a universal tool for both local and global epidemiology, which would be readily comparable. The critical feature of SNP typing is that, since it is DNA sequence based, it provides an unequivocal signature of any bacterial isolate.

1.1.6.4 Evolution of *S.* Typhi

The estimated age of *S.* Typhi, based on MLST analysis, is approximately ~50,000 years old (Kidgell *et al.* 2002). However, this estimation is somewhat imprecise due to the small number of synonymous SNPs discovered by the MLST method (seven genes – 3336 bp) and the uncertainty of the molecular clock rate. Roumagnac *et al.* further dissected the global *S.* Typhi population by the scanning of ~200 gene fragments (88,739 bp) for sequence variations, originally within 105 *S.* Typhi isolates (Roumagnac *et al.* 2006). They discovered 88 biallelic polymorphisms (BiPs) which define 59 haplotypes that form a fully parsimonious tree representative of *S.* Typhi globally. The most recent common ancestor for *S.* Typhi was calculated to have appeared between 10 to 43 thousand years ago based on these BiPs. There is also evidence of signifiant recombination between *S.* Typhi and *S.* Paratyphi A progenitors occurring before the emergence of the common ancestor (Didelot *et al.* 2007). The carrier state is thought to

have influenced the preservation of the ancestral nodes by extant bacteria and it is also believed to have contributed significantly to the global transmission of *S.* Typhi, as observed by this study.

1.2 Plasmid biology

Plasmids are extra-chromosomal DNA molecules capable of autonomous replication inside the host bacterial cell (Lederberg 1952, Marmur *et al.* 1961, Watanabe, Fukasawa 1961, Hayes 1964). Conjugative plasmids are those that can transfer horizontally within and across bacterial genera and species by conjugation. They can also serve as vehicles for mobile elements including transposons and integrons, providing a tool for bacteria to sample a wide array of genes from the mobile gene pool. Plasmids are thus a crucial factor in the rapid dissemination of antibiotic resistant determinants in bacteria.

A plasmid consists of modules of survival and propagation functions, including replication, partitioning, post-segregational killing and conjugative transfer (Thomas 2000). This section describes plasmids in terms of their functional modules with a focus on IncHI1 plasmids in order to understand their role in dissemination of antibiotic resistance and other possible contribution to the biology of the pathogens *S.* Typhi and Paratyphi A.

1.2.1 Replication mechanisms

1.2.1.1 Theta mechanism

The theta replication mechanism is perhaps the most intensively studied mechanism for plasmids from Gram-negative bacteria, including plasmids R1, P1, R6K, ColE1 etc. although this mechanism has also been described for plasmids from Gram-positive bacteria. Theta-type replication starts from an origin of replication (*ori*) and then precedes either uni- or bi-directionally. Theta-type plasmids in general require an *ori*

and a plasmid encoded replication protein (Rep initiator protein). Some plasmids may also need host proteins such as polymerases and host initiation factors for their replication (Giraldo-Suarez *et al.* 1993).

The origins of replication contain sites that are required for the binding of Rep proteins and potentially other host encoded proteins. They may also contain an AT-rich region of direct repeats called iterons where opening of the strands and assembly of host initiation factors occurs (Bramhill, Kornberg 1988). Rep initiator proteins normally bind specifically to DNA sequences in the *ori*. They facilitate the assemble of the replication machinery, which are nucleoprotein complexes including, in some cases, host replication proteins such as DNA polymerase III, helicase and primase. The organisation of the plasmid replication machinery resembles that of the chromosomal replication complex (Bramhill, Kornberg 1988).

The replication process initiates when the Rep protein binds to the *ori*, which leads to the melting of double strand DNA at the iteron, promoting the assembly of the replication machinery complex, the synthesis of a RNA primer and then the synthesis of DNA by covalent extension of the primer (del Solar *et al.* 1998). DNA synthesis of both strands is coupled and occurs continuously on one of them (leading strand) and discontinuously on the other (lagging strand). The replication progresses to completion and terminates by the interaction with the terminus sites.

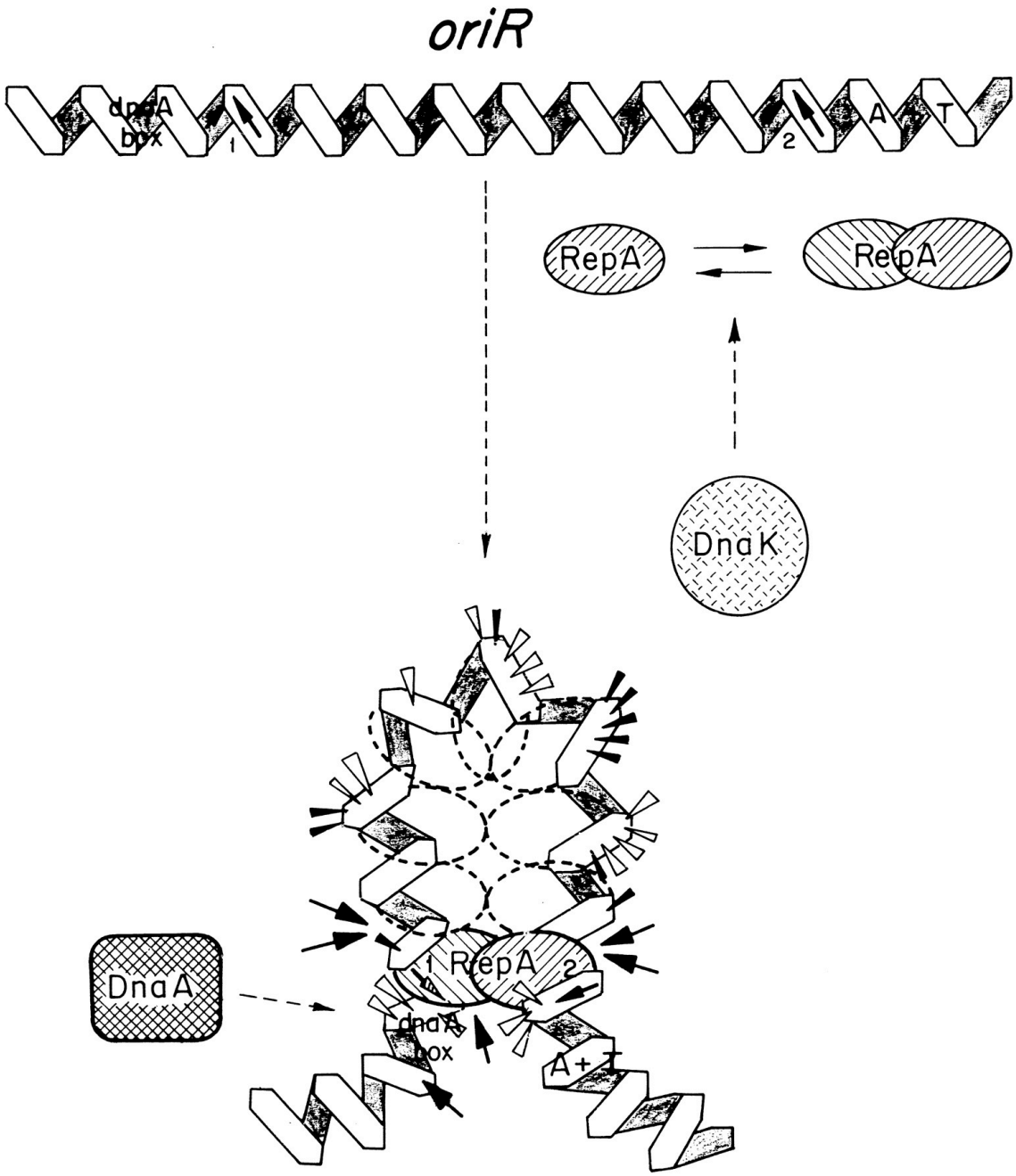


Figure 1-4 RepA-*oriR* complexes in initiation of R1 plasmid replication.
The dimer of RepA binds specifically to the two partially panlindromic sequences (site 1 and 2) to form a DNA loop, exposing the *dnaA* box for the binding of the host DnaA. The DNA loop would be filled afterwards with more RepA molecules that are brought to the complex mainly by protein-protein interactions. Arrowheads indicate DNase I-hypersensitive sites (the size is proportional to the intensity of cleavage), whereas arrows point to strong cleavage sites for hydroxyl radicals. A hypothetical role for DnaK in modulating the aggregation and activation state of RepA dimmers is also shown. Figure is adapted from del Solar *et al.* (1998).

1.2.1.2 Rolling-circle replication

This mechanism of replication involves three elements: an initiator protein (Rep protein), a double stranded origin (*dso*) and a single stranded origin (*sso*). The Rep

proteins of the rolling-circle replicating plasmids have DNA strand transferase activity (Koepsel *et al.* 1985). They bind specifically to the *bind* site and nick supercoiled DNA within an unpaired sequence of the *nic* region (Thomas, Balson & Shaw 1990). Both the *bind* site and the *nic* region are located within the *dso*. The nick leaves a 3'-OH end of the plus strand that is used as a primer for leading strand synthesis, which requires the host replication proteins. The synthesis of the leading strand continues until the replisome reaches the reconstituted *dso*, where the Rep protein makes the second nick to release newly synthesised ssDNA intermediate (Rasooly, Wang & Novick 1994). The ssDNA molecule is then converted to a double stranded plasmid DNA by the host replication proteins starting at the *sso* site (Dempsey, Zhao & Khan 1995).

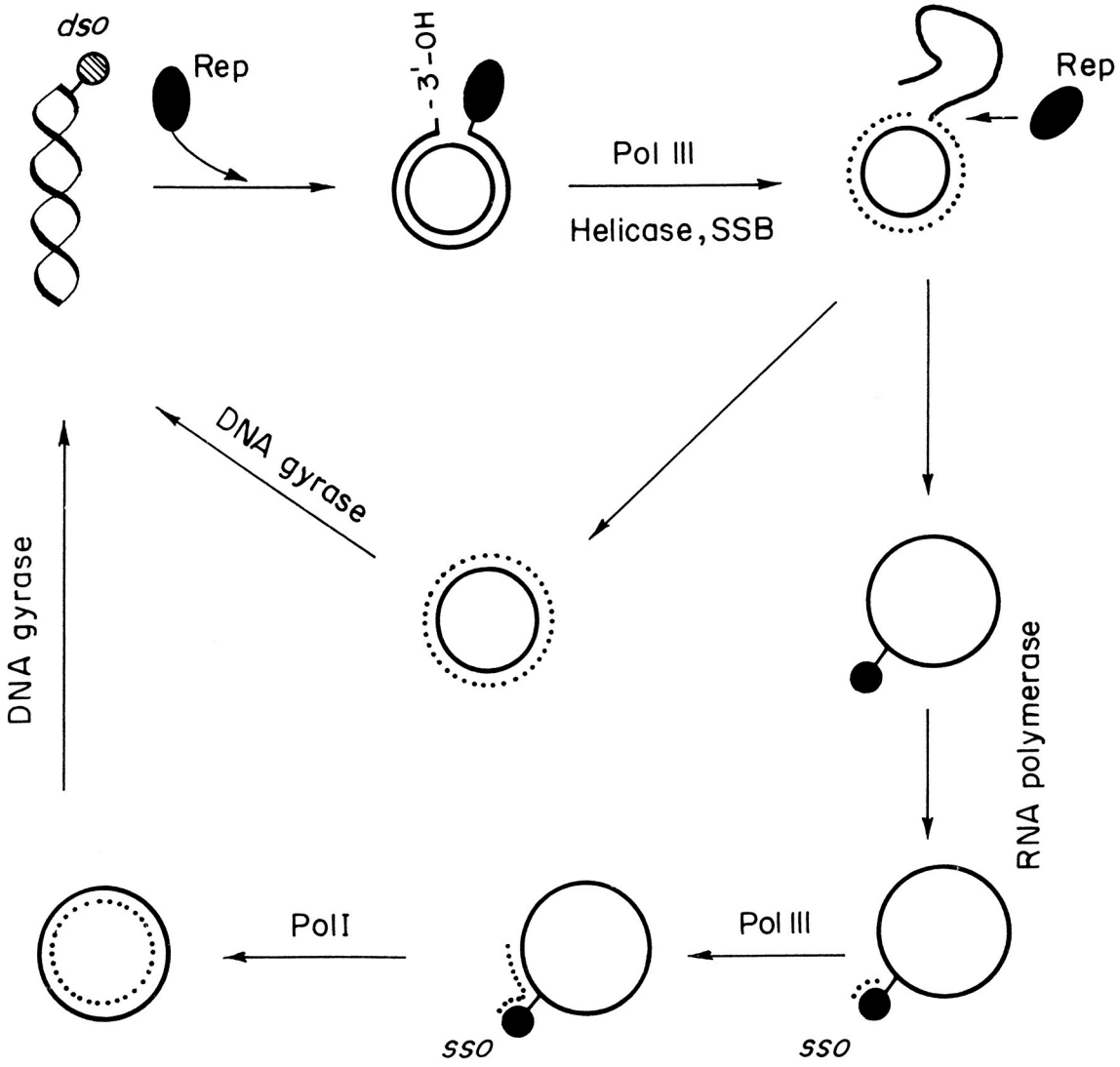


Figure 1-5 Model for rolling-circle replication.
The plasmid-encoded Rep protein recognises the *dso* on supercoiled DNA and introduces a site-specific nick generating a free 3'-OH end. This end is elongated by host proteins as the parental strand is being displaced. When the replication fork reaches the reconstituted *dso*, Rep protein catalyzes a strand transfer reaction, releasing an ssDNA intermediate and a dsDNA molecule with a parental and a newly synthesized (dotted) strand. Lagging-strand synthesis on the ssDNA molecule is initiated at the *sso* signal by the host RNA polymerase. This enzyme would synthesize a short primer RNA, and lagging-strand synthesis is performed by host DNA polymerases. The end products are supercoiled plasmid DNA molecules. Figure is adapted from del Solar *et al.* (1998).

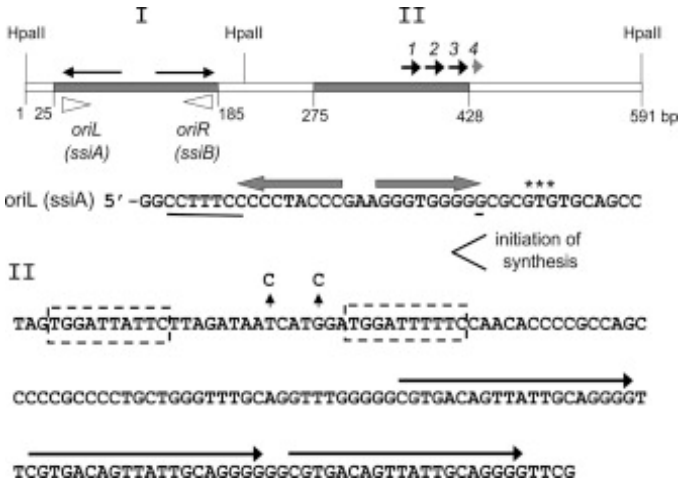
1.2.1.3 Strand displacement replication

Strand displacement replicating plasmids consists of those from the IncQ group, a group of plasmids that have a broadest host-range (Meyer 2009). Three almost identical plasmids from this group, RSF1010, R300B and R1162 from *E. coli, S.* Typhimurium and *Pseudomonas aeruginosa* respectively (Barth, Grinter 1974), have been studied intensively. Their characteristic replication mechanism plays a major role in making IncQ plasmids promiscuous (Sakai, Komano 1996).

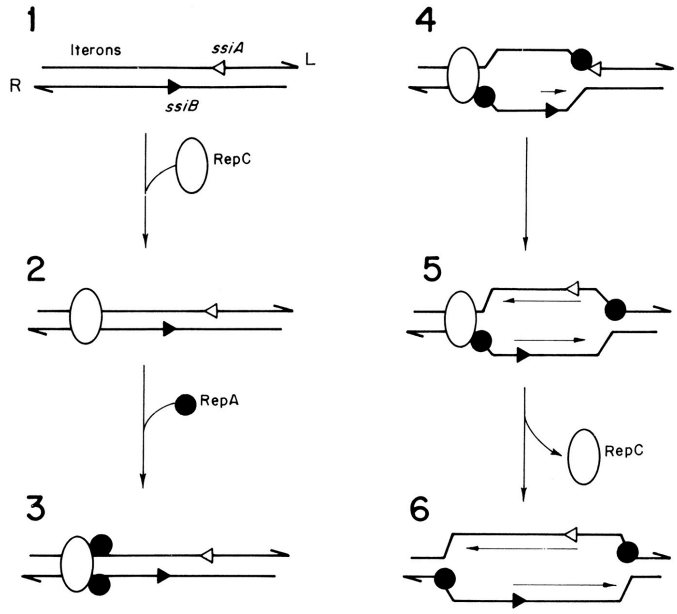
RSF1010 requires an origin of replication (*oriV)* and three plasmid-encoded proteins (RepA – helicase, RepB – primase and RepC – initiator) for replication (Sakai, Komano 1996, Diaz, Staudenbauer 1982). These three proteins are keys for the plasmid's own initiation of replication, thus explaining the broad host-range characteristic is due to an independence in replication from host proteins. The *oriV* consists of two, non-contiguous domains, I and II (Kim, Lin & Meyer 1987). Domain I contains two inverted repeats *ssiA* and *ssiB,* which are specifically recognised by the RepB primase and domain II contains three identical iterons and an AT-rich conserved direct repeats, which is a site for strand separation (Haring *et al.* 1985).

The replication starts with the binding of RepC to the iterons, inducing the strand separation and allowing the entry of RepA helicase to domain II within the AT-rich region (Kim, Meyer 1991). RepA then migrates to domain I and exposes *ssiA* and *ssiB* as single-stranded regions (one *ssi* on each strand). The priming of DNA synthesis at *ssiA* and *ssiB* is catalysed by RepB, follows by the continuous DNA synthesis on each

strand and this results in the displacement of the complementary strand. Replication of this displaced strand is initiated at the exposed *ssi* site (del Solar *et al.* 1998).



(a)



(b)

Figure 1-6 Replication of plasmid RSF1010 by the strand displacement mechanism.
(a) Origin of replication. (a) Domains I and II, containing DNA essential for replication, are indicated by the regions where the double line is filled. The horizontal arrows indicate the location of the inverted repeat and the iterons. DNA containing oriL (ssiA) is shown below. The initiation site for DNA synthesis is indicated by the starred bases and DNA forming the hairpin loop by the filled arrows. Bases contacting the catalytic domain of RepB are underlined. The base sequence of domain II is at the bottom of the figure. The iterons are indicated by the horizontal arrows. Single base-pair mutations affecting replication and inhibiting RepC-induced strand separation are shown, along with the flanking AT-rich, conserved direct repeats (outlined by the dashed lines). (b) Model for initiation of replication by the strand displacement mechanism in plasmid RSF1010 (266). Replication occurs with opposite polarities from two origins (ssiA and ssiB), which are independently used. Interactions between the plasmid-encoded proteins RepC and RepA are indicated. Priming is catalyzed by RepB' (not shown). Thin lines indicate

newly synthesized DNA, with the direction of synthesis indicated by arrowheads. Figures are adapted from del Solar *et al.* (1998) and Meyer (2009).

1.2.2 Maintaining mechanisms

1.2.2.1 Conjugation systems

Bacterial conjugation or the transfer of plasmid DNA from one bacterial cell to another is a major mechanism by which horizontal gene transfer occurs. From the plasmid point of view this is important to ensure the dissemination and persistence of plasmids within and between populations or communities of bacteria in natural settings (Bahl, Hansen & Sorensen 2007, Bahl, Hansen & Sorensen 2009).

The bacterial conjugation system is most recently viewed as the merging of two ancient bacterial systems: the rolling-circle replication (RCR) and the type IV secretion system (T4SS) (Llosa *et al.* 2002). Conjugation initiates when a protein, called relaxase, creates a nick in one of the plasmid strands at the origin of transfer (*oriT*), covalently binding to the transferred strand and, with the help of other tranfer proteins, forms a relaxosome. The relaxosome unwinds the transferred strand (T-strand) from the duplex plasmid and takes part in the conjugative replication of the remaining strand (Frost, Ippen-Ihler & Skurray 1994). This process is very similar to that of the rolling circle replication method (Waters, Guiney 1993). A coupling protein is needed to link the relaxosome with the T4SS secretion machinery (Llosa *et al.* 2002). T4SS then transfer the DNA (and the covalent-link relaxase as proven by Draper *et al.* (2005) to the recipient cells.

A model for conjugal DNA transport, using the plasmid R388 transfer system as a paradigm, is decribed (Llosa *et al.* 2002) (Figure 1-7). In R388, the coupling protein is TrwB and the relaxase protein is TrwC. The T-strand is transported in two mechanistically distinct steps. Firstly the DNA is transported through the T4SS in a passive form, as a tail that is covalently linked to the relaxase TrwC, the active substrate

for the T4SS (Draper *et al.* 2005). Secondly TrwB pumps the T-strand processively into the T4SS. This two-step mechanism explains the need for the coupling protein late in the conjugation process.

Relaxase has been proposed as a pharmaceutical target to stop the spread of conjugative plasmids. Lujan *et al.* used the F-plasmid relaxase structure to identify relaxase inhibitors and test them *in vitro*. The results show that the clinically approved bisphosphonates etidronate (Didronel) and clodronate (Bonefos), but not other bisphosphonate therapeutics, are potently effective at killing F+ cells and preventing conjugative DNA transfer (Lujan *et al.* 2007). Other studies have also identified conjugation inhibitors although their exact inhibitory mechanisms are unknown (Fernandez-Lopez *et al.* 2005). Intracellularly-expressed antibodies (intrabodies) targeting the relaxase active sites also show conjugation inhibitory activities but do not induce cell death. The therapeutic application of these intrabodies however will be difficult due to their biological stability, cell permeability, and pharmacokinetic problems faced by any macromolecular drug (Garcillan-Barcia *et al.* 2007).

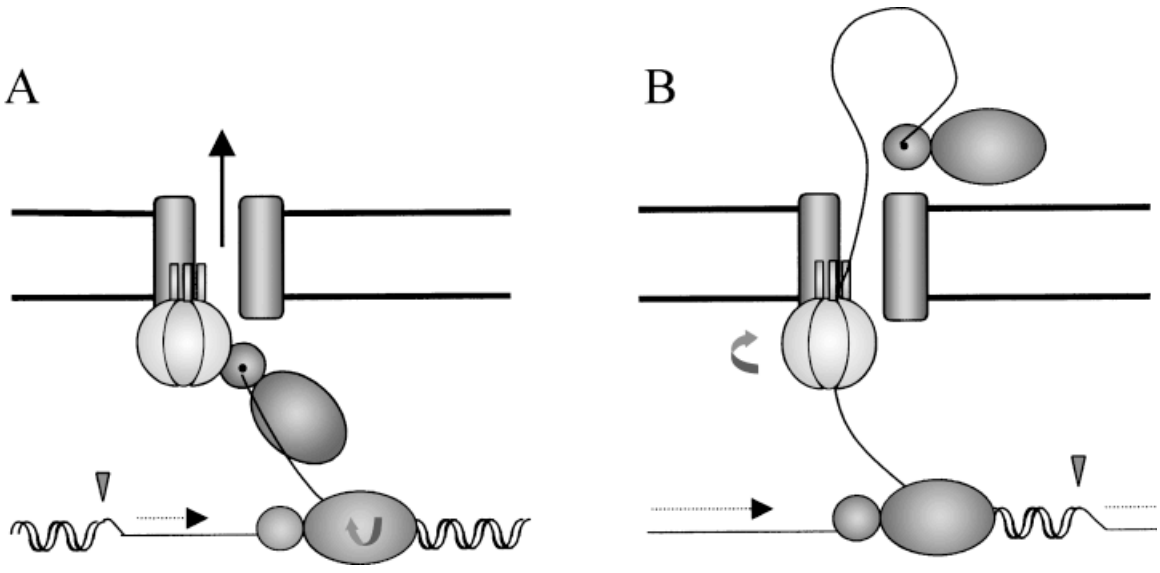


Figure 1-7 A two-step model for conjugal DNA transport.

Horizontal thick black lines represent bacterial membranes, traversed by grey cylinders that represent the T4SS. TrwC is represented as the two-domain circle + oval (relaxase + helicase) shape; TrwB is represented as a hexamer, with an orange-like shape, anchored to the inner membrane. DNA is represented by a thin black line; newly replicated DNA, by a dashed arrow. The vertical arrowhead represents the nic site. Curved arrows indicate postulated motion forces required for DNA movement .A. TrwB is coupling the T4SS and the relaxosome; a TrwC monomer covalently linked to the nicked T-strand is the substrate for T4SS secretion.B. TrwB is pumping out the T-strand as it is displaced from the donor plasmid. Upon reaching the *nic* site for the second time, the TrwC monomer in the donor would perform a second strand-transfer reaction, thus liberating the T-strand. The translocated TrwC monomer would rejoin the two T-strand ends by a reverse cleavage reaction. Figure adapted from Llosa *et al.* (2002).

1.2.2.2 Exclusion systems

The entry exclusion phenomenon was first noticed by Lederberg *et al.* when cells habouring the F sex factor were not available as conjugation recipients (Lederberg, Cavalli & Lederberg 1952). Entry exclusion is now considered an essential feature of conjugative plasmid biology, which plays a direct role in plasmid survival and is essential for the stability of a conjugative plasmid (Garcillan-Barcia, de la Cruz 2008). In the F-plasmid systems, two entry exclusion mechanisms were identified, encoded by two genes: *traT* and *traS* (Achtman, Kennedy & Skurray 1977).

TraT protein resides at the outer membrane of the cells, mediating a surface entry exclusion mechanism that interferes with the initial attachment of a donor bacterium to a potential recipient (Achtman, Kennedy & Skurray 1977). The interactive counter part of TraT in donor cell remains unknown. This type of exclusion is found only in plasmids harbouring F-like T4SSs (Garcillan-Barcia, de la Cruz 2008). The TraS protein is an inner membrane protein in recipient cell that blocks DNA transfer within stable mating pairs. In F-plasmid systems, a donor TraG protein was shown to interact with TraS (Audette *et al.* 2007). Most conjugative elements carry only one gene, the *traS*-like, for entry exclusion.

1.2.2.3 Partition systems

Plasmid partition systems ensure the equal distribution of plasmid copies into daughter cells. There are now four types of partition systems found on plasmids as summarised in Table 1-2. In this section, only the two most intensively studied types (types I and II) are discussed.

Plasmid partition requires three elements: a centromere-like DNA site, which typically consists of several tandem repeats, and two proteins: a motor protein, generally an ATPase, and a centromere-binding protein. Different partition types are classified by their motor proteins. Type I motor protein is a Walker-box ATPase whilst motor protein of type II is a actin-like ATPase. Type I partition systems can be subdivided into type Ia and Ib based on the location of the centromere site: downstream of the *par* operon in Ia and upstream in type Ib.

Table 1-2 Bacterial plasmid partition systems.

Classification	Motor protein	centromere-binding protein	Dynamic characteristic	Representative plasmid	Bacterial host
Type I	ParA/SopA: Walker-box ATPase	ParB/SopB	Oscillation	P1, F	E. coli
Type II	ParM: actin-like ATPase	ParR	Dynamic instability	R1	E. coli
Type III	TubZ: Tubulin-like GTPase	TubR	Treadmilling	pBtoxis	Bacillus thuringiensis
Type IV	Par: Unknown	Par	Unknown	pSK1	Staphylococcus aureus

Data from Schumacher 2008, Moller-Jensen and Gerdes (2007).

To start the partition process, multiple centromere-binding proteins bind to the centromere to form a higher-order nucleoprotein complex called the partition complex. The partition complex then recruits the ATPase to form the segrosome and activates the ATPase activity that drives the partition of plasmids (Schumacher 2008).

The binding of centromere-binding multimers to the centromere site forms a higher-order protein-DNA complex in which DNA topology is drastically changed. In the type I system of the F-plasmid, the centromere DNA is wrapped around a multimeric SopB protein core (Lynch, Wang 1994). The centromere of R1 plasmid binds to ParR to form a U-shaped structure: the DNA folds back on itself nearly completely, including an angle of approximately 150 degrees (Hoischen *et al.* 2008).

The mechanisms by which motor proteins separate plasmid molecules into daughter cells are less clear for the type I Walker-type ATPase than the type II actin-like ATPase. There are several models proposed to explain the partition mechanisms of type I system as reviewed in Schumacher (2008). The consensus of these models is that the segregation of plasmids is mediated by the polymerisation of ParA ATPase. The mechanism of type II partition system is better understood (Schumacher 2008), especially after the *in vivo* visualisation of the partition process by time-lapse fluorescence microscopy (Campbell, Mullins 2007). ParM filament polymerisation is triggered by the interaction with ParR. The binding of ParR to ParM also stabilises the polymer form of ParM. The two plasmids are then pushed to opposite cell poles by this insertional polymerisation of ParM. After reaching the cell poles, the plasmids collide with the cell wall, creating a force to dissociate the ParM polymer from its interaction with ParR, resulting in the destabilisation and depolymerisation of ParM filament (Figure 1-8).

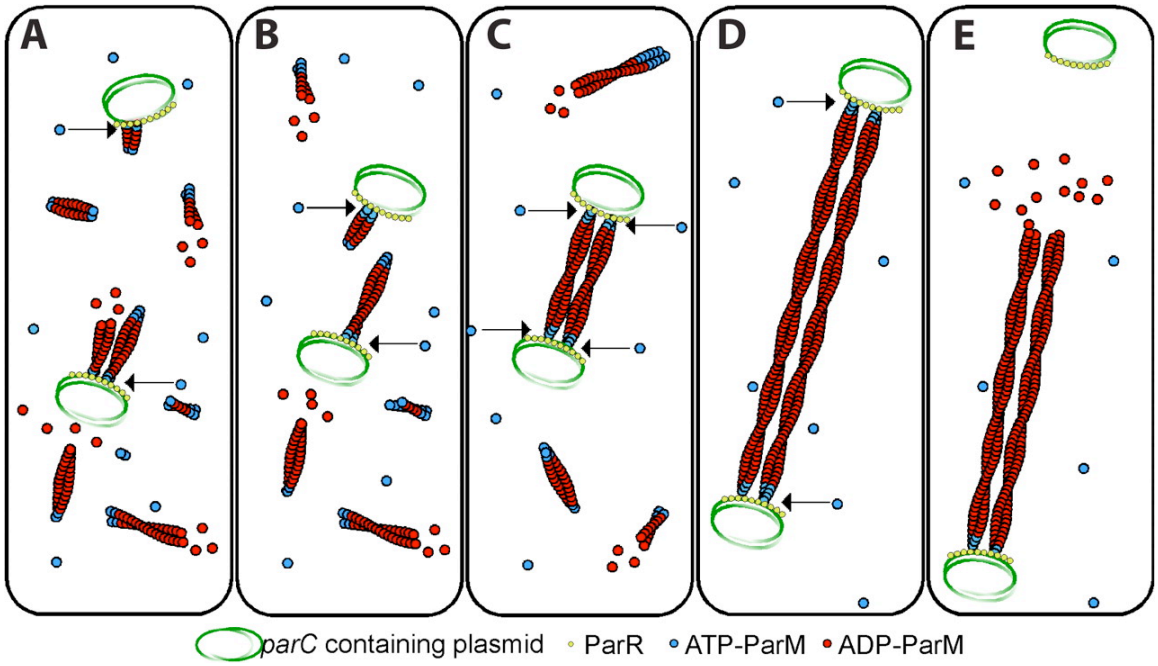


Figure 1-8 Molecular model of plasmid segregation by the R1 par operon (type II partition system). (A) Nucleation of filaments happens in the cell. (B) Plasmids attached with filaments diffuse around the cell until they encounter each other. (C) Filaments are bound at each end by a plasmid, forming a spindle. (D) These stabilized filaments polymerise, pushing the two plasmids to opposite poles. (E) when reaching a pole, the filament dissociates from the plasmid at one end and quickly depolymerise. Figure adapted from Campbell and Mullins (2007).

1.2.2.4 Post-segregational killing systems / addiction systems

Post-segregational killing systems (PSK), also known as addiction systems or toxin-antitoxin (TA) systems, were originally found in plasmids and were proposed to mediate plasmid stability (Gerdes, Rasmussen & Molin 1986). A plasmid carrying a PSK system is able to kill plasmid-free daughter cells, thereby ensuring that a growing bacterial culture predominantly consists of plasmid-containing cells (Gerdes, Rasmussen & Molin 1986). PSK systems have now been identified in many plasmids and bacterial chromosomes (Gerdes, Christensen & Lobner-Olesen 2005). There are two types of PSK systems. Type I toxins are small, hydrophobic proteins that cause damage to the cell membrane. The expression of type I toxin proteins is suppressed by the antitoxin RNAs (antisense or adjacent and divergent to the toxin gene) that bind complementarily to the toxin mRNAs. A type II PSK system consists of a small

antitoxin protein that sequesters the toxin through proteic interaction. There are eight known families of type II PSK systems, members of which are found on both plasmids and chromosomes (Gerdes, Christensen & Lobner-Olesen 2005).

The mechanisms by which the two PSK systems control plasmid stability are similar. The antitoxins (either RNAs or proteins) that neutralise the toxins are metabolically unstable. The new born plasmid-free cells inherit both the toxin and antitoxin but rapidly lose the antitoxins, leaving the toxins to either kill the cells (type I) or stop it from growing (type II).

Cooper and Heinemann have suggested that PSK systems may facilitate competition between plasmids of the same incompatibility group (Cooper, Heinemann 2000). When two plasmids (one is PSK^+ and the other PSK^-) of the same incompatibility group transfer into the same cell, incompatibility results in a high frequency of plasmid mis-segregation. In the cells that have lost the PSK^+ plasmids, the depletion of antitoxin eventually leads to cell death. The PSK^- plasmids in those cells are also effectively eliminated.

The chromosomally encoded PSK systems may also affect the resident plasmids in an "anti-addiction" manner. The PSK system on the chromosome in this case is able to counteract the toxin of the plasmid-encoded system. The daughter cells that do not inherit the plasmids therefore survive the plasmid post-segregational killing (Saavedra De Bast, Mine & Van Melderen 2008). It has also been proposed that these chromosomally encoded PSK systems provide a control mechanism that helps free-living prokaryotes cope with nutritional stress (Gerdes, Christensen & Lobner-Olesen 2005).

1.2.3 Fitness cost

It is generally accepted that bearing a plasmid initially imposes a fitness burden to the bacteria (Saavedra De Bast, Mine & Van Melderen 2008). However, after a few hundred generations of co-existence, the bacterial chromosome and plasmid can evolve into fitness-enhanced organisms in comparison to the ancestral bacteria (either with or without the ancestral plasmid) (Bouma, Lenski 1988, Dionisio *et al.* 2005, Dahlberg, Chao 2003). It has also been suggested that once a conjugative plasmid enters a bacterial population it will remain even if the original selective pressure is withdrawn (Dahlberg, Chao 2003).

1.2.4 Toward the phylogenetic analysis of plasmids isolated from clinical strains

Studies of clinical isolates often report plasmids based on their phenotypic resistance profiles, plasmid profiles (sizes of plasmid(s) as estimated by gel electrophoresis) and less frequently, plasmid RFLP patterns. These methods have already been reviewed in sections 1.1.6.1 and 1.1.6.2. The use of these methods is very useful in the typing of outbreak clinical isolates. They, however, do not give any indication about the relatedness of the plasmids. Some conjugative plasmids can be found in various bacterial species and a quick, reliable method to identify them is needed.

Plasmid incompatibility is defined as "the failure of two co-resident plasmids to be stably inherited in the absence of external selection" (Novick *et al.* 1976). This definition has been used since to define and classify bacterial plasmids (Datta, Hedges 1972). The method to test for plasmid incompatibility involves introduction of a plasmid into a strain carrying another plasmid. These two plasmids must contain different markers in order to monitor their segregation. This method poses several

technical problems including the selection of suitable markers for discriminating between the plasmids and the difficulties in introducing a plasmid into bacterial cells (if the plasmid is non-transmissible or there is an effect from entry exclusion) (Couturier *et al.* 1988).

Couturier *et al.* propose the use of DNA hybridisation as a tool to identify and classify plasmids in a method called replicon typing (Couturier *et al.* 1988). A replicon is the genes and sites on a plasmid that are required for autonomous replication and control. This method overcomes the technical problems in the previous incompatibility testing method and is also capable of identify plasmids containing more than one replicon. However, DNA hybridisation is still labour intensive to perform, especially as ~19 probes have to be used to detect ~19 commonly found replicons from *Enterobacteriaceae*, which limits its application on plasmids from clinical isolates.

Carattoli *et al.* developed a PCR-based technique to replace DNA hydridisation for replicon typing based on the probes described by Couturier *et al.* (Carattoli *et al.* 2005). This method uses multiplex PCR to detect 18 replicons representative of 18 incompatibility groups of plasmids from *Enterobacteriaceae.* This method is quick and easily applicable for clinical isolates and has been used in approximately 30 studies to describe plasmids from clinical isolates since its development in 2005.

While PCR-based replicon typing is a simple way to classify plasmids from clinical isolates, it provides no further resolution for plasmids within one incompatible group. Plasmids have been seen as primitive organisms based on their capability to disseminate through their own replication (Couturier *et al.* 1988, Datta 1985). A method to identify plasmids based on their phylogenetic relationship would undoubtedly provide insights into the dissemination and evolution of many plasmid groups within clinical bacterial

communities, which in turn will help us understand the dissemination of antibiotic resistance.

1.3 IncHI1 plasmids

1.3.1 Plasmids of incompatibility group H

Plasmids of the *Enterobacteriaceae* are categorized into groups based upon their ability (or inability) to co-exist within the same bacterial cell. Plasmids of the same incompatibility (Inc) group are closely related plasmids, with similar replication mechanisms and regulatory circuits of vital plasmid functions, and, thus, cannot survive together in the same bacterial cell (Novick 1987). More than 30 incompatibility groups have been described.

The H incompatibility complex originally comprised of two subgroups, IncH1 and IncH2, based on their similar H-pilus structures (Taylor, Grant 1977). Based on DNA-DNA filter hybridisation, Roussel and Chabbert later proposed a new subgroup, IncH3, for the plasmid MIP233, which showed very poor homology with both IncH1 and IncH2 plasmids despite a strong incompatibility with them (Roussel, Chabbert 1978). After describing a new subgroup, which is called the IncHII group, Bradley *et al.* proposed the nomenclature of the IncH plasmid groups that we still use today. Plasmids previously called IncH1, IncH2 and IncH3 now belong to the IncHI group and are called IncHI1, IncHI2 and IncHI3 (Bradley *et al.* 1982).

All IncHI1 plasmids are large molecules with sizes of 150 kb or more, all are temperature sensitive for conjugative transfer (Taylor, Levine 1980). The IncHI1 group characteristic of thermosensitive transfer (transfer efficiency is optimal at 22-30°C), suggests that IncHI1 plasmids are potential vectors for the dissemination of genes among bacterial species in water and soil environments (Maher, Taylor 1993). It is

therefore somewhat unexpected that IncHI1 plasmids have become established in human restricted *Salmonella* (*S.* Typhi and probably *S.* Paratyphi A).

The complete nucleotide sequences of two IncHI1 plasmids, R27 (180 kb) and pHCM1 (218 kp) have been determined (Sherburne *et al.* 2000, Parkhill *et al.* 2001). R27, the prototype of the IncHI1 plasmids, was first isolated in the UK from *S.* Typhimurium in 1961 (Meynell, Datta 1966). Very similar plasmids have subsequently been recovered from *S.* Typhi. The plasmid pHCM1 originated from *S.* Typhi in Vietnam in 1993 (Wain *et al.* 2003). These two plasmids share a common backbone sequence (>99% identity) comprising of 83% of pHCM1 sequence. This suggests a close evolutionary relationship and also the spread of IncHI1 plasmids from one *Salmonella* serovar to another.

1.3.2 R27 – progenitor of IncHI1 plasmids

The prototypical IncHI1 plasmid is R27, which encodes resistance to tetracycline. This plasmid has been intensively studied for over 20 years and, like all IncHI1 plasmids tested, R27 possesses three replicons, RepHI1A, RepHI1A and RepFIA-like (Couturier *et al.* 1988, Gabant, Chahdi & Couturier 1994). RepHI1A and RepHI1B have been shown to mediate plasmid replication and copy number control (Gabant, Chahdi & Couturier 1994, Newnham, Taylor 1994). Both RepHI1A and RepHI1B are specific for IncI1 plasmids. Minimal replicon analysis suggested that either of the two replicons can efficiently replicate the entire plasmid and the RepFIA-like replicon initiation protein of R27 is unable to stably replicate the entire plasmid under challenge conditions (Gabant, Chahdi & Couturier 1994). This RepFIA-like protein accounts for the one-way compatibility between the HI1 and F plasmids. This is due to a short region of similarity

in the essential replication protein of the F-plasmid (gene E) (Taylor, Chumpitaz & Goldstein 1985, Saul, Lane & Bergquist 1988).

Conjugative transfer of HI1 plasmids is of special scientific interest, not only for the fact that the conjugative machinery has to accommodate the transfer of a very large plasmid, but also in the way conjugation is regulated. There are two transfer regions (Tra1 and Tra2) encoded on R27, which were originally mapped to a R27 restriction endonuclease map by Tn5 and Tn7 insertion mutagenesis (Taylor, Chumpitaz & Goldstein 1985).

The Tra1 region is located between the coordinates 98 and 117 kb on the R27 sequence, comprising of 14 CDSs, 9 of which are essential transfer genes. The CDSs are organised into 3 operons: H, R and F (Figure 1-9) (Alonso *et al.* 2005). The origin of transfer *oriT* was also identified in this region, between the H and R operons. The R and F operons encode the mating pair formation (Mpf) components (H-pilus) (Lawley *et al.* 2002) and showed sequence homology to the IncF transfer system. The H operon contains genes encoding relaxosome proteins (traJ and traI), coupling protein (traG) and other proteins not essential for transfer. The coupling protein was shown to be ancestrally related (low identity) to that of IncP plasmids (Lawley *et al.* 2002).

The Tra2 region is sixty-three kilobases away from Tra1, spanning over ~36 kb and containing 28 CDSs. Except for four CDSs involving in partition function, the remaining CDSs are arranged into 3 operons encoding for Mpf/H-pilus components (AC and AN operons) and entry exclusion proteins (Z operon) (Figure 1-9) (Alonso *et al.* 2005). A total of 11 genes in this region have been demonstrated by mutational and genetic analysis to be essential for conjugative transfer (Lawley *et al.* 2003). Similar to Tra1, this region also exhibits a mosaic of IncF-related (9 Mpf genes) and IncP-related (*trhA* and *trhP*) sequences. In summary, the conjugative transfer system of R27 consists

of 20 essential genes located in 2 separate regions. The Mpf system is of common ancestor to IncF lineage whilst the relaxosome, pilin and peptidase show closer relatedness to IncP plasmids.

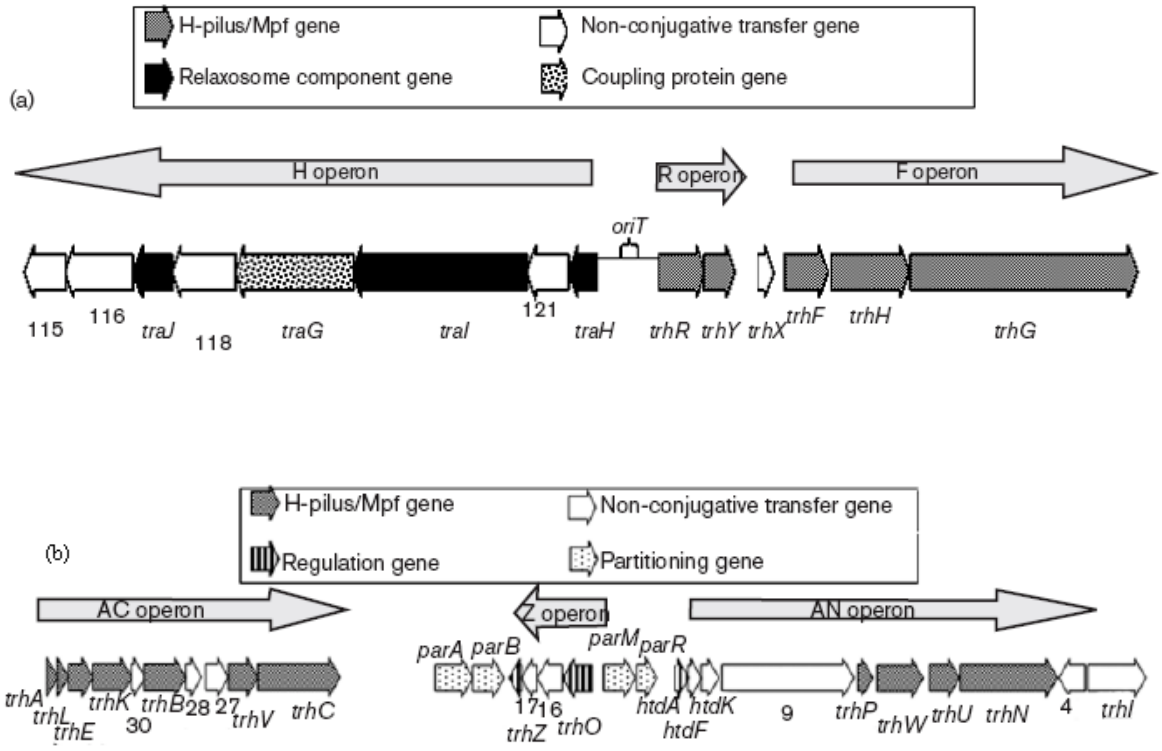


Figure 1-9 Organisation of (a) the tra1 region and (b) the tra2 region.

The CDSs are represented as pattern boxes with arrowheads indicating their orientation. Grey arrows above the CDSs indicate the operon arrangements. Figure adapted from Alonso *et al.* (2005).

1.3.3 Temperature dependent conjugation and H-NS

An interesting phenotype observed with R27 and R27-like plasmids is that they regulate their conjugative transfer in a temperature-dependent manner. This has implications for the transmission and persistence of the resistance phenotype on the typhoidal *Salmonella* population. The thermoregulation of transfer is a multi-level regulatory process involving thermo-dependent mating pair formation and the global regulatory protein H-NS (CDS R0164 in R27) (Ono *et al.* 2005, White-Ziegler, Davis 2009) and thermo-modulator Hha (R0182) (Forns *et al.* 2005). The H pilus is synthesised at

optimal temperature of 27°C by extrusion from the cell surface. When the temperature is increased to 37°C, the morphology of H pili remained stable whilst the formation of mating aggregates was inhibited, suggesting mating pair formation as the first level of thermo-regulation (Maher, Sherburne & Taylor 1993). No pili were observed by transmission electron microscopy at 37°C. H-NS-like and Hha-like proteins encoded on the R27 plasmid repress the transcription of several CDSs in both Tra1 and Tra2 regions at non-permissive temperature (33°C) (Alonso *et al.* 2005, Forns *et al.* 2005). Band shift assays suggest the inhibition mechanism of H-NS involves the interaction with the promoter regions from Tra1 and Tra2 and with the *oriT* sequence. The presence of Hha facilitates the generation of higher-order hetero-oligomers that interact with DNA (Forns *et al.* 2005).

1.3.4 The development of IncHI1 plasmids in S. Typhi

The stable maintenance of IncHI1 plasmids in *S.* Typhi was observed throughout the development of antibiotic resistance in *S.* Typhi. The first *S.* Typhi harbouring an IncH plasmid encoded resistance to chloramphenicol, tetracycline, streptomycin, and sulphonamides was reported after a very large outbreak of typhoid fever in Mexico City (Gangarosa *et al.* 1972). Two other outbreaks of IncHI1-carrying *S.* Typhi also began in the same year in Vietnam and India. The plasmids from the three outbreaks were all of the IncHI1 group and all were approximately the same size, but they differed in the phenotype conferred to the bacterial host in terms of resistance to mercury and the ability to utilise citrate (Smith, Parsell & Green 1978). Analysis of Peruvian IncHI1 plasmids from the 1980s shows that they underwent a process of modular evolution, probably by sequential acquisition of resistance determinants (Taylor, Chumpitaz & Goldstein 1985). The spread of chloramphenicol-resistant *S.* Typhi forced the change in treatment to either co-trimoxazole or ampicillin in several regions, including India and

Vietnam (Butler, Rumans & Arnold 1982). Drug resistance in *S.* Typhi developed through the early 1980s and there was evidence of plasmids of different incompatibility groups harbouring distinct resistant determinants (Ling, Chau 1984). However, there was no single plasmid harbouring resistance to all three first-line drugs (ampicillin, co-trimoxazole and chloramphenicol) until 1988 when an MDR outbreak of typhoid fever occurred in Kashmir, India (Kamili *et al.* 1993). MDR *S.* Typhi was then reported from Egypt in the same year (Mikhail *et al.* 1989), Shanghai in 1988-1989 (Zhang 1991), Qatar in 1988 (Uwaydah *et al.* 1991) and eastern India in 1989 (Anand 1993). Several isolates from these regions were shown to carry MDR plasmids of high molecular weights (150 kb to 185 kb). The MDR *S.* Typhi spread globally and by 1998 IncHI1 plasmids could be isolated from MDR *S.* Typhi worldwide (Hampton *et al.* 1998).

1.3.5 IncHI1 plasmids in S. Paratyphi A

Similar to the situation in *S.* Typhi where MDR are strongly associated with related IncHI1 plasmids, plasmid-mediated MDR was also observed although in rare incidents in *S.* Paratyphi A. Data from India as early as 1977 have suggested plasmid mediated MDR in *S.* Paratyphi A (Paramasivan, Subramanian & Shanmugasundaram 1977). Four strains of drug resistant *S.* Paratyphi A were shown to harbour IncHI plasmids conferring resistance to ampicillin, chloramphenicol, sulfamethoxazole and tetracycline in Bombay in1983 (Rangnekar, Banker & Jhala 1983). Most (75%) MDR *S.* Paratyphi A isolates in Bangladesh were found to be habouring a large transferable plasmid of 140 MDa (~212 kb) in 1992 to 1993 (Hasan *et al.* 1995).

1.3.6 The cost of IncHI1 plasmid in S. Typhi

The IncHI1 plasmids found in *S.* Typhi today have been co-evolving with their bacterial host at least since the 1970s. They are still present in many areas of the world even

though resistance to the antibiotics of choice for enteric fever (fluoroquinolones) is not encoded on IncHI1 plasmids. It is thus possible that the IncHI1 plasmid might confer no detectable cost to *S.* Typhi. This stable co-existence requires a sophisticated integration of plasmid and bacterial gene regulation networks. A plasmid encoded H-NS-like protein has been shown to confer a stealth mechanism through which the A+T-rich R27-like plasmids can enter *Salmonella* hosts with minimal impact on global gene expression patterns (Doyle *et al.* 2007). This strategy has an effect of smoothing the initial entry of extra-chromosomal genetic material, mediating the process of adaptation and integration of this new element into the regulatory network. There is, however, evidence for subtle mechanisms which confer an advantage associated with the presence of an IncHI1 plasmid. The possession of an R27-like plasmid was shown to increase the level of survival inside monocytic cell lines in S. Typhimurium (Doyle *et al.* 2007) and to facilitate higher level of bacteraemia for *S.* Typhi during typhoid fever (Wain *et al.* 1998).

However, there are currently no definitive studies to investigate the fitness cost of IncHI1 in the natural host - *S.* Typhi.

1.4 Mobile genetic elements associated with antibiotic resistant genes

Mobile genetic elements are segments of DNA that encode enzymes and other proteins that mediate the movement of DNA within genomes (intracellular mobility) or between bacterial cells (intercellular mobility) (Frost *et al.* 2005). These elements include plasmids, bacterial phages, integrons and transposons. In this section, I would like to focus on the two main mobile genetic elements that mediate the transfer of DNA intracellularly and hijack conjugative plasmids for their transfer intercellularly: integron

and transposons. An integron on a conjugative plasmid serves as a collecting point for antibiotic resistant genes whilst a transposon is a vehicle for the movement of genes into integrons, hence antibiotic resistant genes, across plasmids (even into chromosomes). Conjugative plasmids then disseminate transposons, integrons and antibiotic resistant genes between bacterial cells.

1.4.1 Transposons

Transposable elements are defined as "specific DNA segments that can repeatedly insert into one or more sites in one or more genomes". This definition was proposed by Robert *et al.* to include many different types of transposable elements that have been discovered (Frost *et al.* 2005). Table 1-3 summarises many of the transposable elements covered by this definition. For the scope of this section, I will focus on the two most commonly found on plasmids: unit transposons and composite transposons.

Table 1-3 Types of transposable elements covered by Robert's definition.

Type of transposable element	**Examples**	**Definition**
Composite transposons	Tn*5*	Flanked by IS elements. The transposase of the IS element is responsible for the catalysis of insertion and excision
Unit transposons	Tn*3*, Tn*554*, Tn*4430*	Typical unit elements encode an enzyme involved in excision and integration (DD(35)E or tyrosine) often a site-specific recombinase or resolvase and one or several accessory (e.g. resistance) genes in one genetic unit
Conjugative transposons (CTns)/integrative conjugative elements (ICEs)	Tn*916*, pSAM2, Tn*5397*	The conjugative transposons (CTns), also known as integrative conjugative elements (ICEs), carry genes for excision, conjugative transfer and for integration within the new host genome. They carry a wide range of accessory genes, including antibiotic resistance
Mobilisable transposons (MTns)/integrative mobilisable elements	Tn*4451*, SGI1, Tn*4555*	The mobilizable transposons (MTns), also known as integrative mobilizable elements (IMEs), can be mobilized between bacterial cells by other ''helper" elements that

(IMEs)		encode proteins involved in the formation of the conjugation pore or mating bridge. The MTns exploit these conjugation pores and generally provide their own DNA processing functions for intercellular transfer and subsequent transposition
Mobile genomic islands	HPI*YPS*, *SCC*mec	Some chromosomally integrated genomic islands encode tyrosine or serine site-specific recombinases that catalyze their own excision and integration but do not harbor genes involved in transfer. They carry genes encoding for a range of phenotypes. The name of a genomic island reflects the phenotype it confers, e.g. pathogenicity islands encode virulence determinants (toxins, adhesins, etc.)
Integrated or transposable prophage	Mu, phiC31	An integrated or transposable prophage is a phage genome inserted as part of the linear structure of the chromosome of a bacterium which is able to excise and insert from and into the genome
Integrated satellite prophage	P4	Bacteriophage genome inserted into that of the host which requires gene products from ‘‘helper” phages to complete its replication cycle
Group I intron	Unnamed group I intron inserted the *td* gene of *E. coli* phage T4.	Small post-transcriptionally splicing (splicing occurs in the pre-mRNA), endonuclease encoding element. Will home to allelic site
Group II intron	Ll.LtrB	Small post-transcriptionally splicing (splicing occurs in the pre-mRNA), restriction endonuclease encoding element
IStron	CdlSt1	Chimeric ribozyme consisting of a group I intron linked to an IS605 like transposase
intein	Unnamed intein inserted in the DNA helicase *DnaB* of *E. coli*	Small post-translational splicing (splicing occurs in the polypeptide), endonuclease encoding element. Will home to allelic site

This table is adapted from Frost *et al.* (2005).

1.4.1.1 Unit transposons

A unit transposon typically consists of flanking inverted repeats, an enzyme involved in its own excision and integration and one or several accessory genes in one unit. An example of a unit transopon is transposon Tn*3*. The transposon Tn*3* represents a large family of transposons characterised by the flanking inverted repeats of 38 bp and two genes, *tnpA* and *tnpR*, required for their transposition. Transposon Tn*3* also carries an ampicillin resistant gene. The transposase TnpA and resolvase TnpR within the Tn*3* family share at least 70% homology to each other.

The transposition of Tn*3* is replicative and involves two steps. The first step involves the transposase TnpA to form a cointegrate intermediate. The donor and recipient DNA are fused together and the Tn*3* is duplicated as a result. The site-specific resolvase TnpR then recognises the two *res* sites on each copy of Tn*3* in the cointegrate, cleaves the DNA, exchanges strands and rejoins them to leave one copy of the Tn*3* on the target DNA (Nollmann, Byron & Stark 2005, McCormick *et al.* 1981).

One notable member of the Tn*3* family is the transposon Tn*21* (which represents the Tn*21* subgroup), an element extensively found in clinical isolates playing a key role in the dissemination of antibiotic resistant genes. The transposon Tn*21* was found originally in plasmid R100 isolated from *Shigella flexneri* in Japan in the late 1950s (Liebert, Hall & Summers 1999). Besides encoding the typical genes for its transposition, transposon Tn*21* harbours a mercury resistant operon, a class 1 integron with an *aadA1* antibiotic resistant gene cassette and two insertion sequences, IS*1326* and IS*1353* (Liebert, Hall & Summers 1999). Transposons of the Tn*21* type are widely found in both environmental and clinical isolates of Gram-negative bacteria. The pHCM1 plasmid in *S.* Typhi encodes a version of transposon Tn*21*.

1.4.1.2 Composite transposons

A composite transposon is typically composed of one or several accessory genes flanked by IS elements. Transposition is mediated by one or other of the IS elements. Examples of composite transposons are Tn*9* and Tn*10* found in the *S.* Typhi plasmid pHCM1. Tn*9* consists of two direct repeats of the insertion sequence IS*1* flanking a region of 1,102 bp encoding a chloramphenicol resistant gene (Alton, Vapnek 1979). Tn*10* is of 9,147 bp in length. It consists of two flanking IS*10* elements , a tetracycline resistant operon (*tetRACD*) and three open reading frames of no known function (Haniford 2006).

The transposition of IS*1* leads to various outcomes including simple insertion of the element at new sites, formation of cointegrate molecules in which the donor replicon is fused to the target by flanking copies of the element, deletion of DNA adjacent to the element, inversion, and circle formation by precise excision (Rouquette, Serre & Lane 2004). The mechanisms behind these different outcomes remain unknown. The transposition of IS*10* and hence Tn*10* is better understood. It represents a non-replicative transposition mechanism, which involves the formation of a transposome. The transposome is a higher-order DNA-protein complex formed by the binding of transposase to the two ends of the transposon. The transposition starts when the transposase recognises the two ends of the transposon, forms the transposome and excises the transposome from donor DNA. The transposome then binds to target DNA and the transposase catalyses the strand transfer reaction to integrate the transposon into recipient DNA. The insertion of Tn*10* creates 9-bp direct repeats of target DNA flanking the transposon as a consequence of the strand transfer reaction (Haniford 2006).

1.4.2 Class one integrons

Integrons are 'assembly' platforms that incorporate exogenous open reading frames by site-specific recombination and convert them to functional genes by ensuring their correct expression (Mazel 2006, Hall, Stokes 1993). Integrons are classified based on the sequence homology of the *intI* gene encoding site-specific recombinases. Class I integrons are most commonly found in clinical isolates and most of the known antibiotic resistant gene cassettes belong to this class (Mazel 2006).

An integron comprises of two parts: the recombination platform and the gene cassette. The recombination platform or the "core" integron includes a site-specific recombinase (integrase) gene (*intI*) and a recombination site (*attI*). The gene cassette usually consists of one or more genes and a second type of recombination site. This second site was originally termed the 59-base element by Hall *et al.* (Hall, Stokes 1993) but the name *attC* (attachment site associated with cassettes) are now preferable in recent publications. The gene cassette is incorporated into the integron by an integrase-mediated site-specific recombination reaction between the *attI* and *attC* sites.

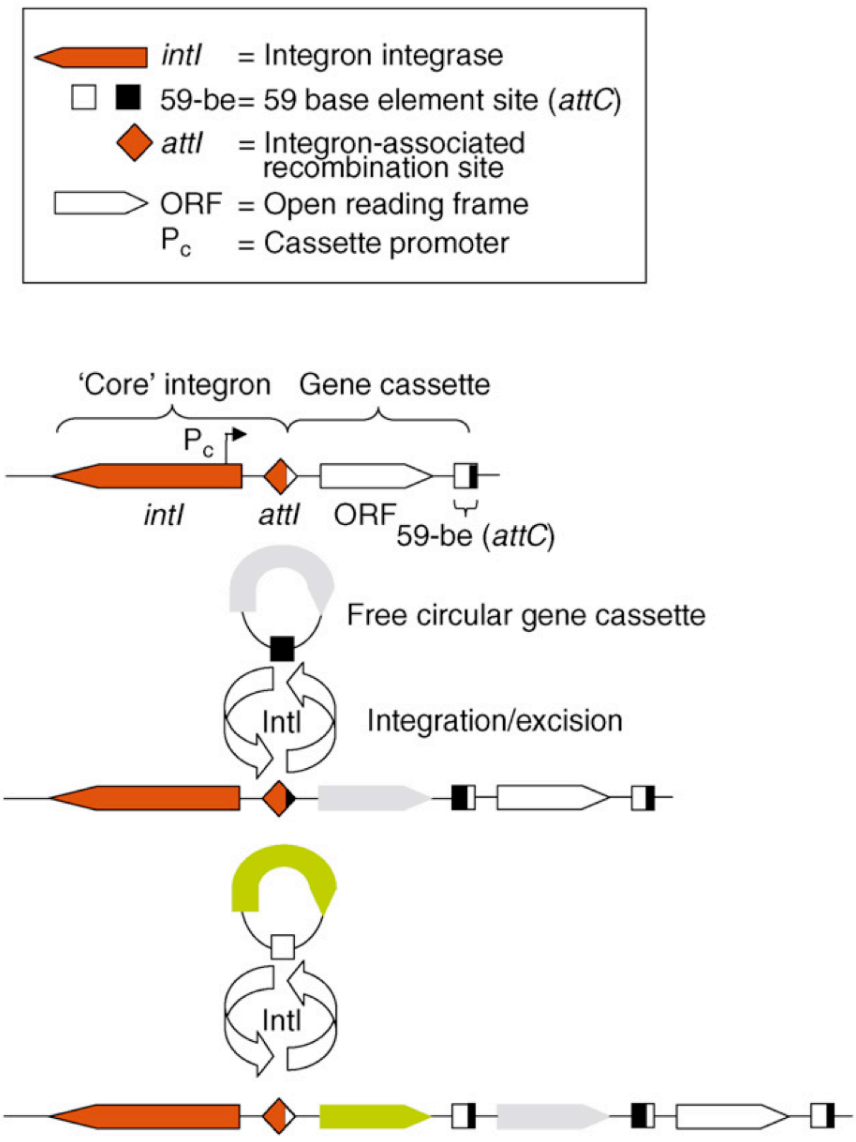


Figure 1-10 The integron-gene cassette site-specific recombination system.
Figure adapted from Boucher *et al.* (2007).

The diversity of IS elements, transposons, integrons and gene cassettes contributes to the dissemination of antibiotic resistant genes via conjugative plasmids between clinical bacterial isolates. Multiple insertion events of these elements into a plasmid and other subsequent events (inversion, deletion, and duplication) mediated by mobile elements result in a mosaic structure containing multiple resistant genes. One such mosaic structure is seen in the *S.* Typhi plasmid pHCM1 (Figure 1-11). The insertion of one transposon into another creates a "super" transposon in which one transfer event of the

outmost transposon might accommodate the transfer of all elements within. This might contribute to the rapid dissemination of multiple resistant determinants although the transfer rate is generally slow unless triggered by stress response pathways (Haniford 2006, Wardle *et al.* 2005).

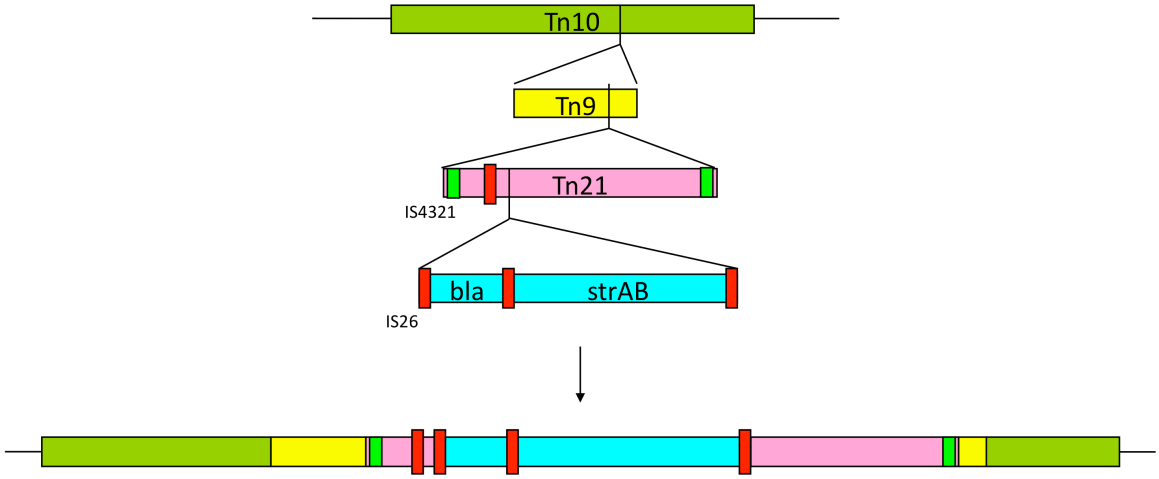


Figure 1-11 The mosaic structure of a resistant region in plasmid pHCM1 demonstrating several acquisition events of mobile elements into the plasmid.
The Tn*10* (dark green) conferring tetracycline resistance was disrupted by the insertion of Tn*9* (yellow) carrying the chloramphenicol resistant gene. Transposon Tn*21* (pink), encoded for mercury resistance and possibly contain the *bla* and *strAB* genes, inserted into Tn9. Other IS elements are also found in the regions: IS4321 (bright green) targeting the inverted repeats of Tn*21*, and IS*26* (red). The IS*26s* flanking the *bla/strAB* genes might also be responsible for the movement of these genes into Tn*21*.

1.5 Aims of the Study

Plasmids are the crucial vehicles for the dissemination of antibiotic determinants in bacterial pathogens. While plasmid biologists have long viewed plasmids as primitive organisms with vertical and horizontal replicative dissemination mechanisms, clinical researchers see plasmids as carriers for antibiotic resistant genes and describe plasmids as features of clinical isolates. As plasmids harbouring MDR genes threaten the use of antibiotics in many infections including *Salmonella*, it has become necessary to investigate the dissemination of plasmids and their stable existence in bacterial population as biological processes of primitive organisms in their own right. More importantly, because of the intra-cellular nature of plasmids, plasmid and bacterial host population should be investigated simultaneously.

In this study, we use clinical isolates of *S.* Typhi and their associated IncHI1 plasmids to explore the plasmid/bacteria relationship. We demonstrate that plasmids, as primitive organisms, have their own evolutionary history and population structure. We believe that the plasmid-bacterial host interaction could be revealed partly by studies on the capability of plasmid to exist with minimum fitness cost to the host and the capability of the host to maintain the plasmid in the most beneficial way. The relationship between plasmids and their hosts could shape the population structures of both. Hence the aims of this study are:

- An investigation of the genetic factors encoded on plasmid as well as host chromosome that are responsible for the stable maintenance of IncHI1 plasmid in *S.* Typhi.
- A dissection of the evolutionary progress and population structure of IncHI1 plasmid

- A comparison of the spatio-temporal changes in the population structures of IncHI1 plasmids and their host *S*. Typhi in order to understand their interdependent relationship.