Analysis of IncHI1 plasmids in *Salmonella enterica* serovar Typhi

by

Minh-Duy Phan

This dissertation is submitted for the degree of Doctor of Philosophy

December 2009

Darwin College

Abstract

Analysis of IncHI1 plasmids in *Salmonella enterica* serovar Typhi

Minh-Duy Phan

Plasmids play an important role in bacterial adaptation and evolution. Plasmids of incompatibility group HI1 (IncHI1) are the major vectors for the global dissemination of multiple antibiotic resistance genes among *Salmonella enterica* serovar Typhi (*S.* Typhi). *S.* Typhi is a human adapted serovar which cause the major human infection: typhoid fever. The majority of cases are in developing countries where sanitation and safe drinking water are inadequate and the true burden of disease is unknown.

This project investigated the genetic factors encoded on plasmid as well as host chromosome that are responsible for the stable maintenance of IncHI1 plasmid in *S.* Typhi. Transposon Directed Insertion-site Sequencing (TraDIS), a novel method that enables the simultaneous assay of every gene in the genome using Illumina next generation sequencing technology, was used to identify a gene set involved in plasmid stability in the bacterial host. The method successfully identified the known stability factor *sfh* among other hypothetical CDSs.

The evolution and population dynamics of IncHI1 plasmids were also studied by adapting Multi-Locus Sequence Typing for IncHI1 plasmids (PMLST). The method defined eight different plasmid sequence-types (PST), clustering into 2 groups. Group 1 was found to consist of plasmids isolated before 1993, whilst group 2 consisted of plasmids isolated after 1993. To obtain greater typing resolution on a larger strain collection, the Illumina GoldenGate SNP-typing platform was used to type both chromosomal and plasmid SNPs for 473 *S.* Typhi strains collected from 45 countries between 1916 and 2007. There is an absolute association of PST6, the predominant plasmid since 1993, with a widespread chromosomal background, H58. This suggests a competitive advantage of the ST6-plasmid/H58-haplotype combination.

In conclusion, this project demonstrates the important impact that resistance plasmids can have on the biology of a major human pathogen.

Acknowledgements

I would like to thank my supervisor Dr. John Wain for giving me the opportunity to complete this PhD. Thank you for all the guidance, discussion, supporting and especially for patience during my writing stage.

I would also like to thank Professor Gordon Dougan for supervising me after John left the Sanger Institute and for the help with my writing up. I would like to thank Dr. David Summers and my thesis committee, Dr. Julian Parkhill and Dr. Keith Turner, for their guidance.

A special thanks to Kathryn Holt, Gemma Langridge, and Keith Turner for their help in the lab and their collaboration in several projects which made up parts of my PhD.

Team 100, especially Satheesh Nair, and team 15 have been very supportive and helped me with many important little things. Thanks for taking great care of me.

I would like to extend my gratitude to all my collaborators, especially Trần Thụy Châu and Christiane Dolecek at the Oxford University Clinical Research Unit in Vietnam, François-Xavier Weill at Institut Pasteur Paris, France, among others including John Albert, Getenet Bevene, Fiona Cooke, Shanta Dutta, Rajni Gaind, Sam Kariuki and Bianca Paglietti, who helped me collect and extract DNA from clinical samples.

It is impossible to list here all the people at the Sanger Institute and elsewhere in the world who have consulted and helped me during my research. I am very grateful for their help.

My friends, Lý, Hiếu, Huy, Nguyên, Long, Quang, Hằng and many others, have been very encouraging and patiently listened to my ranting. Thanks for keeping me sane!

Last but not least, I would like to dedicate this work to my family and my partner who always believe in me and give me lots of encouragement although we have to live so far apart.

Declaration

I hereby declare that this thesis consists of work derived entirely of my own work. Due to the nature of this work, and interdisciplinary nature of biological sciences, it was not possible and impractical to perform all of these techniques, however, it was possible to design all experiments associated with this thesis. Work that was done by other persons is clearly stated in the Materials and Methods section.

This thesis is no longer than 300 pages as required by the School of the Biological Sciences.

Minh-Duy Phan

December 2009

Table of Content

List of Tables

List of Figures

Abbreviations

ALFP	Amplified fragment length polymorphism
CDC	Center for Disease Control and Prevention
CDS	Coding Sequence
Cm	Chloramphenicol
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
HIV	Human Immunodeficiency Virus
Inc	Incompatibility
Km	Kanamycin
MDR	Multi-drug resistance
MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
MLVA	Multiple loci VNTR analysis
Mpf	Mating pare formation
NCCLS	National Committee for Clinical Laboratory Standards
PCR	Polymerase chain reaction
PFGE	Pulse-field gel electrophoresis
PMLST	Plasmid multilocus sequence typing
PSK	Post-segregational killing
PST	Plasmid sequence type
RAPD	Random amplification of polymorphic DNA
RFLP	Restriction fragment length polymorphism
S. Paratyphi A	*Salmonella enterica* serovar Paratyphi A
S. Typhi	*Salmonella enterica* serovar Typhi
SNP	Single nucleotide polymorphism
Tet	Tetracycline
VNTR	Variable number of tandem repeats

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.

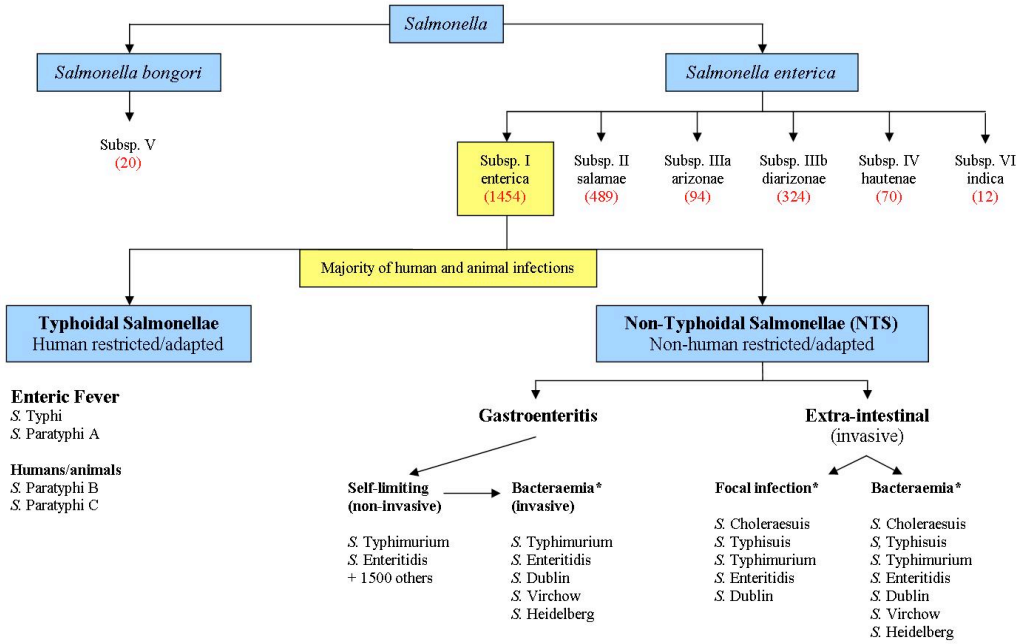


Figure 1-1 Diagram showing the classification of the *Salmonella* genus and its diseases.
Figure adapted from Langridge, Wain & Nair, 2009

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.

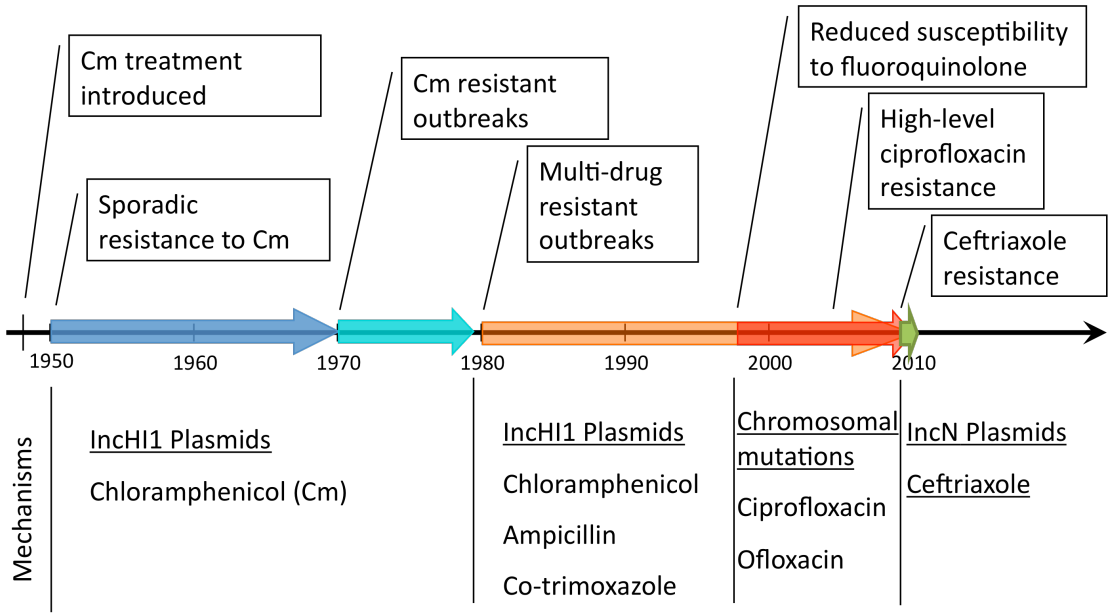


Figure 1-2 Development of antibiotic resistance in *S.* Typhi.

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).

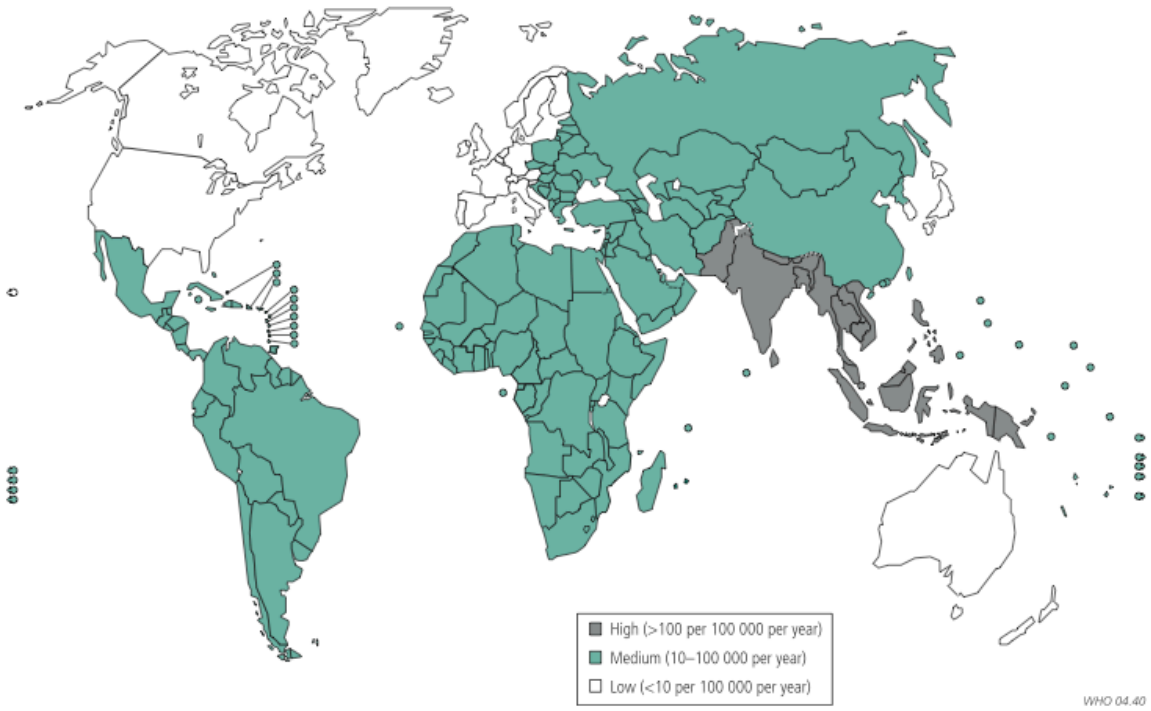


Figure 1-3 Geographical distribution of typhoid fever.
Figure is adapted from Crump, Luby & Mintz (2004).

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.

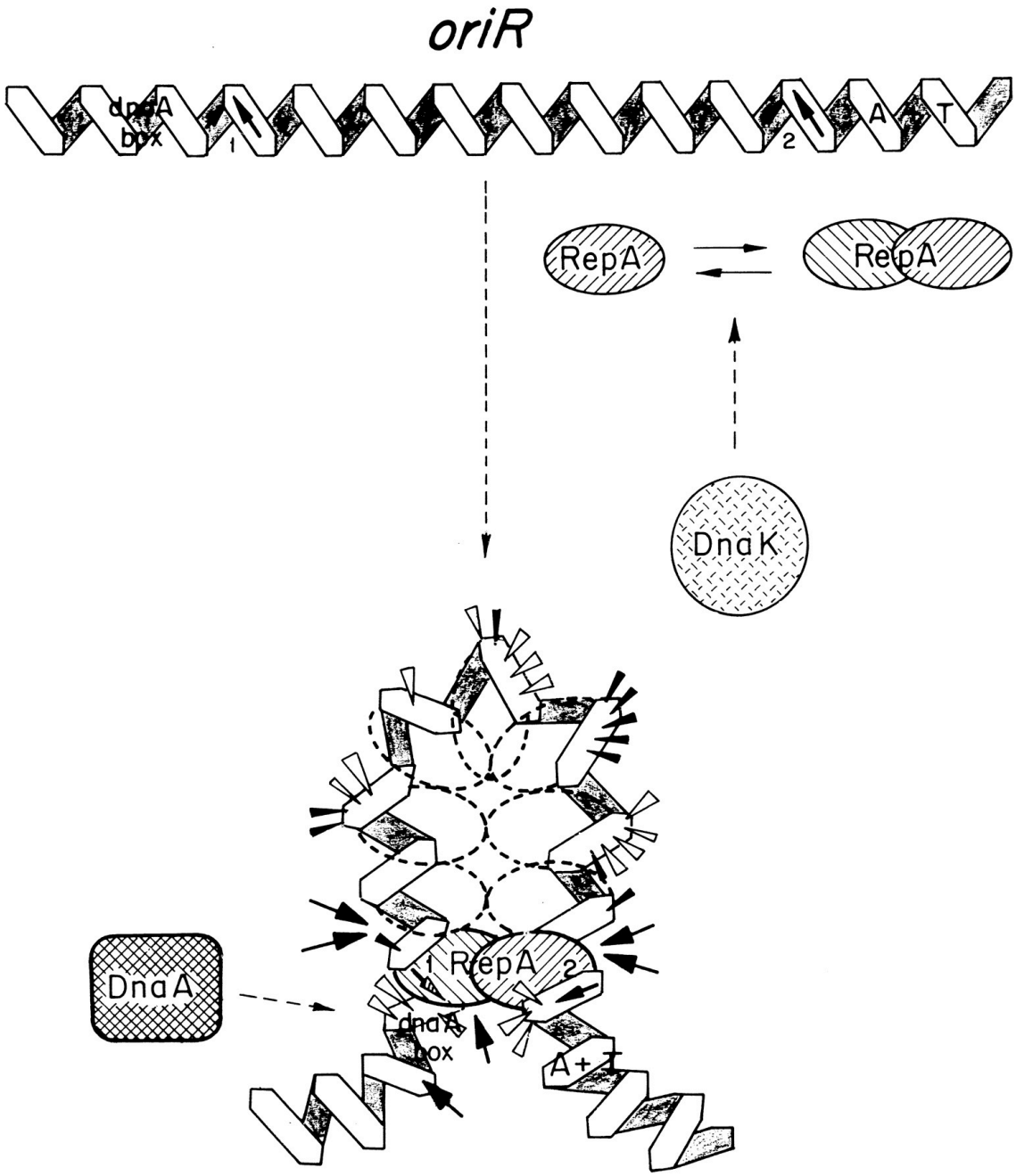


Figure 1-4 RepA-*oriR* complexes in initiation of R1 plasmid replication.
The dimer of RepA binds specifically to the two partially panlindromic sequences (site 1 and 2) to form a DNA loop, exposing the *dnaA* box for the binding of the host DnaA. The DNA loop would be filled afterwards with more RepA 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).

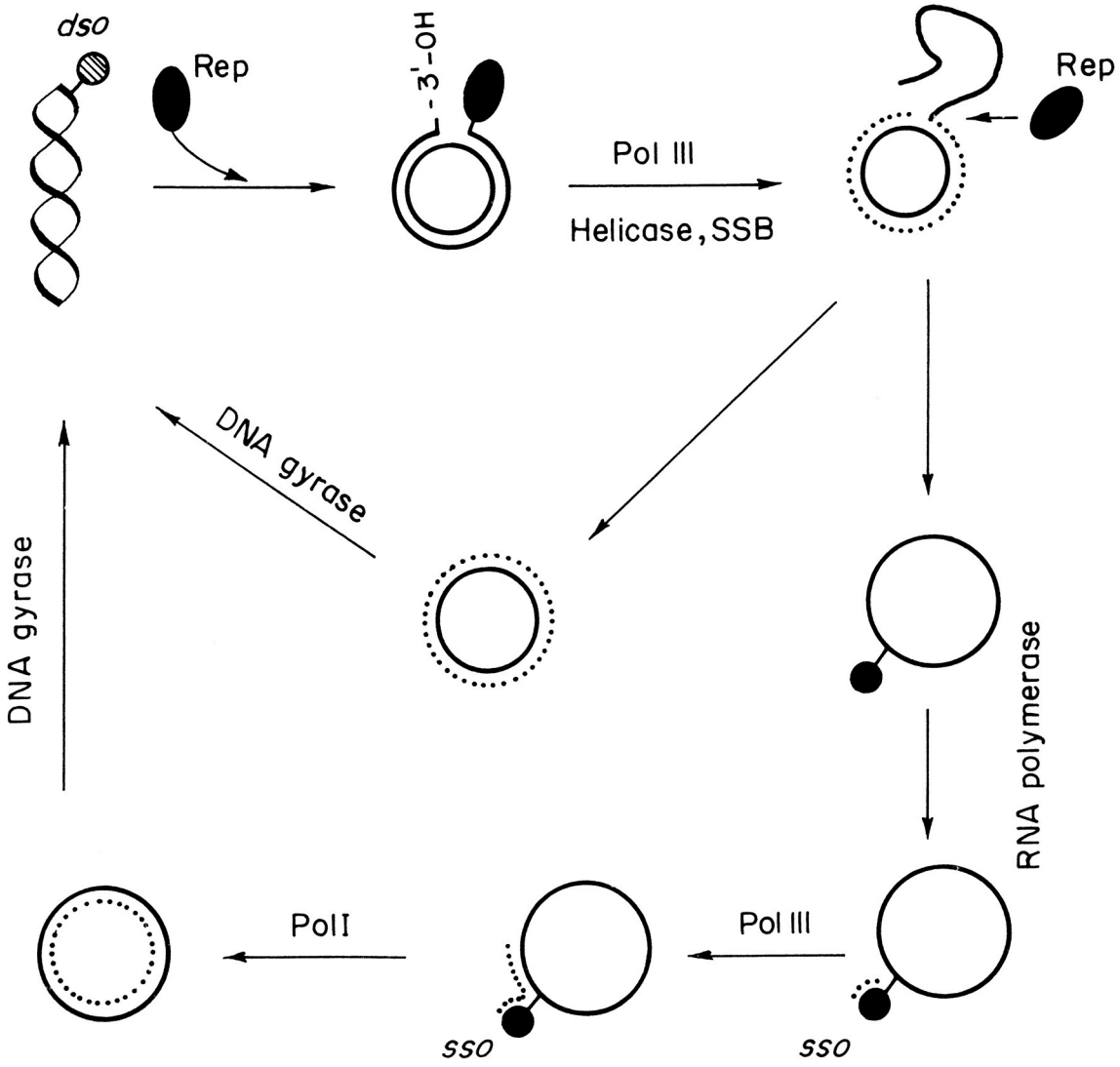


Figure 1-5 Model for rolling-circle replication.
The plasmid-encoded Rep protein recognises the *dso* on supercoiled DNA and introduces a site-specific nick generating a free 3'-OH end. This end is elongated by host proteins as the parental strand is being displaced. When the replication fork reaches 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).

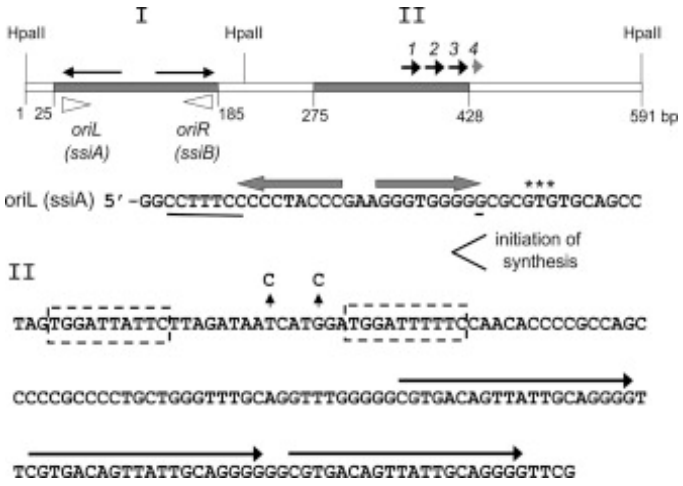
1.2.1.3 Strand displacement replication

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).

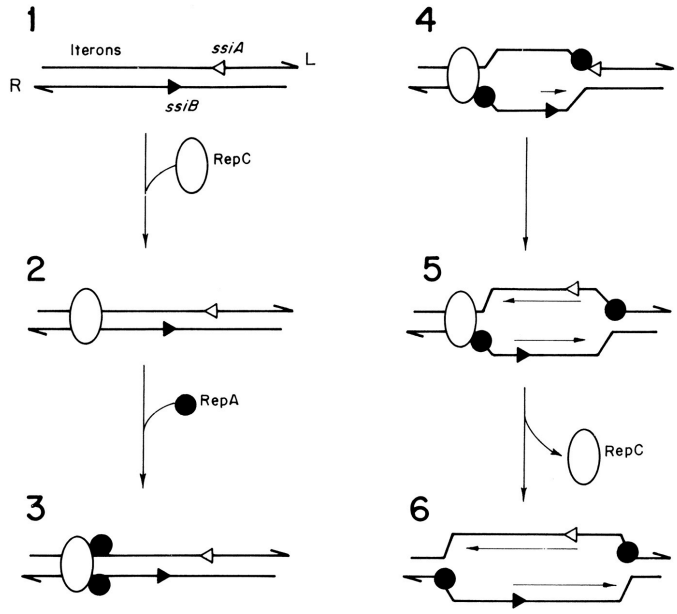
RSF1010 requires an origin of replication (*oriV)* and three plasmid-encoded proteins (RepA – helicase, RepB – primase and RepC – initiator) for replication (Sakai, Komano 1996, Diaz, Staudenbauer 1982). These three proteins are keys for the plasmid's own initiation of replication, thus explaining the 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).

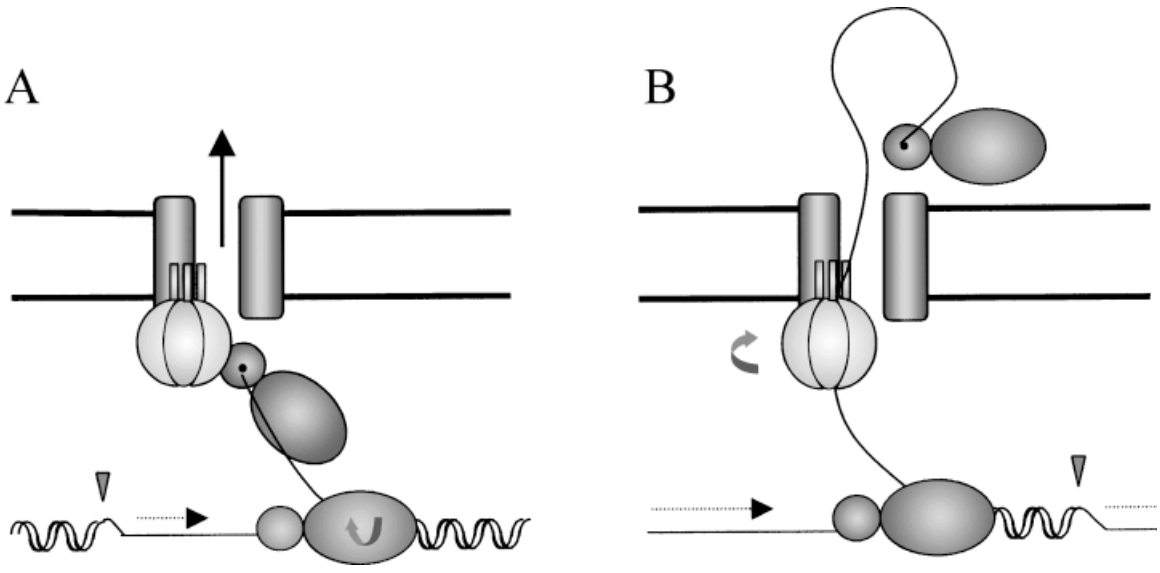


Figure 1-7 A two-step model for conjugal DNA transport.

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).

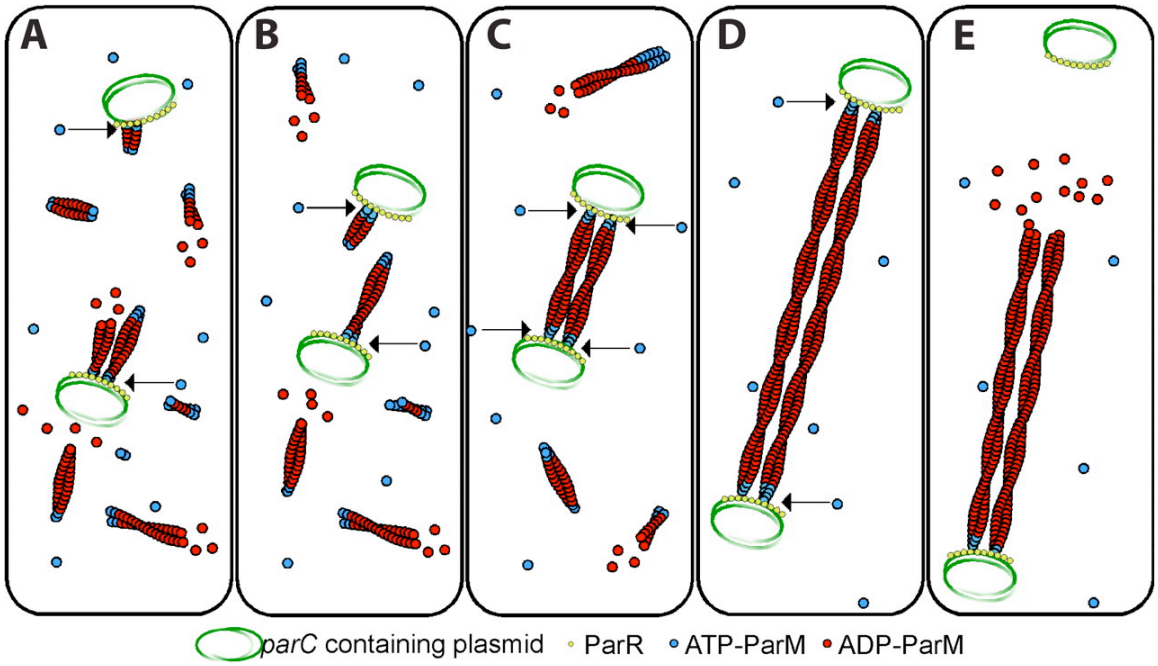


Figure 1-8 Molecular model of plasmid segregation by the R1 par operon (type II partition system). (A) Nucleation of filaments happens in the cell. (B) Plasmids attached with filaments diffuse around the cell until they encounter each other. (C) Filaments are bound at each end by a plasmid, forming a 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.

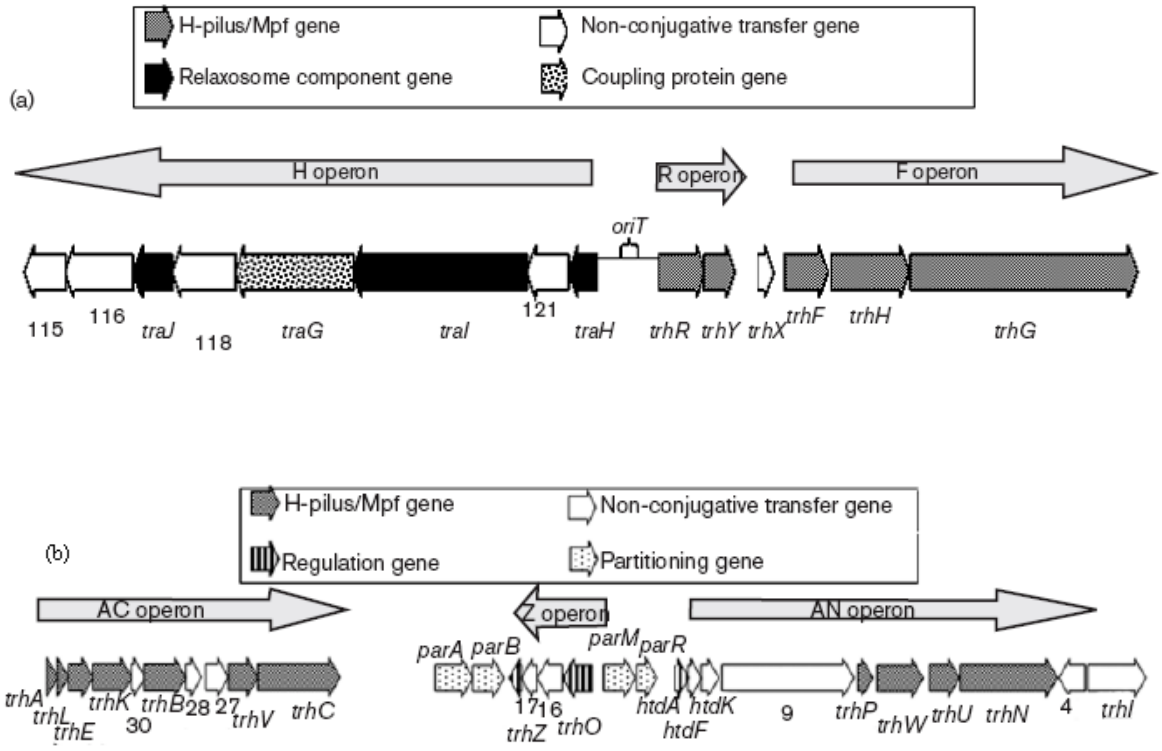


Figure 1-9 Organisation of (a) the tra1 region and (b) the tra2 region.

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.

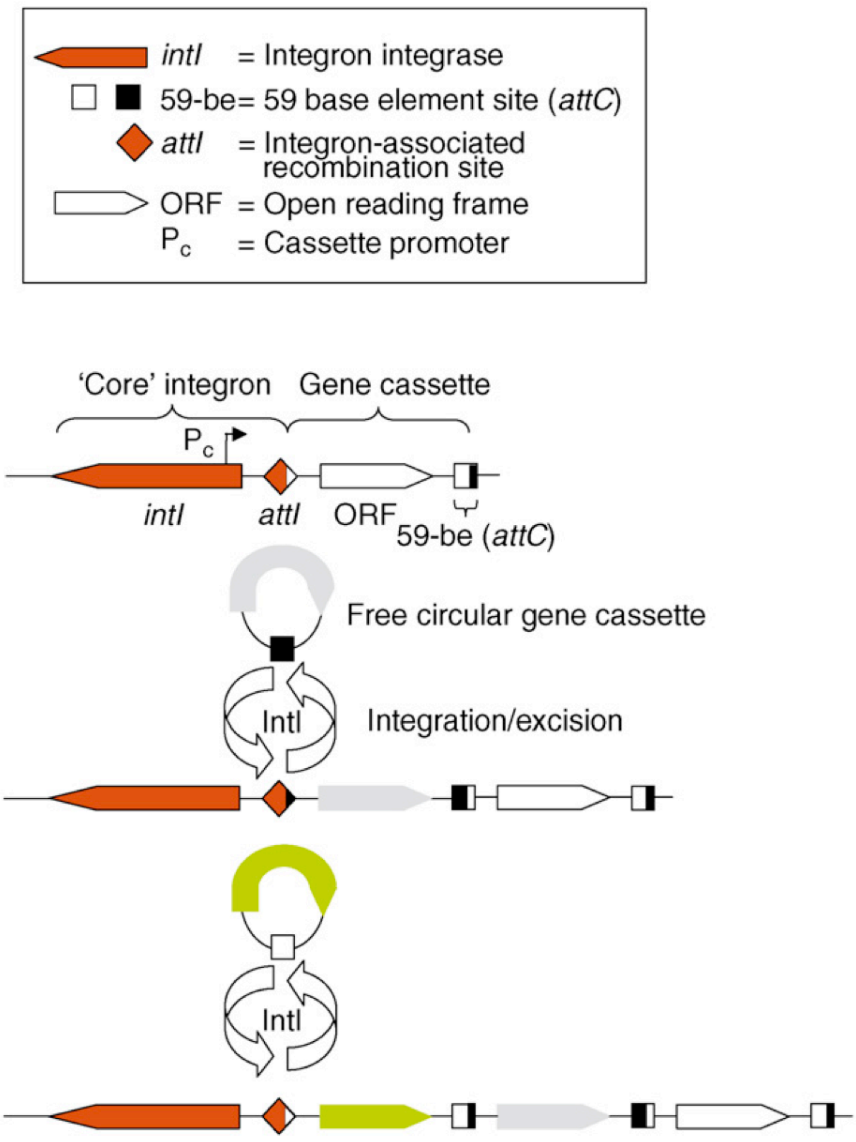


Figure 1-10 The integron-gene cassette site-specific recombination system.
Figure adapted from Boucher *et al.* (2007).

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).

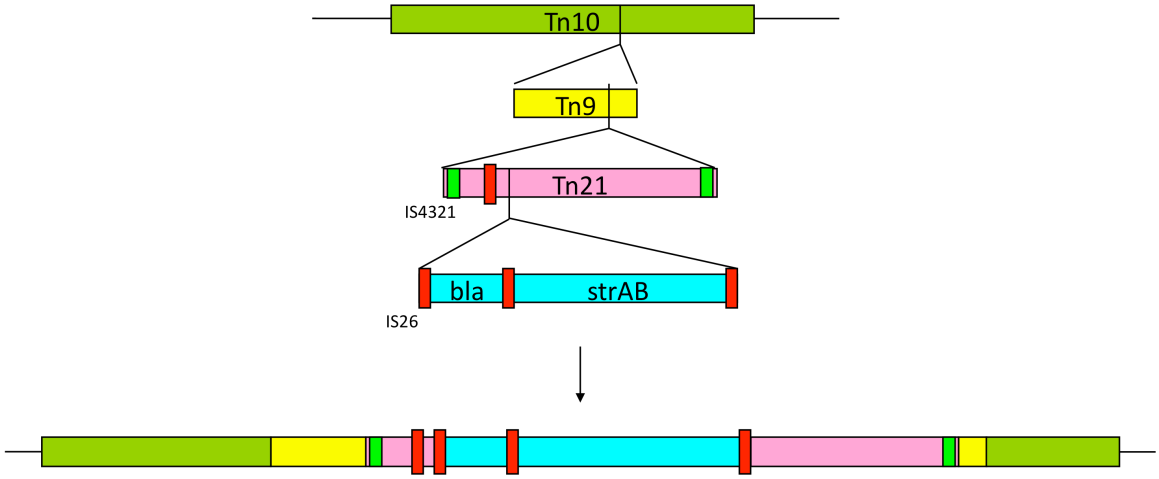


Figure 1-11 The mosaic structure of a resistant region in plasmid pHCM1 demonstrating several acquisition events of mobile elements into the plasmid.
The Tn*10* (dark green) conferring tetracycline resistance was disrupted by the insertion of Tn*9* (yellow) carrying the chloramphenicol resistant gene. 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.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals and reagents

General laboratory chemicals were obtained from either Sigma (Poole, UK) or Oxoid (Basingstoke, UK) unless otherwise stated. Buffers were prepared as aqueous solution in distilled H_20 according to standard methods, and adjusted to the required pH. When required, solutions were sterilized by autoclaving (121°C, for fifteen minutes) or by filtration (pore size 0.43 μm). The following standard buffers were used:

- TE (Tris-EDTA): 10 mM Tris, 1 mM EDTA, pH 8.0
- TAE (Tris-Acetate-EDTA): 40 mM Tris-Acetate (BRL-Gibco), 1 mM EDTA pH 8.0

2.1.2 Plasmid collection

The plasmids pCR2.1-Tn5, pACYA184 and suicide vector pJCB12 were provided by Dr. Keith Turner.

2.1.3 Bacterial strains and isolates

Bacterial isolates are listed in Table 2-1. Stocks of bacteria were stored in 25% glycerol at -80°C. When required a scrape of glyrerol-culture was removed and streaked onto an agar plate. Broth cultures were inoculated with a single colony.

Many of the isolates included in this study were sent to us as genomic DNA by our collaborators.

Table 2-1 Bacterial isolates used in this study

Experiment	Organism	Source
TraDIS	BRD948: attenuated Ty2-derived strain CVD908-	Keith Turner[a]

Experiment	Organism	Source
	htrA with deletion mutations in *aroC*, *aroD* and *htrA* (Tacket *et al.* 1997). WT26 (pHCM1): BRD948 with a point mutation in gyrase A (Ser83Phe) (Turner, Nair & Wain 2006) harbouring pHCM1 plasmid	
sfh knock out	BRD948 with pHCM1 plasmid *E. coli* CC118λpir: araD139 (del)(ara, leu)7697 (del)lacX74 phoA(del)20 galE galK thi rpsE rpoB argE(am) recA1 (lambda)pir (Herrero, de Lorenzo & Timmis 1990)	Keith Turner[a]
PMLST	11 *E. coli* transconjugants harbouring IncHI1 plasmids from Vietnam 5 *E. coli* transconjugants harbouring historical plasmids from 1970s 16 *S.* Paratyphi A from Pakistan 2 *S.* Typhi DNA from Jordan	John Wain[a] Henry Smith[b] Rumina Hasan[c] Bianca Paglietti[d]
SNP typing*	157 *S.* Typhi DNA from Vietnam, Laos, Bangladesh 2 *S.* Typhi from UK 8 *S.* Typhi from India 46 *S.* Typhi DNA and 2 *S.* Paratyphi A from India 4 *S.* Typhi DNA from Jordan 2 *S.* Typhi from Ethiopia 26 *S.* Typhi from Kuwait 22 *S.* Typhi from Kenya 3 *S.* Typhi DNA from Murray's colletion 2 *S.* Typhi DNA from the US 176 *S.* Typhi DNA from various countries (Pasteur's Institute collection) 19 *S.* Typhi DNA from re-sequencing (Holt *et al.* 2008) 8 plasmid control strains from PMLST above	OUCRU[e] Fiona Cooke[f] Rajni Gaind[g] Shanta Dutta[h] Bianca Paglietti[d] Getenet Bevene[i] John Albert[J] Sam Kariuki[k] Derek Pikard[a] CDC[m] François-Xavier Weill[l] Kathryn Holt[a]
Growth curves and plasmid competition	BRD948 BRD948 with pSTY7 plasmid BRD948 with pHCM1 plasmid	Keith Turner[a]

* Note: DNA extracted and sent to us by our collaborators. I extracted the DNA from OUCRU during a two-week visit. See appendix 8.5 for details of each isolate.

[a] The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

[b]Health Protection Agency, Colindale, London, United Kingdom

[c]Aga Khan University Medical Centre, Karachi, Pakistan

[d]Università 23 degli Studi di Sassari, Italy

[e]Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam

[f] Addenbrooke's Hospital, Cambridge, UK

[g] Safdarjung Hospital, New Delhi, India

[h] National Institute Of Cholera And Enteric Diseases, Kolkata, India

[i] Department of Microbiology, Immunology and Parasitology, Jimma University, Ethiopia
[j] Department of Microbiology, Faculty of Medicine, Kuwait University, Kuwait
[k] Centre for Microbiology Research, Kenya Medical Research Institute, Kenya
[l] Institut Pasteur, Laboratoire des Bactéries Pathogènes Entériques, Paris, France
[m] Center for Disease Control and Prevention

2.1.4 Bacterial culture media

Bacteria were routinely cultured in Luria-Bertani (LB) broth (Oxoid) prepared as per manufacturers instructions. One litre of standard Luria Bertani broth (pH 7.0) contains 10g of bacterial tryptone (Oxoid), 5g yeast extract (Oxoid) and 10g NaCl (Sigma). LB agar consisted of LB broth supplemented with 1.5% (w/v) bacto agar. Iso-sensitest, Xylose lysine deoxycholate agar (XLD agar) and Cysteine Lactose Electrolyte Deficient (CLED) agar were all obtained from Oxoid and made as per manufacturers instructions. Culture media was regularly supplemented with antibiotics and aromatic mix (for attenuated *S.* Typhi) (Table 2-2).

Table 2-2 Supplements used in culture media

Supplement	Making Stock	Concentration in media
Chloramphenicol (1000 µg/mL)	Chloramphenicol 40mg; ethanol 20 mL; distilled water to final volume of 40 mL	15 µg/mL
Ampicillin (100 mg/mL)	Ampicillin 3g; add distilled water to final volume of 30 mL	200 µg/mL
Tetracycline (1000 µg/mL)	Tetracycline 40 mg; distilled water to final volume of 40 mL	15 µg/mL
Kanamycin (1000 µg/mL)	Kanamycin 40mg; distilled water to final volume of 40 mL	20 µg/mL
Nalidixic acid (1000 µg/mL)	Nalidixic acid 40 mg; 20 mL water, a few drops of NaOH 1M to dissolve; distilled water to final volume of 40 mL	15 µg/mL
Aromatic Mix (100x)	Phenylalanine 0.4g; tryptophan 0.4g; para aminobenzoic acid 0.1g; dihydro benzoic acid 0.1g; per 100 mL solution	1x

2.1.5 Enzymes

Restrictions endonucleases, T4 ligase, T4 polynucleotide Kinase were purchased from New England Biolab, UK.

Taq DNA polymerases were purchased from Invitrogen and Bioline.

PfuUltra™II Fusion HS DNA Polymerase was from Stratagene.

Deoxytriphosphate nucleotides dATP, dGTP, dCTP and dTTP were purchased as a 10mM mix (Invitrogen, UK) or as dNTPs mix 100 mM total (Bioline, USA).

2.1.6 Oligonucleotides

All oligonucleotide primers used in this study were purchased from Sigma-Genosys (Sigma-Aldrich, UK) and were purified by reverse-phase cartridge purification (RP1). Oligonucleotides were re-suspended to a concentration of 100 µM with tissue culture grade sterile, distilled water (Sigma, UK) and stored at –20°C. Stocks of 5 µM concentrations of primers were made using sterile distilled water and were also stored at -20°C.

2.2 Microbiological methods

2.2.1 Bacterial culture

Broth cultures of bacteria were incubated at 37°C with aeration (200rpm) in an air incubator (5ml LB broth in 20ml tube) unless otherwise stated. Plate cultures were incubated overnight at 37 °C. Bacteria were harvested from small volume cultures (<2 ml) by centrifugation using a bench top centrifuge at 8000 rpm for 3-5 minutes. For larger volumes, centrifugation was performed at 4000 rpm for 10-15 minutes, using a Sorvall centrifuge.

2.2.2 Identification of bacteria

2.2.2.1 Bacterial culture

The identification of *S.* Typhi isolates was confirmed as follows. *S.* Typhi isolates were cultured using XLD and CLED agar and incubated at 37°C overnight. Plates were inspected after 18hrs and then left at room temperature for a further 8hrs.

2.2.2.2 Slide agglutination

Bacterial isolates used in this study were all identified prior to use as *S.* Typhi by slide agglutination using anti-Vi and anti-09 antisera (Murex). Anti-04 rabbit antisera were used as a negative control in all cases (Murex). A fine suspension of *S.* Typhi was made in sterile saline (20 µl) on a microscope slide. An equivalent volume of antisera was added to the suspension and mixed evenly. The slide was gently rotated for approximately 15 – 30 seconds and the presence of agglutination was inspected by eye against a dark background. *S.* Typhi was identified positive by either Vi or 09, and negative by 04.

2.2.2.3 Biochemsitry

In cases that require further confirmation, following identification of *S.* Typhi by culture on XLD and CLED and agglutination with specific antisera, isolates were confirmed as *S.* Typhi by biochemical analysis. Biochemistry was carried out using API 20E test strips (Biomeriuex) as per manufacturers instructions.

2.2.3 Antimicrobial susceptibility testing using disc diffusion

Disc diffusion tests were performed using a modified Kirby-Bauer method and interpreted using zone size standards for members of the Enterobacteriaceae (CLSI, 2003).

S. Typhi isolates were tested against ampicillin, trimethoprim, tetracycline, sulphamethoxazole, chloramphenicol and nalidixic acid. *E. coli* NCTC 10418 (Gift from Fiona Cooke, The Wellcome Trust Sanger Institute, Cambridge, UK) was used as control strains for all antibiotic discs. A suspension of test bacteria with a opacity equivalent to a McFarland 0.5 standard was inoculated onto an Iso-Sentitest agar plate with an exact depth of 4mm using a sterile cotton swab. The following antibiotic discs were applied and the plates were incubated at 37°C overnight; ampicillin (10µg), chloramphenicol (30µg), tetracycline (30µg), trimethoprim (1.25µg), nalidixic acid (30µg) and sulphamethoxazole (25µg). Zone sizes were measured and interpreted (CLSI, 2003). Plates were only read if the inoculum gave growth with confluent colonies.

2.2.4 Growth curves

2.2.4.1 By colony counting method

Bacterial broths were started by the inoculation of 1 ml of overnight cultures into 100 mL LB broth (1 in 100) in a 250 ml flask. A sample of 100 µl was taken out every 0.5 hour (from time point 0 to 3 hours), then at 5 hour, 7 hour and 24 hour. Each sample was serial-diluted 10 times to 10^{-6} and 4 drops of 20µl of every dilution were spotted on LB agar plates and incubate at 37°C overnight before the colonies were counted. The colony forming unit at each time point was calculated taking into account the dilution factor and then plot on the X-Y plot in Excel. The growth of each isolate was done in duplicate.

2.2.4.2 By measurement of OD_{600} by Optima plate reader

Overnight bacterial cultures were measured for absorbance at OD_{600} and then diluted by distilled water to the cell suspension of 0.1 OD_{600} before 1 µl of the cell suspension was

used to inoculate 200 µl LB broth in a well of a 96-well plate. This initial dilution was to ensure uniform starting inoculum in each well. Each strain was inoculated to at least 3 wells (3 biological replicates). The 96-well plate was then covered with an optical seal to allow the measurement of absorbance as light travel from the bottom to the top of each well. The measurement was carried out in the Optima plate reader (BMG Labtech, Germany), which took OD measurement every 15 minutes for the whole incubation time of 24 to 48 hours. The bacteria in the plate were grown inside the plate reader at 37°C with shaking at 300 rpm. The absorbance data were collected and saved in Excel format for further analysis.

2.2.4.3 Calculation of bacterial generation time

The bacterial growth rate during the exponential phase was calculated according to (Monod 1949). Where R is the exponential growth rate (number of division per unit time), x_1 is the number of cells (or the absobance OD_{600}) at time point t_1 and x_2 is the number of cells (or the absobance OD_{600}) at time point t_2, the exponential growth rate is calculated as followed:

$$R = \frac{\log_2 x_2 - \log_2 x_1}{t_2 - t_1}$$

The bacterial generation time G is equal to $1/R$. For *S.* Typhi growing in the condition specified in section 2.2.4.2, the exponential phase was chosen between 3 and 4 hour after inoculation.

2.3 Molecular methods

2.3.1 Genomic and plasmid DNA extraction

2.3.1.1 DNA extraction by Promega kits

Promega's Wizard Genomic DNA purification kit was used to extract genomic DNA for general PCR-based application. Manufacturer's instructions were followed with small modification. Briefly, 1 ml of overnight culture was harvested, the cells were lysed using 600 µl of Nuclei lysis solution at 80°C for 5 min and then RNA was removed by 3µl of RNase st 37°C for 30 minutes. Protein was precipitated by adding 200µl of Protein Precipitation solution, vortex rigorously and incubate on ice for 5 minute. The tube was then centrifuged at maximum speed for 10 to 30 minutes to ensure all protein precipitate was removed from the supernatant. The supernatant was collected in a clean tube and DNA was precipitated by adding 600µl of isopropanol at room temperature. DNA pellet was collected by centrifugation at maximum speed for 5 minutes. DNA pellet was washed by 70% ethanol followed by another centrifugation. The supernatant was then removed and DNA was air dry for 10-30 minute. DNA pellet was then rehydrated by adding 50 µl of TE and store at 4°C overnight.

2.3.1.2 Plasmid extraction by Qiagen kits using QIAGEN-tip 100

The isolation of large plasmids (>150kb) was performed using QIAGEN Plasmid Midi Kit. Manufacturer's instruction was followed with recommendation for low-copy number plasmids, including using 500 mL of bacterial culture for the QIAGEN-tip 100. The plasmid DNA was dissolved in 50 µl of TE and stored at 4°C.

2.3.1.3 PCR product clean up

PCR products were purified using QIAquick PCR Purification Kit as instructed by the manufacturer. PCR products were purified for subsequent enzymatic reactions (ligation, PNK) or sequencing.

2.3.1.4 Determination of DNA concentration

DNA concentration was determined by either the Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific) or using the Quant-iT fluorescence kit (Invitrogen). For the Nanodrop, one µl of sample DNA was applied directly to the machine's pedestal for measurement. The Invitrogen's Quant-iT dsDNA Assay kit – Broad range can be used for the measurement of multiple samples in 96-well plates (using a plate fluorescence meter such as the Optima plate reader – BMG labtech) or single sample using the Qubit quantitation platform (Intitrogen). The measurement uses 1 µl of sample DNA followed the manufacturer's instruction.

2.3.1.5 Agarose gel electrophoresis

Electrophoresis was carried out on 0.5% to 1% agarose gels, depending on estimated size of fragments. Agarose gels were prepared by boiling agarose (Invitrogen molecular biology grade) in TAE buffer. Once cooled, ethidium bromide was added (final concentration 0.5 – 1.0 µg/ml). DNA samples were loaded with one-sixth volume of bromophenol blue loading dye. A constant voltage (80-100 V) was applied across the electrodes, and amplification products were visualised under ultraviolet light after staining with ethidium bromide.

2.3.2 Primer design and PCR

Oligonucleotide primers were designed using the online service Primer3 (http://frodo.wi.mit.edu/primer3/) (Rozen, Skaletsky 2000) with the default primer criteria except for the "Max Tm difference" was 5°C. Primers for the *sfh* knock-out experiment however were designed by the program Vector NTI 9.0 (Invitrogen). For primers of overlap primer extension PCR, the overlapping regions on the primers were manually added and subsequent oligonucleotides were checked for their thermodynamic properties by Vector NTI.

A standard PCR condition was used for general PCR unless otherwise stated. PCRs were performed on a TETRA DNA Engine Peltier Thermal Cycler (MJ Research) with a reaction consisting of 1.2µl of Mango PCR buffer 10x, 1.5 mM of $MgCl_2$, 25µM of each dNTP, 1.25U of Mango Taq (Bioline), 0.3µM of each primer, 1.0µl of DNA template and nuclease free water to the total reaction volume of 12 µl. PCR program for product < 1kb was as followed: 5 min at 94°C, 30 cycles of 15s at 94°C, 15s at 58°C, and 60s at 72°C; and a final extension of 5 min at 72°C.

2.3.3 Plasmid replicon-typing by PCR

Salmonella isolates were examined for the presence of 18 plasmid replicons using three multiplex panels (Table 2-3). Template DNA was extracted using a Promega Wizard Genomic DNA purification kit (section 2.3.1.1). The PCR replicon typing of plasmid was done following the protocol described by Carattoli *et al,* 2005, and Johnson *et al* 2007. In brief, PCRs were performed on a TETRA DNA Engine Peltier Thermal Cycler (MJ Research) with a reaction consisting of 2.5µl of Mango PCR buffer 10x, 1.5 mM of $MgCl_2$, 25µM of each dNTP, 1.25U of Mango Taq (Bioline), 0.5µM of each primer (the

primer list for each panel in Table 2-3), 1.0µl of DNA template and nuclease free water to the total reaction volume of 25µl. Conditions used for PCR were used as follow:
5 min at 94°C; 30 cycles of 30s at 94°C, 30s at 60°C, and 90s at 72°C; and a final extension of 5 min at 72°C. Amplicons were visualized on 1.5% Tris-acetate-EDTA agarose gels alongside a 1-kb ladder (Bioline), and if an amplicon of the expected size was observed, then an isolate was considered positive for the corresponding replicon. Genomic DNA of a *Salmonella* isolate without plasmid was used as a negative control.

Table 2-3 Primers for PCR replicon-typing in three multiplex panels

Multiplex panels	Primer name	Detected replicon (product size - bp)	Sequence
1	K/BFW	B/O (159)	GCGGTCCGGAAAGCCAGAAAAC
1	B/ORV		TCTGCGTTCCGCCAAGTTCGA
1	FICFW	F_{IC} (262)	GTGAACTGGCAGATGAGGAAGG
1	FICRV		TTCTCCTCGTCGCCAAACTAGAT
1	A/CFW	A/C (465)	GAGAACCAAAGACAAAGACCTGGA
1	A/CRV		ACGACAAACCTGAATTGCCTCCTT
1	PFW	P (534)	CTATGGCCCTGCAAACGCGCCAGAAA
1	PRV		TCACGCGCCAGGGCGCAGCC
1	TFW	T (750)	TTGGCCTGTTTGTGCCTAAACCAT
1	TRV		CGTTGATTACACTTAGCTTTGGAC
2	K/BFW	K/B (160)	GCGGTCCGGAAAGCCAGAAAAC
2	KRV		TCTTTCACGAGCCCGCCAAA
2	WFW	W (242)	CCTAAGAACAACAAAGCCCCCG
2	WRV		GGTGCGCGGCATAGAACCGT
2	FIISFW	F_{IIS} (270)	CTGTCGTAAGCTGATGGC
2	FIISRV		CTCTGCCACAAACTTCAGC
2	FIAFW	F_{IA} (462)	CCATGCTGGTTCTAGAGAAGGTG
2	FIARV		GTATATCCTTACTGGCTTCCGCAG
2	FIBFW	F_{IB} (702)	GGAGTTCTGACACACGATTTTCTG
2	FIBRV		CTCCCGTCGCTTCAGGGCATT
2	YFW	Y (765)	AATTCAAACAACACTGTGCAGCCTG
2	YRV		GCGAGAATGGACGATTACAAAACTTT
3	I1FW	I1 (139)	CGAAAGCCGGACGGCAGAA
3	I1RV		TCGTCGTTCCGCCAAGTTCGT
3	F_{repB}FW	Frep (270)	TGATCGTTTAAGGAATTTTG

3	F_{repB}RV		GAAGATCAGTCACACCATCC
3	XFW	X (376)	AACCTTAGAGGCTATTTAAGTTGCTGAT
3	XRV		TGAGAGTCAATTTTTATCTCATGTTTTAGC
3	HI1FW	HI1 (471)	GGAGCGATGGATTACTTCAGTAC
3	HI1RV		TGCCGTTTCACCTCGTGAGTA
3	NFW	N (559)	GTCTAACGAGCTTACCGAAG
3	NRV		GTTTCAACTCTGCCAAGTTC
3	HI2FW	HI2 (644)	TTTCTCCTGAGTCACCTGTTAACAC
3	HI2RV		GGCTCACTACCGTTGTCATCCT
3	L/MFW	L/M (785)	GGATGAAAACTATCAGCATCTGAAG
3	L/MRV		CTGCAGGGGCGATTCTTTAGG

Note: primers were taken from (Carattoli *et al.* 2005).

2.3.4 Long-range PCR

Long-range PCRs were used to investigate the conserved arrangement of regions on IncHI1 plasmids. The primers were designed based on the nucleotide sequence of pHCM1 (NC_003384) (Table 2-4). Twenty 10-kb overlapping PCRs were designed to cover the whole pHCM1 plasmid whilst the sixteen 5-kb PCR focused on region A of the plasmid.

Table 2-4 PCR Primers used for 5kb and 10kb PCR around IncHI plasmids.

Primer name	***Sequence (5'-3')***
HCM1_01F *	GGCGATTTTCTCCAGTTCCTC
HCM1_01R *	ACAGCCTGGAAGTTGAATACC
HCM1_02F *	CATACCGAAACTGTCGGTAG
HCM1_02R *	ATTTTCGGTGGTAGTGTGGC
HCM1_03F *	ATACGGGCAGTGATTCACTG
HCM1_03R *	GCAGCTATAACGTGCAGCTC
HCM1_04F *	AATACGAATGTTCTCCAGTCC
HCM1_04R *	GCATTCGAACGAATACCAGAAG
HCM1_05F *	GAAGTCCAAAACTTGCTCTGG
HCM1_05R *	GTAAGAGCAATCAAGTCGGC
HCM1_06F *	CTTCTCCACAACTGATGCGG
HCM1_06R *	AAGGGATTCTCGACCAGGAC
HCM1_07F *	GCTGACTGAATGTTATCACGC
HCM1_07R *	TATCACCTTTGTAAAGCGGTG
HCM1_08F *	CGCCGAATTCATTCTGTCGG
HCM1_08R *	GCCTTACGTTTTCCAACGAC
HCM1_09F *	TACAGTGCGGAGTTAGGTGG

HCM1_09R *	CAATATCTATAGTACTGTGCGC
HCM1_10bF	GACCTTAATCGTTGTTTTCGGC
HCM1_10bR	GTACTTTTACAGCACCTTCCGC
HCM1_10F *	CTTTTAAAGCAAATGTGGCGC
HCM1_10R *	ATGAGTAAGCGCAGGATGCC
HCM1_11bF	GTCGTAGTGTTTGATTTGGG
HCM1_11bR	TCGACCTTGTTCGGTCATCG
HCM1_11F *	CGTGATATAACCTCCATGACG
HCM1_11R *	TCTACGCTCCACTTATCGAC
HCM1_12bF	TAACGGCTTCGCTGATGCTG
HCM1_12bR	GCCAATCCAAAATGAACAGCG
HCM1_12F *	TGCTCAAAAACAACTTCGAAGG
HCM1_12R *	CTTAGTGAGGTCATATGTCCG
HCM1_13bF	CTTCAGGTCGCTGGATTTGG
HCM1_13bR	TTATGCCAGCAATGTTGTACGC
HCM1_13F *	GATTCTTCCTGACGTCGTTC
HCM1_13R *	CAGTGTTCTGCTTCAGTTTCA
HCM1_14bF	TCACTGGTATGATGTGACGG
HCM1_14bR	TAAATGCCTCAGCCTGCTCG
HCM1_14F *	GTGGTTAAATGCCAAAGGACC
HCM1_14R *	CGTCTTTCATTGCCATACGG
HCM1_15.1F	ATCCCAATGGCATCGTAAAG
HCM1_15.1R	TGGAAAACCGCGAATATCAT
HCM1_15bF	CATTTTCGTGTCGCCCTTAT
HCM1_15bR	TGTTTAACGCTGCCAAGAAT
HCM1_15cF	CAGCAGTTCGGTGATCTTGA
HCM1_15cR	GATACGGGAGGGCTTACCAT
HCM1_16.1F	TTTCGGCATCGTCAACATAA
HCM1_16.1R	GCCGCGGTAAATAGCAATAA
HCM1_16bF	ATACCGTGTAGGCGCTCATC
HCM1_16bR	GTTGTTCCCCTACCTGACGA
HCM1_17bF	AATTTGGTCAATTGGCAAGC
HCM1_17bR	TGAGGGGATCTCTCAGTGCT
HCM1_17F *	TCCCTATCCAAGAAGTGATGC
HCM1_17R *	TTAATTTCTGCGGTGATTGCG
HCM1_18F *	TTTTCTAGCGCCTTACCTGC
HCM1_18R *	CATTACCCAGTCAATTCTGCC
HCM1_19.1F	AGCTCGGTAACGGGGTAGAT
HCM1_19.1R	CTTCGACAGGTGTCAGGTCA
HCM1_20F *	GAACCTGTCTCGCGTTATGG
HCM1_20R *	TGATCAGCTTGCGTCTTACC

* Primers designed by Keith Turner

The long-range PCRs were performed using FideliTaq™ PCR Master Mix (2X) (USB Corporation). Each PCR consisted of 12.5 µl FideliTaq™ PCR Master Mix (2X), 0.5

mM of each primer and nulease-free water to the final volume of 25 µl. Table 2-5 shows the primer combination used for each PCR. The PCR program for 10-kb reaction was as followed: 3 minutes at 94 °C, 30 cycles of 94 °C for 30s, 58 °C for 30 s and 68 °C for 10.5 minutes followed by 5 minutes at 68 °C. For 5-kb reaction, the elongation time was reduced to 6 minutes.

Table 2-5 Primer combinations for long-range PCRs

Reaction number	Primer combination	Predicted product size from pHCM1 sequence
1	HCM1_01F + HCM1_01R	10844
2	HCM1_02F + HCM1_02R	10943
3	HCM1_03F + HCM1_03R	10506
4	HCM1_04F + HCM1_04R	11498
5	HCM1_05F + HCM1_05R	10905
6	HCM1_06F + HCM1_06R	10999
7	HCM1_07F + HCM1_07R	10889
8	HCM1_08F + HCM1_08R	10910
9	HCM1_09F + HCM1_09R	11185
10	HCM1_10F + HCM1_10R	11560
11	HCM1_11F + HCM1_11R	12126
12	HCM1_12F + HCM1_12R	11407
13	HCM1_13F + HCM1_13R	11253
14	HCM1_14F + HCM1_14R	11917
15	HCM1_15.1F + HCM1_15.1R	11753
16	HCM1_16.1F + HCM1_16.1R	11482
17	HCM1_17F + HCM1_17R	10747
18	HCM1_18F + HCM1_18R	11034
19	HCM1_19.1F + HCM1_19.1R	11727
20	HCM1_20F + HCM1_20R	8509
21	HCM1_10F + HCM1_10bR	5121
22	HCM1_10bF + HCM1_10R	7225
23	HCM1_11F + HCM1_11bR	5023
24	HCM1_11bF + HCM1_11R	7225
25	HCM1_12F + HCM1_12bR	5058
26	HCM1_12bF + HCM1_12R	6463
27	HCM1_13F + HCM1_13bR	5155
28	HCM1_13bF + HCM1_13R	6229
29	HCM1_14F + HCM1_14bR	5022
30	HCM1_14bF + HCM1_14R	7017
31	HCM1_15.1F + HCM1_15bR	4600
32	HCM1_15bF + HCM1_15.1R	4337
33	HCM1_16.1F + HCM1_16bR	5905

34	HCM1_16bF + HCM1_16.1R	5707
35	HCM1_17F + HCM1_17bR	5682
36	HCM1_17bF + HCM1_17R	5735

2.3.5 Sequencing and analysis

The PCR product to be sequenced was purified using QIAquick PCR Purification Kit (QIAgen). It was diluted in sterile water to a concentration of 5 ng/µl. The product was then submitted for DNA sequencing, along with sequencing primers at 5 µM concentration using the Big Dye Terminator Cycle Sequencing Kit (version 3.1) from Applied Biosystems. Reactions were run on an ABI 3730 capillary sequencer.

Sequence assembly and analysis was performed using Phrap (de la Bastide, McCombie 2007) and Gap4 (Bonfield, Smith & Staden 1995).

2.3.6 Plasmid multi-locus sequence typing (PMLST)

Thirteen loci were chosen as candidates for PMLST scheme (Table 2-6). These were amplified, sequenced and aligned to determine the variation between 14 plasmids. Six conserved genes were selected for the final PMLST scheme to apply to a set of 36 plasmids. In order to differentiate it from MLST (Kidgell *et al.* 2002, Maiden *et al.* 1998) we have termed the method PMLST. Analysis of PMLST data was carried out using E-burst (http://eburst.mlst.net/).

E-Burst (Feil *et al.* 2004) was used to analyse allelic profiles, determined from the PMLST sequences using the standard MLST approach (Maiden *et al.* 1998). As an alternative approach to analysis, the individual locus sequences were concatenated to give a single representative sequence for each plasmid. Variant bases were identified by aligning these sequences, and recoded into discrete characters for analysis with the MIX

algorithm from the Phylip package (http://evolution.genetics.washington.edu/phylip.html - March 2008).

Table 2-6 Primers for PMLST

Name	*Sequence (5'-3')*	*Gene (bp)*
P43F	CTGGATTCCCCAGAAAAACA	HCM1.043 (570)
P43R	TGAATCACTGCCCGTATCAA	
P54F	CTCCGCCTAGGTGTGTTTGT	HCM1.054 (733)
P54R	CGTAATCGCCGTTTTCTTTG	
P64F	ATGTGACCAACACGGAGACA	HCM1.064 (728)
P64R	CATCGCCTTCCTGATGATCT	
P94F	GGAACTAGCGGGATCATGTG	HCM1.094 (562)
P94R	GTAAGACCCCGCCACTGTTA	
P99F	AGAAAAACGGGGACCTCAGT	HCM1.099 (590)
P99R	GGATTGCTCACGGGAGATTA	
P107F	CAGCATTGAGCATGAAAGGA	HCM1.107 (583)
P107R	TTCGCAGCCTCTTTCAATTT	
P116F	TCTTCACCACGCCATATTCA	HCM1.116 (772)
P116R	GTATCGTCATGCGGGTCTTT	
P177F	CCTGATGGAGCCTTTGACAT	HCM1.177 (464)
P177R	GAATCAGGGTCGATCGAAAA	
P178acF	ACTGAGCTGTTCGCGATTTT	HCM1.178ac (578)
P178acR	GCGGGGTGGTTAATGTCTTT	
P259F	GAACGTAATTCCAGCGGAGA	HCM1.259 (599)
P259R	CGCATTGTTTATGGCTACGA	
P277F	TGTGCTTTACTGCCTGATGG	HCM1.277 (588)
P277R	CGCATGGTTGTTTTGTATCG	
P280F	GGTGTTTGGCGAGTTTAACG	HCM1.280 (576)
P280R	CATCAGGTTGTTAGCCACGA	
P286F	GCCCTTGTTCTGCTTTTCAG	HCM1.286 (713)
P286R	CTCTCCATCAAACGGATGGT	

2.3.7 Generation of sfh knock-out

2.3.7.1 Making competent cells

Bacterial strains *E. coli* CC118λpir and *S.* Typhi BRD948 (pHCM1) were made electro-competent by the following protocol. Overnight culture in LB was used to inoculate 400 mL 2xTY broth (1:100 ratio) and incubated at 37 °C with rigorous shaking (250 rpm)

until the broth reached the OD_{600} of 0.3 to 0.4. The cells were then cooled on ice, harvested and washed with 10% cold (4°C) glycerol. Three more subsequent wash steps were carried out, each time with half of the previous 10% glycerol volume. The cells were finally re-suspended in 1/1000 x vol 10% glycerol and ready to use or stored at -80°C.

2.3.7.2 Creating of vector-construct: pJCB12::(delta)178ac::Km

The *sfh* gene (HCM1.178ac) on plasmid pHCM1 was knocked out based on allelic exchange by homologous recombination between a construct on suicide vector (pJCB12) and the flanking regions of *sfh* gene, swapping the *sfh* on pHCM1 with the kanamycin resistant gene on the construct. The vector-construct was made by fusing PCR products from reaction 1, 2, 3 and 4 together using overlap primer extension PCR. Reaction 1 and 3 amplified the flanking sequences of *sfh* on pHCM1, reaction 2 amplified the kanamycin gene from pCR2.1-Tn5, and reaction 4 amplified the suicide vector pJCB12. Each individual PCR was performed using PfuUltra™II Fusion HS DNA Polymerase (Stratagene); each PCR contained 7.5 µl of PCR buffer (Stratagene), 0.25 mM each dNTP, 0.3 mM each primer, 1.5 µl of PfuUltra™II Fusion HS DNA Polymerase, 1.0 µl of DNA template and water to the final volume of 75 µl. The reaction was aliquoted into 3 separate PCR tubes to run independently on PCR machine before pooled together for subsequent steps. This step is to further minimise the chance of introducing point mutations into the PCR product. PCR was performed on a TETRA DNA Engine Peltier Thermal Cycler (MJ Research) using a PCR program including 2 minutes at 95 °C, 30 cycles of 20s at 95 °C, 20s at 55 °C, 15s (15s per kb) at 72 °C, and the final step of 3 minutes at 72 °C. The PCR products were cleaned by QIAquick PCR Purification Kit to remove primers before using in the next PCR step. Overlap primer

extension PCR was performed using the same PCR conditions with equal molar of DNA products from the two individual PCRs.

Once the cleaned linear PCR product fusing reaction 1, 2, 3 and 4 was obtained, it was phosphorylated by T4 PNK and then ligated by Quick T4 ligase (all from New England Biolab) to circularise to product.

Table 2-7 Primers used to make the *sfh* knock-out

Reaction	Name	Sequence	Template	Product (bp)
1	LF3-F	ATCGCTGCATAAAGGAAGCA	pHCM1	258
	LF4-R	gtttttctaaAAAGCACTGGCCGAACAACT		
2	KmR-15	ccagtgctttTTAGAAAAACTCATCGAGCA	pCR2.1-Tn5	932
	KmR-16	tcttcgtgcgTCTCAAAATCTCTGATGTTACATTG		
3	RF4-F	gattttgagaCGCACGAAGAGTACGAATGTTG	pHCM1	409
	RF1-R	gggcccttctATACCAACACATACGACCCT		
4	Cut1	gtgttggtatAGAAGGGCCCCACTAGTGAC	pJCB12	4119
	Cut2	AGATCTTGCATGCGGGTAAC		
5*	47125	TTCATTATGGTGAAAGTTGGAACC	Vector-construct	1730
	R6K-01	GTGACACAGGAACACTTAACGGC		
6	LF2-F	GAGTAAAAATCCCGAAAGAA	To verify mutant	2107
	HCM1.22	TCATGAACCGCACTGATGAT		

* Primers from reaction 5 were designed by Keith Turner

2.3.7.3 Electroporation and selection

The circularised vector-contruct DNA (1µl) was electroporated into 50µl of competent *E. coli* CC118λpir in 1-mm cuvette using a BioRad GenePulser II set to 1.4 kV, 25 µF and 200 Ω. Cells were re-suspended in 1mL of SOC media (Invitrogen) and incubated at 37 °C for 1 hour before spreading on LB agar plates supplemented with kanamycin and chloramphenicol to select for cells harbouring the vector-construct. The suicide

vector can only replicate in background with λpir. Colonies were picked and checked for positive vector-construct using reaction 5 under standard PCR condition (see 2.3.2.). The positive colonies were isolated and grew on purity plate before stored in 25% glycerol at -80 °C. Plasmid DNA was then extracted to get high concentration of vector-construct plasmid DNA.

The plasmid vector-construct DNA was electroporated into 50µl of competent BRD948 (pHCM1) the same condition. Cells were resuspended in 1mL of SOC media, incubated at 37 °C for 1 hour then plated on LB agar plates supplemented with aromatic mix, kanamycin and chloramphenicol. The final mutant required two homologous recombination events, the first one to incorporate the vector-construct into pHCM1 at *sfh* locus, the second to excise the vector and *sfh* gene, leaving behind the kanamycin gene. The final mutants were selected directly from the colonies grown on the plates using reaction 6. The correct mutant gave the product size of 2kb. That PCR product was then sent to sequencing to confirm the sequence of the mutant. Three mutants were obtained independently from the electroporation into BRD948 (pHCM1) step.

2.3.8 TraDIS

2.3.8.1 Strain

The *S*. Typhi strain used in these experiments is WT26 (pHCM1), a derivative of the attenuated Ty2-derived strain CVD908-*htrA* which has deletion mutations in *aroC*, *aroD* and *htrA* (Tacket *et al.* 1997). WT26 (Turner, Nair & Wain 2006) has a point mutation in *gyrA* conferring reduced susceptibility to fluoroquinolone antibiotics and the multiple antibiotic resistance plasmid, pHCM1, has been introduced. These additions are intended to allow the transposon mutant library to be used for fluoroquinolone resistance and plasmid studies.

2.3.8.2 Preparation of transposomes

The TraDIS transposon is a derivative of EZ-Tn5 <R6Kγori/KAN-2> (Epicenter Biotechnologies, Madison, Wisconsin) with outward oriented T7 and SP6 promoters at each end respectively, and with R6Kγori deleted. The transposon was amplified using oligonucleotides 5'-CTGTCTCTTATACACATCTCCCT-3' and 5'-CTGTCTCTTATACACATCTCTTC-3' with PfuUltra™II Fusion HS DNA Polymerase, (Stratagene, La Jolla, California) and the amplicon was phosphorylated using polynucleotide kinase (New England Biolabs, Hitchin, UK). 400 ng of this DNA were incubated with EZ-Tn5™ transposase (Epicenter Biotechnologies, Madison, Wisconsin) at 37°C for 1h then stored at -20°C.

2.3.8.3 Preparation of bacterial cells for transformation

Bacterial cells for electro-transformation were grown in 2 x TY broth to an OD_{600} of 0.3 – 0.5, then cells were harvested and washed three times in ½ x vol 10% glycerol. Cells were finally resuspended in 1/1000 x vol 10% glycerol and stored at -80°C. 60 μl cells were mixed with 0.2 μl transposomes and electrotransformed in a 2 mm electrode gap cuvette using a BioRad GenePulser II set to 1.4 kV, 25 μF and 200 Ω. Cells were resuspended in 1 mL SOC medium (Invitrogen) and incubated at 37°C for 2h then spread on LB agar supplemented with aromatic mix. After incubation overnight at 37°C, the number of colonies on several plates was estimated by counting a proportion of them, and from this the total number of colonies on all plates was estimated conservatively. Kanamycin resistant colonies were resuspended in sterilised deionised water using a bacteriological spreader.

Normally, ten or more electrotransformations would be performed to generate one batch of mutants. The number of mutants in each batch ranged from estimates of 42,000 to

146,000. From the estimated total number of mutants and using the OD_{600} to estimate the cell concentration in each batch, volumes containing approximately similar numbers of mutants from 13 batches were pooled to create the mutant library mixture estimated to include 1.1 million mutants.

The whole library was created as a join effort between Dr. Keith Turner and me.

2.3.8.4 Transposon library passage

Approximately 2 x 10^9 viable mutants were inoculated into 500 mL LB broth in 1 L flask and grown overnight at 37°C with shaking. Subsequently, 1 mL of this culture was transferred to 500 mL fresh LB broth and similarly grown overnight. This was continued for a total of 6 passages. Genomic DNA was extracted directly from cells harvested from 5 mL of each passage and from approximately 5 x 10^9 cells of the original 1.1 million mutant pool, using tip-100g columns and the genomic DNA buffer set from Qiagen (Crawley, UK).

2.3.8.5 Nucleotide sequencing

Five µg of genomic DNA was fragmented to an average size of 300 bp by Covaris AFA (Quail *et al.* 2008) and Illumina DNA fragment library preparation was performed following the manufacturer's instructions, but using 1.5x the recommended reagent volumes in each step. Ligated fragments were run in a 12 cm 2 % agarose gel in 1 x TBE buffer, at 6 V cm^{-1} without the preceding column clean up step. After 45 minutes, fragments corresponding to an insert size of 250-350 bp were excised, and DNA was extracted from the gel slice without heating (Quail *et al.* 2008). The DNA was quantified on an Agilent DNA1000 chip, following the manufacturer's instructions.

To amplify the transposon insertion sites, 22 cycles of PCR were performed using a transposon-specific forward primer (5'-

AATGATACGGCGACCACCGAGATCTACACCTGAATTACCCTGTTATCCCTAT TTAGGTGAC-3') and a custom Illumina reverse primer (5'-CAAGCAGAAGACGGCATACGAGATCGGTACACTCTTTCCCTACACGACGCT CTTCCGATCT-3'), and 100 ng of DNA fragment library per reaction. Amplified libraries were cleaned up with a QiaQuick PCR product purification column following the manufacturer's instructions, eluted in 30 µl EB, and then quantified by qPCR (Quail *et al.* 2008). The amplified DNA fragment libraries were sequenced on paired or single end Illumina flowcells using an Illumina GAII sequencer, for 36 or 54 cycles of sequencing, using a custom sequencing primer and 2x Hybridization Buffer. This primer was designed such that the first 10 bp of each read was transposon sequence (5' ATCCCTATTTAGGTGACACTATAGAAGAGATGTGTA-3').

The Illumina sequencing was performed by Dr. Daniel Turner.

2.3.8.6 Analysis of nucleotide sequence data

Sequence reads from the Illumina FASTQ files were parsed for 100% identity to the last 10bp of the transposon (`TAAGAGACAG`). Matching sequence reads were stripped of this transposon tag, converted to Sanger FASTQ format and mapped to the *S.* Typhi Ty2 chromosome using Maq version maq-0.6.8 (Li, Ruan & Durbin 2008). The output from the Maq mapview command was used to determine the first nucleotide position to which each read mapped, giving a precise insertion site. The number and frequency of insertions mapping to each nucleotide in the *S.* Typhi genome were then determined for each growth condition. Comparison of these data with gene boundaries defined from the GenBank annotation (Accession number: AE014613) enabled the number of sequence reads and the number of different insertion sites to be determined for every gene. Genes were grouped into functional classes based on the *S.* Typhi CT18 annotation (Parkhill *et al.* 2001). The number of insertions expected per functional class was calculated by

dividing the total number of insertions recovered for a particular growth condition by the summed total of all gene lengths within that class.

The perl scripts for sequence data analysis were written by Gemma Langridge and I.

2.3.8.7 Statistical analyses

Essential genes

As the number of insertion sites for any gene is dependent upon the gene length, the values were made comparable by dividing the number of insertion sites by the gene length to give an "insertion index" for each gene. The distribution of insertion indices is bimodal, corresponding to the essential (mode at 0) and non-essential models. For the original mutant pool and each passage condition, we fitted gamma distributions for the two modes using the R MASS library (Gentleman, Ihaka 1997). Log_2-likelihood ratios (LR) were calculated between the essential and non-essential models for each condition and we called a gene essential if it had a log_2-LR of less than -2, indicating it was at least 4 times more likely according to the essential model than the non-essential model. Genes were assigned 'non-essential' if they had a log_2-LR of greater than 2.

Comparison of culture passages and growth in the presence and absence of bile

For each pair of conditions tested *(A,B)*, we calculated the log_2 fold change ratio $S_{g,A,B}$ in the number of observed reads $n_{g,A}$, $n_{g,B}$ for every gene g as $S_{g,A,B} = log_2 \frac{(n_{g,A}+100)}{(n_{g,B}+100)}$. The correction of 100 reads smooths out the high scores for genes with very low numbers of observed reads. We fitted a normal model to the mode of distribution of $S_{A,B}$, and calculated p-values for each gene according to the fit. After excluding essential genes from the original pool, we considered genes to be important/costly for a particular condition with a log_2 fold change of at least 2, which

corresponds to a 10^{-5} p-value and a 2.5 x 10^{-4} false discovery rate (FDR) according to the normal model.

We calculated the p-value for the distances between insertion sites using F = G/N where G is the number of bases in the genome (4,791,961) and N is the number of unique insert sites (394,921). The p-value for at least X consecutive bases without an insert site is $e^{(-X/F)}$, giving a 5% cut-off at 37 bp and a 1% cut-off at 56 bp.

The R scripts for statistical analysis were written by Leopold Parts.

The analysis scripts for TraDIS were included in appendix 8.4 (on CD).

2.3.9 Genotype array (SNP typing)

2.3.9.1 Array design

We have previously identified over 2000 SNPs in the Typhi chromosome, including 1964 SNPs identified from sequence data on 19 Typhi genomes (Holt *et al.* 2008) and an additional 83 SNPs identified from analysis of 2% of the Typhi chromosome among 180 isolates (Roumagnac *et al.* 2006). We have also previously identified eight SNPs within six genes of the conserved backbone of IncHI1 plasmids (in chapter 4). Using MUMmer (Kurtz *et al.* 2004) we identified an additional 337 SNPs between the conserved backbones of eight IncHI1 plasmid sequences, originally isolated from Typhi, Paratyphi A, Choleraesuis, Typhimurium and *E. coli*. In order to assay as many of these SNPs as possible in clinical Typhi isolates, we designed two custom Illumina GoldenGate arrays. The GoldenGate assay utilizes mega-plex PCR (up to 1536 sets of oligonucleotides per pool) followed by hybridization to custom bead arrays. Due to the PCR step, it is not possible to uniquely target two SNP loci separated by less than 60bp in a single oligonucleotide pool. It is also not possible to assay any SNP locus that lies within 10bp of another SNP, insertion or deletion, as these variants will interfere with

primer binding. For these reasons, 35 (1.8%) of the 1964 Typhi chromosomal SNPs identified in (Holt *et al.* 2008) were not suitable for SNP typing with GoldenGate. Oligonucleotides were designed to target the remaining 1929 SNPs, as well as 72 SNP loci identified in (Roumagnac *et al.* 2006). A total of 473 Typhi isolates were genotyped using the GoldenGate arrays, including 446 multidrug isolates sourced from around the world and 26 control isolates (the 19 isolates sequenced in (Holt *et al.* 2008), and 8 plasmid controls).

2.3.9.2 SNP clustering and quality control

The Illumina GoldenGate assay generates a detection signal for each of two target alleles, at each SNP locus, in each sample. Converting these signals into genotype calls is essentially a two-dimensional clustering problem, for which several algorithms have been developed, including Illuminus (Teo *et al.* 2007). These are mostly aimed at calling genotypes in human and other diploid organisms, where three clusters are expected – two homozygous and one heterozygous. Since Typhi is haploid, we expect only homozygous clusters for chromosomal SNPs, analogous to genotype calling on the human Y chromosome. Using Illuminus to call (haploid) genotypes gave perfect allele calls for 1104 chromosomal SNPs (57%). However, we noticed that many of the Typhi chromosomal SNPs showed evidence of a third cluster with zero signal for both alleles, suggesting that occasional deletions may be common in Typhi. Adapting Illuminus to fit a 'no signal' cluster in place of a heterozygous cluster, which accommodates the possibility of deleted target loci, improved this to 1402 SNPs (71%) with perfect allele calls. Thus we used the genotype clustering generated by the adapted version of Illuminus, referred to hereafter as Illuminus-P to assign SNP alleles to each Typhi isolate. Similarly, isolates that lack the IncHI1 plasmid can be expected to form a 'no signal' cluster at each IncHI1 plasmid SNP locus.

Each of the 19 sequenced strains had previously been assigned alleles at each of the chromosomal SNP loci based on sequencing data (Holt *et al.* 2008). Analysis of GoldenGate data for these sequenced isolates was used to determine whether each SNP was (a) assayed successfully, (b) clustered accurately and (c) truly polymorphic as expected from sequence data in (Holt *et al.* 2008). Each SNP assay was considered successful if it generated signals of reasonable strength that were able to be clustered. For 1402 SNPs, alleles assigned by Illuminus-P clustering of GoldenGate data agreed with all those expected from sequence data, and these were considered high quality SNP assays for downstream analysis. For 19 SNPs (~1%), GoldenGate analysis found no evidence of the derived allele (based on manual inspection of the signal plots in addition to Illuminus-P clustering). These SNPs (listed in Appendix 8.7) were considered most likely to be genuinely nonpolymorphic sites, representing false positives in our earlier SNP calling from sequence data and were not included in downstream analysis. For a further 46 SNPs, GoldenGate analysis found evidence of the derived allele, but not in all strains which had been assigned the derived allele from sequence data. Signal plots for these SNPs were manually inspected and assessed to be good quality signals and accurately clustered. We therefore conclude that these loci are truly polymorphic, but our earlier sequence-based allele assignments contained some errors which can now be corrected by the GoldenGate analysis. Thus 1448 SNPs (75% of those designed) were considered successful GoldenGate assays of polymorphic loci identified previously from sequence analysis. We manually inspected signal plots to assess the GoldenGate assays of 72 SNPs defined in a different study (RAMSEY, EDWARDS 1961), of which 60 (83%) were of high quality. Thus the phylogenetic analysis of experimental Typhi isolates presented in this study is based on 1508 SNP loci distributed randomly int the Typhi chromosome. While the high rate (25%) of failed assays means a reduction in

resolution, the SNP loci concerned are distributed evenly within the phylogenetic tree defined by the complete SNP set. Thus failures in the design, signal generation or clustering do not bias the distribution of SNPs that were assayed successfully with GoldenGate.

A total of 294 IncHI1 SNPs were included in the GoldenGate arrays, and their genotypes called using Illuminus-P followed by a heuristic to better identify the third 'no signal' cluster (which correlates with absence of the plasmid). IncHI1 plasmid SNPs were validated in the same way as chromosomal SNPs, by comparing alleles from GoldenGate assays with sequence data from which the SNPs were originally identified. Sequencing and genotyping alleles matched perfectly for 200 SNPs (68.0%), and these loci were used for the remainder of the study. A total of 218 SNPs designed to assess the presence or absence of resistance genes and specific IncHI1 sequences were included on the GoldenGate arrays, and their genotypes called using Illuminus-P and the heuristic. Here the 'no signal' cluster implies absence of the target sequence, which may be due to absence of the entire plasmid (for IncHI1-specific sequences, if no other IncHI1 targets are detected) or absence of the specific locus (if most other IncHI1 targets are detected). Note that resistance genes may be present on plasmids of a different type, or potentially integrated into the chromosome, and so are not always associated with the presence of IncHI1 sequences. Perfect matches were obtained between sequence and genotyping data for 119 of these loci (54.6%). This provides reasonable coverage of resistance genes and insertion sequences, as well as several deletions characterised earlier by comparative analysis of the three finished plasmid sequences pHCM1, pAKU_1 and R27.

Two SNPs specific to Paratyphi A were included on the GoldenGate array, in order to identify erroneously serotyped isolates. The SNPs were validated by typing five

Paratyphi A control isolates, which gave distinct allele signals from the Typhi control isolates at the two Paratyphi A-specific loci. Alleles were determined for 89% of Typhi chromosomal SNPs in the Paratyphi A strains, resulting in these strains clustering at the root of the Typhi phylogenetic tree.

The array design and SNP calling were entirely the work of Kathryn Holt as part of her PhD.

2.3.9.3 Phylogenetic analysis

SNP alleles assigned by Illuminus-P for 1508 chromosomal loci were concatenated to give a single haplotype string for each strain. The alignment of these haplotype strings was used as input for phylogenetic analysis (appendix 8.6). Initially, the alignment was analysed using ModelTest which suggested a general time reversible (GTR) model provided the most appropriate phylogenetic model for this data. SNP typing with the Illumina GoldenGate only provides genetic information at the specific assayed loci; in the present study these were mostly loci determined by whole genome comparison of 19 Typhi strains. Note that here branch lengths reflect genetic divergence only at the assayed SNP loci. We used RAxML (Stamatakis, Hoover & Rougemont 2008) to fit maximum likelihood phylogenetic trees to the chromosomal allele data using the GTR model. The same procedure was repeated for the 200 IncHI1 plasmid SNPs. Phylogenetic analysis of IncHI1 plasmid SNP alleles was performed only for the plasmid positive isolates. The final chromosomal and plasmid trees, are each the best-scoring maximum likelihood topology out of 1000 bootstrapped tree topologies, with optimised branch lengths, and were displayed using Dendroscope (Huson *et al.* 2007).

2.3.10 PCR to detect mobile elements

Mobile elements was detected by the following reactions in Table 2-8. Standard PCR conditions were used. For different product sizes, the elongation times were adjusted by the synthesis speed of 1 minute per 1kb.

Table 2-8 Primers to detect mobile elements

Reaction	Primer name	Sequence	Product (bp)
G	GF	GATGGAGAAGAGGAGCAACG	989
	GR	TTCGTTCCTGGTCGATTTTC	
H	HF	GTGCTGTGGAACACGGTCTA	1598/271
	HR	TCATCAACGCTTCCTGAATG	
I	IF	ACGAAAGGGGAATGTTTCCT	1490/163
	IR	CGAGTGGGAATCCATGGTAG	
J	JF	CAAAATGTTCTTTACGATGCC	2200
	JR	CCAGACAGGAAAACGCTCA	
K	KF	CTGTGCCGAGCTAATCAACA	1314
	KR	ACGAAAGGGGAATGTTTCCT	
L	LF	TTTTAAATGGCGGAAAATCG	1872
	LR	GCCAGTCTTGCCAACGTTAT	
M	MF	GGGCGAAGAAGTTGTCCATA	2196
	MR	ATTCGAGCAAAACCATGGAA	
N	NF	CGGGATGAAAAATGATGCTT	2181
	NR	GGTCGGTGCCTTTATTGTTG	
O	OF	GCGTACAAAAGGCAGGTTTG	1823
	OR	GCTTGATGATGTGGCGAATA	
P	PF	TGGTCGGTGCCTTTATTGTT	4190/1899
	PR	GGGCGTCAGAGACTTTGTTC	
Q	QF	TTCGCCCGATATAGTGAAGG	1924
	QR	CTAACGCCGAAGAGAACTGG	

Note:

- G – Insertion site of *bla/sul/str* into Tn*21*;
- H, I – Insertion site of Tn*21* into Tn*9* (left and right boundary), product 1: pHCM1 insertion site, product 2: pAKU1 insertion site;
- J, K – insertion site of Tn*9* to pAKU1 backbone
- M – insertion site of Tn9 to pHCM1 backbone;
- L – insertion site of Tn9 to Tn*10* in pHCM1, product 1: pHCM1 insertion site; product 2: a band of different size;
- N – insertion site of Tn*10* in pHCM1;
- O, P – insertion site of Tn*10* in pAKU1; in P, product 1: pMAK1 insertion site; product 2: pAKU1
- Q – insertion site of the second *strAB* genes in pAKU1.

2.3.11 Plasmid competition

The two plasmids pHCM1 and pSTY7 was competed in BRD948 background in 3 serial passages. Overnight cultures of BRD948 (pHCM1) and BRD948 (pSTY7) were diluted to 10^{-4} and 50 µl of each strain (1:1 ratio) was inoculated into LB broth supplemented with aro mix and chloramphenicol. It was incubate at 37 °C with shaking overnight. The new passage was started by the same procedure. At time point 0 (after the first inoculation), 1, 2, 3 and 4 days, samples were taken, diluted and plated on LB agar supplemented with aromatic mix. Sixty-four colonies were randomly picked from the plates and tested by PCR to detect the plasmid type.

The colony PCR used standard conditions with three primers HCM1.DF 5'-CGATTTGTGAAGTTGGGTCA-3', HCM.DR2 5'-CAACCTGGGCAGGTGTAAGT-3' and HCM.DR3 5'-TTCGTTACGTGTTCATTCCA-3'. Colonies with pHCM1 gave a product of 511bp and pSTY7 gave a product of 285bp.

3 Fundamental factors for plasmid stability

3.1 Introduction

In this genomic era, vast amounts of DNA sequence data are being generated. However, the rate of processing these data into information about biological functions is lagging. This situation calls for the development of high throughput methods for the simultaneous functional analysis of multiple genes within genomes. Defining the genes that are essential under specific conditions is of importance for defining the basic materials of synthesis biology and identifying potential targets for new antimicrobial agents. A gene can be defined as essential under given conditions when it is impossible to obtain the knockout of that gene in the condition investigated – one such condition is life itself, i.e. the gene is essential for the survival of the bacterium. Several experimental approaches have been used to define essential gene lists for bacterial isolates including single-gene deletion (Baba *et al.* 2006, de Berardinis *et al.* 2008, Kobayashi *et al.* 2003), ordered or random global transposon mutagenesis (Akerley *et al.* 2002, Hutchison *et al.* 1999, Salama, Shepherd & Falkow 2004, Sassetti, Boyd & Rubin 2003), antisense RNA inhibition (Ji *et al.* 2001) and trapping lethal insertions (Knuth *et al.* 2004).

Random transposon mutagenesis has been the method of choice for many studies because of its speed and cost effectiveness. However, the major drawback of these methods is the possibility of missing essential genes due to (a) sub-saturation knock-out of the whole genome by the transposon and (b) inaccuracy in identifying transposon insertion sites. Problems can also be encountered if the given transposon is too specific in terms of target sequence selection. The majority of transposon mutant libraries contain only a few thousands mutants per genome (Hutchison *et al.* 1999, Salama,

Shepherd & Falkow 2004, Sassetti, Boyd & Rubin 2003), which account for only a fraction of the genes in a given genome, inevitably lead to the missing of essential genes by chance. Signature–tagged mutagenesis (STM) (Hensel *et al.* 1995), transposon-site hybridisation (TraSH) (Sassetti, Boyd & Rubin 2001) and transposon-mediated differential hybridisation (TMDH) (Chaudhuri *et al.* 2009) are transposon-based mutagenesis methods that make use of PCR and hybridisation on microarray respectively for identifying of transposon insertion sites. Although these methods allow simultaneous investigation of genome-wide transposon insertion sites, they are all sub-optimal due to the numbers of transposons that can be located and the inherent inaccuracy of microarray to identify transposon insertion sites. Transposon insertion can also be used to investigate the role of single genes in the stability of single copy plasmids but new methods are needed to address current technological disadvantages and improve the reliability of identifying essential gene function.

Plasmids are extra-chromosomal DNA molecules capable of autonomous replication within their host cells. The genes on plasmids are therefore normally believed to be non-essential to the host. However, genes that contribute to the stable maintenance of a plasmid within a bacterial cell are of great interest not only to the understanding of plasmid biology but also to the discovery of novel drug targets to limit the transmission of antibiotic resistant plasmids.

IncHI1 plasmids have become strongly associated with *S.* Typhi after the introduction of chemotherapy for typhoid fever (WOODWARD, SMADEL 1948, Wain *et al.* 2003, Wain, Kidgell 2004, Wain *et al.* 2003). The *sfh* gene on IncHI1 plasmids has been shown to play a role in their stability in *S.* Typhimurium by reducing the regulatory disruption caused by the presence of the large plasmid (Doyle *et al.* 2007, Doyle, Dorman 2006). However, our knowledge on many IncHI1 plasmid encoded genes is

otherwise very limited; 43% of the genes are still annotated as encoding a "hypothetical protein".

In this chapter, we used a novel random global mutagenesis method called transposon-directed insertion-site sequencing (TraDIS) to investigate the genes important for IncHI1 plasmid stability inside *S.* Typhi during growth in rich media. By using TraDIS, we combined the use of a large transposon insertion mutant library of over one million mutants, with Illumina (formerly known as Solexa) sequencing technology to identify insertion sites with the accuracy of a single base-pair. This technique allows the precise identification of essential genes on the chromosome of *S.* Typhi and, with knowledge of those essential genes, those which are of important for plasmid stability. Here, the work is aimed at identifying which genes are likely to have an important role in plasmid stability.

3.2 Results

3.2.1 The generation of a one million mutant library

The mutant library was generated using a Tn*5*-derived transposon carrying a kanamycin resistant gene (see 2.3.8.2). PCR amplicons of the transposon were coupled with commercial transposases before being electroporated into an attenuated strain of *S.* Typhi Ty2, WT26 harbours the pHCM1 plasmid (see Table 2-1) (Figure 3-1). Transposon inserted mutants were grown on selective media with kanamycin before being collected into pools. The final transposon mutant library contained an estimated 1.1×10^9 individual mutants. The optimisation of the protocol for transposon insertion was undertaken with the assistance of Keith Turner and the library was generated as part of this PhD.

One aliquot of the mutant library was used to investigate plasmid stability (Figure 3-2).

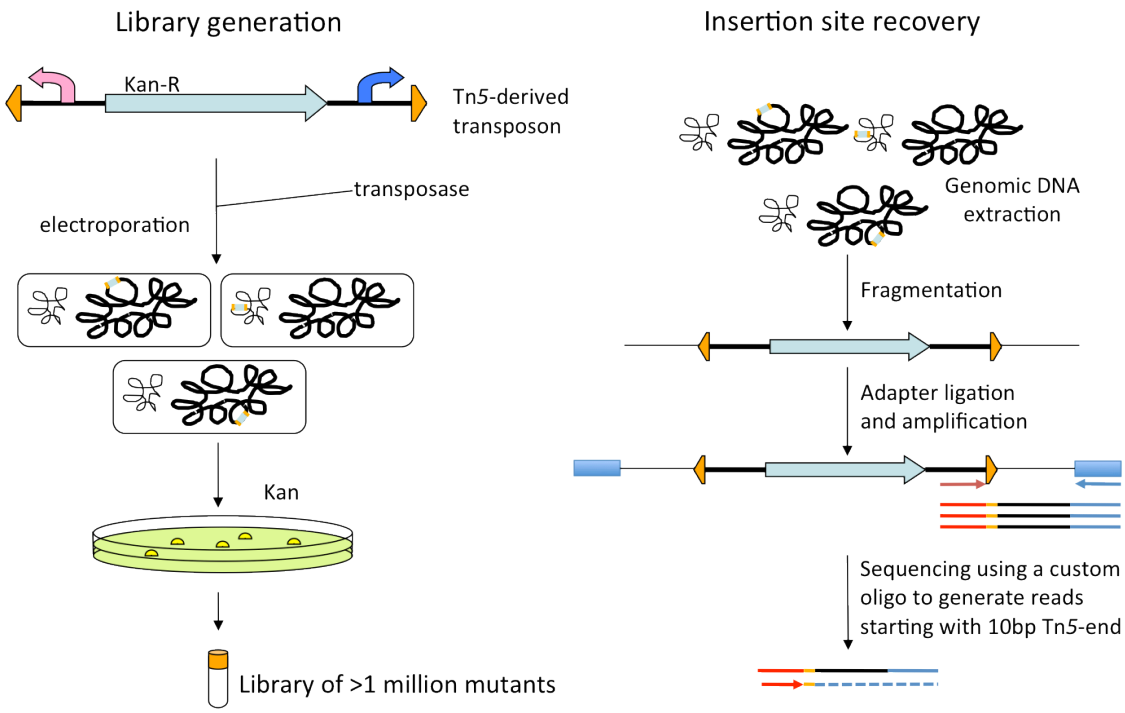


Figure 3-1 Schematic illustration of TraDIS steps

3.2.2 Experimental design

Plasmid pHCM1 encodes resistance to chloramphenicol. To investigate genes involved in stability of IncHI1 plasmids in S. Typhi, the TraDIS transposon mutant pool harbouring pHCM1 was grown in LB broth either supplemented with chloramphenicol (CmP) or without chloramphenicol (non-CmP) for six overnight culture passages (equal to approximately 60 cell generations, Figure 3-2).

Mutation by transposon insertion into any chromosomal gene that is required for stable plasmid inheritance will result in plasmid loss following passage, rendering the bacterial cell chloramphenicol sensitive. In the cultures supplemented with chloramphenicol this will result in loss of those mutants from the mutant pool, but no loss of such mutants will occur from the unsupplemented cultures. Thus, such genes may be identified by having transposon insertions when grown without chloramphenicol, but significantly fewer insertions when grown in its presence.

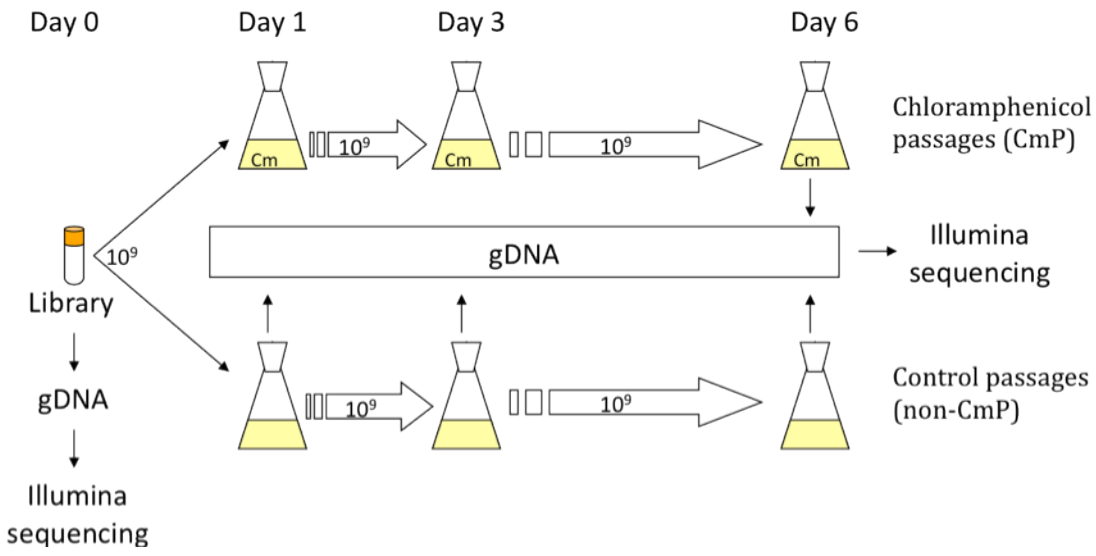


Figure 3-2 Selection assay using the mutant library to investigate plasmid stability

Disruption of plasmid-encoded genes required for stable plasmid inheritance will result in loss of the plasmid following passage. Such genes would therefore be expected to show few insertions regardless of the presence or absence of chloramphenicol. Mutations in genes from post-segregational killing systems would result in the killing of plasmid-free cells instead of just the loss of plasmid. This, however, would also lead to decrease in number of insertions into those genes in both passage conditions.

3.2.3 Identification of insertion sites from the library by Illumina sequencing

DNA samples were extracted from the mutant libraries on day 0 (the initial library), day 1, day 3 and day 6 of the control passages (non-CmP). One DNA sample was obtained from the CmP passaged cells on day 6. Illumina sequencing of DNA prepared from these samples were performed by the sequencing group (Daniel Turner) at WTSI. Briefly, fragmented DNA was sequenced using paired end adaptors and transposon

specific primers. This gave 10bp of transposon sequence to serve as a tag for reads that were transposon-directed.

Each Illumina sequencing lane produced between 1.8 to 6.5 million reads, almost 90% of which contained the 10bp sequence tag. The plasmid specific sequence from each tagged read was then mapped to the reference sequences (NC_004631 for Ty2 and NC_003384 for pHCM1) to identify up to 294,588 insertion sites. The Perl scripts for sequence data manipulation (Minh-Duy Phan and Gemma Langridge) are included in Appendix 8.4.

To identify maximum unique insertion sites from the mutant library, samples from day 0 and day 6 were sequenced on 4 and 5 Illumina lanes respectively, producing up to 12 million reads (day 6, combining of 5 lanes). Figure 3-3 shows the linear increase in reads when combining lanes from the same samples whilst the number of insertion sites reaches saturation after 3 lanes. To balance the maximum unique insertion sites identified and the sequencing cost, it was decided to sequence two lanes for each sample. Thus, the insertion sites identified from this point onwards are from two sequencing lanes (with the exception of the data from day 0 which used 4 lanes).

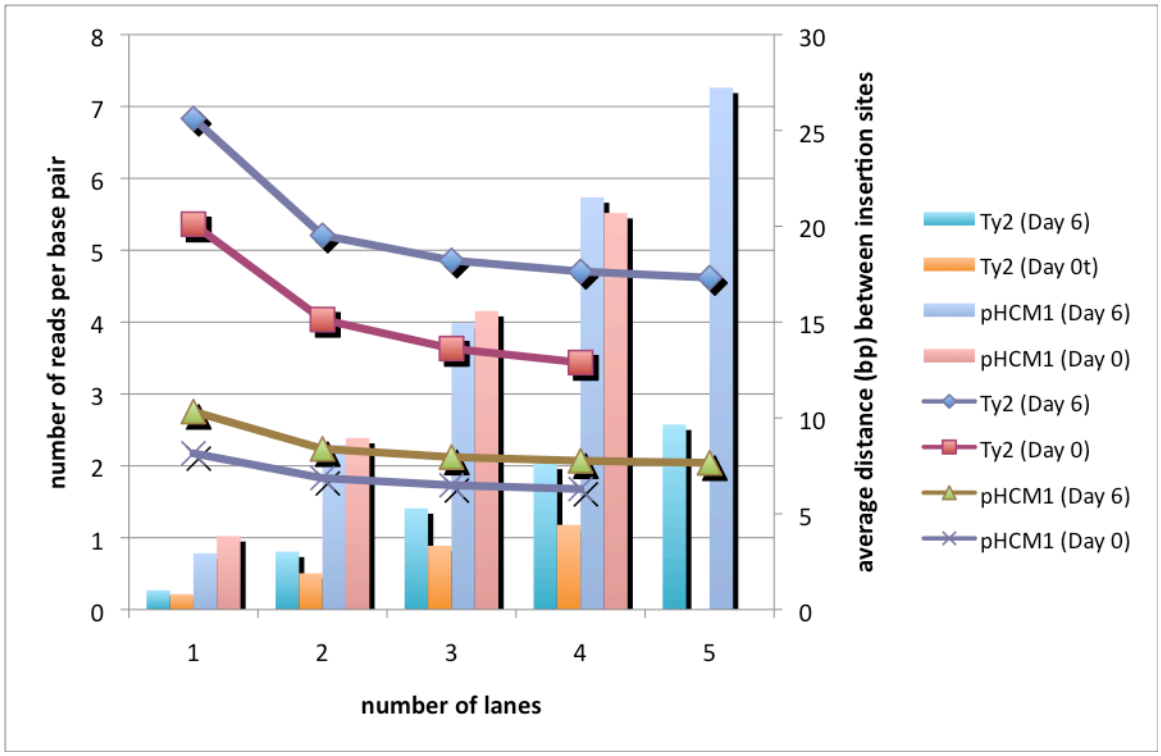
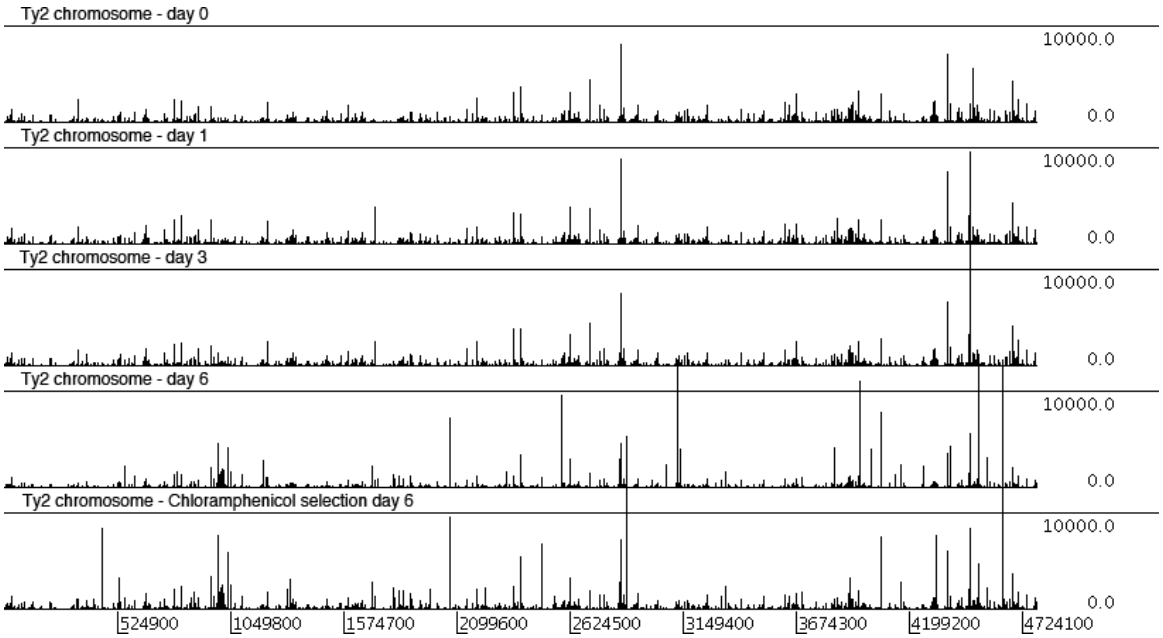


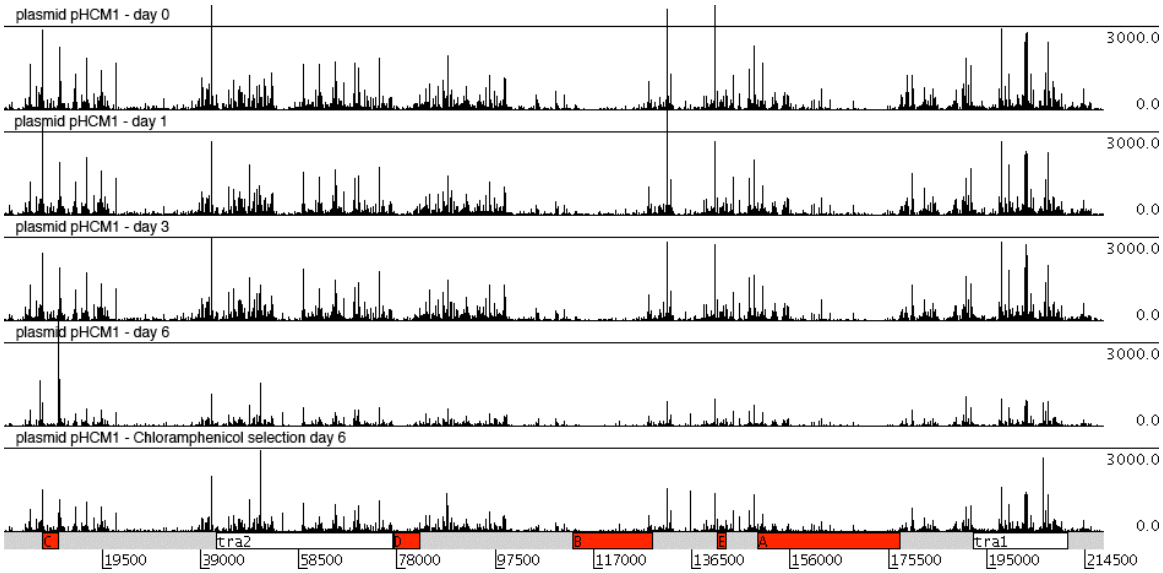
Figure 3-3 Saturation of the transposon insertion sites.
The increase in number of sequencing reads used to map to reference genome leads to near saturation of transposon insertion sites after three lanes
The columns show number of sequence reads per base pair (left axis); the lines show average distance between insertion sites (right axis).

The general distribution of insertion sites across the genome is shown in Figure 3-4. Overall, the insertion sites were well-spread across the genome with some occasional hotspots. Day 0, 1 and 3 showed a similar distribution whilst day 6 and CmP6 were quite different with the reduction in number of reads at the majority of insertion sites. Some insertion sites in day 6 however showed a marked increase in the number of reads. The plasmid plots showed clear cold spots, some of which correspond to the position of mobile elements on the plasmids. Closer inspection of these cold spots revealed some insertion sites, the number of which is still high enough to conclude that very few genes on the plasmid are completely protected (no insertions).

The probability of any gene being missed by transposon insertion was also calculated: The probability of mis-identifying the shortest gene in Ty2 genome, *hisL* (23bp), as essential was 0.186.



(a)



(b)

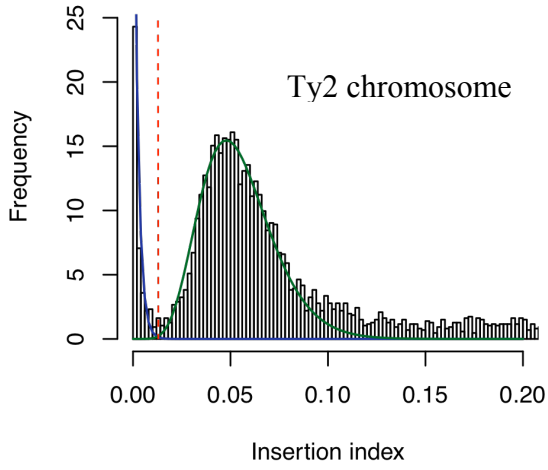
Figure 3-4 Frequency and distribution of transposon directed insert-site sequence reads across the genome of Ty2 (a) and pHCM1 plasmid (b) over time.
The red regions in (b) are those in pHCM1 but not in R27 plasmid, an earlier incHI1 plasmid. The x-axis shows nucleotide position within the genome, the y-axis shows number of reads mapped to each insertion sites. The maximum number of reads shown for Ty2 is 10,000 and for pHCM1 is 3,000.

3.2.4 Essential genes and genes require for long-term survival

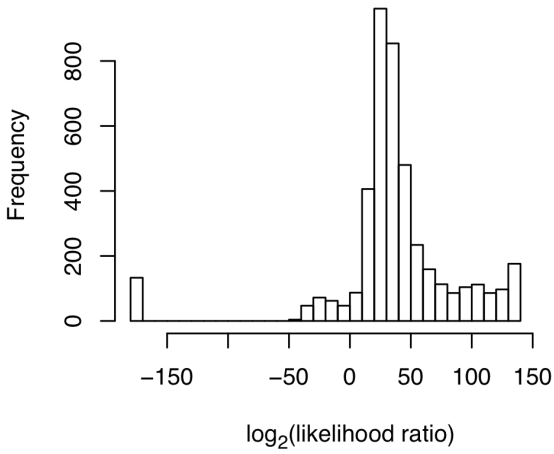
In order to compare the level of insertions across genes of different length, the data were normalised by dividing the number of unique insertion sites within any gene by the gene length to give an insertion index. A frequency distribution of insertion index for all the annotated genes on the Ty2 chromosome gives a clear bimodal distribution (Figure 3-5a). The leftmost peak includes genes with 0 or very low number of insertions. Transposon insertions into these genes were probably lethal to the cells or are required for cellular growth hence their corresponding mutants did not survive or were greatly diminished in the library pool. The rightmost peak represents genes with tolerance to transposon insertions. The gene knock out is either neutral or even advantageous to cellular growth which allows the mutants to survive or even thrive within the pool. This bimodal distribution allowed us to calculate the likelihood ratio of any gene to be on the leftmost peak i.e. essential to the bacteria. The histogram of $\log_2$ likelihood ratio ($\log_2$LR) for all the genes on Ty2 chromosome from day 0 shows a clear cluster at -175 which represents essential genes with no tolerance to transposon insertions (Figure 3-5b). A $\log_2$LR of -2 corresponds to the lowest point between the peaks of the bimodal distribution for all samples investigated. We therefore chose $\log_2$LR of -2 as the global cut-off for essentiality (at which point a gene is four times more likely to belong to the "essential" peak). A $\log_2$LR of 2 was chosen as the cut-off for non-essentiality (four times more likely to be non-essential). Genes with a $\log_2$LR between -2 and 2 could not be assigned as either essential or non-essential.

The insertion index for all the annotated genes on pHCM1 however does not fall into a bimodal distribution (Figure 3-5c) This may be due to a number of factors including the small number of genes, the non-essential nature of plasmid genes and perhaps the redundancy of genes caused by a copy number effect. It is therefore not possible to

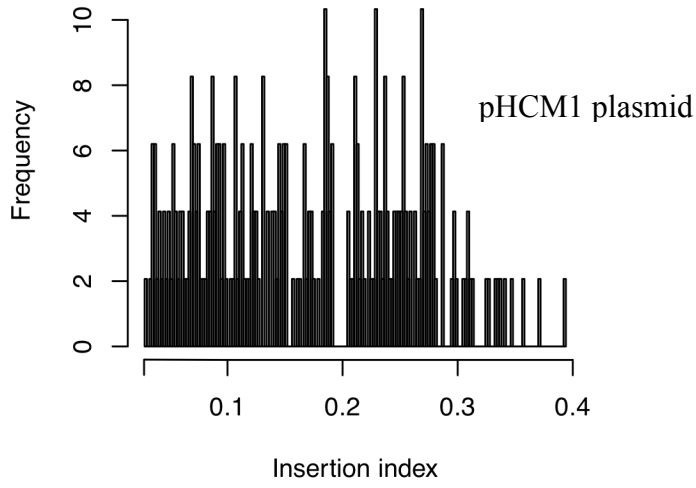
calculate a cut-off for essentiality of plasmid genes. The data however still gives us a ranking list of genes that show low tolerance to transposon insertions suggesting their contribution cellular growth and/or plasmid stability.



(a)



(b)



(c)

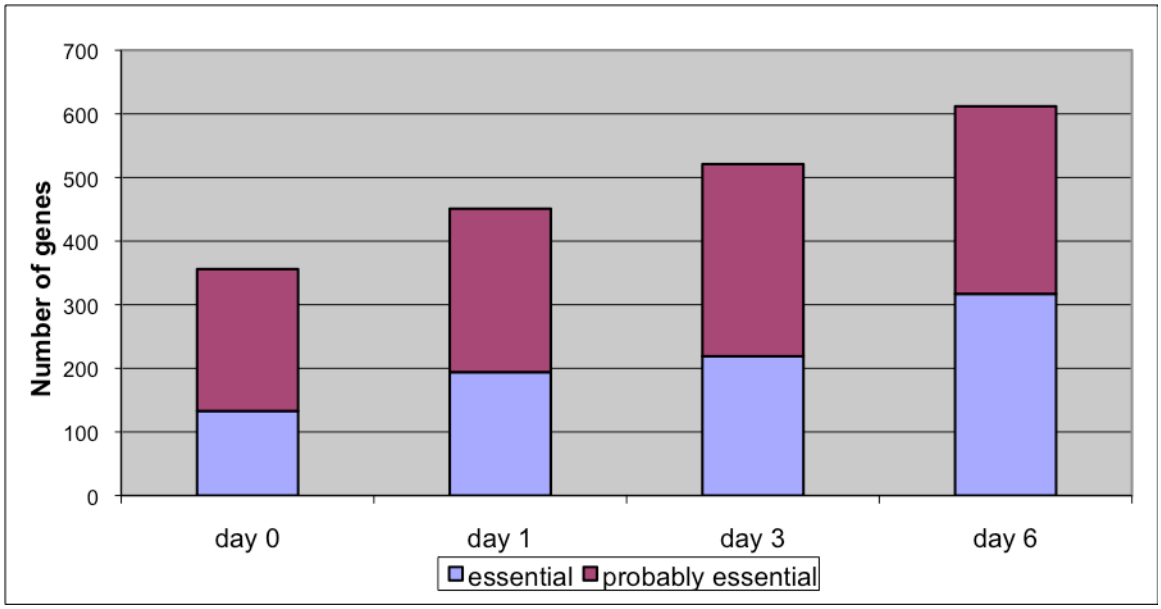
Figure 3-5 Identification of essential genes
(a) Frequency distribution of insertion index (the red dotted vertical line represents the cut-off selected to distinguish between essential and non-essential genes) and (b) Frequency distribution of $\log_2$ likelihood ratio of genes on Ty2 chromosome from day 0 sample; (b) Frequency distribution of insertion index of genes on pHCM1 plasmid on day 0.

From the day 0 sample, 4301 out of 4323 genes on Ty2 chromosome (99.49%) could be assigned to a specific group. Of 356 genes on the chromosome that were protected from transposon insertion 133 had no insertions (essential) and a further 223 genes with

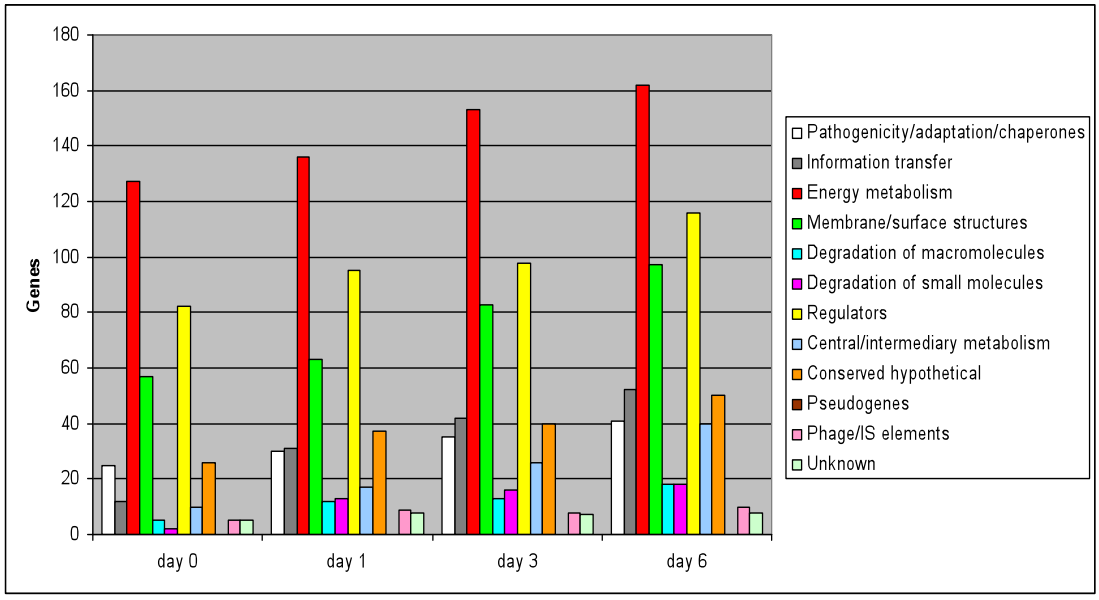
log_2LR lower than -2 (probably essential) (Figure 3-6a). In addition 22 genes could not be confidently assigned as essential or non-essential due to their log_2LR of between -2 and 2, leaving 3945 (91%) non-essential genes at day 0.

In rich, non-selective media the number of essential and probably essential genes increased over time (Figure 3-6a) from 356 on day 0 to 612 on day 6 (see appendix 8.1 for full list of essential genes from day 0). This gradual dropping out of mutants from the library pool likely highlights gene disruptions that do not have an immediately lethal effect on the cell. Such mutants, however, adversely affect cellular growth so that eventually cell death occurs after several generations possibly because the cell is unfit to compete with others and is thus driven out of the pool. Together these are the genes required for long-term survival of the bacterial cells and, as such, cannot be considered as directly involved in plasmid stability. This is important as the plasmid stability experiment was carried out over 6 days and thus 3711 (85.8%) of chromosomal genes could be tested for their effect on plasmid stability.

The long-term survival genes were investigated to assess if they share the same or similar function (Figure 3-6b). The functional classes were assigned based on the CT18 genome annotation of corresponding Ty2 homologues. Genes from the energy metabolism class account for the majority of essential genes. The number of essential genes in all functional classes increases overtime. However, no particular functional classes are significantly associated with genes required for long-term survival.



(a)



(b)

Figure 3-6 The number of essential and probably essential genes for survival in rich, non-selective media at different time points during 6 day passages.
(a) The number of Ty2 essential genes (genes without insertions) and probably essential genes (genes with insertions but have $\log_2$LR lower than -2) at different time points during 6 day passages (b) The number of essential genes within functional classes.

3.2.5 Chloramphenicol resistant and plasmid stability genes on Ty2 chromosome

Plasmid stability is defined as a measure of the likelihood with which a plasmid is inherited by daughter cells at cell division (Nordstrom, Austin 1989). However the term can be used more loosely as the collective results of different mechanisms to ensure the stable maintenance of a plasmid in a bacterial population. Plasmid encoded machineries have predominantly been the focus of researchers who study plasmid stability. However, there have been suggestions of chromosome-plasmid interaction and co-evolution that lead to enhanced fitness of the host cells and hence the stable maintenance of plasmids in the population (Dionisio *et al.* 2005, Lenski, Simpson & Nguyen 1994). However, it is rare to find reports about chromosome-encoded genes that affect plasmid stability. By using a saturated insertion mutant library in combination with long-term passages in selective and non-selective media, we attempted to look for candidates on the Ty2 chromosome that might contribute to the stable inheritance of IncHI1 plasmid in an *S.* Typhi population.

Insertion mutations in plasmid stability gene(s) on the chromosome would be gradually decreased in the library pool overtime under the presence of chloramphenicol as a selective agent for plasmid positive cells. The comparison between non-Cm passage day 6 and CmP day 6 highlighted the mutants that disappear in CmP day 6. Mutants in essential genes defined previously were not considered as candidates for plasmid stability. Care should be taken though to interpret the data because the CmP also selects for genes on the chromosome that contribute to survival in the presence of Cm independent of plasmid mediated resistance.

Three measurements were calculated for each gene: $\log_2$ read ratio ($\log_2$RR) (the ratio of reads in day 6 of non-Cm and Cm passage), the probability of a gene having more reads

in non-Cm than in Cm passage, and the real difference in number of insertion sites within a gene. A list of genes showing high insertion site difference and high probablitity (>0.98) of difference in the two passages is shown in Table 3-1 and the visual comparisons of them are in Figure 3-7.

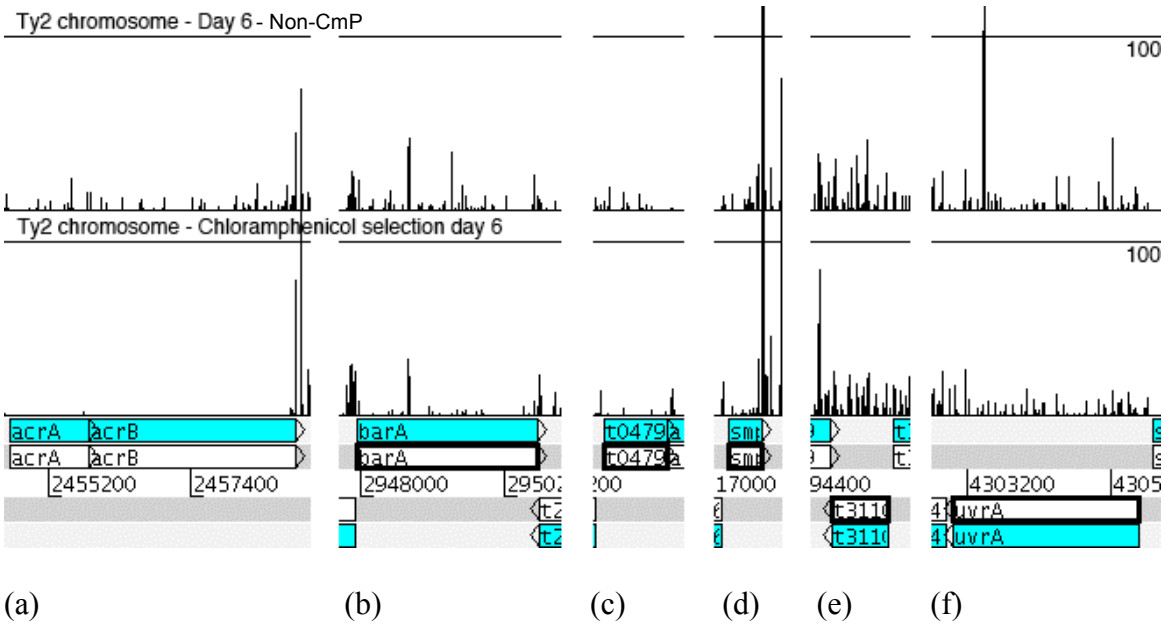


Figure 3-7 Genes on chromosome potentially contributing to survival in chloramphenicol
Artemis plots comparing the insertion sites and their frequency of genes in non-CmP day 6 and CmP day 6; (a) *acrAB,* efflux pump operon; (b) *barA,* encoded a sensor protein; (c) t0479, predicted N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase; (d) *smpB,* tmRNA-binding protein; (e) t3110, predicted extradiol ring-cleavage dioxygenase, class III enzyme, subunit B; (f) *uvrA,* nucleotide excision repair protein subunit A.

The gene with the lowest read difference, t3110 - a conserved hypothetical protein, has 2.7 times more reads in CmP day 6 as in non-CmP day 6 ($Log_2LR = 1.93$). This gene is conserved (with lowest nucleotide identity of 97%) across many serovars within subspecies *enterica* (Typhi, Paratyphi A, B, C, Newport, Choleraesuis etc.). The latest prediction for the function of this gene is extradiol ring-cleavage dioxygenase, class III enzyme, subunit B, which involves in oxidation reduction activity and plays a key role in degradation of aromatic compounds (Interpro entry IPR004183). This is interesting

because chloramphenicol, the selective agent in this experiment, does have an aromatic ring in its structure.

Two genes in a single operon encode an efflux pump (*acrAB)* and show the highest log_2LR (first, 3.04 and third, 1.64 in the list). This pump has been shown to significantly contribute to multiple-antibiotic resistance phenotype in *E. coli* (Okusu, Ma & Nikaido 1996) and *S.* Typhimurium (Piddock *et al.* 2000). This suggests a clear link between *acrAB* mediated resistance and chloramphenicol in the media. The role of *acrAB* in Cm resistance in a Cm acetyltransferase expressing strain has been previously reported in *E. coli* (Potrykus, Baranska & Wegrzyn 2002) and so verifies the assay.

BarA (Log_2LR = 2.37) is a sensor protein that plays a global response regulatory role in cell division, carbon metabolism, iron metabolism and pili formation (Sahu *et al.* 2003). This gene is also well conserved across many serovars of *S. enterica* (nucleotide identity of 98% or more) and also across many *Enterobacteriacae*. BarA belongs to a two-component signal-transduction system: BarA-SirA (Altier *et al.* 2000). In *S.* Typhimurium, BarA-SirA activates SPI1 genes, including the type III secretion system and its effector proteins (Sips), in response to high salt concentration (300mM NaCl) (Mizusaki *et al.* 2008).

The updated annotation of t0479 predicts the gene to encode a N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase, involves in methylation of ribosomal protein L3 (STY2617, NP_456926.1).

SmpB is a tmRNA-binding protein which binds to SsrA RNA to mediate the addition of a short peptide tag to the C-terminus of the partially synthesized polypeptide chain for degradation. SmpB knockout results in phage development defects and failure to tag protein translated from defective mRNA (Karzai, Susskind & Sauer 1999). A SmpB-SsrA mutant in *Yersinia pseudotuberculosis* suffered severe deficiencies in expression

and secretion of *Yersinia* virulence effector proteins, resulting in avirulent phenotype and inability to proliferate in macrophages (Okan, Bliska & Karzai 2006). An SmpB deletion mutant in *S.* Typhimurium affects the expression of 189 proteins in the bacterium proteome, rendering the mutant avirulent and defective in intramacrophage proliferation (Ansong *et al.* 2009). Deletion of *ssrA* in cyanobacterium *Synechocystis* sp. strain PCC6803 results in mutants that are not viable in the presence of the protein synthesis inhibitors chloramphenicol, lincomycin, spiramycin, tylosin, erythromycin, and spectinomycin at low doses that do not significantly affect the growth of wild-type cells (de la Cruz, Vioque 2001). This hyper-sensitivity phenotype of *ssrA* suggests that a SmpB knock-out would also show sensitivity to chloramphenicol.

UvrA is a nucleotide excision repair protein subunit A (Selby, Sancar 1990) involved in DNA damage repair, such as UV radiation damage. This subunit recognises damage DNA and delivers subunit UvrB to the damage site. UvrC then recognises UvrB-damage DNA complex for the excision and repair of the damage DNA.

Table 3-1 Top genes on the chromosome contributing to the survival in chloramphenicol passages

Non-CmP6 Total inserts	Non-CmP6 Total reads	Gene length	Sys ID	Name	CmP6 Total inserts	CmP6 Total reads	$\log_2$(RR)	probability	Insert site diff	Function
67	804	3131	t2385	acrB	3	10	3.04	1	64.00	acriflavin resistance protein B
71	1200	2738	t2867	barA	37	151	2.37	0.9999	34.00	sensor protein
21	220	1175	t2384	acrA	1	3	1.64	0.9981	20.00	acriflavin resistance protein A precursor
23	259	914	t0479	-	8	17	1.62	0.9978	15.00	conserved hypothetical protein
33	562	464	t2642	smpB	23	114	1.63	0.9980	10.00	SsrA (tmRNA)-binding protein
49	1091	812	t3110	-	39	340	1.44	0.9891	10.00	conserved hypothetical

Non-CmP6 Total inserts	Non-CmP6 Total reads	Gene length	Sys ID	Name	CmP6 Total inserts	CmP6 Total reads	log_2(RR)	probability	Insert site diff	Function
										protein
93	1976	2807	t4160	uvrA	83	446	1.93	0.9999	10.00	excision nuclease subunit A

Note: see appendix 8.2 for the full list of genes.

3.2.6 Plasmid mediated cell death and plasmid stability genelist

In order to identify plasmid borne candidate genes for plasmid stability the changes in the insertion index (the number of reads per base pair) of each gene from day 0 to day 6 were compared. Non-CmP and CmP conditions should give a similar gene list because disruptions in stability genes would cause plasmid loss in both conditions. Table 3-2 shows the most significant candidates identified from non-Cm and Cm passages. The level of significance was measured by log_2 of read ratio (log_2RR) and the probability of reads in a gene from day 6 being higher than those in day 0. We used *hok*, a member of hok/sok toxin/antitoxin system involved in the post segregational killing of plasmid free cells to ensure the stable inheritance of plasmid in the population, as a known marker and considered genes with a log_2RR higher than that of *hok* to be candidates for plasmid stability.

Table 3-2 Top plasmid gene candidates for plasmid stability recovered from control and Cm passages after 6 days

Non-CmP day 6 against day 0

Total inserts	Total reads	Gene length	Systematic ID	Name	Day 0 reads	log_2RR	Probability	Gene function
23	**848**	**608**	**HCM1.243**	***tetR***	**5851**	**-2.65017**	**1**	tetracycline repressor protein
61	**686**	**989**	**HCM1.87**	***parA***	**3724**	**-2.28248**	**1**	putative plasmid partition protein
63	**1034**	**1235**	**HCM1.86**	***parB***	**3265**	**-1.56919**	**1**	putative plasmid partition protein
14	**369**	**125**	**HCM1. 141ac**	**-**	**619**	**-0.6164**	**0.999919**	hypothetical protein
79	**7911**	**386**	**HCM1. 178ac**	***sfh***	**10039**	**-0.33986**	**0.996986**	putative DNA-binding protein

Total inserts	Total reads	Gene length	Systematic ID	Name	Day 0 reads	log_2RR	Probability	Gene function
4	69	38	HCM1.166c	-	102	-0.25733	0.992663	putative aminoglycoside acetyltransferase
23	**322**	**437**	**HCM1.128**	**-**	**401**	**-0.24757**	**0.991894**	putative membrane protein
10	274	122	HCM1.53	-	329	-0.19794	0.986795	hypothetical protein
18	231	374	HCM1.145	-	261	-0.12517	0.974434	hypothetical protein
58	**2688**	**368**	**HCM1.245c**	**-**	**2793**	**-0.05334**	**0.953917**	hypothetical protein
18	**669**	**260**	**HCM1.130**	**-**	**697**	**-0.0516**	**0.95329**	hypothetical protein
62	**1558**	**572**	**HCM1.125**	**-**	**1618**	**-0.05129**	**0.953178**	putative membrane protein
23	**4312**	**341**	**HCM1.45**	**-**	**4403**	**-0.02945**	**0.944704**	hypothetical protein
61	**1706**	**695**	**HCM1.277**	**-**	**1743**	**-0.02925**	**0.944623**	putative periplasmic protein
11	**275**	**140**	**HCM1.290c**	***hok***	**281**	**-0.02290**	**0.941937**	putative stable plasmid inheritance protein

CmP day 6 against day 0

Total inserts	Total reads	Gene length	Systematic ID	Name	Day 0 reads	log_2RR	Probablitity	Gene function
11	187	641	HCM1.206	*cat*	4781	-4.08805	1	chloramphenicol acetyltransferase
27	**417**	**608**	**HCM1.243**	***tetR***	**5851**	**-3.5249**	**1**	tetracycline repressor protein
55	**258**	**989**	**HCM1.87**	***parA***	**3724**	**-3.41705**	**1**	putative plasmid partition protein
74	**481**	**1235**	**HCM1.86**	***parB***	**3265**	**-2.534**	**1**	putative plasmid partition protein
81	**4009**	**386**	**HCM1. 178ac**	***sfh***	**10039**	**-1.30306**	**0.997981**	putative DNA-binding protein
48	1548	251	HCM1.124	-	3373	-1.07547	0.979964	hypothetical protein
230	7431	1016	HCM1.92	-	15677	-1.06691	0.978417	putative plasmid stability/partition protein
97	1522	659	HCM1.93	-	3164	-1.00887	0.965039	hypothetical protein
30	**2144**	**341**	**HCM1.45**	**-**	**4403**	**-1.00481**	**0.963893**	hypothetical protein
65	**783**	**572**	**HCM1.125**	**-**	**1618**	**-0.96024**	**0.949159**	putative membrane protein
18	113	332	HCM1.182	-	314	-0.95877	0.948603	hypothetical protein
12	**272**	**125**	**HCM1. 141ac**	**-**	**619**	**-0.95068**	**0.945448**	hypothetical protein
152	3217	854	HCM1.106c	-	6256	-0.93824	0.940296	putative lipoprotein
29	234	302	HCM1.246c	-	536	-0.92917	0.936311	hypothetical protein
20	**326**	**260**	**HCM1.130**	**-**	**697**	**-0.90372**	**0.924011**	hypothetical protein
131	4487	761	HCM1.190	-	8470	-0.90174	0.922982	hypothetical protein
58	831	1202	HCM1.183	-	1630	-0.89391	0.918816	hypothetical protein
102	1609	467	HCM1.269	-	3047	-0.88082	0.911467	hypothetical protein
70	**910**	**695**	**HCM1.277**	**-**	**1743**	**-0.86770**	**0.903616**	putative periplasmic protein
64	**1488**	**368**	**HCM1.245c**	**-**	**2793**	**-0.86535**	**0.902161**	hypothetical protein
60	4772	302	HCM1.24c	-	8700	-0.85298	0.894220	hypothetical protein
115	2581	989	HCM1.100	*trhU*	4730	-0.84925	0.891731	plasmid transfer protein
62	2182	416	HCM1.95	*htdF*	3985	-0.84003	0.885415	putative periplasmic

Total inserts	Total reads	Gene length	Systematic ID	Name	Day 0 reads	log_2RR	Probablitity	Gene function
								protein
104	1104	986	HCM1.209c	-	2047	-0.83448	0.881486	putative transposase
24	**181**	**437**	**HCM1.128**	**-**	**401**	**-0.83424**	**0.881309**	putative membrane protein
18	294	323	HCM1.44	-	600	-0.82915	0.877626	hypothetical protein
107	4160	341	HCM1.199c	-	7416	-0.81911	0.870108	hypothetical protein
116	2881	689	HCM1.61c	-	5126	-0.80991	0.862949	hypothetical protein
38	440	251	HCM1.112	-	846	-0.80888	0.862131	hypothetical protein
566	17887	2663	HCM1.77	*trhC*	31095	-0.79436	0.850245	plasmid transfer protein
10	**121**	**140**	**HCM1.290c**	***hok***	**281**	**-0.78574**	**0.842879**	putative stable plasmid inheritance protein

Genes in bold appear in both non-CmP and CmP gene list.

Chloramphenicol acetyltransferase gene *cat* was highly protected in CmP as oppose to non-CmP (Table 3-2 and Figure 3-8). This enzyme inactivates Cm by covalently binding one or two acetyl groups to the hydroxyl groups on the Cm molecule. The known mechanism of the product of *cat* to confer resistance to Cm is our positive confirmation that the passage worked in selection against mutations within the *cat* gene.

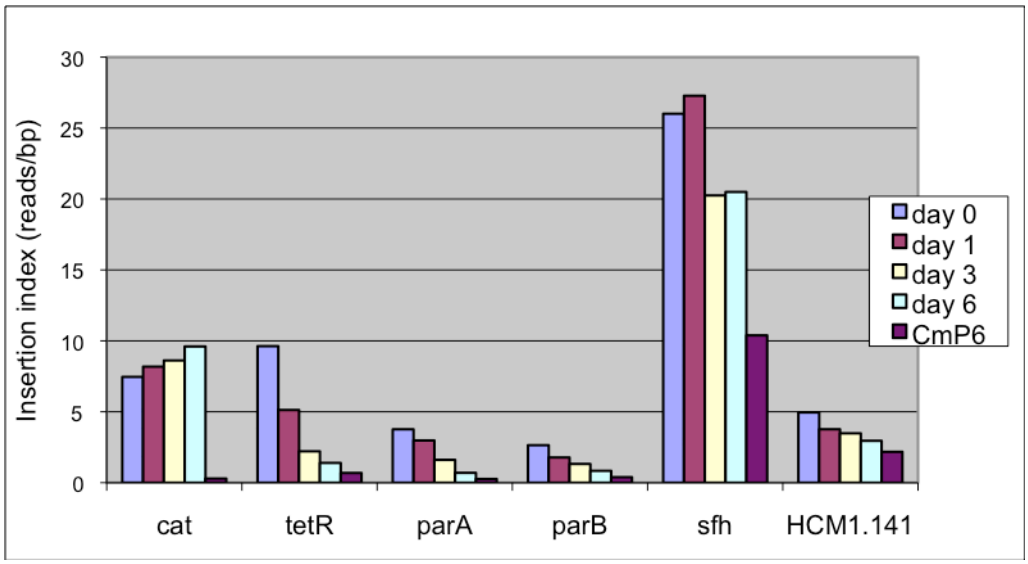
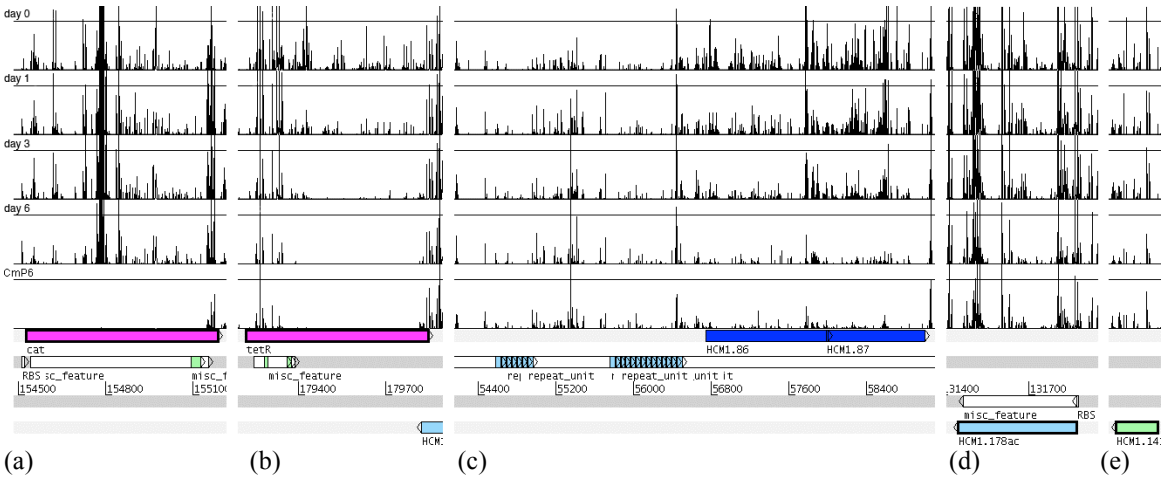


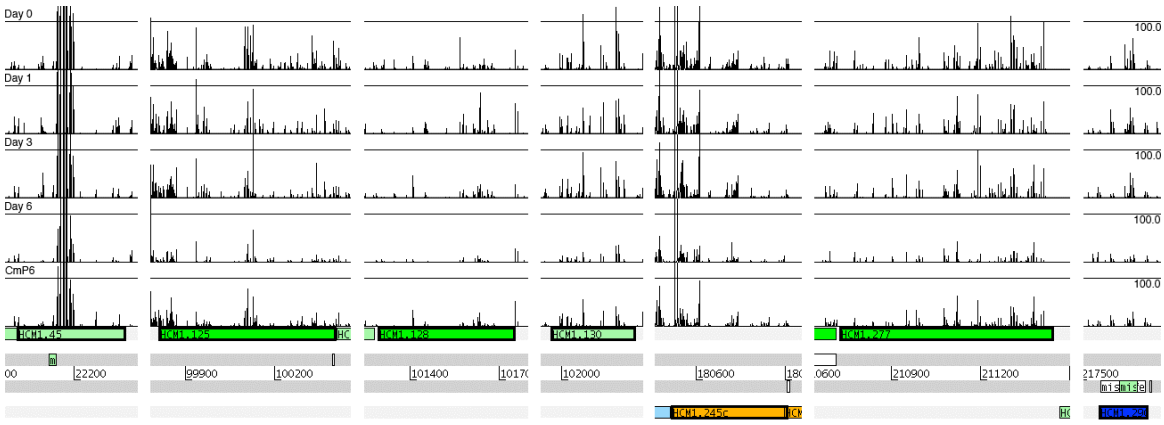
Figure 3-8 Changes in insertion index (number of reads per base pair) of candidate genes across passages and in the chloramphenicol selection passage at day 6 (CmP6)

The changes in insertion index (i.e. number of reads per bp) for several top genes are shown in Figure 3-8. With the exception of *cat*, mutants in other genes gradually

decrease overtime and were lowest in CmP. Care should be taken in interpretation of these data because although a pattern of a gradual decrease in mutants of a gene indicates good candidates for plasmid stability, it may also indicate survival mutants; insertions in plasmid genes that adversely affect cellular growth would also slowly disappear from the passages. One example of this plasmid mediated killing phenomenon is perhaps the *tetR* repressor. The Tet operon on Tn*10* is well characterised: TetR is a tetracycline dependent regulator and *tetA* encodes a metal-tetracycline/H^+ antiporter (Yamaguchi, Someya & Sawai 1992). The expression of *tetA* in the absence of tetracycline causes partial collapse of the membrane potential, arrest of growth and killing of the cells (Eckert, Beck 1989). The repressor *tetR* is also protected in experiments to select spontaneous mutations on the Tet operon within Tn*10* encoded on a multicopy plasmid (Moyed, Bertrand 1983). The data presented here confirms that a pattern of a gradual decrease in mutants of a gene indicates good candidates for plasmid stability but may also indicate insertions in plasmid genes that adversely affect cellular growth. Our plasmid stability gene list in fact likely contains a mixture of real plasmid stability genes and genes, which when knocked out, cause plasmid associated killing of the host cell.



(a) (b) (c) (d) (e)



(f)

Figure 3-9 Changes in number of insertion in top candidates for plasmid stability genes
(a) chloramphenicol resistant gene (*cat*); (b) repressor of tetracycline resistant operon (*tetR*); (c) partition region including *parAB*; (d) *hns*-like gene HCM1.178ac; (e) hypothetical gene HCM1.141ac. (f) seven other candidates for plasmid stability. The graphs were drawn with window size 1, maximum read of 100.

The involvement of *parA* and *parB* in IncHI1 plasmid stability has been shown previously (Lawley, Taylor 2003). The double deletion of *parA* and *parM,* the minor partition genes, resulted in the integration of R27 plasmid into the chromosome (Lawley, Taylor 2003).

```
>lcl|35551 Sfh_shigella
Length=134

 Score =  257 bits (657),  Expect = 4e-74, Method: Compositional matrix adjust.
 Identities = 132/134 (98%), Positives = 132/134 (98%), Gaps = 0/134 (0%)

Query  1    MSEALKSLNNIRTLRAQGRELPLEILEELLEKLSVVVEERRQEESSKEAELKARLEKIES  60
            MS ALKSLNNIRTLRAQGRELPLEILEELLEKLSVVVEERRQEESSKEAELKARLEKIES
Sbjct  1    MSGALKSLNNIRTLRAQGRELPLEILEELLEKLSVVVEERRQEESSKEAELKARLEKIES  60

Query  61   LRQLMLEDGIDPEELLSSFSAKSGAPKKVREPRPAKYKYTDVNGETKTWTGQGRTPKALA  120
            LRQLMLEDGIDPEELLS FSAKSGAPKKVREPRPAKYKYTDVNGETKTWTGQGRTPKALA
Sbjct  61   LRQLMLEDGIDPEELLSPFSAKSGAPKKVREPRPAKYKYTDVNGETKTWTGQGRTPKALA  120

Query  121  EQLEAGKTLDDFLI  134
            EQLEAGKTLDDFLI
Sbjct  121  EQLEAGKTLDDFLI  134

(a)
```

```
>lcl|54641 Hns_CT18
Length=137

 Score =  141 bits (355),  Expect = 4e-39, Method: Compositional matrix adjust.
 Identities = 82/135 (60%), Positives = 99/135 (73%), Gaps = 1/135 (0%)

Query  1    MSEALKSLNNIRTLRAQGRELPLEILEELLEKLSVVVEERRQEESSKEAELKARLEKIES  60
            MSEALK LNNIRTLRAQ RE  LE LEE+LEKL VVV ERR+EES+  AE++ R  K++
Sbjct  1    MSEALKILNNIRTLRAQARECTLETLEEMLEKLEVVVNERREEESAAAAEVEERTRKLQQ  60

Query  61   LRQLMLEDGIDPEELLSSFSAKSGAPKKVREPRPAKYKYTDVNGETKTWTGQGRTPKALA  120
             R++++ DGIDP ELL+S +A     K  R  RPAKY Y D NGETKTWTGQGRTP  +
Sbjct  61   YREMLIADGIDPNELLNSMAAAKSGTKAKRAARPAKYSYVDENGETKTWTGQGRTPAVIK  120

Query  121  EQL-EAGKTLDDFLI  134
            + + E GK L+DFLI
Sbjct  121  KAMEEQGKQLEDFLI  135

(b)



>lcl|35767 StpA_CT18
Length=133

 Score =  140 bits (354),  Expect = 5e-39, Method: Compositional matrix adjust.
 Identities = 77/134 (57%), Positives = 101/134 (75%), Gaps = 1/134 (0%)

Query  1    MSEALKSLNNIRTLRAQGRELPLEILEELLEKLSVVVEERRQEESSKEAELKARLEKIES  60
            M+  L++LNNIRTLRA  RE  +++LEE+LEK  VV +ERR+EE  ++ +L  + EKI +
Sbjct  1    MNLMLQNLNNIRTLRAMAREFSIDVLEEMLEKFRVVTKERREEEELQQRQLAEKQEKINA  60

Query  61   LRQLMLEDGIDPEELLSSFSAKSGAPKKVREPRPAKYKYTDVNGETKTWTGQGRTPKALA  120
              +LM  DGI+PEEL +  SA   + KK R+PRPAKY++TD NGE KTWTGQGRTPK +A
Sbjct  61   FLELMKADGINPEELFAMDSAMPRSAKK-RQPRPAKYRFTDFNGEEKTWTGQGRTPKPIA  119

Query  121  EQLEAGKTLDDFLI  134
            + L AGK+LDDFLI
Sbjct  120  QALAAGKSLDDFLI  133

(c)
```

Figure 3-10 Pair-wise comparisons of Sfh protein on pHCM1 with its homologues
Comparison of pHCM1 Sfh against (a) Sfh protein from *Shigella flexneri* 2a 2457T, (b) Hns protein from *S.* Typhi CT18 and (c) StpA protein from *S.* Typhi CT18. The comparison was run using BLAST for protein (blastp).

The *sfh* gene also contributes to plasmid stability by silencing plasmid genes to minimise the interference to chromosomal gene regulation (Doyle *et al.* 2007, Banos *et al.* 2009). The Sfh protein, the third member of H-NS-like protein family, was first reported in *Shigella flexneri* 2a 2457T to be encoded on an R27-like plasmid, (Beloin *et al.* 2003). Two other H-NS-like proteins are H-NS and StpA (Dorman, Hinton & Free 1999). All three proteins were also found in *S.* Typhi CT18 harbouring pHCM1 plasmid. The homology of these proteins is shown in Figure 3-10. The DNA binding profile of Sfh to promoters of virulence genes and to DNA curvature (similar to the

binding of H-NS and StpA) suggests its role in regulating virulence genes and the interaction of these paralogues in a complex regulatory network within the cell (Beloin *et al.* 2003). It was also shown that the Sfh in R27 interacts with Hha to thermo-regulate the conjugation of IncHI1 plasmid (Alonso *et al.* 2005, Forns *et al.* 2005). The Hha protein in pHCM1 (HCM1.135), however, was not identified in our experiment as contributing factor to the stable plasmid inheritance. This might be due to the fact that our passages were performed at non-permissive temperature (37°C) for plasmid conjugation.

The remaining genes in our list are good candidates for further investigation, especially those genes highlighted from both passage conditions (Table 3-2 and Figure 3-9). These include one putative periplasmic protein (HCM1.277), two putative membrane proteins (HCM1.125 and HCM1.128) and four hypothetical proteins (HCM1.45, 130, 141ac and 245c). HCM1.277 belongs to the nuclease-related domain (NERD) superfamily. HCM1.125 encodes a potential ribonucleotide-diphosphate reductase subunit alpha domain (PRK07632). With the exception of HCM1.245c, which shares close similarity to proteins in other plasmids, all these genes are unique for IncHI1 plasmids.

3.2.7 Growth curves of sfh knock-out

Previous evidence for the involvement of *sfh* in the silencing of plasmid genes to avoid the disruption of chromosomal regulation is based on experiments in *S.* Typhimurim (Doyle *et al.* 2007, Banos *et al.* 2009). Although similar IncHI1 plasmids have evolved in *S.* Typhi, it is possible that there are unique interactions between pHCM1 and *S.* Typhi. We generated *sfh* deletion mutants in pHCM1 to further investigate the plasmid-chromosome interaction in *S.* Typhi. The *sfh* deletion was generated by an allelic exchange strategy (Turner, Nair & Wain 2006). The strategy was designed to use homologous recombination to swap the *sfh* gene with a kanamycin resistant marker.

The mutant genotype was then confirmed by sequencing. Three Δ*sfh* mutants were generated independently as biological replicates.

This section presents the growth curves of Δ*sfh* mutants in comparison with the wild type (Figure 3-11). There is no significant difference in the observed growth rates between the mutants and wildtype strains. The effect of Δ*sfh* on cellular growth is perhaps too subtle to be detected in the growth curves. Other methods such as competitive growth, long-term plasmid stability assay or gene expression analysis are needed to characterise these mutants. Data from Doyle et al (2007) suggests that the knock out of *sfh* in *S.* Typhi might also reduce the relative fitness of the mutant whilst enhance the level of survival in macrophage.

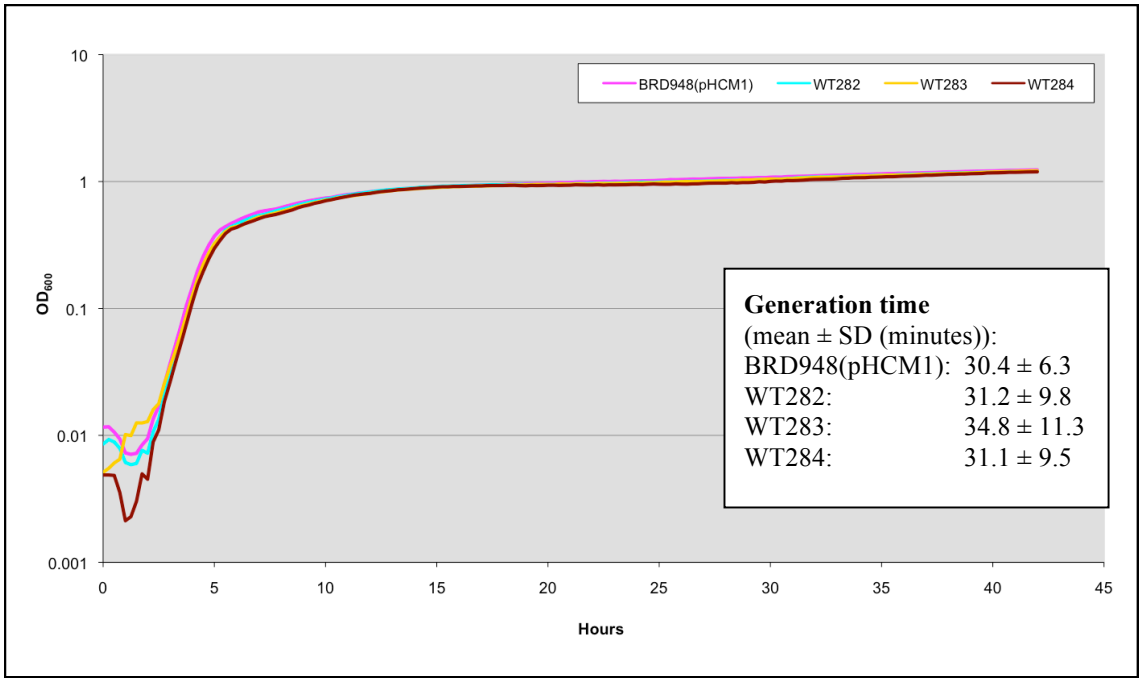


Figure 3-11 Growth curves of Δ*sfh* mutants (WT282, WT283 and WT284) and wildtype strains
(See Methods section 2.2.4.3 for the calculation of generation time)

3.3 Discussion

Transposon insertion mutagenesis is the method of choice for genome wide high throughput investigation of genetic essentiality due to its speed and cost effectiveness. However, major drawbacks of this method involve the capability to generate a saturated

mutant library and the ability to accurately identify each insertion site. In this chapter, we presented a novel method to simultaneously and precisely identify a very large number of transposon insertion sites using Illumina sequencing. The actual sequence of every insertion site provides unprecedented clear signals for each transposon insertion in a semi-quantitative manner. Our mutant library consisted of more than 1 million mutants, which proved essentially to be close to saturation for the method employed. Our transposon-directed insertion-site sequencing (TraDIS) method has therefore addressed some of the disadvantages of previous transposon mutagenesis methods.

TraDIS should be easily applicable to other transposon libraries without the need to re-generate them. One simply needs to re-design PCR primers and sequencing primers for Illumina sequencing for the new transposon. One bottleneck of this method, however, is the generation of a large mutant library. Optimisation of a suitable transposon for a particular organism is of importance to achieve the high frequency of randomised mutagenesis.

The semi-quantitative data enabled us to compare not only the location of insertion events but also the frequency of each event. This is particularly useful in the discovery of long-term survival genes over serial passages. The gradual disappearance of certain mutants over time identifies any genes contributing to survival or fitness. This is of particular importance in assessing plasmid stability because the copy number of plasmids in some cells may be more than one and so knocking out only one copy of a multi-copy gene may not affect the plasmids stability. The measurement of the comparative “success” of each mutant identifies genes that are costly to cellular growth (in the conditions tested) and the deletions of such genes are therefore beneficial (data not shown). The ability to follow the dynamic changes of almost every mutant overtime could also be powerful in multi-stage experimental designs such as cell adhesion and

invasion assays. One mutant library can therefore be used to investigate a range of the biological features associated with bacterial cells.

We used TraDIS to identify plasmid stability genes on both the pHCM1 plasmid and the *S.* Typhi Ty2 chromosome. Two parallel serial passages of up to 6 days in rich media, one with Cm and one without, were used as negative selection for plasmid stability mutants. For chromosome-encoded plasmid stability genes, non-Cm passages were also used to set a baseline for essential genes and long-term survival genes. The genes identified were subtracted in the data from the Cm passages. For plasmid-encoded plasmid stability genes the two passages were effectively duplicates and generated similar gene lists ranked by the difference in insertion index for each gene between day 0 and day 6. The chromosome-encoded genes identified by our assays are a mixture of Cm resistance associated genes and stability genes. The mutants highlighted in Cm passages are those that rendered the cells unfit to remain after 6 days of serial culture in rich media (LB) broth supplemented with Cm. Based on literature information, it is most likely that *acrA* and *acrB* are involved in Cm resistance (Okusu, Ma & Nikaido 1996, Piddock *et al.* 2000). It is known that the action of chloramphenicol acetyl transferase (Cat) alone leads to the depletion of intracellular acetyl coenzyme A, hence the *acrAB* efflux pump system may compensate for Cat activity (Potrykus, Baranska & Wegrzyn 2002). This phenomenon however appears to be in a background specific for *E. coli* strain CM2555 which has a dysfunctional *acrA* gene (Potrykus, Baranska & Wegrzyn 2002), suggesting that in other *E. coli* there might be another pump (*acrEF*) contributing to Cm resistance. In our *S.* Typhi background however, it appears that *acrAB* is the sole efflux pump system for Cm resistance. SmpB is a tmRNA that plays an important role in the degradation of partially synthesized polypeptide chain (Karzai, Susskind & Sauer 1999). There is also evidence suggesting that SmpB deletion might

cause hypersensitivity to Cm (de la Cruz, Vioque 2001). We also discovered a hypothetical protein t3110 that has never been reported as involved in Cm resistance. This gene is predicted to encode an extradiol ring-cleavage dioxygenase class III enzyme, which potentially has a role in degradation of aromatic compounds, of which Cm is one. Experimental evidence is still needed but our preliminary conclusion for this gene is that it is involved in antibiotic resistance.

Two genes *barA* and *uvrA* have known functions (Sahu *et al.* 2003, Selby, Sancar 1990) but none of these suggested a role in either Cm resistance or plasmid stability. The t0479 gene is also not previously predicted to be involved in plasmid stability. We can conclude based on our assay that *barA, uvrA* and t0479 are candidates for plasmid stability genes encoded on Ty2 chromosomes.

We were able to rank the plasmid stability candidate genes based on the decrease of mutants between day 0 and day 6 ($\log_2$(read ratio)). It is however difficult to define a cut-off for plasmid stability because we have no prior knowledge of how such mutant differences should be accounted for. Stability and copy number may both be contributing factors. Our statistical analysis can only provide a measure of how significant a difference is but not how likely a gene is to be responsible for stability. The *hok* gene on pHCM1 is similar to the host-killing gene on plasmid R1, which has been shown to contribute to the maintenance of plasmids (Gerdes, Rasmussen & Molin 1986). It is therefore very likely that pHCM1 *hok* gene contributes to the stable maintenance of this plasmid. We thus used this gene as a phenotypic cut-off. Genes that show higher level of difference than *hok* are more likely to contribute to plasmid stability.

Because any mutants that de-stabilise the plasmid would cause plasmid loss in both non-Cm and Cm passages, the plasmid stability genes were identified as the genes

showed significant decreased in mutants between day 0 and day 6 in both conditions. Non-Cm and Cm passages were in this case considered as two replicates. It is worth noting that the difference is bigger in CmP than in non-CmP for the same gene. For example, $\log_2$RR of *sfh* in non-CmP is -0.33986 whilst in CmP is -1.30306. This means the number of *sfh* mutants in CmP decreases more rapidly than in non-CmP. The presence of chloramphenicol is likely to have attributed to this because of competition during growth; with no plasmid free cells (killed by chloramphenicol) plasmid positive cells could grow to higher densities, or during the sequencing reactions, chromosomal DNA in plasmid free cells may have diluted plasmid DNA.

Any plasmid mutants that are lethal to the cells were also selected by this analysis. TetR mutants are potentially an example of plasmid mediated cell death. The disruptions of *tetR*, the *tetA* repressor, would result in constitutive over expression of *tetA*. The presence of TetA, a proton antiporter, in the absence of tetracycline causes loss of membrane potential resulting in cell death (Eckert, Beck 1989).

A literature search of other top genes on the list did not provide alternative evidence to support their role in plasmid mediated bacterial cell death. We therefore believed that they are candidates for plasmid stability genes. Apart from *tetR*, the two partition genes *parA* and *parB* showed most significant decrease in their mutants after 6 days in both conditions. The *parA* gene encodes a Walker-type ATPase similar to those in P1/F plasmids and *parB* encodes a DNA-binding protein that binds to the centromere region. This partitioning module has been shown to contribute significantly to IncHI1 plasmid stability, especially in condition causing slow growth (Lawley, Taylor 2003). The identification of *parAB* partitioning module proves that our method is picking up plasmid stability genes.

Other genes that were highlighted include HCM1.178ac (*sfh*), HCM1.45, HCM1.125, HCM1.141ac, HCM1.130, HCM1.277, HCM1.245c and HCM1.128. The *sfh* is an *hns*-like gene on IncHI1 plasmids that has been shown to play an important role in minimising the bacterial fitness cost by minimising the regulatory disruption caused by the presence of a large plasmid (Doyle *et al.* 2007, Doyle, Dorman 2006, Banos *et al.* 2009). HCM1.125 and HCM1.128 encodes two putative membrane proteins with the predicted signal peptides and transmembrane domains. HCM1.277 carries a nuclease-related domain (NERD) superfamily, which suggests a role in DNA processing and this may have nuclease function (IPR011528). The remaining genes encode hypothetical proteins. These genes are all conserved within IncHI1 plasmids. Thus, our assay has successfully identified candidate genes for plasmid stability along with genes known to be involved in chloramphenicol resistance and plasmid mediated cell death.

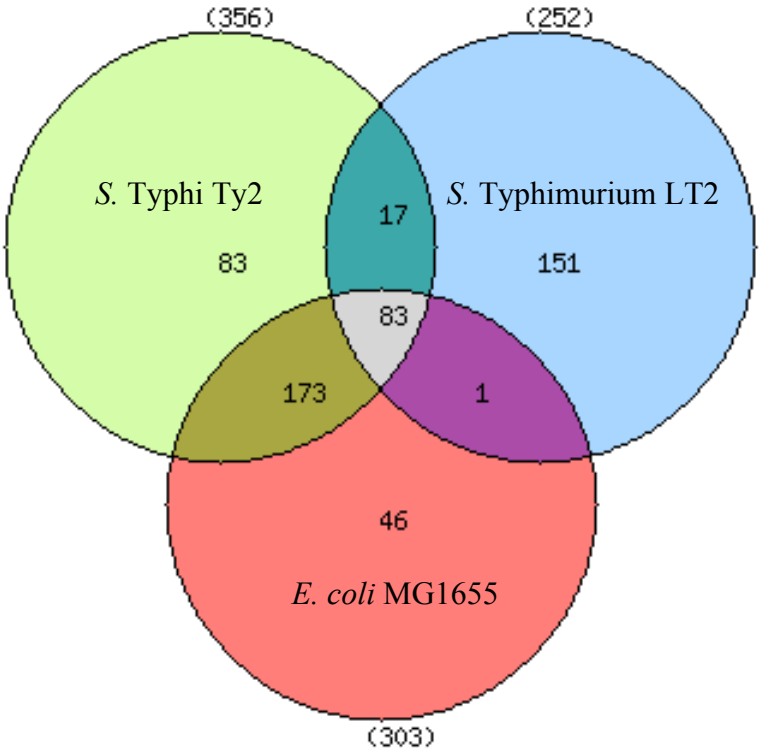


Figure 3-12 Comparing essential genes of *S.* Typhi, *S.* Typhimurium and *E. coli*
The essential genes of *S.* Typhi are taken from this study, *S.* Typhimurium from Knuth *et al.* (2004) and *E. coli* from Baba *et al.* (2006)

It is encouraging that our method provides high enough resolution to precisely discern and trace the dynamic changes of any mutant of interest among million of other mutants. Our list of 356 essential genes is compatible with other studies (Baba *et al.* 2006, Knuth *et al.* 2004, Zhang, Lin 2009, Zhang, Zhang 2008) (Figure 3-12). There are 256 genes shared between our S. Typhi essential gene candidates (70%) and *E. coli* (84%) (Baba *et al.* 2006). Surprisingly, only 40% of the essential genes in *S.* Typhimurium LT2 (100 genes) (Knuth *et al.* 2004) are also in our list. It is worth noting that the methods to identify essential genes are different in these three studies. Our passage data also suggests that a minimal bacterium with full fitness competency and long-term survival might need a larger set of genes, one of approximately 600 genes. Our plasmid stability gene candidates have provided insightful information to prioritise future research on this field.

4 Plasmid Multi-Locus Sequence Typing for IncHI1 plasmids in *S.* Typhi

4.1.1 Introduction

Plasmids are horizontally acquired extra-chromosomal DNA molecules that represent the major depots and acquisition routes for antibiotic resistant genes in many bacterial species (Sorensen *et al.* 2005). Resistance plasmids are generally large (50-300 kb) and consist of distinct functional backbone regions, composed of genetic modules that enable functions such as autonomous replication and stable maintenance inside bacteria (Summers, Beton & Withers 1993). Within this core backbone plasmids can accumulate DNA insertions and mobile elements such as integrons (Hall, Stokes 1993) or simple or composite transposons (Salyers *et al.* 1995), which may carry multiple antibiotic resistant genes. The dynamic interactions between the plasmid backbone, the mobile elements and the bacterial host can shape both bacterial and plasmid evolution and coevolution. Since the introduction of widespread antimicrobial therapy in the 1950s, selective pressure has selected for bacteria expressing resistance to the antibiotics used, hence there is co-selection for plasmids harbouring individual and subsequently multiple antibiotic resistance genes. Indeed the serial capture of mobile elements by plasmids has played a major role in the evolution of plasmid encoded MDR phenotypes across many bacterial species.

Significant research effort has, therefore, been focused on the characterisation of the antibiotic resistance genes, the cargo on the plasmid backbones. Consequently, comparatively, little is known about the significance and type of changes that can occur within the plasmid backbone, which is more likely to reflect the longer term evolutionary history of the plasmid than the horizontally acquired antibiotic resistance

genes. Some studies have suggested that a plasmid and its bacterial host can co-evolve over time in batch culture to reduce the initial fitness cost caused by the presence of plasmid in the cell (Dionisio *et al.* 2005, Dahlberg, Chao 2003). In the absence of new horizontally acquired sequences, this observation suggests a potential contributing role for the plasmid backbone in this evolution.

Since the early 1970s antibiotic resistance has become established in *Salmonella enterica* serovar Typhi, the cause of human typhoid (Wain, Kidgell 2004). The first resistant *S.* Typhi isolate was described in 1950, two years after chloramphenicol was first used for treating typhoid fever (Woodward, Smadel 1948, Colquhoun, Weetch 1950). However, it took 22 more years before chloramphenicol resistance became a major clinical problem with the first major outbreak of chloramphenicol resistant typhoid fever in 1972 (Olarte, Galindo 1973) Thereafter, the global spread of chloramphenicol resistant *S.* Typhi occurred very rapidly and over the next two decades further antibiotic resistance genes were acquired to generate multiple drug resistant (MDR) *S.* Typhi (Wain, Kidgell 2004). Although occasionally IncA/C and IncI plasmids are found (Datta, Olarte 1974, Datta, Richards & Datta 1981), it is striking that the majority of R plasmids from *S.* Typhi belong to the HI1 incompatibility group. Thus, IncHI1 plasmids appear to have evolved a stable relationship with the *S.* Typhi bacterial host so that even in the absence of obvious antibiotic selection, a proportion of the isolates maintain the plasmid (Dutta *et al.* 2005).

In Vietnam during the year 1993 and 1996, IncHI1 plasmids were isolated from *S.* Typhi causing outbreaks and sporadic typhoid cases. The authors observed a diversity of MDR IncHI1 plasmids in *S.* Typhi isolated before 1996, initially discriminated on the basis of seven different *Hind*III-digested RFLP patterns, with pattern 1 being the most common. Significantly, this trend of diversity changed after 1996 whereby a single

pattern 7 type became predominant whilst plasmids of other patterns seemed to disappear (Wain *et al.* 2003). It was hypothesised that a significant event in IncHI1 plasmid evolution had been identified, in which one type of plasmid evolved that drove the less fit plasmids of the same family out of the population.

Following the description of seven RLFP patterns within IncHI1 plasmid population, a gene content variation exhibited by these plasmids was investigated by Clare Kidgell (PhD thesis) using a DNA microarray. Figure 4-1 shows data covering the presence or absence of different genes on the various plasmids investigated; genes which were found to be present in all plasmids tested were defined as core. These core genes include several involving replication, maintenance, as well as many hypothetical genes of unknown function. This microarray data also highlighted 5 variable regions (region A to E, Figure 4-1), of which two (region A and B) are hotspots for the incorporation of mobile elements and antibiotic resistance genes.

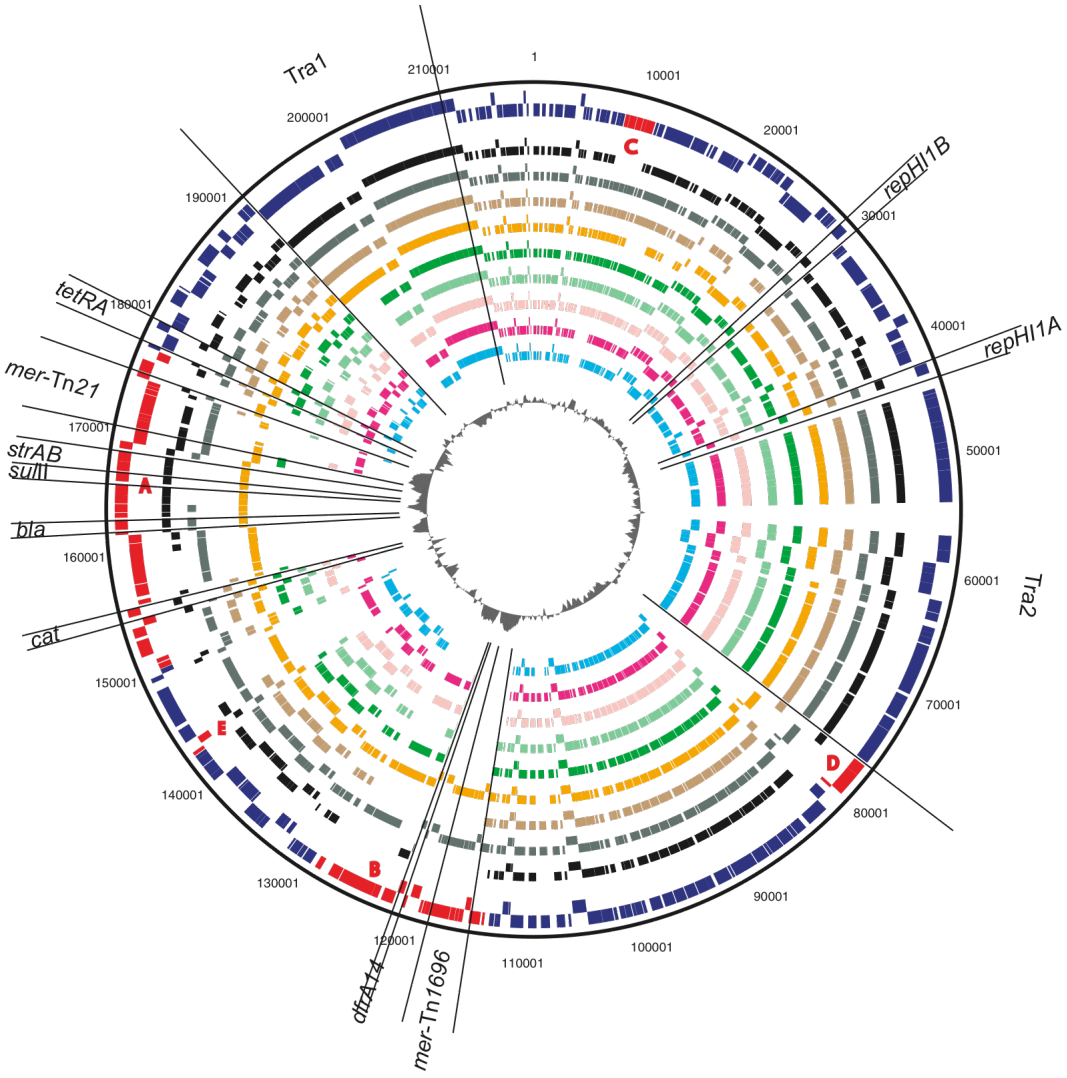


Figure 4-1 Genetic variation in IncHI1 plasmids from *S.* Typhi as determined by microarray analysis.
Predicted coding sequences (CDS) are defined in the sequence of pHCM1 in the outer ring (NC_003384). The inner concentric rings each represent the CDS predicted as present or absent in each of the different plasmid by microarray analysis. The order, from the inside is as follows: blue R27 (UK, 1961), magenta 40R181 (Mexico, 1972), pink 40R344 (India 1972), light green 40R311 (Thailand 1970's), dark green 42R917 (Vietnam 1970's), yellow pSTY1 (Vietnam 1993), sand pSTY4 (Vietnam 1993), grey pSTY6 (Vietnam 1993), black pSTY7 (Vietnam, 1996), and red and blue pHCM1 (Vietnam 1993) on the outside. The CDS of pHCM1 are marked blue when present and red when absent from the R27 sequence (NC_002305). Data from Kidgell (PhD thesis).

In this chapter, we extend the previous studies to investigate our hypothesis that a new plasmid type is spreading in the *S.* Typhi population by analysing data from a global collection of IncHI1 plasmids isolated from *S.* Typhi over the past several decades. A combination of long-range PCR and sequence analysis was also used to check for any conservation of gene order. We subsequently devised a molecular typing scheme called

Plasmid Multi-Locus Sequence Typing (PMLST) to monitor the evolution of the conserved plasmid backbones.

4.1.2 Results

4.1.2.1 Plasmid collection

A selected set of 14 previously described IncHI1 plasmids was exploited in these studies (Wain *et al.* 2003) (Table 4-1). The plasmids were chosen in an attempt to capture as much of the potential diversity accumulated to date (e.g. by prior evidence of genetic variation, location and time of isolation). Eight of these plasmids represent the seven previously described RFLP types of IncHI1 plasmids. The five other plasmids were originally isolated from historical *S.* Typhi strains associated with the first antibiotic resistant typhoid outbreaks in the 1970s from India, Mexico, Thailand and Vietnam. Two completely sequenced plasmids, R27 (Sherburne *et al.* 2000) and pHCM1 (Parkhill *et al.* 2001), were also included in the analysis.

A total of 36 IncHI1 plasmids (22 additional plasmids) were also tested for PMLST (Table 4-3 List of plasmids for PMLST and their sequence types).

Table 4-1 Plasmids used in this chapter.

Host	Plasmid RFLP[c]	Plasmid	Resistance	Date of isolation, Location
Control [a]	Not done	R27	T	1961, UK
40R181 [a]	Not done	40R181	CSSuT	1972, Mexico
40R344 [a]	Not done	40R344	CSSuT	1972, India
42R917 [a]	Not done	42R917	CSSuT	1970's, Vietnam
44R311 [a]	Not done	44R311	CSSuT	1970's, Thailand
44R315 [a]	Not done	44R315	ACSSuT	1970's, Thailand

Host	Plasmid RFLP[c]	Plasmid	Resistance	Date of isolation, Location
Ty3	RFLP2	pSTY2	ACSSxtT	1992, Vietnam
CT18	RFLP1	pHCM1	ACSSxtT	1993, Vietnam
Ty10	RFLP1	pSTY1	ACSSxtT	1993, Vietnam
Ty49	RFLP3	pSTY3	ACSSxtT	1993, Vietnam
Ty24	RFLP4	pSTY4	ACSSxtT	1993, Vietnam
Ty39	RFLP5	pSTY5	ACSSxtT	1993, Vietnam
Ty55	RFLP6	pSTY6	ACSSxtT	1993, Vietnam
9541	RFLP7	pSTY7	ACSSxtT	1996, Vietnam

Key:

A: ampicillin, C: chloramphenicol, S: streptomycin,
Sxt: trimethoprim-sulphamethoxazole, Su:sulphathiazole, T: tetracycline.
[a] supplied as an *E. coli* K12 transconjugants by Henry Smith, HPA, Colindale, UK
[b] DNA received only
[c] RFLP patterns from previos study (Wain *et al.* 2003)

4.1.2.2 Syntenic regions of IncHI1 plasmids identified by long-range overlapping PCR

A set of oligonucleotide primers was used to amplify overlapping fragments around the whole pHCM1 plasmid. These studies confirmed the plasmid structure (data not shown). Similar PCR reactions were performed using DNA prepared from the other IncHI1 plasmids. Thus regions similar to pHCM1 should generate similar DNA fragments, whereas divergent regions should produce either no product or a product of different size. In addition, since each PCR product was designed to overlap with each adjacent one, the presence of several adjoining PCR products among plasmids indicates that the genes within share the same genetic order.

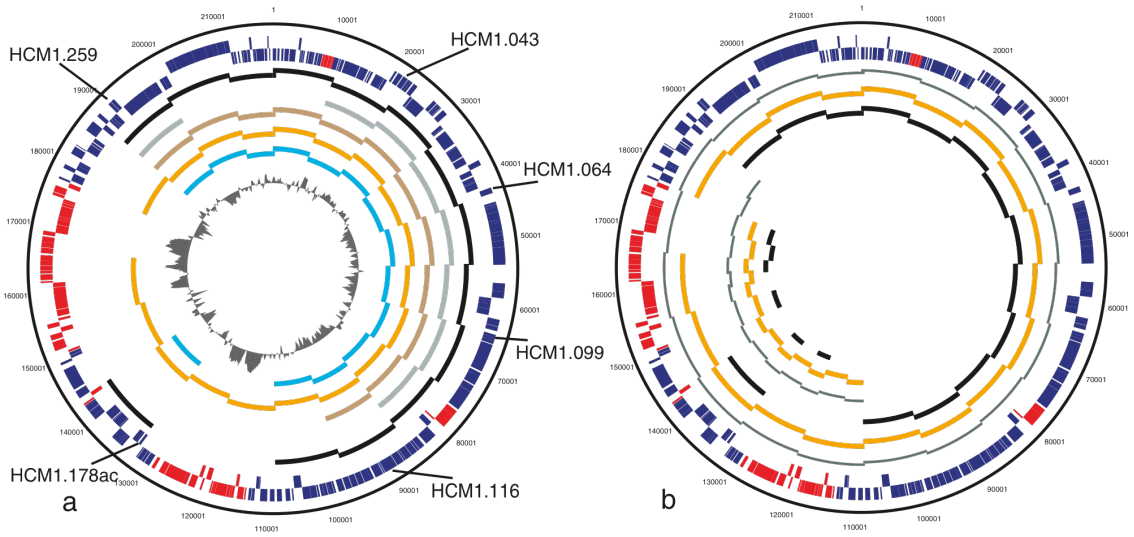


Figure 4-2 Syntenic regions present on eachIncHI1 plasmids.
2a: Long range PCR of ~10Kb amplicons around IncHI1 plasmids of *S.* Typhi. The products generated from R27 (inner grey ring) are predicted from the DNA sequence and the primer binding sites are represented at either end of the grey bars on the inner ring. The order of plasmids from the inside is: R27, pSTY1, pSTY4, pSTY6, pST7 and pHCM1 is the outside ring. The outer most ring shows the positions of the six PMLST loci on pHCM1.
2b: PCR analysis comparing the two 'most successful' IncHI1 plasmids of *S.* Typhi using 5Kb amplicons within regions that do not give products with 10Kb PCR primers. The thin grey lines represent the predicted products from the pHCM1 sequence. The order from the inside is: 5kb fragments of pSTY7, pSTY1 and pHCM1, 10kb fragments of pSTY7, pSTY1 and pHCM1.

In the initial studies, 20 oligonucleotide primer pairs were designed to amplify overlapping 10 kb fragments covering the whole of pHCM1 (Figure 4-2a). Thereafter, for regions in which variability was indicated, oligonucleotides for overlapping 5 kb fragments were designed (Figure 4-2b). PCR reactions for oligonucleotide pairs 2 to 8 and 18 were positive for all plasmids that have been analysed using the 10 kb PCR primers. These PCR-positive regions cover CDS encoding plasmid maintenance and transfer functions in pHCM1. Three adjacent reactions (1, 19 and 20 in Figure 2a) were negative in pSTY6. Further PCR analysis using primers spanning the whole region confirmed a deletion of 10-15kb (data not shown). The regions that did not amplify in the other IncHI1 plasmids are those encoding the antibiotic resistance-determining regions in pHCM1. This suggests that the IncHI1 plasmids possess an almost invariable core of genes that code for plasmid maintenance and transfer (and possibly other

functions), the order of which is generally maintained among the plasmids. The variable regions on these plasmids encode the antibiotic resistance determinants.

4.1.2.3 Plasmid Multi-locus Sequence Typing

The presence of conserved regions among these plasmids provides a potential basis for a molecular typing scheme. Consequently, we next applied the principle of MLST (Maiden *et al.* 1998) to this group of plasmids (PMLST). The criteria for choosing PMLST loci are: i. they are present in all plasmids in the set (based on microarray data), ii. The regions are evenly distributed on the plasmid syntenic regions and iii. they do not belong to any known mobile elements or antibiotic resistance genes.

To initiate this phase of the study, thirteen candidate loci were sequenced and analysed using a small set of 14 plasmids (Table 4-1) to evaluate their usefulness in the PMLST scheme. Four loci (HCM1.054, 094, 107, and 277) were excluded from the PMLST as they showed no variation. Three more loci (HCM1.177, 280c and 286) showed inconsistent PCR amplification among the plasmids analysed and were therefore excluded. The final PMLST set therefore included 6 loci: HCM1.043, 064, 099, 116, 178ac and 259. Following sequencing in a set of 36 plasmids, two to three allelic variants of each locus were detected, the combinations of which define eight unique plasmid sequence types (PSTs) (Table 4-3). It is worth noting here the remarkable level of conservation of this core backbone sequence in these plasmids.

Table 4-2 List of candidate PMLST genes and their annotations

Coding sequences	Size (bp)	PMLST sequence (bp)	Gene function
HCM1.043	777	356	Hypothetical protein
HCM1.054	882		*repA*, RepHI1B replication initiation protein,

Coding sequences	Size (bp)	PMLST sequence (bp)	Gene function
			an IncHI1 specific replication protein.
HCM1.064	876	527	*repA2*, RepHI1A replication initiation protein, an IncHI1 specific replication protein
HCM1.094	453		*htdA*, IncHI1 transfer repressor
HCM1.099	1509	417	*trhW*, one of 9 *trh* genes (*trhALEKBVCPW*) essential for H-pilus production
HCM1.107	1857		*trhI*, *Salmonella* Typhi putative ATP-dependent helicase
HCM1.116	2013	491	Hypothetical protein
HCM1.177	711		Hypothetical protein
HCM1.178ac	405	403	Probable DNA-binding protein, contains Pfam match to H-NS histone family
HCM1.259	516	393	Hypothetical protein, contains Pfam match to Transglycosylase SLT domain
HCM1.277	714		Possible periplasmic protein
HCM1.280c	546		Hypothetical protein
HCM1.286	1206		Possible DNA-binding protein

Figure 4-3 shows the relationship between 36 IncHI1 plasmids based on their sequence type (PST), determined using eBURST (http://eburst.mlst.net/) (Feil *et al.* 2004, Spratt *et al.* 2004). The eBURST program was configured to group PSTs sharing 5 out of 6 identical loci, resulting in two groups and a singleton PST. Group 1 contains PST1, PST2, PST3 and PST4 whilst group 2 includes PST6, PST7 and PST8. Plasmid R27 was assigned PST5 and stood as a singlet, sharing no more than four identical loci with any PST in either group. PST1 includes pHCM1 and four other plasmids, all isolated from sporadic typhoid cases in Vietnam in 1993. Plasmids from PST2 were isolated earlier in the 1970s from Thailand (40R311 and 40R315) and India (40R344) except for pSTY4 which was isolated in Vietnam in 1993. PST3 and PST4 have only one member

each, isolated in the 1970s from Mexico (40R181) and Vietnam (42R917), respectively. In group 2, PST6 contains plasmids from Vietnam and Jordan, all isolated after 1993. PST7 consists of 13 plasmids isolated from *S.* Paratyphi A from Karachi, Pakistan (2002 – 2004). The *S.* Paratyphi A plasmid pAKU1, which has recently been sequenced, also belongs to this PST. PST8 contains four other *S.* Paratyphi A plasmids from Karachi (2003 – 2004) and differs from PST7 by the deletion of locus HCM1.116.

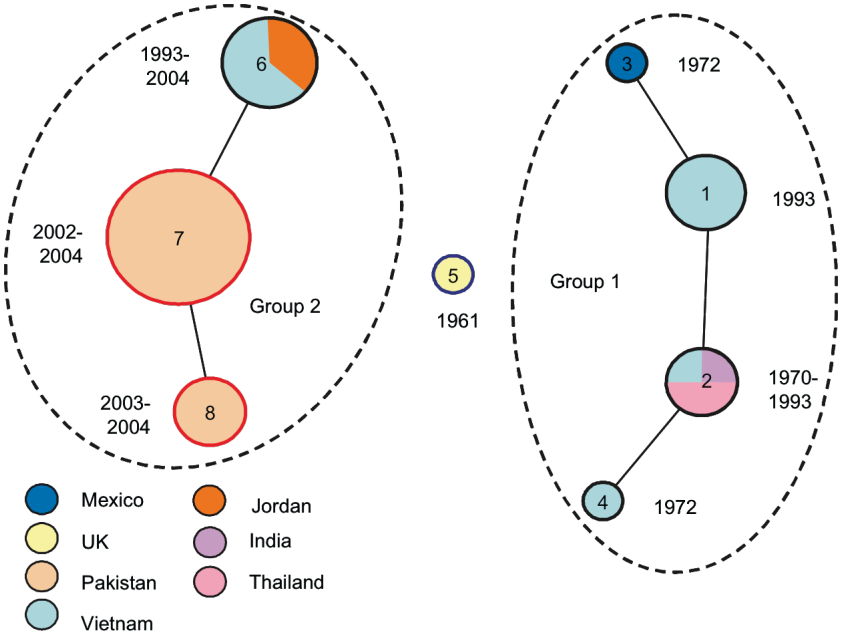


Figure 4-3 Modified eBURST diagram for plasmid multi-locus sequence typing.
The numbers inside the circles represent the plasmid sequence types (PSTs). Lines connecting circles represent single locus variants. The dotted circles represent groups of related plasmids. Plasmids in group 1 are all from before 1993 whereas plasmids from group 2 are from 1993-2004. The singlet PST is the first described IncHI1 plasmid R27.

Although these plasmids were grouped purely by sequence data, it is worth noting that group 1 contains plasmids isolated prior to 1994, whilst group 2 includes only plasmids from 1993 onwards. This chronological division is interesting and could be a consequence of competition between plasmids in the two groups that subsequently lead to a replacement of group 1 by group 2. This phenomenon could be driven by an unkbown selective advantage acquired by the ancestor of the group 2 plasmids.

Table 4-3 List of plasmids for PMLST and their sequence types

Plasmid ID	*Year of isolation*	*Origin*	*Sequence Type*	*PMLST alleles*					
				043	**064**	**099**	**116**	**178ac**	**259**
pHCM1	1993	Vietnam	1	1	1	1	1	1	1
pSTY2	1993	Vietnam	1	1	1	1	1	1	1
pSTY3	1993	Vietnam	1	1	1	1	1	1	1
pSTY5	1993	Vietnam	1	1	1	1	1	1	1
KKG28	1993	Vietnam	1	1	1	1	1	1	1
pSTY4	1993	Vietnam	2	1	2	1	1	1	1
40R344	1972	India	2	1	2	1	1	1	1
44R311	1970s	Thailand	2	1	2	1	1	1	1
44R315	1970s	Thailand	2	1	2	1	1	1	1
40R181	1972	Mexico	3	1	1	1	1	1	3
42R917	1972	Vietnam	4	1	2	3	1	1	1
R27	1961	UK	5	3	2	1	2	1	2
pSTY6	1993	Vietnam	6	3	2	2	2	2	3
pSTY7	1996	Vietnam	6	3	2	2	2	2	3
JCT61	1993	Vietnam	6	3	2	2	2	2	3
K11632	1997	Vietnam	6	3	2	2	2	2	3
KCT51	1994	Vietnam	6	3	2	2	2	2	3
pST661	2004	Jordan	6	3	2	2	2	2	3
pST721	2004	Jordan	6	3	2	2	2	2	3
pAKU1	2002	Pakistan	7	3	2	2	2	1	3
SPA-1308	2004	Pakistan	7	3	2	2	2	1	3
SPA-1464	2004	Pakistan	7	3	2	2	2	1	3
SPA-247	2002	Pakistan	7	3	2	2	2	1	3
SPA-251	2002	Pakistan	7	3	2	2	2	1	3
SPA-275	2002	Pakistan	7	3	2	2	2	1	3
SPA-287	2002	Pakistan	7	3	2	2	2	1	3
SPA-335	2002	Pakistan	7	3	2	2	2	1	3
SPA-416	2003	Pakistan	7	3	2	2	2	1	3
SPA-444	2003	Pakistan	7	3	2	2	2	1	3
SPA-460	2003	Pakistan	7	3	2	2	2	1	3
SPA-510	2003	Pakistan	7	3	2	2	2	1	3
SPA-842	2003	Pakistan	7	3	2	2	2	1	3
SPA-1074	2004	Pakistan	8	3	2	2	0	1	3
SPA-1326	2004	Pakistan	8	3	2	2	0	1	3
SPA-568	2003	Pakistan	8	3	2	2	0	1	3
SPA-688	2003	Pakistan	8	3	2	2	0	1	3

4.1.2.4 PMLST allele accession numbers:

We have deposited the DNA sequences of PMLST alleles in NCBI under these accession numbers: HCM1.043 allele 1 - FJ183728, HCM1.043 allele 3 - FJ183729, HCM1.064 allele 1 - FJ183730, HCM1.064 allele 2 - FJ183731, HCM1.099 allele 1 -

FJ183732, HCM1.099 allele 2 - FJ183733, HCM1.099 allele 3 - FJ183734, HCM1.116 allele 1 - FJ183735, HCM1.116 allele 2 - FJ183736, HCM1.178ac allele 01 - FJ183737, HCM1.178ac allele 02 - FJ183738, HCM1.259 allele 01 - FJ183739, HCM1.259 allele 02 - FJ183740 and HCM1.259 allele 03 - FJ183741.

4.1.2.5 Composite transposon Tn*6062*

DNA genome sequence comparison between a plasmid belonging to group 1 (pHCM1) from *S.* Typhi (CT18) and a group 2 plasmid (pAKU1) from *S.* Paratyphi A (AKU_12601) identified a novel mobile element inserted into the middle of the Tra2 region of pAKU1. This region spanned 4,967 bp from nucleotide position 69376 to 74342 in pAKU1. The annotated structure of this element suggests this is a composite transposon. Two 9-bp direct repeats (CACCTTCAA) flanking the whole element are evidence of duplication via insertion of this element into pAKU1. The two IS*1* at the ends of this element suggest the ability to self-transpose using the transposase from one of the IS*1* and their inverted repeats. This composite transposon was registered at Tn Designation and Database (http://www.ucl.ac.uk/eastman/tn/) (Roberts *et al.* 2008) under the name Tn*6062*.

Tn*6062* carries two CDSs, SPAP0105 and SPAP0106. The first CDS is a hypothetical protein containing 4 putative transmembrane domains and a signal peptide sequence. The second CDS encodes a putative membrane transporter with high similarity (99.85% similarity over 675 amino acids) to a protein in found in *E. coli* HU734 called BetU (Ly *et al.* 2004), which is a member of the betaine-carnitine-choline transporter family, predicted to possess 12 transmembrane helices. Similar transporters have been proposed to contribute to osmoregulation in *E. coli*. The *betU* gene in HU734 is also flanked by two IS elements but these differ from those in pAKU1 (IS*911* and IS*600* in HU734).

The fact that a *betU* gene is presents in one third of *E. coli* (Ly *et al.* 2004) and has been acquired by different IS elements suggest it may confer a selective advantage to the host.

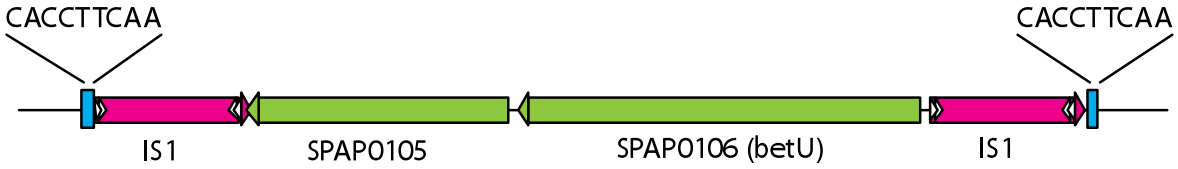


Figure 4-4 Schematic structure of Tn*6062*
Blue: 9-bp direct repeat; Pink: IS*1*; Green: coding sequence.

PCR primers were designed to detect this composite transposon in the plasmid collection. Out of 36 plasmids tested, 24 were positive for this element. Strikingly, there is an absolute association of Tn*6062* with group 2 plasmids (Figure 4-5), suggesting the acquisition of this element by a common ancestor of plasmids from group 2.

4.1.2.6 Differences within the conserved regions

At the time of these studies, there were three fully sequenced plasmids in the collection under study, each of which represents one group on the PMLST map. Sequence comparison between pHCM1 and R27 defined five variable regions, two of which (region A and B) encode for antibiotic resistance determinants and mobile elements whilst the other three (region C, D and E) were actually situated within the conserved region defined by PCR and do not show any obvious evidence of being horizontally acquired. The genes in region C, D and E were present in pHCM1 and absent in both R27 and pAKU1. The sequence contexts surrounding these three regions are similar in R27 and pAKU1. Another feature that lies within the conserved region is the composite transposon Tn*6062*. This transposon was inserted in the middle of the transfer region of

pAKU1 (right after *trhC* in pHCM1). PCR assays were designed to screen for the presence and absence of these regions in the whole set of plasmids. To avoid false negatives, these assays were designed such that a large PCR product was predicted when the region is present, and a smaller PCR product when the region is absent.

Figure 4-5 shows the presence and absence of regions C, D, E and Tn*6062* in each plasmid along with the phylogenetic tree of plasmids of all PST. Tn*6062* was specific to plasmids of group 2, which suggests the singleton plasmid R27 may be more closely related to group 1 than group 2. Regions C, D and E were present in all five plasmids of PST1 (group 1). The remaining plasmids of group 1 were negative for regions C and E (except for 40R181). In group 2, regions C, D and E were generally absent. Region D failed to amplify in the four plasmids of PST8 and two from PST7, suggesting further sequence changes may have occurred which affected the binding sites of PCR primers.

The acquisition or loss of these regions can therefore be inferred within the plasmid lineages represented by the phylogenetic tree. Tn*6062* was most likely acquired by a common ancestor of group 2 plasmids, whilst region C and E were more likely lost within group 1, after the divergence of PST1. Region D came into a plasmid from group 1 which later on became the predecessor of PST1 plasmids.

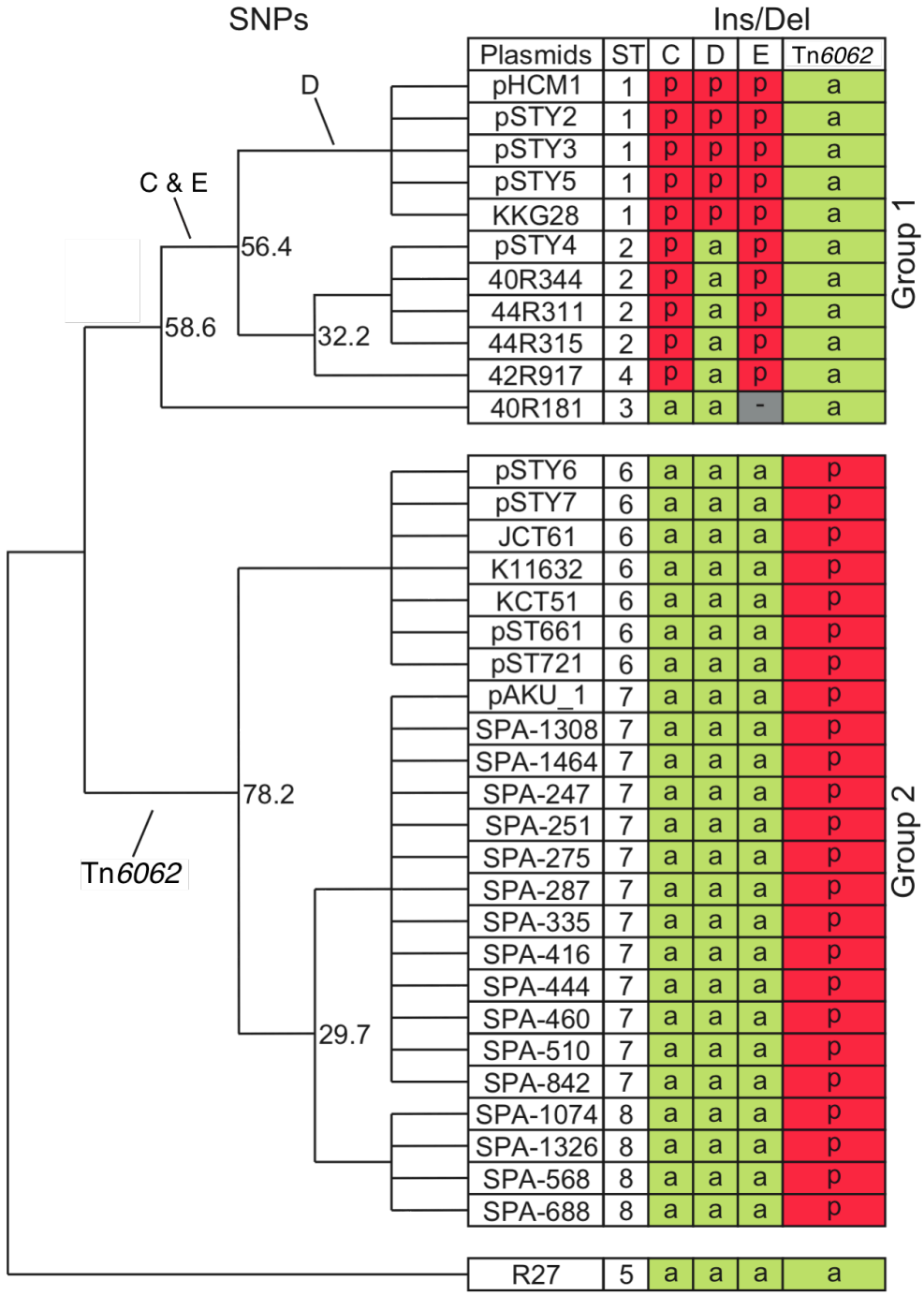


Figure 4-5 IncHI1 plasmids phylogenetic tree and observed insertion/deletion events
The phylogenetic tree was built based on the SNPs identified within PMLST sequences using maximum parsimony for discrete character data (Phylip software package), bootstrap values shown from 1000 re-sampling. The insertion/deletion events were determined by PCR, "a" for absent, "p" present, and "–" negative PCR.

The data showing the absence of the regions C, D and E in 24 plasmids from group 2 including pAKU1) as well as the absence in R27 suggest that R27 and pAKU1 are more closely related to each other than to pHCM1. Combining the eBURST grouping with

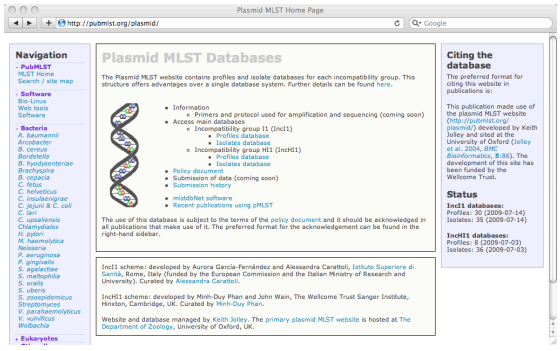
the phylogenetic tree and PCR results, we conclude that plasmids of group 2 (post-1993) did not evolve directly from group 1 (pre-1993) but belong to two different lineages.

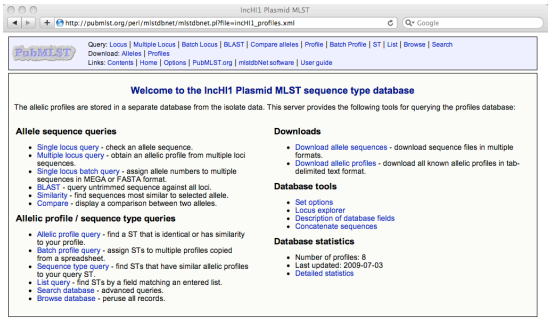
4.1.2.7 Online database for PMLST

The PMLST scheme proved to be a useful tool for the typing of IncHI1 plasmids. One advantage of this method is the ease by which data can be shared among researchers, enabling the meta-analysis of multi-laboratory projects. To provide a central platform for publishing and sharing data, we have collaborated with Keith Jolley (Department of Zoology, University of Oxford, UK) to develop a database and website for handling IncHI1 PMLST data. This website was built on the database program developed specifically for MLST (Jolley, Chan & Maiden 2004).

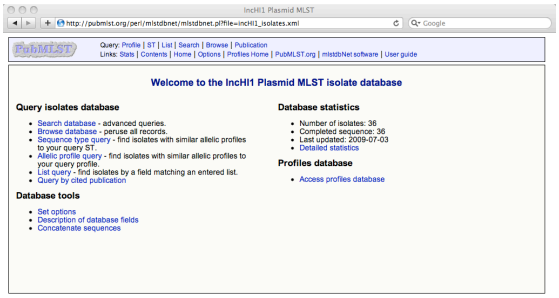
The PMLST information is organised into two interlinked databases: a profile database and an isolate database. The Profile database contains allelic sequences of each locus and plasmid sequence types. The Isolate database provides information on each isolate/plasmid including year of isolation, geographical origin and the corresponding plasmid sequence type. Each database can be found on a separate webpage providing various ways to retrieve information as well as tools to analyse the data.

a.





b.



c.

Figure 4-6 Screenshots of PMLST website.
a. PMLST homepage, b. Profile database page, c. Isolate database page

4.1.3 Discussion

Plasmids of incompatibility group HI1 have been shown to be strongly associated with drug resistance in *S.* Typhi (Wain, Kidgell 2004). This association has existed since the first report of a chloramphenicol resistant typhoid outbreak in 1972 (Paniker, Vimala 1972), and *S.* Typhi harbouring these plasmids have now spread globally (Hampton *et al.* 1998). The RFLP typing of IncHI1 plasmids from Vietnam previously revealed seven RFLP types and a change from RFLP pattern 1 to pattern 7 after 1993 (Wain *et al.* 2003). This data raised questions about the evolutionary processes influencing this particular group of plasmids. It was initially hypothesised that R27 was a direct ancestor of pHCM1 (RFLP pattern 1), which in turn gave rise to plasmids of RFLP pattern 7 (including pSTY7) which predominate after 1993. In order to test this hypothesis of plasmid evolution, we collected IncHI1 plasmids isolated between 1970 and 2004 and analysed their associated variation in much greater detail. Through investigation by

overlapping PCR and sequence comparison combined with DNA microarray data available, we identified both core conserved genes and highly variable resistance-associated regions.

In an attempt to elucidate phylogenetic relationships between the IncHI1 plasmids of *S.* Typhi, we sequenced six core genes from each of the IncHI1 plasmids, an approach based on multilocus sequence typing (MLST) (Maiden *et al.* 1998). All plasmids from the 1970s, and those circulating in *S.* Typhi in Vietnam in the early 1990's, were grouped together (PMLST group 1). We therefore infer they are directly related and represent the expansion of a successful plasmid backbone in the *S.* Typhi population over a period of 20 years. A recognisably distinct plasmid core, PMLST group 2, was consistently present after 1993 suggesting successful competition with those already present in the *S.* Typhi population.

While the earlier RFLP study suggested a new plasmid type arose and spread in Vietnam in the mid 1990's, the origin of the plasmid could not be determined. Contrary to our initial hypothesis of step-wise evolution from R27, to pHCM1, to pSTY7, PMLST analysis suggests that pSTY7 (PMLST group 2) and pHCM1 (PMLST group 1) belong to distinct plasmid lineages. This is also supported by whole-plasmid sequence comparisons of R27, pHCM1 and pAKU1. Thus, the change in the predominant plasmid type in Vietnam is best explained by the acquisition and spread of a distinct plasmid type (group 2), rather than the clonal expansion of a particular variant already present in the *S.* Typhi population. Comparison to the broader collection of plasmids suggests that this replacement of plasmid types in 1993 was not unique in Vietnam, but a global phenomenon.

The typing data of both the *S.* Typhi host and the associated plasmid is needed to investigate whether the predomination of group 2 plasmids is due to the spread of a

particular host strain carrying the plasmid or the spread of the plasmid itself into different hosts. There is limited data on the background strains of *S.* Typhi circulating in Vietnam during these periods. Some authors have shown that multiple RFLP types of *S.* Typhi had become MDR between 1993 – 1997 (Connerton *et al.* 2000), whilst other suggest clonal expansion of single *S.* Typhi strain between 1995 – 2002 (Le *et al.* 2004). Unfortunately several of the plasmids analysed in this study were analysed as transconjugants in *E. coli* and so it was not possible to type the background *S.* Typhi hosts.

A novel composite transposon Tn*6062* is present exclusively in plasmids of PMLST group 2. This transposon carries a transporter protein, BetU, belonging to the betaine-carnitine-choline transporter family. BetU has osmoregulatory activity in *E. coli* (Ly *et al.* 2004). The distinct surrounding genomic location between the *betU* in *E. coli* and that in IncHI1 plasmids suggest independent acquisition and a possible common selective advantage. The exact function of *betU*, or Tn*6062* as a whole, in *S.* Typhi remains to be investigated. It is possible that this element might contribute to the survival of *S.* Typhi in high osmolarity conditions, hence enhancing their ability to spread leading to the predominant of Tn*6062*-carrying IncHI1 plasmids.

IncHI1 plasmids are conjugated with a higher frequency (3-4 logs) at ambient temperatures (27ºC and below) than at *in vivo* temperature (37ºC) (R. Curtiss III, personal communication) (Maher, Taylor 1993, Smith 1974). This property, combined with the fact that contaminated water is the main route for *S.* Typhi transmission (Bhan, Bahl & Bhatnagar 2005), therefore suggest that the plasmids may spread more efficiently in contaminated water at ambient temperature.

RFLP is a potentially generally useful method for typing that could be applied to other plasmid groups. Several plasmids in this study that had the same PMLST, had very

similar accessory gene content, but gave different RFLP patterns (Wain *et al.* 2003). The variation detected by RFLP indicates that the accumulating diversity of IncHI1 plasmids in *S.* Typhi decreased after 1994 (Wain *et al.* 2003). Attempts to draw trees to describe phylogenetic relationships between the different RFLP types however failed. This is because the variation detected by RFLP is generated by several genetic mechanisms with different effects: point mutation, rearrangement, restriction protection mechanisms, and insertion/deletion. For this study PMLST, whilst only applicable to IncHI1 plasmids, was used to define plasmid groups based on the accumulation of mutations in 'non-selected' regions of the plasmid core. The accumulation of these mutations is likely a function of time and this typing scheme therefore represents a quantitative estimate of variation which can be related to the ancestry of the plasmids.

It is intriguing that IncHI1 plasmids, which can be easily transferred to *E. coli* at high frequency *in vitro*, are found almost exclusively in two *S. enterica* serovars, Typhi and Paratyphi A. Plasmid fitness cost could play a role in maintaining IncHI1 plasmids in *Salmonella*. This may involve the integration/interaction of plasmid and chromosomal regulatory networks, such as those mediated by H-NS (Doyle *et al.* 2007). In *S.* Typhimurium there is a detectable fitness cost of harbouring IncHI1 plasmids (Doyle *et al.* 2007) and so it is possible that in *E. coli* there is also a cost. Potential selective advantages to the bacterial host harbouring IncHI1 plasmids have also been demonstrated: such as an increased level of survival inside monocytic cell lines in *S.* Typhimurium (Doyle *et al.* 2007) or higher levels of *S.* Typhi bacteraemia during typhoid fever (Wain *et al.* 1998) or possibly enhanced tolerance to the conditions faced during the infection cycle, as with *betU*.

Taken together this analysis shows that the IncHI1 plasmids of *S.* Typhi contain highly conserved as well as truly dynamic regions. This is reminiscent of the situation found in

E. coli K1 F-like plasmids (Mercer *et al.* 1984). All the IncHI1 plasmids studied shared a conserved backbone into which several acquisitions of different antibiotic resistance and other accessory genes have occurred.

4.1.4 Conclusions

The adaptation of the *S.* Typhi bacterial host to acquire and retain resistance plasmids, and the subsequent evolution of this bacterium-plasmid combination, has occurred over a 20-30 year time frame. It is likely that pressure from chloramphenicol treatment in the 1970's drove the selection of a linked group of genes capable of survival in the *S.* Typhi host. The presence of such linked groups is supported by our results which identified the conserved backbone of IncHI1 plasmids. This backbone has evolved little but point mutations accumulated in these conserved genes can be used to discern subtly different plasmid lineages as demonstrated by our newly developed PMLST typing scheme. By PMLST we showed for the first time the population structure and the dynamic evolutionary process of IncHI1 plasmids. The two lineages defined by PMLST corresponded perfectly to time of plasmid isolation, suggesting that competition between lineages resulted in a global replacement of group 1 by group 2 plasmids. Antibiotic pressure concurrently drove the evolution of IncHI1 plasmids by acquisition of accessory genes into this backbone at certain regions where disruption of DNA can occur without significantly affecting backbone plasmid functions.

In summary, the selective pressure from antimicrobial chemotherapy has forced *S.* Typhi to gain resistance plasmids by facilitating the access of this major pathogen to a wide variety of accessory genes.

Further investigation is required in order to identify the molecular mechanisms responsible for the success of PMLST group 2 plasmids. The mechanism of spread of these plasmids is also of interest as it may shed light upon the transfer route of antibiotic

resistance, whether it is the synergistic spread of a bacterium-plasmid clone or the emergence of a more competitive plasmid. Interrogation of the association between *S*. Typhi strains and IncHI1 plasmid types is vital to answer the above question. The findings suggest a possible approach to destabilise IncHI1 plasmids in *S*. Typhi, perhaps by a specific small molecule, rendering them susceptible once again to first line antibiotics.

5 The spread of IncHI1 plasmids in *S.* Typhi

5.1 Introduction

Monitoring of epidemiology has always been of vital importance to the management of infectious diseases. Consequently, significant effort has been spent on developing methods for the typing of *S.* Typhi isolates. For example, a Vi phage typing method was available as early as 1938 (Craigie, Yen 1938) and although this method is still in use today, there are more than 100 serovar Typhi phage types, it also has limitations. There is an increasing number of Vi-negative or non-typable *S.* Typhi strains (Le *et al.* 2007) and the approach is generally restricted to reference laboratories. More robust methods such as ribotyping (Altwegg, Hickman-Brenner & Farmer 1989) and pulse-field gel electrophoresis (PFGE) (Thong *et al.* 1994) (both using restriction enzyme digestion and electrophoresis) were introduced in the late 1980s and have been used intensively to study outbreak and sporadic strains of *S.* Typhi. Whilst these two methods are discriminatory and are very useful in short-term or local epidemiology (Navarro *et al.* 1996), they can sometimes be misleading for global epidemiology. One inherent disadvantage of electrophoresis-based methods is the difficulty in comparing results between laboratories. Attempts have been made to standardise the protocols and to create a central database for PFGE (www.pulsenetinternational.org).

Different typing methods are needed for long-term epidemiology and population structure studies. Multilocus enzyme electrophoresis (MLEE) (Selander *et al.* 1990) and multilocus sequence typing (MLST) (Maiden *et al.* 1998) are more appropriate for this purpose as they both use slowly accumulating genetic variation (likely to be neutral variation) for type discrimination. MLST was developed to make use of the MLEE concept whilst providing the advantages of DNA sequencing: more variations per locus

can be detected, high throughput platforms can be used and the results can be compared readily between laboratories.

The concept of using sequence data for typing has great advantages in standardisation and sharing data. With the rapid development of sequencing and related technologies, it is now possible to develop typing methods with increased discriminatory power that can be used in both local outbreaks and global epidemiology across international collaborations. This is of vital importance for *S.* Typhi, which is a monophyletic pathogen with very little genome variation. Roumagnac et al (Roumagnac *et al.* 2006) studied a global collection of *S.* Typhi and discovered 88 SNPs in 199 loci of 88.7 kb of DNA which defined 59 haplotypes. Resquencing of 19 Typhi strains from different nodes of the *S.* Typhi haplotree revealed more SNPs, providing the base information for the developing of a new SNP typing scheme (Holt *et al.* 2008).

In this chapter we describe the use of SNP typing on both plasmids and the *S.* Typhi host to investigate whether there is any association of plasmids with background strains. To do this we used 200 SNPs on the plasmid and 1508 SNPs on the chromosome probed using a high throughput SNP detection platform - The Illumina Golden Gate (see methods section 2.3.9).

5.2 Results

A total of 473 isolates were genotyped using the Illumina GoldenGate arrays, including 446 isolates sourced from around the world and 27 control isolates (19 isolates sequenced in Holt *et al.* (2008**)** and 8 *E. coli* transconjugants harbouring IncHI1 plasmids of known plasmid sequence types (PSTs) in chapter 4). A complete list of all isolates is included in Appendix 8.5. A summary of genotyping results is shown in Table 5-1 in which isolates were grouped by their relationship on the phylogenetic tree. (see 5.2.3 and Figure 5-3).

The GoldenGate arrays were designed using 1929 chromosomal SNPs identified from sequence data on 19 *S.* Typhi genomes (Holt *et al.* 2008) and 72 SNPs identified from analysis of the *S.* Typhi chromosome among 180 isolates (Roumagnac *et al.* 2006). We included 8 SNPs within six genes of the conserved backbone of IncHI1 plasmids (Chapter 4) and an additional 294 IncHI1 SNPs identified by comparing the conserved backbones of eight IncHI1 plasmid sequences, originally isolated from *S.* Typhi, *S.* Paratyphi A, *S.* Choleraesuis, *S.* Typhimurium and *E. coli.* A total of 218 SNPs designed to assess the presence or absence of resistance genes and specific IncHI1 sequences were also included on the GoldenGate arrays.

SNPs calling from array signals and quality control using the 19 genomes have identified high quality call SNPs among those included on the arrays. Thus the analysis of experimental *S.* Typhi isolates presented in this study is based on 1508 SNP loci distributed randomly in the *S.* Typhi chromosome, 200 SNP loci on the IncHI1 plasmid and 119 SNP loci for the presence or absence of resistance genes and specific IncHI1 sequences (Appendix 8.8).

Table 5-1 Summary of strain collection categorised by haplogroups*

Haplogroup	n	Country	Year (n)
H1	**11**		
	1	India	2007 (1)
	1	Indonesia	1976 (1)
	2	Laos	2000 (2)
	7	Vietnam	1967 (2), 1993 (1), 2000 (2), 2001 (1), 2004 (1)
H29	**12**		
	1	Algeria	1999 (1)
	1	Benin	2004 (1)
	1	China	2002 (1)
	1	Ethiopia	2006 (1)
	1	Guinea	1999 (1)
	3	Indonesia	1983 (1), 1985 (1), 2003 (1)
	2	Morocco	1999 (1), 2000 (1)
	1	Togo	2003 (1)
	1	Tunisia	1961 (1)

Haplogroup	n	Country	Year (n)
H42	**22**		
	1	Bangladesh	1998 (1)
	1	Congo	1967 (1)
	14	India	2003 (1). 2004 (3), 2005 (8), 2007 (2)
	3	Kuwait	2006 (3)
	1	Peru	2000 (1)
	2	US	2004 (2)
H50a	**53**		
	1	Algeria	1966 (1)
	1	Bangladesh	2000 (1)
	2	Cameroon	1958 (1), 1966 (1)
	1	Cape Verde	2000 (1)
	4	Chad	1962 (1), 1963 (3)
	1	Congo (Zaire)	1976 (1)
	1	Ethiopia	2006 (1)
	1	French Guiana	2002 (1)
	5	India	1975 (1), 1977 (2), 2003 (1), 2005 (1)
	5	Ivory Coast	1967 (4), 2001 (1)
	1	Kenya	1998 (1)
	3	Madagascar	1967 (1), 1980 (1), 1997 (1)
	3	Mali	1999 (1), 2000 (1), 2001 (1)
	2	Mexico	1972 (1), 1998 (1)
	2	Morocco	1999 (2)
	1	Nigeria	2002 (1)
	2	Peru	1981 (1), 1983 (1)
	6	Senegal	1962 (2), 1966 (1), 1967 (1), 1998 (1), 2001 (1)
	2	Thailand	2002 (2)
	1	Togo	2001 (1)
	1	Turkey	1997 (1)
	5	Vietnam	1959 (1), 1967 (1), 1973 (2), 1995 (1)
	2	Unknown	Unknown (2)
H50b	**15**		
	3	Algeria	1966 (2), 2000 (1)
	1	Cameroon	1966 (1)
	1	Chile	1976 (1)
	1	China	2000 (1)
	1	India	2000 (1)
	1	Indonesia	2001 (1)
	1	Kuwait	2006 (1)
	1	Mexico	1998 (1)
	2	Peru	1981 (2)
	2	Tunisia	1961 (1), 1999 (1)
	1	Unknown	1939 (1)

Haplogroup	n	Country	Year (n)
H52	**40**		
	2	Algeria	1966 (1), 1999 (1)
	6	Chad	1963 (2), 1964 (4)
	1	Comoros	1999 (1)
	3	Congo	1966 (1), 1967 (2)
	1	France	1973 (1)
	1	Indonesia	2000 (1)
	1	Kuwait	2006 (1)
	1	Lebanon	2001 (1)
	3	Madagascar	1965 (2), 1966 (1)
	1	Mauritania	1998 (1)
	3	Morocco	1958 (1), 1966 (1), 1967 (1)
	1	Russia	1916 (1)
	4	Senegal	1962 (2), 1967 (1), 2001 (1)
	1	Sri Lanka	2001 (1)
	3	Tunisia	1961 (1), 1966 (1), 1978 (1)
	5	Vietnam	1967 (1), 1972 (2), 1973 (1), 2002 (1)
	3	Unknown	Unknown (3)
H58	**293**		
	3	Bangladesh	2003 (1), unknown (2)
	1	Cambodia	2003 (1)
	1	Central Africa	2004 (1)
	1	Chad	1964 (1)
	43	India	2000 (2), 2002 (1), 2003 (9), 2004 (10), 2005 (17), 2006 (2), 2007 (2)
	4	Jordan	2004 (4)
	22	Kenya	2004 (2), 2005 (12), 2006 (7), 2007 (1)
	21	Kuwait	2006 (21)
	2	Laos	2001 (1), 2002 (1)
	1	Morocco	2003 (1)
	2	Nepal	1999 (1), 2003 (1)
	1	Sri Lanka	2002 (1)
	189	Vietnam	1995 (1), 1996 (1), 1997 (4), 1998 (10), 1999 (3), 2000 (2), 2001 (7), 2002 (94), 2003 (12), 2004 (31), 2005 (22), unknown (2)
	2	UK	1995 (1), 2005 (1)
H81	**19**		
	1	Algeria	1999 (1)
	1	Angola	2001 (1)
	3	Cameroon	1958 (1), 1998 (1), 2001 (1)
	4	Congo (Zaire)	1966 (1), 1967 (1), 1976 (1), 2000 (1)
	1	India	2002 (1)
	3	Ivory Coast	1965 (1), 1967 (1), 2002 (1)
	3	Morocco	1999 (1), 2000 (1), 2005 (1)

Haplogroup	n	Country	Year (n)
	1	Peru	1998 (1)
	1	Philippines	2003 (1)
	1	Senegal	1999 (1)
N/A **	**8**		
	2	India	1970s (2)
	1	Pakistan	2003 (1)
	1	Mexio	1972 (1)
	1	Thailand	1970s (1)
	3	Vietnam	1972 (1), 1993 (1), 1996 (1)
Total	**473**		

* Haplogroups are defined as a group of closely related isolates based on phylogenetic analysis (see section 5.2.3 for more details)

** These 8 strains are transconjugants (*E. coli*) habouring IncHI1 plasmids of known PSTs

5.2.1 IncHI1 plasmid SNP typing

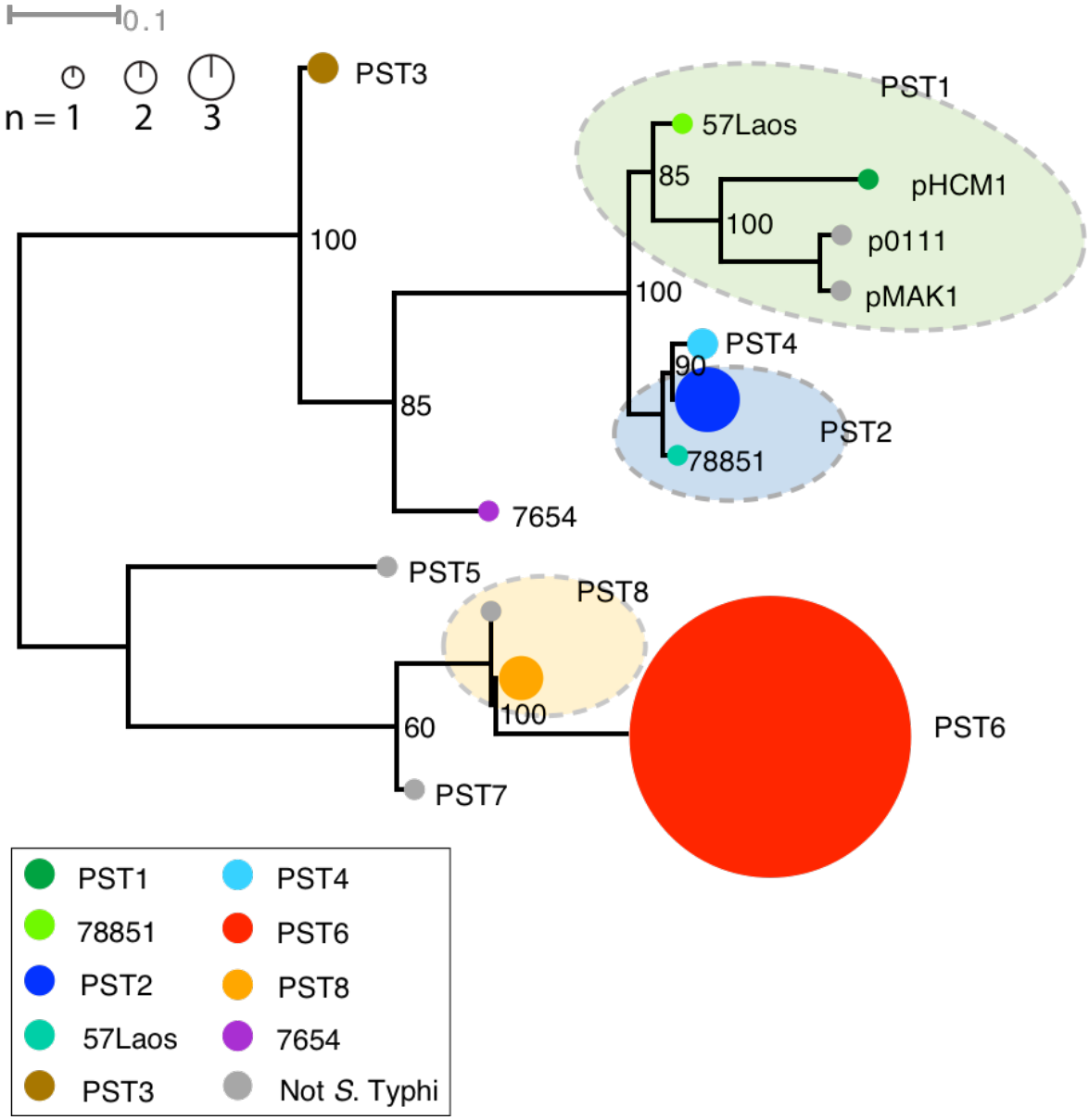


Figure 5-1 Relationship of plasmid sequence types
Demonstrated by a maximum likelihood phylogenetic tree drawn using SNP data from the Golden Gate platform. The PST numbers denotes the sequence types define by PMLST. Bootstrap values after 1000 iterations is shown.

SNP alleles assigned for 200 IncHI1 SNP loci were concatenated to give a single plasmid haplotype string for each plasmid. The alignment of these plasmid haplotype strings was used as input for phylogenetic analysis. We used RAxML (Stamatakis

2006) to fit maximum likelihood phylogenetic trees to the plasmid allele data using the GTR model.

The SNP data supported PMLST in that the same branching was seen by both methods and the plasmid STs clustered on the tree. We found three new SNP types compared to the PSTs but they were rare in our collection: 57Laos, 7654, 78851. This showed that PMLST gives good discrimination of plasmid types and detected most of the variation present. The previously defined group 1 including PST1, 2, 3 and 4 (by PMLST in chapter 4) was now split into a cluster of PST1, 2 and 4 whilst PST3 was placed elsewhere. Plasmids of group 2 (PST6, 7 and 8) remained clustered together. PST5 and the isolate 7654 were distant from others.

The plasmid ST1 included four different SNP types, two of which are from *E. coli* (pO111) and *S.* Choleraesuis (pMAK1) respectively. This by far was the most diverse group of IncHI1 plasmids in terms of both sequence differences and host range. Plasmids related to pHCM1, from PST1 were common in Vietnam before 1993 (Wain *et al.* 2003). PST2 was divided by GoldenGate data into two subtypes, one was the previously known as PST2 and the other a rare type represented by only one isolate from Tunisia in 1978.

There were three Peruvian *S.* Typhi plasmids that were placed in PST8, a group that previously consisted of only *S.* Paratyphi A plasmids. Although these 3 plasmids belonged to a subgroup of their own, this was the best example of *S.* Typhi and *S.* Paratyphi A sharing a closely related backbone.

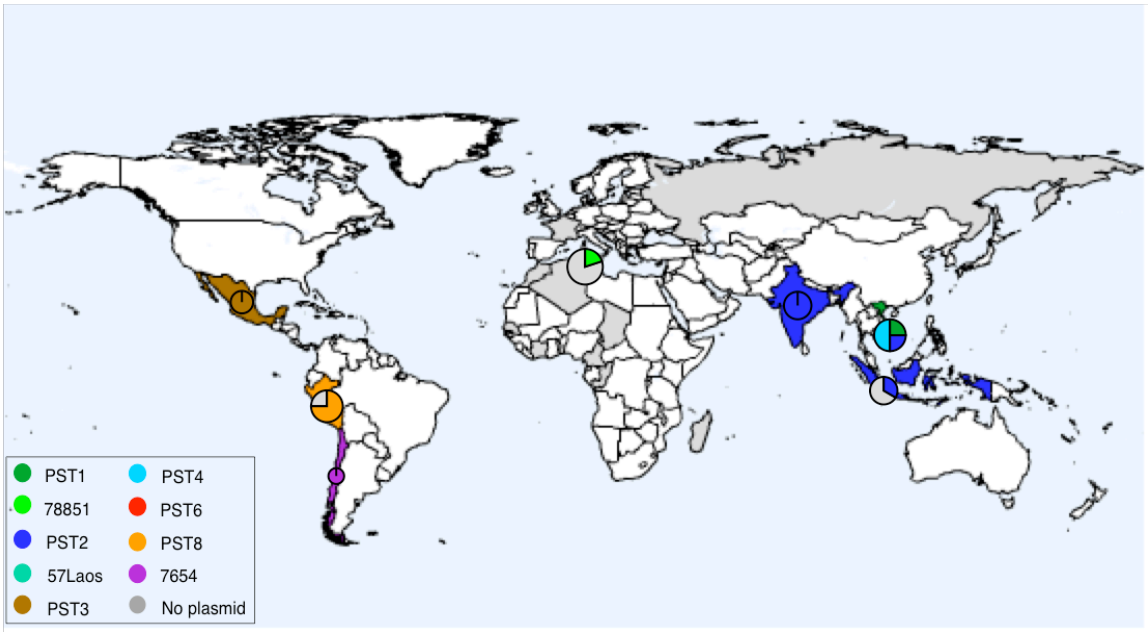
PST6 had the largest number of plasmids in our study. Although there were 180 isolates in this group, they are identical in all SNP loci investigated. All PST6 plasmids in this strain collection were isolated after 1995. This suggested a recent clonal expansion of PST6 plasmids, which supported our conclusion in chapter 4.

Table 5-2 Number of plasmids isolated before and after 1993 *

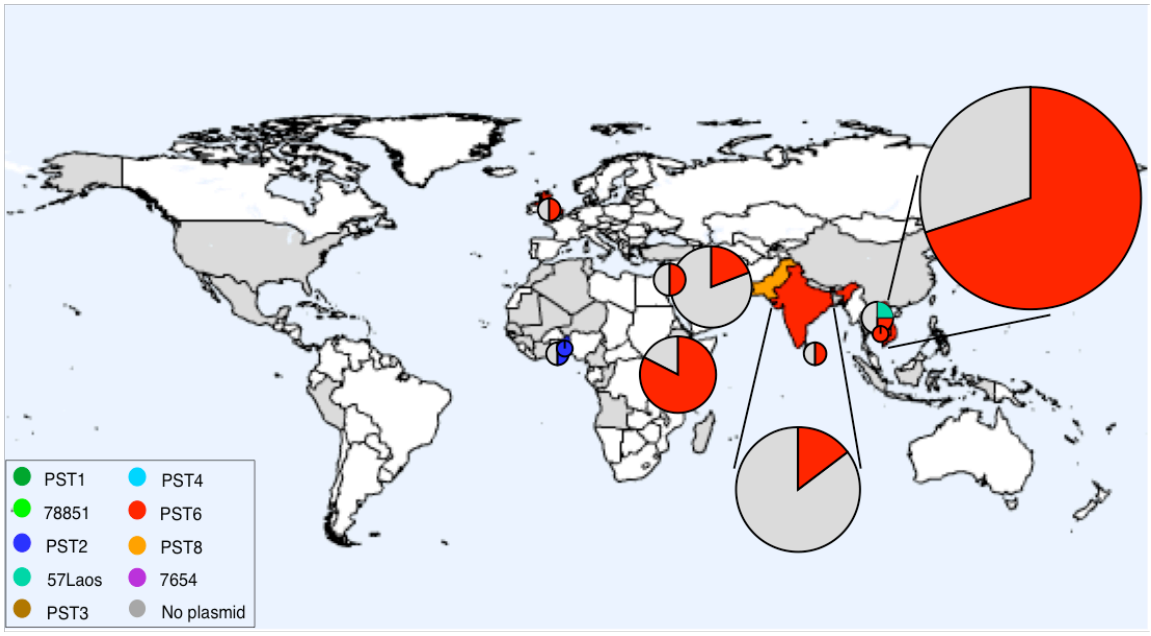
	Before 1993	After 1993	Total
7654	1	0	1
PST1	2	0	2
57Laos	0	1	1
PST2	5	3	8
78851	1	0	1
PST3	2	0	2
PST4	2	0	2
PST6	0	180	180
PST8	3	0	3
No plasmid	66	197	263
Total	82	381	463

- Not include non-Typhi isolates and isolates without year of isolation.

The global distribution of various plasmid types in the time periods before 1993 and after 1993 is shown in Figure 5-2. Although the number of plasmids in each country before 1993 was small, there were already 7 out of 9 *S.* Typhi plasmid SNP types represent globally. This trend dramatically changed after 1993: only 4 plasmid SNP types were found, two of which were new types (the rare type 57Laos and PST6). Plasmid of PST6 predominated in recent years. Nearly half of strains isolated after 1993 harboured PST6 plasmids, whereas only 4 isolates contained plasmids of other types.



(a)



(b)

Figure 5-2 Spatiotemporal distribution of plasmid STs.
Pie charts show the number (size of circle) and proportion of plasmid STs. (a) is before 1993 and (b) is after 1993.

5.2.2 Mobile elements on IncHI1 plasmids

Whole plasmid sequencing has shown that there are similar transposons and composite transposon structure shared between many of the IncHI1 plasmids: a Tn*21*-like structure, Tn*9* carrying *cat* gene and Tn*10* carrying *tet* operon (Holt *et al.* 2007). Various PCRs were designed to examine the insertion sites of these mobile elements in different plasmids. The primers were designed based on mobile element structure and previously characterised arrangements on a PST1 plasmid (pHCM1) and a PST7 plasmid (pAKU1).

The large collection of plasmids arranged into their evolutionary history on a phylogenetic tree provided background to investigate the acquisition of these mobile elements by IncHI1 plasmids over time. The PCR data correlating with the acquisition of mobile elements on several IncHI1 plasmids are shown in Table 5-3.

The PCR results varied between plasmids from different PST and even between plasmids from the same PST. This demonstrated the dynamic of mobile element acquisition events and their structures and/or arrangements on IncHI1 plasmids.

The PCRs for insertion of *bla/sul/str* gene cassette into Tn*21* were positive in only 3 plasmids isolated after 1996 suggesting that this insertion was a recent event. H and I in combination showed the structure of a composite transposon consisting of Tn*21*-like element inserted within Tn*9* (see Figure 1-11 for the mosaic structure of the composite transposons in pHCM1). The insertion sites of the Tn*21*-like element into Tn*9* are exactly the same in pHCM1 and pAKU1 (Holt *et al.* 2007). However, the subsequent insertion of IS*4321* into the two inverted repeats of Tn*21* resulted in longer PCR products for pAKU1 (product 2). The results of H and I suggested that the insertion of Tn*21*-like element into Tn*9* happened quite early (since 1972). J, K, L and M demonstrated two independent acquisition of Tn*9* into IncHI1 plasmids: the insertion of

Tn*9* into pHCM1 (L, M) and pAKU1 (J, K). Altogether, the Tn*9*/Tn*21*-like structure was present in a majority of plasmids investigated except for 7654, 752507 and 81918. The resistance *bla/sul/str* genes however might be integrated into this Tn*9*/Tn*21*-like structure at later time to form the full composite transposon structure as seen in pHCM1 and pAKU1. There was evidence for at least three independent acquisitions of the Tn*9*/Tn*21*-like structure into three different locations on the IncHI1 backbone: the pHCM1 position, the pAKU1 position and the Mexican plasmid position (721258 and 40R181).

N, O and P in combination assess the insertion site of Tn*10* into IncHI1 plasmids. Tn*10* was truncated in pHCM1 and this event was detected by reaction N, which showed that this event is probably specific for pHCM1. We also observed at least two independent acquisitions of Tn*10* in IncHI1 plasmid: the pMAK1 position and the pAKU1 position. A second acquisition of *strAB* seemed to be specific for PST8 plasmids (reaction Q).

A broader view of all mobile element insertion investigated defined three general groups of plasmids. A pHCM1-like group with similar Tn*9* and Tn*10* insertion site included plasmids of PST1, 2 and 4. A pAKU1-like group includes PST6 and 8. Isolate 7654 stood on it own and the two PST3 plasmid clustered together.

Table 5-3 The presence and structure of mobile elements on various plasmid types

Isolates	Year	Country	Hap-group	PST	G	H	I	J	K	L	M	N	O	P	Q	U
7654	1976	Chile	H50b							2						
pHCM1	1993	Vietnam	H1	1	1	1	1			1	1	1				
57Laos	2000	Laos	H1	1		2	2			1	1			1		
761406	1976	Indonesia	H1	2		2				1	1			1		
78851	1978	Tunisia	H52	2		2				1	1			1		
77303	1977	India	H50a	2		2				1	1			1		
721907	1972	Vietnam	H52	2		2				1	1			1		
77302	1977	India	H50a	2		2				1	1			1		
46845	2004	Benin	H29	2		2				2	1					

Isolates	Year	Country	Hap-group	PST	G	H	I	J	K	L	M	N	O	P	Q	U
752507	1975	India	H50a	2										1		
34747	2003	Togo	H29	2		2				1	1					
44R311	1972	Thailand	N/A	2		2				1	1			1		
721258	1972	Mexico	H50a	3		2	2									
40R181	1972	Mexico	N/A	3		2	2									
731102	1973	Vietnam	H52	4		2				1	1			1		
42R917	1972	Vietnam	N/A	4		2				1	1			1		
ISP-04-06979	2004	Central Africa	H58	6		2	2	1	1				1	2		1
E03-9804	2003	Nepal	H58	6	1	2	2	1	1				1	2		1
ISP-03-07467	2003	Morocco	H58	6		2	2	1	1				1	2		1
pSTY7	1996	Vietnam	N/A	6	1	2	2	1	1				1	2		1
81863	1981	Peru	H50a	8		2	2	2	1				1	2	1	1
81918	1981	Peru	H50b	8									1	2	1	
81424	1981	Peru	H50b	8		1	1		2				1	2	1	1
SPA568	2003	Pakistan	N/A	8		2	2						1	2	1	1

* PCR to detect (1 denotes a positive product unless stated otherwise – see Methods section 2.3.10):
G – Insertion site of bla/sul/str into Tn21;
H, I – Insertion site of Tn21 into Tn9 (left and right boundary), product 1: pHCM1 insertion site, product 2: pAKU1 insertion site;
J, K – insertion site of Tn9 to pAKU1 backbone; product 1: pAKU1 insertion site; product 2: a band of different size
M – insertion site of Tn9 to pHCM1 backbone;
L – insertion site of Tn9 to Tn10 in pHCM1, product 1: pHCM1 insertion site; product 2: a band of different size;
N – insertion site of Tn10 in pHCM1;
O, P – insertion site of Tn10 in pAKU1; in P, product 1: pMAK1 insertion site; product 2: pAKU1
Q – insertion site of the second *strAB* genes in pAKU1.
U – GoldenGate signals for the presence of *betU* gene within Tn*6062*.

The composite transposon Tn*6062* described in chapter 4 (section 4.1.2.5) showed absolute association with PST6. The GoldenGate signals specific for *betU* gene within Tn*6062* indicated the presence of this transposon in all 180 PST6 plasmids in our collection (see Appendix 8.8). Two of the three Peruvian PST8s were also positive for Tn*6062*.

The data in general suggested that the acquisition of mobile elements happened independently of the evolution of the plasmid backbone because the transposons were inserted at different points on the tree. However, some events might have happened early in the evolutionary history of IncHI1 plasmids, resulting in the clustering of

various PSTs into roughly three groups, which corresponded to the phylogenetic grouping using SNP typing data.

5.2.3 The distribution of plasmids in relation to S. Typhi haplogroups worldwide

The simultaneous genotyping of both the *S.* Typhi backbone and the IncHI1 plasmids allowed us to investigate whether there was any link between the spatiotemporal distribution of plasmids and the spatiotemporal distribution of haplogroups of *S.* Typhi. To facilitate a simpler investigation of spatiotemporal distribution of *S.* Typhi backbone, the chromosomal SNP types were grouped into haplogroups based on their placements on the phylogenetic trees. Isolates from the same haplogroup were thus closely related to each other. The major haplotypes (defined by Rougmanac's SNPs (Roumagnac *et al.* 2006)) in each haplogroup were used to name the corresponding haplogroups. Figure 5-3 shows the haplogroups on the maximum likelihood phylogenetic trees built by the same method as in section 5.2.1 overlaid with the correspondent plasmid types.

The association of the plasmid STs with haplogroup (Figure 5-3) showed that there is a very strong link between haplogroup H58 and PST6. Other PSTs scattered in various haplogroups with up to 4 plasmid types in one haplogroup (H52). PST2 plasmids were found in three haplogroups (H52, H29 and H50a).

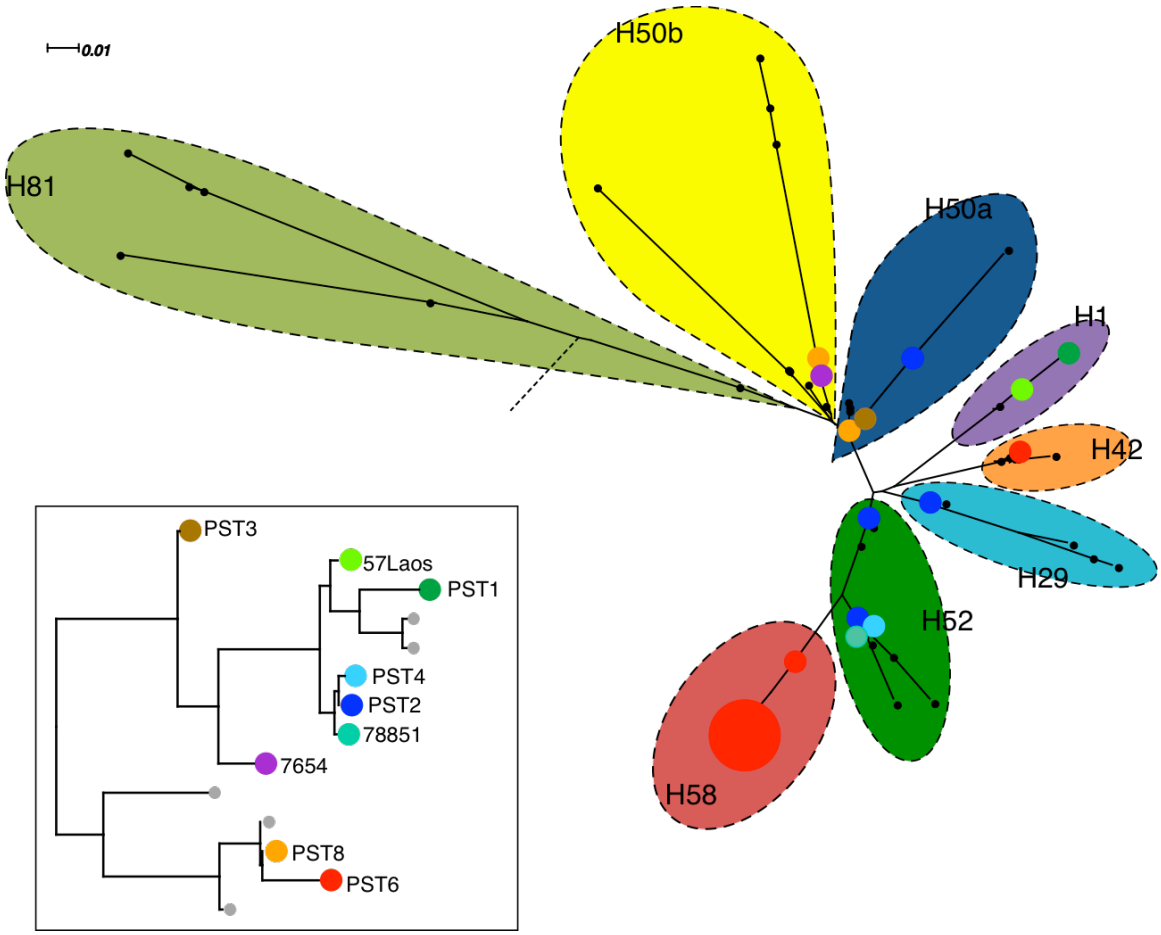
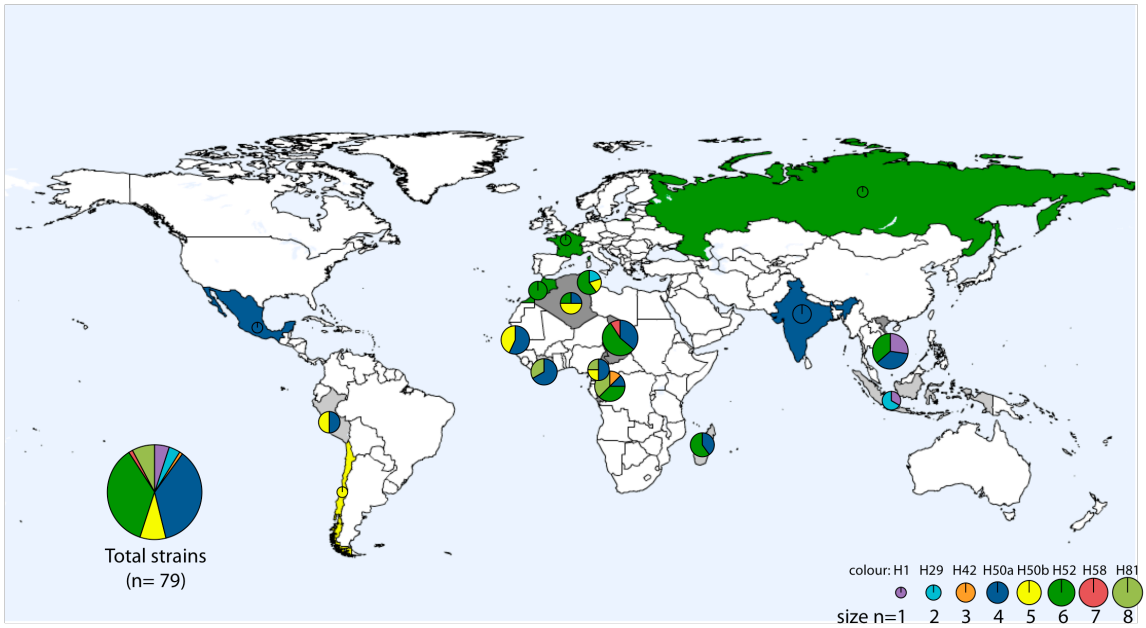
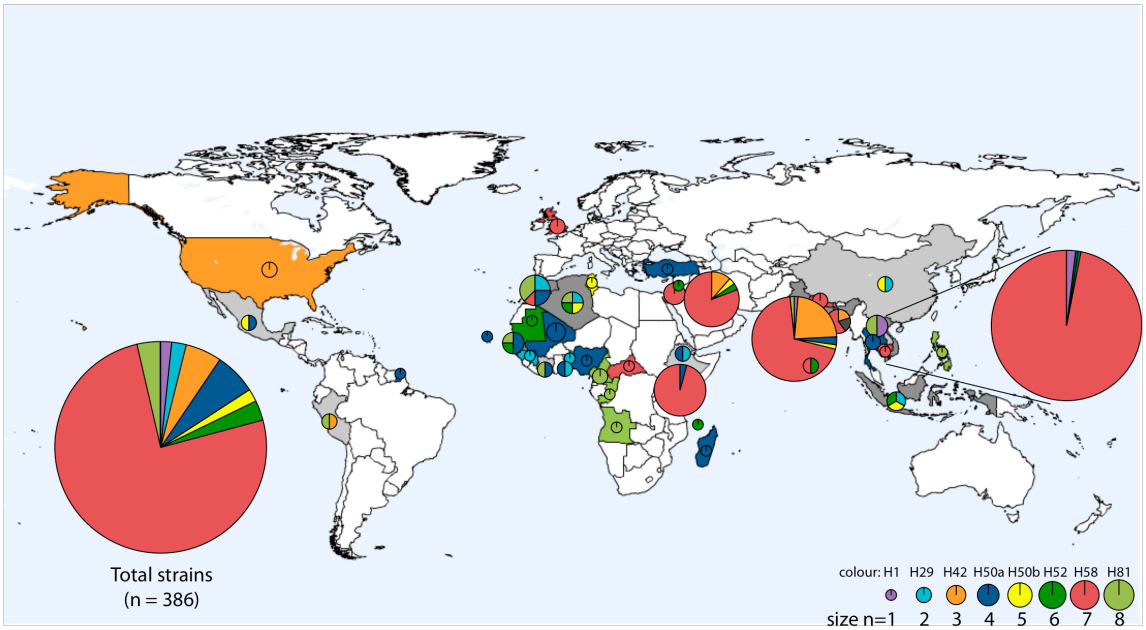


Figure 5-3 The association of plasmid type and strain background.
Tree drawn from chromosomal SNPs (colours represent the different haplogroups in all subsequent figures.). Internal circles represent the different plasmid STs (same colours as those in the plasmid tree inlet). All the branching shown has bootstrap value of 100 after 1000 iterations.

Also most of the other plasmid STs were on internal nodes of the phylogenetic tree and so were lost as the *S.* Typhi evolved. Recently emerged haplotypes (on the ends of branches) of *S.* Typhi do not have a plasmid except for *S.* Typhi H58 with plasmid PST6 and *S.* Typhi H1 with plasmid PST1. We then looked at the spatiotemporal distribution of haplogroups of *S.* Typhi and we saw that H58 has emerged in a similar fashion to plasmid type PST6 (Figure 5-4).



(a)



(b)

Figure 5-4 Spatiotemporal distribution of haplogroups of *S.* Typhi
(a) before 1993 and (b) after 1993.

The predominant haplogroups before 1993 were H50a and H52. African countries seemed to have the most diverse haplogroups. H1 were found only in Southeast Asia. The situation has changed after 1993 however. Whilst most haplogroups from before

1993 were still present afterwards, a newly emerged haplogroup H58 has become predominant, spreading across Asia to the Middle East, Africa and Europe.

5.2.4 PST6 is linked with H58 haplotype

Some *S.* Typhi haplogroups had a single but others were capable of harbouring different plasmid types (Figure 5-5). The majority of haplogroups had less than 15% of isolates harbouring a plasmid; H81 did not have any plasmids. Haplogroup H58 on the contrary had more than 60% of isolates containing only one type of plasmid: PST6.

Whilst PST6 seemed to be predominantly restricted to H58 haplogroups, PST2 plasmids were in many *S.* Typhi haplogroups (H1, H29, H50a, H52). There was only one non-H58 isolate harbouring PST6. This isolate, SDG1197, was an H42 isolated in 2005 from India. India was the country with the largest number of circulating haplogroups. Whether this mixture of strains from different backgrounds provide the environment for the spread of PST6 to haplogroups other than H58 remains to be determined.

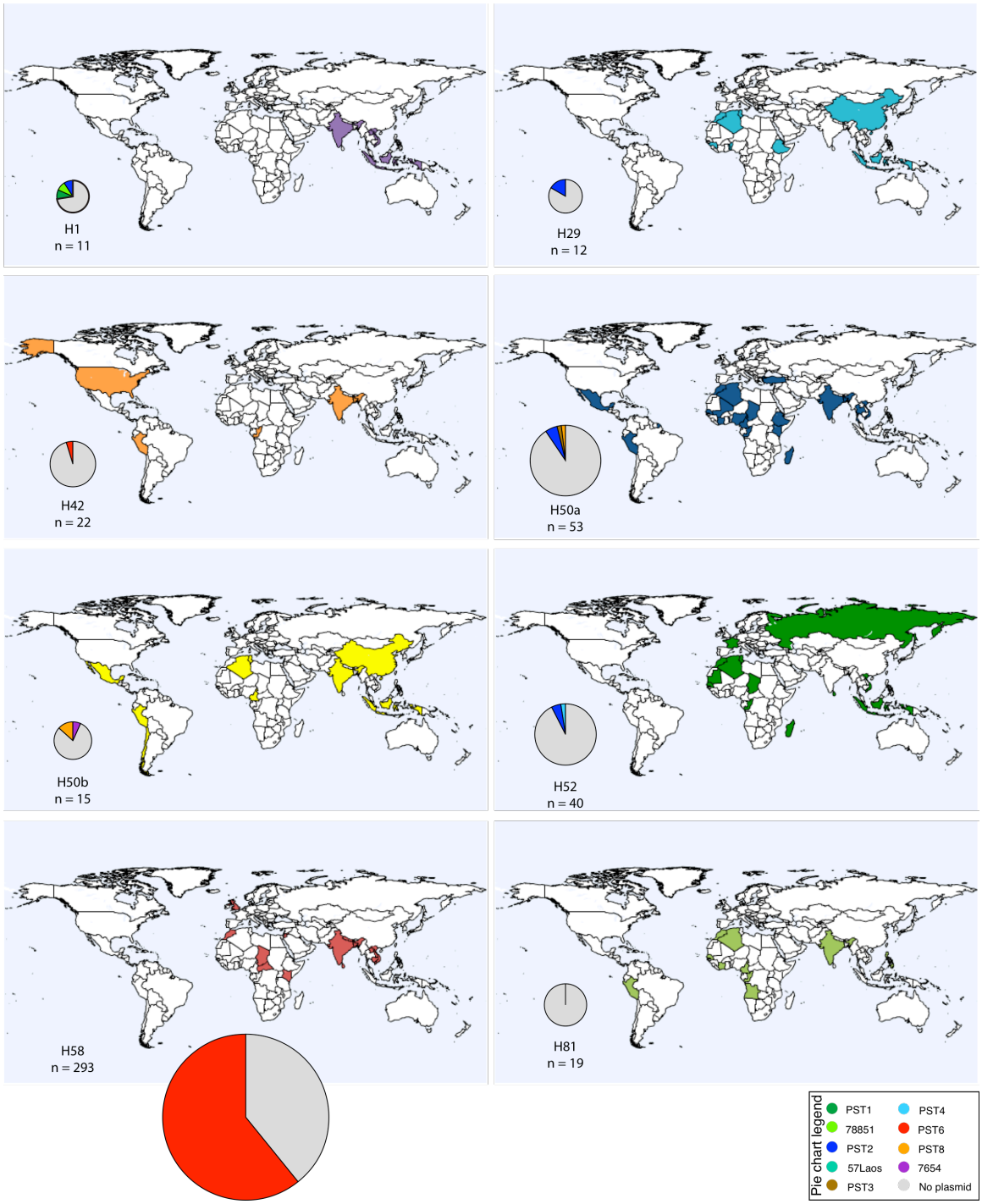


Figure 5-5 The distribution of individual haplogroups, and their plasmids, across the world. Colours within the map represent the distribution of the haplotype stated underneath the pie chart. The pie chart represents the number of isolates (size) and the colour the relative numbers of plasmid types.

5.2.5 Plasmid competition

In this study, PST6 completely predominated over plasmids of other types in recent years. We would like to investigate whether this predomination was the result of PST6

conferring more advantage or inflicting lower fitness cost to the background strain than other plasmid types. Thus, two plasmids, pHCM1 (PST1) and pSTY7 (PST6), were transferred independently into an attenuated *S.* Typhi background BRD948 (see Methods section 2.3.11) derived from Ty2, an H10 (haplogroup H52) strain. The growth of the two plasmid-containing strains and the plasmid-free background strains in rich media (LB media) was measured (Figure 5-6). No significant differences were observed, suggesting the presence of PST1 or PST6 plasmid in this background caused no or small (out of detection range of this method) cost to the bacterial host.

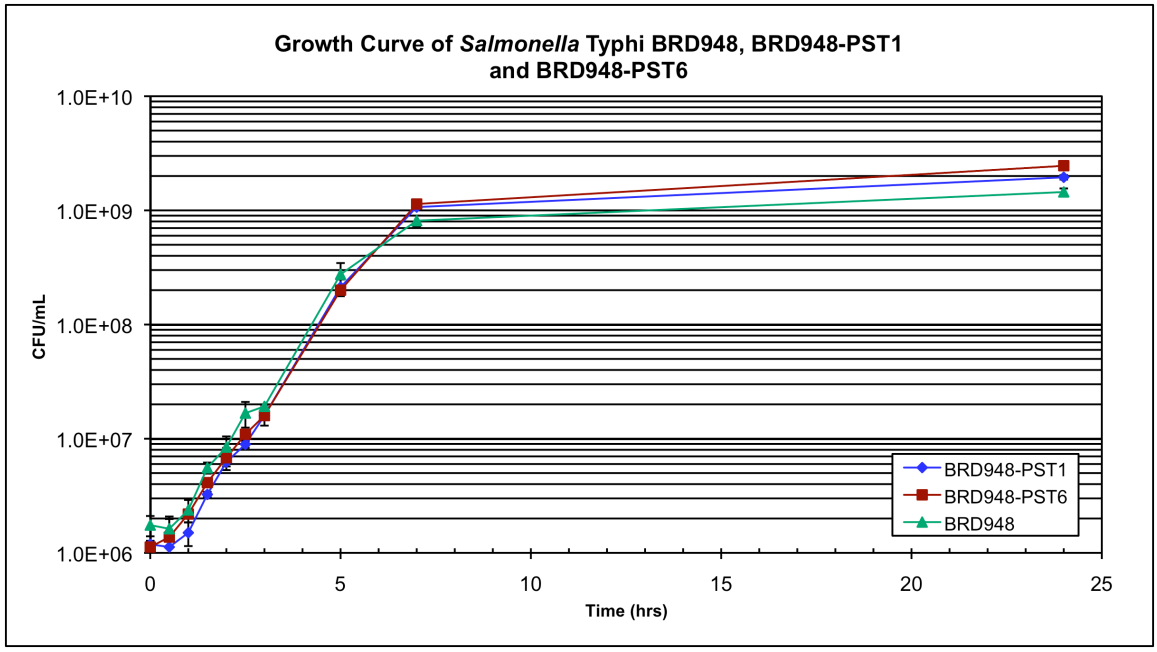


Figure 5-6 Comparing the growth of *S.* Typhi harbouring PST1 and PST6 plasmids.

Competitive growth is a better method to compare the fitness of two bacterial strains. Here we let the two plasmid-containing BRD948 strains mentioned above compete (1:1 inoculation ratio) in LB broth with chloramphenicol for 4 overnight serial passages. Samples were taken after each passage to grow on LB agar plates and colonies were randomly picked for colony-PCR to detect the plasmid type.

The results show that the *S.* Typhi Ty2 derivative (haplogroup H52) with a PST1 plasmid grows better in laboratory media than the same *S.* Typhi with a PST6 plasmid (Table 5-4). We can conclude that in this background PST6 failed to outcompete PST1 plasmid, contrary to the PST6 predomination situation observed from our data.

Table 5-4 Changes in the ratios of the two plasmids in competition assays

Time (days)	Replicate 1 (PST1:PST6 ratio)	Replicate 2 (PST1:PST6 ratio)	Replicate 3 (PST1:PST6 ratio)	Average (PST1:PST6 ratio)
0	1:0.78	1:1.21	1:1.03	1:0.99
1	1:0.86	1:0.78	1:0.68	1:0.77
2	1:0.97	1:0.55	1:0.63	1:0.70
3	1:0.32	1:0.33	1:0.36	1:0.34
4	1:0.19	1:0.39	1:0.32	1:0.29

5.3 Discussion

SNP typing is an expansion of the previous PMLST study trying to capture the population structure of IncHI1 plasmids in *S.* Typhi using more SNPs on a more diverse collection of strains. Interestingly, SNP typing of *S.* Typhi plasmids using 200 SNPs gave only 3 more subtypes compared to PMLST using 8 SNPs in 6 genes. These new subtypes were not particularly useful for typing because they were very rare in our collection of *S.* Typhi strains. For IncHI1 plasmids, our PMLST scheme captured most of the diversity. This shows that IncHI1 plasmids in *S.* Typhi are a closely related group and expanding the detecting SNP from 8 to 200 just increases the plasmid types from 8 to 11. Our finding is in concordance with the PMLST study: PST6 is the newly emerged (after 1993) plasmid type that is spreading globally.

While the investigation of plasmid backbones provided insights to IncHI1 plasmid vertical progression, information on how mobile elements were acquired was of great interest to understand the development of multiple drug resistance phenotypes. Based

on identifying the insertion sites of mobile elements on plasmids, plasmids clustered into three main groups that generally supported the branching of plasmid phylogenetic tree. The same insertion sites were identified on Tn*9* and Tn*10* in PST1, 2 and 4 indicating this group share a common ancestral plasmid, which acquired Tn*9* (conferring chloramphenicol resistance) and Tn*10* (conferring tetracycline resistance) then subsequently passed them on to all plasmids of this group. PST1, 2 and 4 all belong to PMLST group 1 defined in chapter 4. A different set of Tn*9* and Tn*10* insertion sites in PST6 and PST8 (PMLST group 2) suggests a different common ancestral plasmid for this group. PST3 is central on the SNP tree (Figure 5-1), suggesting this is an older set. However, only two plasmids of the PST3 type was analysed so this tentative conclusion requires further investigation. PST3 acquired Tn*9* and Tn*10* into different sites compared to the other plasmids.

Tn*6062* was previously reported in chapter 4 as a mobile element specific for group 2 plasmids. This was still held true for all 180 PST6 plasmids and two out of three PST8 plasmids. We were unable to investigate whether Tn*6062* was acquired into the same position on PST6 and PST8 plasmids because our PCR gave non-specific products due to the presence of multiple copies of IS*1* on both the plasmid and the chromosome. The strong association of Tn*6062* and PST6 suggested that there might be an advantage for PST6 plasmid or its H58 bacterial host to carry this composite transposon.

There were three PST8 plasmids isolated from Peru in 1981. Two of them were assigned to haplogroup H50b and one to H50a. These were the closet plasmid type to PST6. The mobile element profiles of these three were slightly different with 81918 likely to be the oldest strain for which plasmid had not acquired Tn*9*. All three harboured *strAB* genes similar to those in *S.* Paratyphi A plasmids. This suggested that these might be the progenitor of the IncHI1 plasmids in *S.* Paratyphi A.

Our study is the first to report the simultaneous typing of IncHI1 plasmids and their *S.* Typhi hosts with high discriminatory power at a global scale. Although the same principle was previously applied in many epidemiological studies for local outbreaks, the drawbacks of electrophoresis-based technology (PFGE, ribotyping, RFLP plasmid profiling) reduce its usefulness in global epidemiology studies. A total of 26 subtypes (combined type of *Xba*I PFGE type-*Spe*I PFGE type-ribotype) were assigned to isolates of H58 haplogroups from Vietnam between 1996 and 2004 (Le *et al.* 2007). Such fine subtyping is good for tracing local outbreak isolates but fails to call attention to the fact that these subtypes are very closely related in evolutionary terms.

Our SNP typing data highlight the rapid global expansion of the H58 haplogroup and its strong association with PST6 IncHI1 plasmid since 1993. The association pattern of H58 haplogroup and PST6 contradicts what is observed before 1993 where multiple plasmid types present in multiple haplogroups. The outbreak of a single *S.* Typhi clone (ribotype 3a) harbouring a single RFLP type plasmid was observed in Vietnam from 1995 to 2002 by Le et al (Le *et al.* 2004) and we believe that this could have been associated with a H58 with PST6 plasmid. Although many H58 *S.* Typhi might have lost plasmids or do not have plasmids, the plasmid harbouring H58 can be found in many countries, suggesting the initial presence of H58/PST6 in each area with subsequent simultaneous loss of plasmid. Plasmid free H58 isolates from the past ten years predominantly harbour a chromosomal IS*1* element which was most likely acquired as a consequence of plasmid carriage (Holt, personal communication). Thus, even antibiotic sensitive and plasmid free H58 harbour a genetic scar of plasmid carriage.

Although the H58 haplogroup started expanding only in the past fifteen years, our data showed the presence of one H58 isolate from Chad in 1964. This is in ageement with

Roumagnac's study (Roumagnac *et al.* 2006) in which several isolates of the H58 haplogroup from African countries were isolated in the late 1950s and the 1960s. The presence of plasmid-free H58 strain since the 1960s and a close relative of PST6 plasmids in the 1980s seemed to provide the right components for the emergence of H58/PST6 strain later in the 1990s.

All PST6 plasmids in this study had identical SNP type over 200 loci indicating a clonal expansion of this group. The plasmid expansion might happen in two ways: vertically following the expansion of the background strain or horizontally by conjugation to other plasmid free backgrounds. We have strong evidence to suggest PST6 plasmids spread vertically. It is very rare to find PST6 in a background other than haplogroup H58 (only one H42 isolate harboured a PST6 plasmid). This suggests that a selective advantage occurred when the combination of background and plasmid is matched. This advantage is not transferable. Indeed, our competition assays showed that PST6 does not confer any advantage enabling it to outcompete the PST1 plasmid in a non-H58 background (Ty2 straind from haplotype H10). It is thus unlikely that PST6 spread via conjugation to plasmid free cells but more likely the spread of an H58/PST6 combination.

Ideally we would like to apply our SNP typing method to a representative global collection of *S.* Typhi isolates. Keeping this in mind, we tried to obtain a wide range of isolates from many countries isolated at various time points. While we are able to capture the general spatiotemporal distribution of different haplogroups and plasmid types, our method of collecting samples could introduce a certain level of bias to our study as we disregard the original purpose of strain collection at each laboratory. Under-representation of numbers of isolates in many countries are still a big confounding factor. The scarcity of historical strains is also an obstacle to fully understanding the plasmid – background strain association in the past.

While it is clear that the H58 haplogroup is spreading, we avoid comparing the scale of expansion of H58 between countries because the strain collection was biased by the large number of isolates from Vietnam. Any SNP typing method relies on a known set of SNP loci discovered by re-sequencing. The inherent disadvantage of this typing method is that it would fail to pick up newly emerged isolate if it carries mutations on loci not covered by the typing assay.

In conclusion, IncHI1 plasmids are surviving in *S.* typhi predominantly on a specific bacterial host genetic background. The molecular basis of the H58 haplogroups and PST6 should be investigated further.

6 Final Discussion

Conjugative plasmids are important vehicles for the horizontal gene transfer events that shape the evolution of prokaryotic organisms. In clinical settings, the rapid dissemination of antibiotic resistant determinants is mediated by the combination of integrona as assembly platforms, transposons as intracellular transfer vehicles, and conjugative plasmids as vehicles for intercellular transfer. Through one conjugative event, a large array of antibiotic resistant genes can be transferred across bacterial species or genera. Bacteria such as *S.* Typhi have gained resistance to most antibiotics available, even third generation cephalosporins, potentially leaving azithromycin the only choice for treatment in many areas.

The majority of studies on outbreaks of drug resistant *S.* Typhi were performed in developing world and occasionally in developed world when a traveller brought back the infection. Techniques such as PFGE and plasmid profiling are widely used and are great tools to study local outbreaks. However it becomes difficult to tell whether a strain brought back by a traveller is the same outbreak strain from a remote village in, say, India. Attempts have been made to address the capability to compare strain identities across laboratories, either by standardising current methods (PulseNet or EnterNet) or develop new methods (MLST, SNP typing). For the purposes of studying global epidemiology and evolution, even standardised PFGE shows its drawbacks as it is more likely to overestimate the distance between isolates. Multilocus sequence typing, developed as a tool to study global epidemiology and evolution, undoubtedly serves us better than PFGE for these purposes. SNP typing also has an overwhelm advantage over PFGE: the readiness to compare strain identities. However, it will take time before laboratories in developing countries can routinely use newer method like MLST, mostly

because of the limited availability of sequencing facilities, which are becoming faster and cheaper.

Unfortunately, the studies of plasmids in clinical settings are lagging behind. RFLP for plasmids, probably the equivalent of PFGE for bacterial chromosome with similar drawbacks, are not even routinely performed, and the current MLST methods predominantly focus only on chromosomal sequences. The only tool for the identification and classification of plasmids are replicon typing, or more recently, the PCR-based version of replicon typing. This method groups plasmids into incompatibility groups without actually performing the classical incompatibility tests. The PCR replicon typing method has recently gained popularity due to its ease of use and interpretation. The Inc groups are however too broad for the close investigation of plasmid transfer, as well as the transfer of the antibiotic resistant determinants they carry. While *S.* Typhi are endemic in developing countries such as India, Pakistan etc., only a few studies perform RFLP and even less use PCR replicon typing. Our understanding of plasmid transmission and drug resistance transmission in *S.* Typhi is therefore very limited.

Evidence has suggested the co-evolution of IncHI1 plasmids in *S.* Typhi backgrounds since the introduction of chemotherapy for the treatment of typhoid fever. Plasmids of IncHI1 are the main carriers of antibiotic resistant genes in *S.* Typhi. Given the situation, I investigated the genetic factors that contribute to stabilise the plasmid-background co-existence. Second, realising the lack of tools for the investigation of plasmids in general and IncHI1 plasmids in particular, I proposed to develop a new typing tool for IncHI1 plasmids. And lastly, I demonstrated how the simultaneous typing of both the plasmid and the background strain is useful to study the transmission of plasmids within a bacterial population.

The co-evolution of a conjugative plasmid and its bacterial host over several hundred generations potentially results in a fitter strain suggesting the plasmid and the bacterial chromosome have found a way to better co-exist. Plasmids IncHI1 have been isolated from *S.* Typhi for nearly 40 years. Studies on the global regulator H-NS have suggested a mechanism for an IncHI1 plasmid to minimise its adverse effect on a *S.* Typhi chromosome. I present in chapter 3 a high-throughput method to screen for novel genes responsible for the stable inheritance of plasmid IncHI1 in *S.* Typhi.

Transposon-directed insertion site sequencing (TraDIS) is a method that combines the commonly used Tn*5* transposon to generate random mutation libraries with a next generation sequencing technique from Illumina (formerly Solexa) to produce a high-resolution screening. I applied this method to an attenuated *S.* Typhi strain harbouring an IncHI1 plasmid and used serial passages as the selective pressure to screen for candidate genes that might contribute to plasmid stability. I identified several candidate genes on the chromosome and on the plasmids. Whilst most of the chromosomal candidates are more likely to have a role in helping the bacteria survive in the presence of chloramphenicol, the antibiotic used to select for plasmids, seven plasmid-encoded hypothetical genes, along with known genes for plasmid stability such as *parA*, *parB, sfh* and *hok,* were identified as candidates for plasmid stability. A Δ*sfh* plasmid was generated and the *S.* Typhi strain harbouring this mutant plasmid shows no growth defects compared to the same *S.* Typhi background with a wild-type plasmid. This result suggests that the effect of plasmid stability genes can be subtle. To follow up with this work, a set of carefully designed experiments, including long-term experiments and plasmid competition assays, is needed to further access the impact of plasmid stability genes. The seven candidate genes described here should be systematically knocked out and subjected to such experiments for evaluating their role in plasmid stability.

TraDIS is a novel method to take advantage of the next generation sequencing technology to simultaneously identify with high confident and high accuracy the insertion sites of more than one million mutants. This will take all the previously applications using random mutant libraries to a new level of power by using more mutants in one experiment. The applications range for this kind of library-based sequencing is huge, including, but not limiting to, defining the essential gene sets, defining genes required for the survival in different niches, defining genes required for adhesion and invasion of host cells or genes required for infection and transmission in animal models. Potential drawbacks to applying this method to other organisms is the availability of a suitable transposon with a high enough transfer rate for the generation of a large library of a million mutants. Care should also be taken in the interpretation of the results as genes that are not directly involved in the measured phenotype might be co-selected. Our example of such a situation is the co-selection of chloramphenicol resistant genes on the chromosome in the assay intended to select for plasmid stability genes because of the presence of chloramphenicol in the media.

In providing tools to investigate the population structure and evolution of IncHI1 plasmids, I applied the concept of MLST to plasmid. Six conserved genes were chosen for the IncHI1 plasmid typing scheme called IncHI1 plasmid multilocus sequencing typing (PMLST). The scheme involves amplifying and sequencing of six loci on the plasmid for their identification and classification. A collection of IncHI1 plasmid investigated using PMLST revealed two main groups of IncHI1 plasmids: group 1 consists of older plasmids isolated before 1993 and group 2 of more recent plasmids isolated after 1993. The disappearance of group 1 plasmids after 1993 suggests a replacement of group 1 by group 2 plasmids in current *S.* Typhi population. The mechanisms for this replacement remains unclear at this point but we are able to think

of two possibilities: the spread of group two plasmids to different *S.* Typhi backgrounds, driving the older plasmids out of the bacterial population – a direct plasmid competition; or the spread of a particular strain carrying a group 2 plasmids – bacterial competition, plasmid most likely to contribute to the success of this competition.

Our concept of using MLST for plasmids was also independently used by García-Fernández *et al.* to IncI1 plasmids (Garcia-Fernandez *et al.* 2008). This on one hand shows the usefulness of this method to study the evolution of plasmids but on the other hand highlights a disadvantage: the need for a separate gene set for each plasmid Inc group due to the phylogenetic distance and high level of sequence variations. Recently, Garcillan-Barcia describes the use of relaxase gene to study the evolution and classification of all conjugative and mobile plasmids (Garcillan-Barcia, Francia & de la Cruz 2009). This might provide a generic tool for the identification and typing of all plasmids but their level of discrimination between closely related plasmid requires further investigation.

In order to answer the question about the mechanism by which group 2 replaces group 1 plasmids, I made use of a newly developed SNP typing scheme to genotype *S.* Typhi chromosome and IncHI1 plasmid simultaneously. A previous study by Holt *et al.* describing the whole genome sequences of 19 *S.* Typhi isolates selected from different branches of *S.* Typhi phylogenetic tree has identified a set of SNPs (on both chromosome and plasmid) from these sequences. The SNPs were then used to develop a SNP typing method based on the Illumina GoldenGate platform. The application of this typing method on a collection of 473 *S.* Typhi isolates revealed striking data on the movement of IncHI1 plasmid within *S.* Typhi populations and highlights the emergence of a background-plasmid combination, H58/PST6, that is spreading globally. By

simultaneously typing of plasmid and chromosome, we are able to identify a strong association of a group 2 plasmid type, PST6, with the background H58. This suggests PST6 displays background specificity and that the spread of this plasmid type is the result of the successful spread of a single background strain type. This plasmid-background specificity has never been reported previously. It would be interesting to better understand the distinct contribution of the plasmid and chromosome to the success of this H58/PST6 strain. We were able to a culture based experiment to compete an older plasmid from group 1 (PST1) and PST6 in an identical Ty2-derived background (H10). The results were in apparent contrast with what is observed in nature: PST1 outcompetes PST6 in H10 background. This might suggest that PST6 requires an H58 background to show its beneficial effect, although many other factors could play a role in this artificial experiment. Further investigation is needed to fully understand the relationship of PST6 and the specific background H58.

Our SNP typing method has proved that simultaneous tracing of plasmid and background strains is very useful to provide insights to the spread of plasmids and their accompanied resistant genes in a pathogen bacterial population. This data will provide information for the management of drug resistance in the future.

Unlike MLST which is a variation discovery method where newly emerged strains will be added to the database, SNP typing uses a defined set of known SNPs for the typing of isolates which means that newly emerged strains carrying new SNP loci will be missed by SNP typing. This drawback can be addressed by periodical addition of new SNP loci to the current typing set. This requires the continuing use of SNP discovery methods such as MLST and re-sequencing.

In conclusion, there is still much to learn about the interactions between a plasmid and its host's chromosome in terms of their stable co-existence. Many interactions are likely

to be subtle and need specially designed experiments to investigate. Although the role of plasmids in the dissemination of antibiotic resistance has been established and widely known, detailed reports on plasmids from clinical isolates are rare due to technological limitations. This study presents the details of PMLST, a typing and classification scheme for IncHI1 plasmids, and a SNP typing scheme to simultaneously trace plasmids and their host genetic backgrounds. Future work will assess the utility of these approaches to demonstrate their value to study the evolution and transmission of plasmids in clinical isolates within the context of their bacterial host population.

Publications arising from this Thesis

Gemma C. Langridge*, Minh-Duy Phan*, Daniel J. Turner*, Timothy T. Perkins, Leopold Parts, Jana Haase, Ian Charles, Duncan J. Maskell, Sarah E. Peters, Gordon Dougan, John Wain, Julian Parkhill and A. Keith Turner. 2009, "Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants", *Genome Research*, vol. 19, no. 12, pp. 2308-2316.

* Co-first authors.

Phan, M. D., Kidgell, C., Nair, S., Holt, K. E., Turner, A. K., Hinds, J., Butcher, P., Cooke, F. J., Thomson, N. R., Titball, R., Bhutta, Z. A., Hasan, R., Dougan, G. and Wain, J. 2009, "Variation in Salmonella enterica serovar typhi IncHI1 plasmids during the global spread of resistant typhoid fever", *Antimicrob. Agents Chemother*, vol. 53, no. 2, pp. 716-727.

Phan, M.D., Wain, J. 2008, "IncHI1 plasmids, a dynamic link between resistance and Pathogenicity", *J. Infect. Developing Countries*, vol. 2, no. 4, pp. 272-278.

Holt, K. E., Thomson, N. R., Wain, J., Phan, M. D., Nair, S., Hasan, R., Bhutta, Z. A., Quail, M. A., Norbertczak, H., Walker, D., Dougan, G. and Parkhill, J. 2007, "Multidrug- resistant Salmonella enterica serovar paratyphi A harbors IncHI1 plasmids similar to those found in serovar typhi", *J. Bacteriol*, vol. 189, no. 11, pp. 4257-4264.

7 References

1. Achtman, M. 2008, "Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens", *Annual Review of Microbiology,* vol. 62, pp. 53-70.
2. Achtman, M., Kennedy, N. & Skurray, R. 1977, "Cell--cell interactions in conjugating Escherichia coli: role of traT protein in surface exclusion", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 74, no. 11, pp. 5104-5108.
3. Adachi, T., Sagara, H., Hirose, K. & Watanabe, H. 2005, "Fluoroquinolone-resistant *Salmonella* Paratyphi A", *Emerging Infectious Diseases,* vol. 11, no. 1, pp. 172-4.
4. Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N. & Mekalanos, J.J. 2002, "A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, no. 2, pp. 966-971.
5. Alonso, G., Baptista, K., Ngo, T. & Taylor, D.E. 2005, "Transcriptional organization of the temperature-sensitive transfer system from the IncHI1 plasmid R27", *Microbiology,* vol. 151, no. 11, pp. 3563-3573.
6. Altier, C., Suyemoto, M., Ruiz, A.I., Burnham, K.D. & Maurer, R. 2000, "Characterization of two novel regulatory genes affecting *Salmonella* invasion gene expression", *Molecular microbiology,* vol. 35, no. 3, pp. 635-646.
7. Alton, N.K. & Vapnek, D. 1979, "Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn*9*", *Nature,* vol. 282, no. 5741, pp. 864-869.
8. Altwegg, M., Hickman-Brenner, F.W. & Farmer, J.J.,3rd 1989, "Ribosomal RNA gene restriction patterns provide increased sensitivity for typing *Salmonella* typhi strains", *The Journal of infectious diseases,* vol. 160, no. 1, pp. 145-149.
9. Anand, A.C. 1993, "The anatomy of an epidemic (the final report on an epidemic of multidrug resistant enteric fever in eastern India)", *Tropical Gastroenterology,* vol. 14, no. 1, pp. 21-7.
10. Ansong, C., Yoon, H., Porwollik, S., Mottaz-Brewer, H., Petritis, B.O., Jaitly, N., Adkins, J.N., McClelland, M., Heffron, F. & Smith, R.D. 2009, "Global systems-level analysis of Hfq and SmpB deletion mutants in Salmonella: implications for virulence and global protein translation", *PloS one,* vol. 4, no. 3, pp. e4809.
11. Audette, G.F., Manchak, J., Beatty, P., Klimke, W.A. & Frost, L.S. 2007, "Entry exclusion in F-like plasmids requires intact TraG in the donor that recognizes its cognate TraS in the recipient", *Microbiology,* vol. 153, no. 2, pp. 442-451.
12. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. & Mori, H. 2006, "Construction of *Escherichia*

coli K-12 in-frame, single-gene knockout mutants: the Keio collection", *Molecular systems biology,* vol. 2, pp. 2006.0008.

13. Bahl, M.I., Hansen, L.H. & Sorensen, S.J. 2009, "Persistence mechanisms of conjugative plasmids", *Methods in molecular biology (Clifton, N.J.),* vol. 532, pp. 73-102.

14. Bahl, M.I., Hansen, L.H. & Sorensen, S.J. 2007, "Impact of conjugal transfer on the stability of IncP-1 plasmid pKJK5 in bacterial populations", *FEMS microbiology letters,* vol. 266, no. 2, pp. 250-256.

15. Banos, R.C., Vivero, A., Aznar, S., Garcia, J., Pons, M., Madrid, C. & Juarez, A. 2009, "Differential regulation of horizontally acquired and core genome genes by the bacterial modulator H-NS", *PLoS genetics,* vol. 5, no. 6, pp. e1000513.

16. Barth, P.T. & Grinter, N.J. 1974, "Comparison of the deoxyribonucleic acid molecular weights and homologies of plasmids conferring linked resistance to streptomycin and sulfonamides", *Journal of Bacteriology,* vol. 120, no. 2, pp. 618-630.

17. Beloin, C., Deighan, P., Doyle, M. & Dorman, C.J. 2003, "*Shigella flexneri* 2a strain 2457T expresses three members of the H-NS-like protein family: characterization of the Sfh protein", *Molecular genetics and genomics : MGG,* vol. 270, no. 1, pp. 66-77.

18. Bhan, M.K., Bahl, R. & Bhatnagar, S. 2005, "Typhoid and paratyphoid fever", *Lancet,* vol. 366, no. 9487, pp. 749-62.

19. Bonfield, J.K., Smith, K. & Staden, R. 1995, "A new DNA sequence assembly program", *Nucleic acids research,* vol. 23, no. 24, pp. 4992-4999.

20. Boucher, Y., Labbate, M., Koenig, J.E. & Stokes, H.W. 2007, "Integrons: mobilizable platforms that promote genetic diversity in bacteria", *Trends in microbiology,* vol. 15, no. 7, pp. 301-309.

21. Bouma, J.E. & Lenski, R.E. 1988, "Evolution of a bacteria/plasmid association", *Nature,* vol. 335, no. 6188, pp. 351-352.

22. Bradley, D.E., Hughes, V.M., Richards, H. & Datta, N. 1982, "R plasmids of a new incompatibility group determine constitutive production of H pili", *Plasmid,* vol. 7, no. 3, pp. 230-238.

23. Bramhill, D. & Kornberg, A. 1988, "A model for initiation at origins of DNA replication", *Cell,* vol. 54, no. 7, pp. 915-918.

24. Brenner, F.W., Villar, R.G., Angulo, F.J., Tauxe, R. & Swaminathan, B. 2000, "*Salmonella* nomenclature", *Journal of clinical microbiology,* vol. 38, no. 7, pp. 2465-2467.

25. Butler, H. & Ragoussis, J. 2008, "BeadArray-based genotyping", *Methods in molecular biology (Clifton, N.J.),* vol. 439, pp. 53-74.

26. Butler, T., Linh, N.N., Arnold, K. & Pollack, M. 1973, "Chloramphenicol-resistant typhoid fever in Vietnam associated with R factor", *Lancet,* vol. 302, no. 7836, pp. 983-985.

27. Butler, T., Rumans, L. & Arnold, K. 1982, "Response of typhoid fever caused by chloramphenicol-susceptible and chloramphenicol-resistant strains of

Salmonella typhi to treatment with trimethoprim-sulfamethoxazole", *Review of Infectious Diseases,* vol. 4, no. 2, pp. 551-61.

28. Campbell, C.S. & Mullins, R.D. 2007, "In vivo visualization of type II plasmid segregation: bacterial actin filaments pushing plasmids", *The Journal of cell biology,* vol. 179, no. 5, pp. 1059-1066.

29. Capoor, M.R., Nair, D., Walia, N.S., Routela, R.S., Grover, S.S., Deb, M., Aggarwal, P., Pillai, P.K. & Bifani, P.J. 2009, "Molecular analysis of high-level ciprofloxacin resistance in *Salmonella enterica* serovar Typhi and *S.* Paratyphi A: need to expand the QRDR region?", *Epidemiology and infection,* vol. 137, no. 6, pp. 871-878.

30. Carattoli, A., Bertini, A., Villa, L., Falbo, V., Hopkins, K.L. & Threlfall, E.J. 2005, "Identification of plasmids by PCR-based replicon typing", *Journal of Microbiological Methods,* vol. 63, no. 3, pp. 219-228.

31. Chaudhuri, R.R., Allen, A.G., Owen, P.J., Shalom, G., Stone, K., Harrison, M., Burgis, T.A., Lockyer, M., Garcia-Lara, J., Foster, S.J., Pleasance, S.J., Peters, S.E., Maskell, D.J. & Charles, I.G. 2009, "Comprehensive identification of essential *Staphylococcus aureus* genes using Transposon-Mediated Differential Hybridisation (TMDH)", *BMC genomics,* vol. 10, no. 1, pp. 291.

32. Clinical and Laboratory Standards Institute. 2003, "Performance Standards for Antimicrobial Disk Susceptibility Tests - Approved Standard - Eight Edition" , vol. 23 no. 1.

33. Colquhoun, J. & Weetch, R.S. 1950, "Resistance to chloramphenicol developing during treatment of typhoid fever", *Lancet,* vol. 2, no. 6639, pp. 621-623.

34. Connerton, P., Wain, J., Hien, T.T., Ali, T., Parry, C., Chinh, N.T., Vinh, H., Ho, V.A., Diep, T.S., Day, N.P.J., White, N.J., Dougan, G. & Farrar, J.J. 2000, "Epidemic typhoid in Vietnam: molecular typing of multiple-antibiotic-resistant *Salmonella enterica* serotype Typhi from four outbreaks", *Journal of Clinical Microbiology,* vol. 38, no. 2, pp. 895-897.

35. Cooper, T.F. & Heinemann, J.A. 2000, "Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 97, no. 23, pp. 12643-12648.

36. Couturier, M., Bex, F., Bergquist, P.L. & Maas, W.K. 1988, "Identification and classification of bacterial plasmids", *Microbiol. Mol. Biol. Rev.,* vol. 52, no. 3, pp. 375-395.

37. Craigie, J. & Yen, C.H. 1938, "The demonstration of types of *B. typhosus* by means of preparation of type II Vi phage. ", *Can. Public Health J.,* vol. 29, pp. 448-496.

38. Crump, J.A., Luby, S.P. & Mintz, E.D. 2004, "The global burden of typhoid fever", *Bull World Health Organ,* vol. 82, no. 5, pp. 346-353.

39. Dahlberg, C. & Chao, L. 2003, "Amelioration of the cost of conjugative plasmid carriage in *Eschericha coli* K12", *Genetics,* vol. 165, no. 4, pp. 1641-1649.

40. Datta, N. 1985, "Plasmids as organisms", *Basic life sciences,* vol. 30, pp. 3-16.

41. Datta, N. & Hedges, R.W. 1972, "R factors identified in Paris, some conferring gentamicin resistance, constitute a new compatibility group", *Annales de l'Institut Pasteur,* vol. 123, no. 6, pp. 849-852.

42. Datta, N., Richards, H. & Datta, C. 1981, "*Salmonella* typhi in vivo acquires resistance to both chloramphenicol and co-trimoxazole", *Lancet,* vol. 1, no. 8231, pp. 1181-1183.

43. Datta, N. & Olarte, J. 1974, "R factors in strains of *Salmonella* typhi and *Shigella dysenteriae* 1 isolated during epidemics in Mexico: classification by compatibility", *Antimicrobial Agents and Chemotherapy,* vol. 5, no. 3, pp. 310-317.

44. de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G., Richez, C., Durot, M., Kreimeyer, A., Le Fevre, F., Schachter, V., Pezo, V., Doring, V., Scarpelli, C., Medigue, C., Cohen, G.N., Marliere, P., Salanoubat, M. & Weissenbach, J. 2008, "A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1", *Molecular systems biology,* vol. 4, pp. 174.

45. de la Bastide, M. & McCombie, W.R. 2007, "Assembling genomic DNA sequences with PHRAP", *Current protocols in bioinformatics / editoral board, Andreas D.Baxevanis ...[et al.],* vol. Chapter 11, pp. Unit11.4.

46. de la Cruz, J. & Vioque, A. 2001, "Increased sensitivity to protein synthesis inhibitors in cells lacking tmRNA", *RNA (New York, N.Y.),* vol. 7, no. 12, pp. 1708-1716.

47. del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M. & Diaz-Orejas, R. 1998, "Replication and control of circular bacterial plasmids", *Microbiol. Mol. Biol. Rev.,* vol. 62, no. 2, pp. 434-464.

48. Dempsey, L.A., Zhao, A.C. & Khan, S.A. 1995, "Localization of the start sites of lagging-strand replication of rolling-circle plasmids from gram-positive bacteria", *Molecular microbiology,* vol. 15, no. 4, pp. 679-687.

49. Diaz, R. & Staudenbauer, W.L. 1982, "Replication of the broad host range plasmid RSF1010 in cell-free extracts of *Escherichia coli* and *Pseudomonas aeruginosa*", *Nucleic acids research,* vol. 10, no. 15, pp. 4687-4702.

50. Didelot, X., Achtman, M., Parkhill, J., Thomson, N.R. & Falush, D. 2007, "A bimodal pattern of relatedness between the *Salmonella* Paratyphi A and Typhi genomes: convergence or divergence by homologous recombination?", *Genome research,* vol. 17, no. 1, pp. 61-68.

51. Dionisio, F., Conceição, I.C., Marques, A.C.R., Fernandes, L. & Gordo, I. 2005, "The evolution of a conjugative plasmid and its ability to increase bacterial fitness", *Biology Letters,* vol. 1, no. 2, pp. 250–252.

52. Dorman, C.J., Hinton, J.C. & Free, A. 1999, "Domain organization and oligomerization among H-NS-like nucleoid-associated proteins in bacteria", *Trends in microbiology,* vol. 7, no. 3, pp. 124-128.

53. Doyle, M. & Dorman, C.J. 2006, "Reciprocal transcriptional and posttranscriptional growth-phase-dependent expression of *sfh*, a gene that encodes a paralogue of the nucleoid-associated protein H-NS", *J. Bacteriol.,* vol. 188, no. 21, pp. 7581-7591.

54. Doyle, M., Fookes, M., Ivens, A., Mangan, M.W., Wain, J. & Dorman, C.J. 2007, "An H-NS-like stealth protein aids horizontal DNA transmission in bacteria", *Science,* vol. 315, no. 5809, pp. 251-2.

55. Draper, O., Cesar, C.E., Machon, C., de la Cruz, F. & Llosa, M. 2005, "Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 45, pp. 16385-16390.

56. Dutta, S., Sur, D., Manna, B., Bhattacharya, S.K., Deen, J.L. & Clemens, J.D. 2005, "Rollback of *Salmonella enterica* serotype Typhi resistance to chloramphenicol and other antimicrobials in Kolkata, India", *Antimicrobial Agents and Chemotherapy,* vol. 49, no. 4, pp. 1662-1663.

57. Eckert, B. & Beck, C.F. 1989, "Overproduction of transposon Tn*10*-encoded tetracycline resistance protein results in cell death and loss of membrane potential", *Journal of Bacteriology,* vol. 171, no. 6, pp. 3557-3559.

58. Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P. & Spratt, B.G. 2004, "eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data", *J Bacteriol,* vol. 186, no. 5, pp. 1518-30.

59. Fernandez-Lopez, R., Machon, C., Longshaw, C.M., Martin, S., Molin, S., Zechner, E.L., Espinosa, M., Lanka, E. & de la Cruz, F. 2005, "Unsaturated fatty acids are inhibitors of bacterial conjugation", *Microbiology (Reading, England),* vol. 151, no. Pt 11, pp. 3517-3526.

60. Forns, N., Banos, R.C., Balsalobre, C., Juarez, A. & Madrid, C. 2005, "Temperature-dependent conjugative transfer of R27: role of chromosome- and plasmid-encoded Hha and H-NS proteins", *J. Bacteriol.,* vol. 187, no. 12, pp. 3950-3959.

61. Frost, L.S., Ippen-Ihler, K. & Skurray, R.A. 1994, "Analysis of the sequence and gene products of the transfer region of the F sex factor", *Microbiological reviews,* vol. 58, no. 2, pp. 162-210.

62. Frost, L.S., Leplae, R., Summers, A.O. & Toussaint, A. 2005, "Mobile genetic elements: the agents of open source evolution", *Nature reviews.Microbiology,* vol. 3, no. 9, pp. 722-732.

63. Gabant, P., Chahdi, A.O. & Couturier, M. 1994, "Nucleotide sequence and peplication characteristics of RepHI1B: a replicon specific to the IncHI1 plasmids", *Plasmid,* vol. 31, no. 2, pp. 111-120.

64. Gabriel, S., Ziaugra, L. & Tabbaa, D. 2009, "SNP genotyping using the Sequenom MassARRAY iPLEX platform", *Current protocols in human genetics / editorial board, Jonathan L.Haines ...[et al.],* vol. Chapter 2, pp. Unit 2.12.

65. Gangarosa, E.J., Bennett, J.V., Wyatt, C., Pierce, P.E., Olarte, J., Hernandes, P.M., Vazquez, V. & Bessudo, D. 1972, "An epidemic-associated episome?", *J Infect Dis,* vol. 126, no. 2, pp. 215-8.

66. Garbarg-Chenon, A., Vu Thien, H., Labia, R., Ben-Yaghlane, H., Godard, V., Deny, P., Bricout, F. & Nicolas, J.C. 1989, "Characterization of a plasmid coding for resistance to broad-spectrum cephalosporins in *Salmonella*

typhimurium", *Drugs under experimental and clinical research,* vol. 15, no. 4, pp. 145-150.

67. Garcia-Fernandez, A., Chiaretto, G., Bertini, A., Villa, L., Fortini, D., Ricci, A. & Carattoli, A. 2008, "Multilocus sequence typing of IncI1 plasmids carrying extended-spectrum beta-lactamases in *Escherichia coli* and *Salmonella* of human and animal origin", *The Journal of antimicrobial chemotherapy,* vol. 61, no. 6, pp. 1229-1233.

68. Garcillan-Barcia, M.P. & de la Cruz, F. 2008, "Why is entry exclusion an essential feature of conjugative plasmids?", *Plasmid,* vol. 60, no. 1, pp. 1-18.

69. Garcillan-Barcia, M.P., Francia, M.V. & de la Cruz, F. 2009, "The diversity of conjugative relaxases and its application in plasmid classification", *FEMS microbiology reviews,* vol. 33, no. 3, pp. 657-687.

70. Garcillan-Barcia, M.P., Jurado, P., Gonzalez-Perez, B., Moncalian, G., Fernandez, L.A. & de la Cruz, F. 2007, "Conjugative transfer can be inhibited by blocking relaxase activity within recipient cells with intrabodies", *Molecular microbiology,* vol. 63, no. 2, pp. 404-416.

71. Gentleman, R. & Ihaka, R. 1997, "The R Project for Statistical Computing", *Available at http://www.r-project.org.,* .

72. Gerdes, K., Rasmussen, P.B. & Molin, S. 1986, "Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 83, no. 10, pp. 3116-3120.

73. Gerdes, K., Christensen, S.K. & Lobner-Olesen, A. 2005, "Prokaryotic toxin-antitoxin stress response loci", *Nature Reviews Microbiology,* vol. 3, no. 5, pp. 371-382.

74. Giraldo-Suarez, R., Fernandez-Tresguerres, E., Diaz-Orejas, R., Malki, A. & Kohiyama, M. 1993, "The heat-shock DnaK protein is required for plasmid R1 replication and it is dispensable for plasmid ColE1 replication", *Nucleic acids research,* vol. 21, no. 23, pp. 5495-5499.

75. Hall, R.M. & Stokes, H.W. 1993, "Integrons: novel DNA elements which capture genes by site-specific recombination", *Genetica,* vol. 90, no. 2-3, pp. 115-132.

76. Hampton, M.D., Ward, L.R., Rowe, B. & Threlfall, E.J. 1998, "Molecular fingerprinting of multidrug-resistant *Salmonella enterica* serotype Typhi", *Emerg Infect Dis,* vol. 4, no. 2, pp. 317-20.

77. Haniford, D.B. 2006, "Transpososome dynamics and regulation in Tn*10* transposition", *Critical reviews in biochemistry and molecular biology,* vol. 41, no. 6, pp. 407-424.

78. Haring, V., Scholz, P., Scherzinger, E., Frey, J., Derbyshire, K., Hatfull, G., Willetts, N.S. & Bagdasarian, M. 1985, "Protein RepC is involved in copy number control of the broad host range plasmid RSF1010", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 82, no. 18, pp. 6090-6094.

79. Hasan, R., Zafar, A., Abbas, Z., Mahraj, V., Malik, F. & Zaidi, A. 2008, "Antibiotic resistance among *Salmonella enterica* serovars Typhi and Paratyphi A in Pakistan (2001-2006)", *Journal of infection in developing countries,* vol. 2, no. 4, pp. 289-294.

80. Hasan, R., Cooke, F.J., Nair, S., Harish, B.N. & Wain, J. 2005, "Typhoid and paratyphoid fever", *Lancet,* vol. 366, no. 9497, pp. 1603-4.

81. Hasan, Z., Rahman, K.M., Alam, M.N., Afroza, A., Asna, Z.H., Ghosh, P.K. & Alam, N. 1995, "Role of a large plasmid in mediation of multiple drug resistance in *Salmonella* typhi and paratyphi A in Bangladesh", *Bangladesh Medical Research Council bulletin,* vol. 21, no. 1, pp. 50-54.

82. Hayes, W. (ed) 1964, *The Genetics of Bacteria and their Viruses*, 2nd edn, Blackwell.

83. Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E. & Holden, D.W. 1995, "Simultaneous identification of bacterial virulence genes by negative selection", *Science (New York, N.Y.),* vol. 269, no. 5222, pp. 400-403.

84. Herrero, M., de Lorenzo, V. & Timmis, K.N. 1990, "Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria", *Journal of Bacteriology,* vol. 172, no. 11, pp. 6557-6567.

85. Hoischen, C., Bussiek, M., Langowski, J. & Diekmann, S. 2008, "*Escherichia coli* low-copy-number plasmid R1 centromere *parC* forms a U-shaped complex with its binding protein ParR", *Nucleic acids research,* vol. 36, no. 2, pp. 607-615.

86. Holt, K.E., Parkhill, J., Mazzoni, C.J., Roumagnac, P., Weill, F.X., Goodhead, I., Rance, R., Baker, S., Maskell, D.J., Wain, J., Dolecek, C., Achtman, M. & Dougan, G. 2008, "High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi", *Nature genetics,* vol. 40, no. 8, pp. 987-993.

87. Holt, K.E., Thomson, N.R., Wain, J., Phan, M.D., Nair, S., Hasan, R., Bhutta, Z.A., Quail, M.A., Norbertczak, H., Walker, D., Dougan, G. & Parkhill, J. 2007, "Multidrug-resistant *Salmonella enterica* serovar paratyphi A harbors IncHI1 plasmids similar to those found in serovar typhi", *Journal of Bacteriology,* vol. 189, no. 11, pp. 4257-4264.

88. Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M. & Rupp, R. 2007, "Dendroscope: An interactive viewer for large phylogenetic trees", *BMC bioinformatics,* vol. 8, pp. 460.

89. Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O. & Venter, J.C. 1999, "Global transposon mutagenesis and a minimal *Mycoplasma* genome", *Science (New York, N.Y.),* vol. 286, no. 5447, pp. 2165-2169.

90. Islam, A., Butler, T., Nath, S.K., Alam, N.H., Stoeckel, K., Houser, H.B. & Smith, A.L. 1988, "Randomized treatment of patients with typhoid fever by using ceftriaxone or chloramphenicol", *The Journal of infectious diseases,* vol. 158, no. 4, pp. 742-747.

91. Jaques, P.H. & Van der Heyde, H. 1972, "Combination of trimethoprim and sulphamethoxazole in typhoid fever", *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde,* vol. 46, no. 11, pp. 281-284.

92. Ji, Y., Zhang, B., Van, S.F., Horn, Warren, P., Woodnutt, G., Burnham, M.K. & Rosenberg, M. 2001, "Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA", *Science (New York, N.Y.),* vol. 293, no. 5538, pp. 2266-2269.

93. Jolley, K.A., Chan, M.S. & Maiden, M.C. 2004, "mlstdbNet - distributed multi-locus sequence typing (MLST) databases", *BMC bioinformatics,* vol. 5, pp. 86.

94. Joshi, S. & Amarnath, S.K. 2007, "Fluoroquinolone resistance in *Salmonella* typhi and S. paratyphi A in Bangalore, India", *Trans R Soc Trop Med Hyg,* vol. 101, no. 3, pp. 308-10.

95. Kamili, M.A., Ali, G., Shah, M.Y., Rashid, S., Khan, S. & Allaqaband, G.Q. 1993, "Multiple drug resistant typhoid fever outbreak in Kashmir Valley", *Indian J Med Sci,* vol. 47, no. 6, pp. 147-51.

96. Kapil, A., Sood, S., Reddaiah, V.P., Das, B. & Seth, P. 1997, "Paratyphoid fever due to *Salmonella enterica* serotype Paratyphi A", *Emerging Infectious Diseases,* vol. 3, no. 3, pp. 407.

97. Karzai, A.W., Susskind, M.M. & Sauer, R.T. 1999, "SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA)", *The EMBO journal,* vol. 18, no. 13, pp. 3793-3799.

98. Kauffmann, F. 1950, *The diagnosis of Salmonella types,* Springfield, Ill.

99. Kidgell, C., Reichard, U., Wain, J., Linz, B., Torpdahl, M., Dougan, G. & Achtman, M. 2002, "*Salmonella* typhi, the causative agent of typhoid fever, is approximately 50,000 years old", *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases,* vol. 2, no. 1, pp. 39-45.

100. Kim, Y.J., Lin, L.S. & Meyer, R.J. 1987, "Two domains at the origin are required for replication and maintenance of broad-host-range plasmid R1162", *Journal of Bacteriology,* vol. 169, no. 12, pp. 5870-5872.

101. Kim, Y.J. & Meyer, R.J. 1991, "An essential iteron-binding protein required for plasmid R1162 replication induces localized melting within the origin at a specific site in AT-rich DNA", *Journal of Bacteriology,* vol. 173, no. 17, pp. 5539-5545.

102. Knuth, K., Niesalla, H., Hueck, C.J. & Fuchs, T.M. 2004, "Large-scale identification of essential *Salmonella* genes by trapping lethal insertions", *Molecular microbiology,* vol. 51, no. 6, pp. 1729-1744.

103. Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S.C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L.C., Danchin, A., Debarbouille, M., Dervyn, E., Deuerling, E., Devine, K., Devine, S.K., Dreesen, O., Errington, J., Fillinger, S., Foster, S.J., Fujita, Y., Galizzi, A., Gardan, R., Eschevins, C., Fukushima, T., Haga, K., Harwood, C.R., Hecker, M., Hosoya, D., Hullo, M.F., Kakeshita, H., Karamata, D., Kasahara, Y., Kawamura, F., Koga, K., Koski, P., Kuwana, R., Imamura, D., Ishimaru, M., Ishikawa, S.,

Ishio, I., Le Coq, D., Masson, A., Mauel, C., Meima, R., Mellado, R.P., Moir, A., Moriya, S., Nagakawa, E., Nanamiya, H., Nakai, S., Nygaard, P., Ogura, M., Ohanan, T., O'Reilly, M., O'Rourke, M., Pragai, Z., Pooley, H.M., Rapoport, G., Rawlins, J.P., Rivas, L.A., Rivolta, C., Sadaie, A., Sadaie, Y., Sarvas, M., Sato, T., Saxild, H.H., Scanlan, E., Schumann, W., Seegers, J.F., Sekiguchi, J., Sekowska, A., Seror, S.J., Simon, M., Stragier, P., Studer, R., Takamatsu, H., Tanaka, T., Takeuchi, M., Thomaides, H.B., Vagner, V., van Dijl, J.M., Watabe, K., Wipat, A., Yamamoto, H., Yamamoto, M., Yamamoto, Y., Yamane, K., Yata, K., Yoshida, K., Yoshikawa, H., Zuber, U. & Ogasawara, N. 2003, "Essential Bacillus subtilis genes", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 100, no. 8, pp. 4678-4683.

104. Koepsel, R.R., Murray, R.W., Rosenblum, W.D. & Khan, S.A. 1985, "Purification of pT181-encoded repC protein required for the initiation of plasmid replication", *The Journal of biological chemistry,* vol. 260, no. 14, pp. 8571-8577.

105. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S.L. 2004, "Versatile and open software for comparing large genomes", *Genome biology,* vol. 5, no. 2, pp. R12.

106. Langridge, G., Wain, J. & Nair, S. 2009, "Invasive salmonellosis in humans" in *EcoSal - Escherichia coli and Salmonella: Cellular and Molecular Biology.*, eds. A. Böck, R. Curtiss III, J.B. Kaper, et al, ASM Press, Washington, DC., .

107. Lawley, T.D., Gilmour, M.W., Gunton, J.E., Standeven, L.J. & Taylor, D.E. 2002, "Functional and mutational analysis of conjugative transfer region 1 (Tra1) from the IncHI1 plasmid R27", *J. Bacteriol.,* vol. 184, no. 8, pp. 2173-2180.

108. Lawley, T.D., Gilmour, M.W., Gunton, J.E., Tracz, D.M. & Taylor, D.E. 2003, "Functional and mutational analysis of conjugative transfer region 2 (Tra2) from the IncHI1 plasmid R27", *J. Bacteriol.,* vol. 185, no. 2, pp. 581-591.

109. Lawley, T.D. & Taylor, D.E. 2003, "Characterization of the double-partitioning modules of R27: correlating plasmid stability with plasmid localization", *J. Bacteriol.,* vol. 185, no. 10, pp. 3060-3067.

110. Le, T.A., Fabre, L., Roumagnac, P., Grimont, P.A., Scavizzi, M.R. & Weill, F.X. 2007, "Clonal expansion and microevolution of quinolone-resistant *Salmonella enterica* serotype Typhi in Vietnam from 1996 to 2004", *Journal of clinical microbiology,* vol. 45, no. 11, pp. 3485-3492.

111. Le, T.A., Lejay-Collin, M., Grimont, P.A., Hoang, T.L., Nguyen, T.V., Grimont, F. & Scavizzi, M.R. 2004, "Endemic, epidemic clone of *Salmonella enterica* serovar Typhi harboring a single multidrug-resistant plasmid in Vietnam between 1995 and 2002", *J Clin Microbiol,* vol. 42, no. 7, pp. 3094-9.

112. Lederberg, J. 1952, "Cell genetics and hereditary symbiosis", *Physiological Reviews,* vol. 32, pp. 403-430.

113. Lederberg, J., Cavalli, L.L. & Lederberg, E.M. 1952, "Sex compatibility in *Escherichia coli*", *Genetics,* vol. 37, no. 6, pp. 720-730.

114. Lenski, R.E., Simpson, S.C. & Nguyen, T.T. 1994, "Genetic analysis of a plasmid-encoded, host genotype-specific enhancement of bacterial fitness", *J. Bacteriol.,* vol. 176, no. 11, pp. 3140-3147.

115. Li, H., Ruan, J. & Durbin, R. 2008, "Mapping short DNA sequencing reads and calling variants using mapping quality scores", *Genome research,* vol. 18, no. 11, pp. 1851-1858.

116. Liebert, C.A., Hall, R.M. & Summers, A.O. 1999, "Transposon Tn*21*, flagship of the floating genome", *Microbiology and molecular biology reviews : MMBR,* vol. 63, no. 3, pp. 507-522.

117. Ling, J. & Chau, P.Y. 1984, "Plasmids mediating resistance to chloramphenicol, trimethoprim, and ampicillin in *Salmonella* typhi strains isolated in the Southeast Asian region", *J Infect Dis,* vol. 149, no. 4, pp. 652.

118. Liu, Y., Lee, M.A., Ooi, E.E., Mavis, Y., Tan, A.L. & Quek, H.H. 2003, "Molecular typing of *Salmonella enterica* serovar Typhi isolates from various countries in Asia by a multiplex PCR assay on variable-number tandem repeats", *Journal of clinical microbiology,* vol. 41, no. 9, pp. 4388-4394.

119. Llosa, M., Gomis-Ruth, F.X., Coll, M. & Cruz, F.d.l. 2002, "Bacterial conjugation: a two-step mechanism for DNA transport", *Molecular Microbiology,* vol. 45, no. 1, pp. 1-8.

120. Lujan, S.A., Guogas, L.M., Ragonese, H., Matson, S.W. & Redinbo, M.R. 2007, "Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 104, no. 30, pp. 12282-12287.

121. Ly, A., Henderson, J., Lu, A., Culham, D.E. & Wood, J.M. 2004, "Osmoregulatory systems of *Escherichia coli*: identification of betaine-carnitine-choline transporter family member BetU and distributions of BetU and TrkG among pathogenic and nonpathogenic isolates", *J. Bacteriol.,* vol. 186, no. 2, pp. 296-306.

122. Lynch, A.S. & Wang, J.C. 1994, "Use of an inducible site-specific recombinase to probe the structure of protein-DNA complexes involved in F plasmid partition in *Escherichia coli*", *Journal of Molecular Biology,* vol. 236, no. 3, pp. 679-684.

123. MADDOCK, C.R. 1962, "Ampicillin in the treatment of typhoid fever", *Lancet,* vol. 1, no. 7235, pp. 918.

124. Maher, D., Sherburne, R. & Taylor, D.E. 1993, "H-pilus assembly kinetics determined by electron microscopy", *J. Bacteriol.,* vol. 175, no. 8, pp. 2175-2183.

125. Maher, D. & Taylor, D.E. 1993, "Host range and transfer efficiency of incompatibility group HI plasmids", *Can J Microbiol,* vol. 39, no. 6, pp. 581-7.

126. Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. & Spratt, B.G. 1998, "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms", *Proc Natl Acad Sci U S A,* vol. 95, no. 6, pp. 3140-5.

127. Mandal, S., Mandal, M.D. & Pal, N.K. 2004, "Reduced minimum inhibitory concentration of chloramphenicol for *Salmonella enterica* serovar Typhi", *Indian J Med Sci,* vol. 58, no. 1, pp. 16-23.

128. Marmur, J., Rownd, R., Falkow, S., Baron, L.S., Schildkraut, C. & Doty, P. 1961, "The nature of intergeneric episomal infection", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 47, pp. 972-979.

129. Maskey, A.P., Day, J.N., Tuan, P.Q., Thwaites, G.E., Campbell, J.I., Zimmerman, M., Farrar, J.J. & Basnyat, B. 2006, "*Salmonella enterica* serovar Paratyphi A and *S. enterica* serovar Typhi cause indistinguishable clinical syndromes in Kathmandu, Nepal", *Clin Infect Dis,* vol. 42, no. 9, pp. 1247-1253.

130. Mazel, D. 2006, "Integrons: agents of bacterial evolution", *Nat Rev Micro,* vol. 4, no. 8, pp. 608-620.

131. McCormick, M., Wishart, W., Ohtsubo, H., Heffron, F. & Ohtsubo, E. 1981, "Plasmid cointegrates and their resolution mediated by transposon Tn3 mutants", *Gene,* vol. 15, no. 2-3, pp. 103-118.

132. Mercer, A.A., Morelli, G., Heuzenroeder, M., Kamke, M. & Achtman, M. 1984, "Conservation of plasmids among *Escherichia coli* K1 isolates of diverse origins", *Infection and immunity,* vol. 46, no. 3, pp. 649-657.

133. Mermin, J.H., Villar, R., Carpenter, J., Roberts, L., Samaridden, A., Gasanova, L., Lomakina, S., Bopp, C., Hutwagner, L., Mead, P., Ross, B. & Mintz, E.D. 1999, "A massive epidemic of multidrug-resistant typhoid fever in Tajikistan associated with consumption of municipal water", *The Journal of infectious diseases,* vol. 179, no. 6, pp. 1416-1422.

134. Meyer, R. 2009, "Replication and conjugative mobilization of broad host-range IncQ plasmids", *Plasmid,* vol. 62, no. 2, pp. 57-70.

135. Meynell, E. & Datta, N. 1966, "The relation of resistance transfer factors to the F-factor (sex-factor) of *Escherichia coli* K12", *Genetical research,* vol. 7, no. 1, pp. 134-140.

136. Mikhail, I.A., Haberberger, R.L., Farid, Z., Girgis, N.I. & Woody, J.N. 1989, "Antibiotic-multiresistant *Salmonella* typhi in Egypt", *Trans R Soc Trop Med Hyg,* vol. 83, no. 1, pp. 120.

137. Mizusaki, H., Takaya, A., Yamamoto, T. & Aizawa, S. 2008, "Signal pathway in salt-activated expression of the Salmonella pathogenicity island 1 type III secretion system in *Salmonella enterica* serovar Typhimurium", *Journal of Bacteriology,* vol. 190, no. 13, pp. 4624-4631.

138. Mohanty, S., Renuka, K., Sood, S., Das, B.K. & Kapil, A. 2006, "Antibiogram pattern and seasonality of *Salmonella* serotypes in a North Indian tertiary care hospital", *Epidemiology and Infection,* vol. 134, no. 5, pp. 961-966.

139. Moller-Jensen, J. & Gerdes, K. 2007, "Plasmid segregation: spatial awareness at the molecular level", *The Journal of cell biology,* vol. 179, no. 5, pp. 813-815.

140. Monod, J. 1949, "The growth of bacterial cultures", *Annual Review of Microbiology,* vol. 3, no. 1, pp. 371-394.

141. Mortimer, C.K., Peters, T.M., Gharbia, S.E., Logan, J.M. & Arnold, C. 2004, "Towards the development of a DNA-sequence based approach to serotyping of *Salmonella enterica*", *BMC microbiology,* vol. 4, pp. 31.

142. Moyed, H.S. & Bertrand, K.P. 1983, "Mutations in multicopy Tn*10 tet* plasmids that confer resistance to inhibitory effects of inducers of *tet* gene expression", *Journal of Bacteriology,* vol. 155, no. 2, pp. 557-564.

143. Nair, S., Schreiber, E., Thong, K.L., Pang, T. & Altwegg, M. 2000, "Genotypic characterization of *Salmonella* typhi by amplified fragment length polymorphism fingerprinting provides increased discrimination as compared to pulsed-field gel electrophoresis and ribotyping", *Journal of microbiological methods,* vol. 41, no. 1, pp. 35-43.

144. Navarro, F., Llovet, T., Echeita, M., Coll, P., Aladuena, A., Usera, M. & Prats, G. 1996, "Molecular typing of *Salmonella enterica* serovar Typhi", *Journal of clinical microbiology,* vol. 34, no. 11, pp. 2831-2834.

145. Newnham, P.J. & Taylor, D.E. 1994, "Molecular analysis of RepHI1A, a minimal replicon of the IncHI1 plasmid R27", *Molecular microbiology,* vol. 11, no. 4, pp. 757-768.

146. Nollmann, M., Byron, O. & Stark, W.M. 2005, "Behavior of Tn*3* resolvase in solution and its interaction with res", *Biophysical journal,* vol. 89, no. 3, pp. 1920-1931.

147. Nordstrom, K. & Austin, S.J. 1989, "Mechanisms that contribute to the stable segregation of plasmids", *Annual Review of Genetics,* vol. 23, pp. 37-69.

148. Novick, R.P. 1987, "Plasmid incompatibility", *Microbiol Rev,* vol. 51, no. 4, pp. 381-95.

149. Novick, R.P., Clowes, R.C., Cohen, S.N., Curtiss, R.,3rd, Datta, N. & Falkow, S. 1976, "Uniform nomenclature for bacterial plasmids: a proposal", *Bacteriological Reviews,* vol. 40, no. 1, pp. 168-189.

150. Ochiai, R.L., Wang, X., von Seidlein, L., Yang, J., Bhutta, Z.A., Bhattacharya, S.K., Agtini, M., Deen, J.L., Wain, J., Kim, D.R., Ali, M., Acosta, C.J., Jodar, L. & Clemens, J.D. 2005, "*Salmonella* paratyphi A rates, Asia", *Emerg Infect Dis,* vol. 11, no. 11, pp. 1764-6.

151. Octavia, S. & Lan, R. 2009, "Multiple-locus variable-number tandem-repeat analysis of *Salmonella enterica* serovar Typhi", *Journal of clinical microbiology,* vol. 47, no. 8, pp. 2369-2376.

152. Octavia, S. & Lan, R. 2007, "Single-nucleotide-polymorphism typing and genetic relationships of *Salmonella enterica* serovar Typhi isolates", *Journal of clinical microbiology,* vol. 45, no. 11, pp. 3795-3801.

153. Okan, N.A., Bliska, J.B. & Karzai, A.W. 2006, "A Role for the SmpB-SsrA system in *Yersinia pseudotuberculosis* pathogenesis", *PLoS pathogens,* vol. 2, no. 1, pp. e6.

154. Okusu, H., Ma, D. & Nikaido, H. 1996, "AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants", *Journal of Bacteriology,* vol. 178, no. 1, pp. 306-308.

155. Olarte, J. & Galindo, E. 1973, "*Salmonella* typhi resistant to chloramphenicol, ampicillin, and other antimicrobial agents: atrains isolated during an extensive typhoid fever epidemic in Mexico", *Antimicrobial Agents and Chemotherapy,* vol. 4, no. 6, pp. 597-601.

156. Ono, S., Goldberg, M.D., Olsson, T., Esposito, D., Hinton, J.C. & Ladbury, J.E. 2005, "H-NS is a part of a thermally controlled mechanism for bacterial gene regulation", *The Biochemical journal,* vol. 391, no. Pt 2, pp. 203-213.

157. Paniker, C.K. & Vimala, K.N. 1972, "Transferable chloramphenicol resistance in *Salmonella* typhi", *Nature,* vol. 239, no. 5367, pp. 109-110.

158. Paramasivan, C.N., Subramanian, S. & Shanmugasundaram, N. 1977, "Antimicrobial resistance and incidence of R factor among *Salmonella* isolated from patients with enteric fever and other clinical conditions in Madras, India (1975-1976)", *The Journal of infectious diseases,* vol. 136, no. 6, pp. 796-800.

159. Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T., Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S., Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. & Barrell, B.G. 2001, "Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18", *Nature,* vol. 413, no. 6858, pp. 848-52.

160. Parry, C., Wain, J., Chinh, N.T., Vinh, H. & Farrar, J.J. 1998, "Quinolone-resistant *Salmonella* typhi in Vietnam", *Lancet,* vol. 351, no. 9111, pp. 1289.

161. Parry, C.M. 2004, "Typhoid fever", *Current infectious disease reports,* vol. 6, no. 1, pp. 27-33.

162. Parry, C.M. & Beeching, N.J. 2009, "Treatment of enteric fever", *BMJ (Clinical research ed.),* vol. 338, pp. b1159.

163. Parry, C.M. & Threlfall, E.J. 2008, "Antimicrobial resistance in typhoidal and nontyphoidal salmonellae", *Current opinion in infectious diseases,* vol. 21, no. 5, pp. 531-538.

164. Parry, C.M. 2004, "The treatment of multidrug-resistant and nalidixic acid-resistant typhoid fever in Viet Nam", *Trans R Soc Trop Med Hyg,* vol. 98, no. 7, pp. 413-22.

165. Pfeifer, Y., Matten, J. & Rabsch, W. 2009, "*Salmonella enterica* serovar Typhi with CTX-M β-lactamase, Germany [letter]", *Emerg Infect Dis [serial on the Internet],*

166. Piddock, L.J., White, D.G., Gensberg, K., Pumbwe, L. & Griggs, D.J. 2000, "Evidence for an efflux pump mediating multiple antibiotic resistance in *Salmonella enterica* serovar Typhimurium", *Antimicrobial Agents and Chemotherapy,* vol. 44, no. 11, pp. 3118-3121.

167. Pokharel, B.M., Koirala, J., Dahal, R.K., Mishra, S.K., Khadga, P.K. & Tuladhar, N.R. 2006, "Multidrug-resistant and extended-spectrum beta-lactamase (ESBL)-producing *Salmonella enterica* (serotypes Typhi and Paratyphi A) from blood isolates in Nepal: surveillance of resistance and a

search for newer alternatives", *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases,* vol. 10, no. 6, pp. 434-438.

168. Pontali, E., Feasi, M., Usiglio, D., Mori, M. & Cassola, G. 2008, "Imported typhoid fever with hepatitis from Bangladesh: a case of delayed response to ceftriaxone?", *Journal of travel medicine,* vol. 15, no. 5, pp. 366-368.

169. Popoff, M.Y., Bockemuhl, J. & Gheesling, L.L. 2004, "Supplement 2002 (no. 46) to the Kauffmann-White scheme", *Research in microbiology,* vol. 155, no. 7, pp. 568-570.

170. Potrykus, J., Baranska, S. & Wegrzyn, G. 2002, "Inactivation of the *acrA* gene is partially responsible for chloramphenicol sensitivity of *Escherichia coli* CM2555 strain expressing the chloramphenicol acetyltransferase gene", *Microbial drug resistance (Larchmont, N.Y.),* vol. 8, no. 3, pp. 179-185.

171. Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H. & Turner, D.J. 2008, "A large genome center's improvements to the Illumina sequencing system", *Nature methods,* vol. 5, no. 12, pp. 1005-1010.

172. RAMSEY, C.H. & EDWARDS, P.R. 1961, "Resistance of *Salmonellae* isolated in 1959 and 1960 to tetracyclines and chloramphenicol", *Applied Microbiology,* vol. 9, pp. 389-391.

173. Rangnekar, V.M., Banker, D.D. & Jhala, H.I. 1983, "Antimicrobial resistance and R-plasmids of *Salmonella* paratyphi A isolated at Bombay", *The Indian journal of medical research,* vol. 77, pp. 5-9.

174. Rasooly, A., Wang, P.Z. & Novick, R.P. 1994, "Replication-specific conversion of the *Staphylococcus aureus* pT181 initiator protein from an active homodimer to an inactive heterodimer", *The EMBO journal,* vol. 13, no. 21, pp. 5245-5251.

175. Roberts, A.P., Chandler, M., Courvalin, P., Guedon, G., Mullany, P., Pembroke, T., Rood, J.I., Smith, C.J., Summers, A.O., Tsuda, M. & Berg, D.E. 2008, "Revised nomenclature for transposable genetic elements", *Plasmid,* vol. 60, no. 3, pp. 167-173.

176. Rotimi, V.O., Jamal, W., Pal, T., Sovenned, A. & Albert, M.J. 2008, "Emergence of CTX-M-15 type extended-spectrum beta-lactamase-producing *Salmonella* spp. in Kuwait and the United Arab Emirates", *Journal of medical microbiology,* vol. 57, no. Pt 7, pp. 881-886.

177. Roumagnac, P., Weill, F.X., Dolecek, C., Baker, S., Brisse, S., Chinh, N.T., Le, T.A., Acosta, C.J., Farrar, J., Dougan, G. & Achtman, M. 2006, "Evolutionary history of *Salmonella* typhi", *Science (New York, N.Y.),* vol. 314, no. 5803, pp. 1301-1304.

178. Rouquette, C., Serre, M.C. & Lane, D. 2004, "Protective role for H-NS protein in IS*1* transposition", *Journal of Bacteriology,* vol. 186, no. 7, pp. 2091-2098.

179. Roussel, A.F. & Chabbert, Y.A. 1978, "Taxonomy and epidemiology of gram-negative bacterial plasmids studied by DNA-DNA filter hybridization in formamide", *J Gen Microbiol,* vol. 104, no. 2, pp. 269-76.

180. Rowe, B., Ward, L.R. & Threlfall, E.J. 1997, "Multidrug-resistant *Salmonella* typhi: a worldwide epidemic", *Clinical infectious diseases: an official*

publication of the Infectious Diseases Society of America, vol. 24 Suppl 1, pp. S106-9.

181. Rozen, S. & Skaletsky, H. 2000, "Primer3 on the WWW for general users and for biologist programmers", *Methods Mol Biol,* vol. 132, pp. 365-86.

182. Saavedra De Bast, M., Mine, N. & Van Melderen, L. 2008, "Chromosomal toxin-antitoxin systems may act as antiaddiction modules", *Journal of Bacteriology,* vol. 190, no. 13, pp. 4603-4609.

183. Sahu, S.N., Acharya, S., Tuminaro, H., Patel, I., Dudley, K., LeClerc, J.E., Cebula, T.A. & Mukhopadhyay, S. 2003, "The bacterial adaptive response gene, *barA*, encodes a novel conserved histidine kinase regulatory switch for adaptation and modulation of metabolism in *Escherichia coli*", *Molecular and cellular biochemistry,* vol. 253, no. 1-2, pp. 167-177.

184. Sakai, H. & Komano, T. 1996, "DNA replication of IncQ broad-host-range plasmids in gram-negative bacteria", *Bioscience, biotechnology, and biochemistry,* vol. 60, no. 3, pp. 377-382.

185. Salama, N.R., Shepherd, B. & Falkow, S. 2004, "Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*", *Journal of Bacteriology,* vol. 186, no. 23, pp. 7926-7935.

186. Salyers, A.A., Shoemaker, N.B., Stevens, A.M. & Li, L.Y. 1995, "Conjugative transposons: an unusual and diverse set of integrated gene transfer elements", *Microbiological reviews,* vol. 59, no. 4, pp. 579-590.

187. Sassetti, C.M., Boyd, D.H. & Rubin, E.J. 2003, "Genes required for mycobacterial growth defined by high density mutagenesis", *Molecular microbiology,* vol. 48, no. 1, pp. 77-84.

188. Sassetti, C.M., Boyd, D.H. & Rubin, E.J. 2001, "Comprehensive identification of conditionally essential genes in mycobacteria", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 98, no. 22, pp. 12712-12717.

189. Saul, D., Lane, D. & Bergquist, P.L. 1988, "A replication region of the IncHI plasmid, R27, is highly homologous with the RepFIA replicon of F", *Molecular microbiology,* vol. 2, no. 2, pp. 219-225.

190. Schumacher, M.A. 2008, "Structural biology of plasmid partition: uncovering the molecular mechanisms of DNA segregation", *The Biochemical journal,* vol. 412, no. 1, pp. 1-18.

191. Selander, R.K., Beltran, P., Smith, N.H., Helmuth, R., Rubin, F.A., Kopecko, D.J., Ferris, K., Tall, B.D., Cravioto, A. & Musser, J.M. 1990, "Evolutionary genetic relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers.", *Infection and immunity,* vol. 58, no. 7, pp. 2262-2275.

192. Selander, R.K., Musser, J.M., Caugant, D.A., Gilmour, M.N. & Whittam, T.S. 1987, "Population genetics of pathogenic bacteria", *Microbial pathogenesis,* vol. 3, no. 1, pp. 1-7.

193. Selby, C.P. & Sancar, A. 1990, "Structure and function of the (A)BC excinuclease of *Escherichia coli*", *Mutation research,* vol. 236, no. 2-3, pp. 203-211.

194. Shangkuan, Y.H. & Lin, H.C. 1998, "Application of random amplified polymorphic DNA analysis to differentiate strains of *Salmonella* typhi and other *Salmonella* species", *Journal of applied microbiology,* vol. 85, no. 4, pp. 693-702.

195. Sherburne, C.K., Lawley, T.D., Gilmour, M.W., Blattner, F.R., Burland, V., Grotbeck, E., Rose, D.J. & Taylor, D.E. 2000, "The complete DNA sequence and analysis of R27, a large IncHI plasmid from *Salmonella* typhi that is temperature sensitive for transfer", *Nucleic Acids Res,* vol. 28, no. 10, pp. 2177-86.

196. Smith, H.W. 1974, "Letter: Thermosensitive transfer factors in chloramphenicol-resistant strains of *Salmonella* typhi", *Lancet,* vol. 2, no. 7875, pp. 281-282.

197. Smith, H.W., Parsell, Z. & Green, P. 1978, "Thermosensitive H1 plasmids determining citrate utilization", *J Gen Microbiol,* vol. 109, no. 2, pp. 305-11.

198. Sorensen, S.J., Bailey, M., Hansen, L.H., Kroer, N. & Wuertz, S. 2005, "Studying plasmid horizontal transfer in situ: a critical review", *Nature reviews.Microbiology,* vol. 3, no. 9, pp. 700-710.

199. Spratt, B.G., Hanage, W.P., Li, B., Aanensen, D.M. & Feil, E.J. 2004, "Displaying the relatedness among isolates of bacterial species -- the eBURST approach", *FEMS microbiology letters,* vol. 241, no. 2, pp. 129-134.

200. Stamatakis, A. 2006, "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models", *Bioinformatics (Oxford, England),* vol. 22, no. 21, pp. 2688-2690.

201. Stamatakis, A., Hoover, P. & Rougemont, J. 2008, "A rapid bootstrap algorithm for the RAxML Web servers", *Systematic Biology,* vol. 57, no. 5, pp. 758-771.

202. Sukhnanand, S., Alcaine, S., Warnick, L.D., Su, W.L., Hof, J., Craver, M.P., McDonough, P., Boor, K.J. & Wiedmann, M. 2005, "DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella enterica* serotypes", *Journal of clinical microbiology,* vol. 43, no. 8, pp. 3688-3698.

203. Summers, D.K., Beton, C.W. & Withers, H.L. 1993, "Multicopy plasmid instability: the dimer catastrophe hypothesis", *Mol Microbiol,* vol. 8, no. 6, pp. 1031-8.

204. Tacket, C.O., Sztein, M.B., Losonsky, G.A., Wasserman, S.S., Nataro, J.P., Edelman, R., Pickard, D., Dougan, G., Chatfield, S.N. & Levine, M.M. 1997, "Safety of live oral *Salmonella* typhi vaccine strains with deletions in *htrA* and *aroC aroD* and immune response in humans", *Infection and immunity,* vol. 65, no. 2, pp. 452-456.

205. Taylor, D.E., Chumpitaz, J.C. & Goldstein, F. 1985, "Variability of IncHI1 plasmids from *Salmonella* typhi with special reference to Peruvian plasmids encoding resistance to trimethoprim and other antibiotics", *Antimicrob. Agents Chemother.,* vol. 28, no. 3, pp. 452-455.

206. Taylor, D.E. & Grant, R.B. 1977, "Incompatibility and surface exclusion properties of H1 and H2 plasmids", *J. Bacteriol.,* vol. 131, no. 1, pp. 174-178.

207. Taylor, D.E. & Levine, J.G. 1980, "Studies of temperature-sensitive transfer and maintenance of H incompatibility group plasmids", *J Gen Microbiol,* vol. 116, no. 2, pp. 475-84.

208. Teo, Y.Y., Inouye, M., Small, K.S., Gwilliam, R., Deloukas, P., Kwiatkowski, D.P. & Clark, T.G. 2007, "A genotype calling algorithm for the Illumina BeadArray platform", *Bioinformatics (Oxford, England),* vol. 23, no. 20, pp. 2741-2746.

209. Thomas, C.D., Balson, D.F. & Shaw, W.V. 1990, "In vitro studies of the initiation of staphylococcal plasmid replication. Specificity of RepD for its origin (oriD) and characterization of the Rep-ori tyrosyl ester intermediate", *The Journal of biological chemistry,* vol. 265, no. 10, pp. 5519-5530.

210. Thomas, C.M. 2000, "Paradigms of plasmid organization", *Mol Microbiol,* vol. 37, no. 3, pp. 485-91.

211. Thong, K.L., Cheong, Y.M., Puthucheary, S., Koh, C.L. & Pang, T. 1994, "Epidemiologic analysis of sporadic *Salmonella* typhi isolates and those from outbreaks by pulsed-field gel electrophoresis", *Journal of clinical microbiology,* vol. 32, no. 5, pp. 1135-1141.

212. Threlfall, E.J., Fisher, I.S.T., Berghold, C., Gerner-Smidt, P., Tschape, H., Cormican, M., Luzzi, I., Schnieder, F., Wannet, W., Machado, J. & Edwards, G. 2003, "Trends in antimicrobial drug resistance in *Salmonella enterica* serotypes Typhi and Paratyphi A isolated in Europe, 1999-2001", *International Journal of Antimicrobial Agents,* vol. 22, no. 5, pp. 487-491.

213. Threlfall, E.J. & Frost, J.A. 1990, "The identification, typing and fingerprinting of *Salmonella*: laboratory aspects and epidemiological applications", *The Journal of applied bacteriology,* vol. 68, no. 1, pp. 5-16.

214. Tindall, B.J., Grimont, P.A., Garrity, G.M. & Euzeby, J.P. 2005, "Nomenclature and taxonomy of the genus *Salmonella*", *International Journal of Systematic and Evolutionary Microbiology,* vol. 55, no. Pt 1, pp. 521-524.

215. Trung, N.D., Suthisarnsuntorn, U., Kalambaheti, T., Wonglumsom, W. & Tunyong, W. 2007, "Antimicrobial susceptibility patterns and phage types of *Salmonella* typhi from Vietnam", *The Southeast Asian journal of tropical medicine and public health,* vol. 38, no. 3, pp. 487-492.

216. Turner, A.K., Nair, S. & Wain, J. 2006, "The acquisition of full fluoroquinolone resistance in *Salmonella* Typhi by accumulation of point mutations in the topoisomerase targets", *The Journal of antimicrobial chemotherapy,* vol. 58, no. 4, pp. 733-740.

217. Uwaydah, A.K., Matar, I., Chacko, K.C. & Davidson, J.C. 1991, "The emergence of antimicrobial resistant *Salmonella* typhi in Qatar: epidemiology and therapeutic implications", *Trans R Soc Trop Med Hyg,* vol. 85, no. 6, pp. 790-2.

218. Vongsthongsri, U. & Tharavanij, S. 1980, "Susceptibility of *Salmonella* typhi to chloramphenicol, ampicillin and cotrimoxazole", *The Southeast Asian journal of tropical medicine and public health,* vol. 11, no. 2, pp. 256-261.

219. Wain, J., Hoa, N.T., Chinh, N.T., Vinh, H., Everett, M.J., Diep, T.S., Day, N.P., Solomon, T., White, N.J., Piddock, L.J. & Parry, C.M. 1997, "Quinolone-

resistant *Salmonella* typhi in Viet Nam: molecular basis of resistance and clinical response to treatment", *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America,* vol. 25, no. 6, pp. 1404-1410.

220. Wain, J., Diem Nga, L.T., Kidgell, C., James, K., Fortune, S., Song Diep, T., Ali, T., O Gaora, P., Parry, C., Parkhill, J., Farrar, J., White, N.J. & Dougan, G. 2003, "Molecular analysis of incHI1 antimicrobial resistance plasmids from *Salmonella* serovar Typhi strains associated with typhoid fever", *Antimicrob Agents Chemother,* vol. 47, no. 9, pp. 2732-9.

221. Wain, J., Diep, T.S., Ho, V.A., Walsh, A.M., Hoa, N.T.T., Parry, C.M. & White, N.J. 1998, "Quantitation of Bacteria in Blood of Typhoid Fever Patients and Relationship between Counts and Clinical Features, Transmissibility, and Antibiotic Resistance", *J. Clin. Microbiol.,* vol. 36, no. 6, pp. 1683-1687.

222. Wain, J. & Kidgell, C. 2004, "The emergence of multidrug resistance to antimicrobial agents for the treatment of typhoid fever", *Transactions of the Royal Society of Tropical Medicine and Hygiene,* vol. 98, no. 7, pp. 423-430.

223. Wardle, S.J., O'Carroll, M., Derbyshire, K.M. & Haniford, D.B. 2005, "The global regulator H-NS acts directly on the transpososome to promote Tn*10* transposition", *Genes and Development,* vol. 19, no. 18, pp. 2224.

224. Watanabe, T. & Fukasawa, T. 1961, "Episome-mediated transfer of drug resistance in Enterobacteriaceae. I. Transfer of resistance factors by conjugation", *Journal of Bacteriology,* vol. 81, pp. 669-678.

225. Waters, V.L. & Guiney, D.G. 1993, "Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication", *Molecular microbiology,* vol. 9, no. 6, pp. 1123-1130.

226. White-Ziegler, C.A. & Davis, T.R. 2009, "Genome-wide identification of H-NS-controlled, temperature-regulated genes in *Escherichia coli* K-12", *Journal of Bacteriology,* vol. 191, no. 3, pp. 1106-1110.

227. Woods, C.W., Murdoch, D.R., Zimmerman, M.D., Glover, W.A., Basnyat, B., Wolf, L., Belbase, R.H. & Reller, L.B. 2006, "Emergence of *Salmonella enterica* serotype Paratyphi A as a major cause of enteric fever in Kathmandu, Nepal", *Transactions of the Royal Society of Tropical Medicine and Hygiene,* vol. 100, no. 11, pp. 1063-1067.

228. Woodward, T.E. & Smadel, J.E. 1948, "Preliminary report on the beneficial effect of chloromycetin in the treatment of typhoid fever", *Annals of Internal Medicine,* vol. 29, no. 1, pp. 131-134.

229. Yamaguchi, A., Someya, Y. & Sawai, T. 1992, "Metal-tetracycline/H+ antiporter of *Escherichia coli* encoded by transposon Tn*10*. The role of a conserved sequence motif, GXXXXRXGRR, in a putative cytoplasmic loop between helices 2 and 3", *The Journal of biological chemistry,* vol. 267, no. 27, pp. 19155-19162.

230. Zhang, C.T. & Zhang, R. 2008, "Gene essentiality analysis based on DEG, a database of essential genes", *Methods in molecular biology (Clifton, N.J.),* vol. 416, pp. 391-400.

231. Zhang, L. 1991, "Mechanism of multiresistant *Salmonella* typhi", *Zhonghua Yi Xue Za Zhi,* vol. 71, no. 6, pp. 314-7, 24.

232. Zhang, R. & Lin, Y. 2009, "DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes", *Nucleic acids research,* vol. 37, Database issue, pp. D455-8.

8 Appendices

8.1 Essential genes from day 0

For full list of genes from day 0, 1, 3 and 6, see folder Appendix 1 in the CD.

Total inserts	Total reads	Gene length	Sys.ID	Name	Pvalue essential	Log likelihood ratio	Pvalue of no hit
0	0	920	t0047	*ribF*	1.00	-174.93	6.14E-30
0	0	2816	t0048	*ileS*	1.00	-174.93	3.85E-90
0	0	482	t0049	*lspA*	1.00	-174.93	4.96E-16
0	0	1748	t0126	*ftsI*	1.00	-174.93	3.14E-56
0	0	1469	t0127	*murE*	1.00	-174.93	2.27E-47
0	0	1064	t0129	*mraY*	1.00	-174.93	1.64E-34
0	0	1298	t0130	*murD*	1.00	-174.93	6.11E-42
0	0	1457	t0133	*murC*	1.00	-174.93	5.46E-47
0	0	1244	t0136	*ftsA*	1.00	-174.93	3.16E-40
0	0	806	t0214	*dapD*	1.00	-174.93	2.56E-26
0	0	833	t0218	*tsf*	1.00	-174.93	3.55E-27
0	0	707	t0219	*pyrH*	1.00	-174.93	3.56E-23
0	0	1178	t0221	*dxr*	1.00	-174.93	3.94E-38
0	0	437	t0228	*fabZ*	1.00	-174.93	1.33E-14
0	0	770	t0229	*lpxA*	1.00	-174.93	3.56E-25
0	0	3464	t0232	*dnaE*	1.00	-174.93	1.02E-110
0	0	941	t0233	*accA*	1.00	-174.93	1.32E-30
0	0	1337	t0258	*pssA*	1.00	-174.93	3.53E-43
0	0	362	t0280	*acpS*	1.00	-174.93	3.2E-12
0	0	1100	t0333	*gcpE*	1.00	-174.93	1.18E-35
0	0	1454	t0337	-	1.00	-174.93	6.8E-47
0	0	1109	t0376	*dapE*	1.00	-174.93	6.12E-36
0	0	1997	t0431	*ligA*	1.00	-174.93	3.9E-64
0	0	134	t0535	-	1.00	-174.93	0.0000556
0	0	2267	t0587	*nrdA*	1.00	-174.93	1.04E-72
0	0	650	t0662	*folE*	1.00	-174.93	2.3E-21
0	0	530	t0931	*pgsA*	1.00	-174.93	1.48E-17
0	0	1715	t0968	*argS*	1.00	-174.93	3.51E-55
0	0	1754	t0976	*aspS*	1.00	-174.93	2.03E-56
0	0	929	t1096	*prsA*	1.00	-174.93	3.18E-30
0	0	833	t1097	*ipk*	1.00	-174.93	3.55E-27
0	0	605	t1098	*hemM*	1.00	-174.93	6.17E-20
0	0	1238	t1099	*hemA*	1.00	-174.93	4.91E-40
0	0	1064	t1100	*prfA*	1.00	-174.93	1.64E-34
0	0	809	t1189	*nadE*	1.00	-174.93	2.05E-26

Total inserts	Total reads	Gene length	Sys.ID	Name	Pvalue essential	Log likelihood ratio	Pvalue of no hit
0	0	1910	t1213	*thrS*	1.00	-174.93	2.26E-61
0	0	524	t1214	*infC*	1.00	-174.93	2.3E-17
0	0	179	t1215	*rpmI*	1.00	-174.93	0.00000207
0	0	338	t1216	*rplT*	1.00	-174.93	1.85E-11
0	0	26	t1217	*pheM*	1.00	-174.93	0.149432419
0	0	965	t1218	*pheS*	1.00	-174.93	2.29E-31
0	0	2369	t1219	*pheT*	1.00	-174.93	6.01E-76
0	0	1256	t1317	*tyrS*	1.00	-174.93	1.32E-40
0	0	212	t1378	-	1.00	-174.93	0.000000186
0	0	1151	t1621	-	1.00	-174.93	2.84E-37
0	0	683	t1702	-	1.00	-174.93	2.06E-22
0	0	623	t1720	*tmk*	1.00	-174.93	1.65E-20
0	0	716	t1725	*fabG*	1.00	-174.93	1.84E-23
0	0	1382	t1934	*asnS*	1.00	-174.93	1.31E-44
0	0	1304	t1942	*mukF*	1.00	-174.93	3.94E-42
0	0	728	t1946	*kdsB*	1.00	-174.93	7.66E-24
0	0	164	t1947	-	1.00	-174.93	0.0000062
0	0	1730	t1950	*msbA*	1.00	-174.93	1.17E-55
0	0	596	t1973	*lolA*	1.00	-174.93	1.19E-19
0	0	200	t1980	*infA*	1.00	-174.93	0.000000446
0	0	2564	t2219	*leuS*	1.00	-174.93	3.87E-82
0	0	1094	t2228	*mrdB*	1.00	-174.93	1.83E-35
0	0	1367	t2324	*cysS*	1.00	-174.93	3.93E-44
0	0	626	t2372	*adk*	1.00	-174.93	1.33E-20
0	0	452	t2446	*ribH*	1.00	-174.93	4.45E-15
0	0	560	t2540	*gmhA*	1.00	-174.93	1.66E-18
0	0	713	t2601	*dnaQ*	1.00	-174.93	2.29E-23
0	0	29	t2621	-	1.00	-174.93	0.120001994
0	0	530	t2630	*yfjA*	1.00	-174.93	1.48E-17
0	0	230	t2631	*rpsP*	1.00	-174.93	4.98E-08
0	0	167	t2727	*csrA*	1.00	-174.93	0.00000498
0	0	2612	t2728	*alaS*	1.00	-174.93	1.16E-83
0	0	461	t2830	*ygbB*	1.00	-174.93	2.3E-15
0	0	857	t2911	*lgt*	1.00	-174.93	6.14E-28
0	0	1080	t2959	*prfB*	1.00	-174.93	5.1E-35
0	0	719	t3094	*plsC*	1.00	-174.93	1.48E-23
0	0	1874	t3102	*parE*	1.00	-174.93	3.14E-60
0	0	344	t3126	*folB*	1.00	-174.93	1.19E-11
0	0	995	t3128	-	1.00	-174.93	2.55E-32
0	0	1727	t3130	*dnaG*	1.00	-174.93	1.46E-55
0	0	209	t3166	-	1.00	-174.93	0.000000231
0	0	1319	t3211	-	1.00	-174.93	1.32E-42
0	0	239	t3220	*rpmA*	1.00	-174.93	2.58E-08
0	0	293	t3221	*rplU*	1.00	-174.93	4.97E-10

Total inserts	Total reads	Gene length	Sys.ID	Name	Pvalue essential	Log likelihood ratio	Pvalue of no hit
0	0	953	t3222	*ispB*	1.00	-174.93	5.5E-31
0	0	707	t3236	-	1.00	-174.93	3.56E-23
0	0	410	t3261	*rplM*	1.00	-174.93	9.59E-14
0	0	1025	t3289	*mreB*	1.00	-174.93	2.84E-33
0	0	452	t3294	*accB*	1.00	-174.93	4.45E-15
0	0	527	t3311	*hemG*	1.00	-174.93	1.85E-17
0	0	1460	t3319	*yigC*	1.00	-174.93	4.38E-47
0	0	737	t3327	*ubiE*	1.00	-174.93	3.97E-24
0	0	1046	t3464	*hemE*	1.00	-174.93	6.13E-34
0	0	347	t3475	*rplL*	1.00	-174.93	9.59E-12
0	0	122	t3680	*rpmH*	1.00	-174.93	0.000133734
0	0	1382	t3681	*dnaA*	1.00	-174.93	1.31E-44
0	0	437	t3787	*dut*	1.00	-174.93	1.33E-14
0	0	1205	t3788	*dfp*	1.00	-174.93	5.48E-39
0	0	218	t3790	*rpmB*	1.00	-174.93	0.00000012
0	0	461	t3793	*coaD*	1.00	-174.93	2.3E-15
0	0	650	t3952	*ftsE*	1.00	-174.93	2.3E-21
0	0	1088	t3981	*asd*	1.00	-174.93	2.84E-35
0	0	356	t4057	*rpsL*	1.00	-174.93	4.97E-12
0	0	452	t4058	*rpsG*	1.00	-174.93	4.45E-15
0	0	293	t4064	*rpsJ*	1.00	-174.93	4.97E-10
0	0	284	t4067	*rplW*	1.00	-174.93	9.6E-10
0	0	803	t4068	*rplB*	1.00	-174.93	3.18E-26
0	0	260	t4069	*rpsS*	1.00	-174.93	5.55E-09
0	0	314	t4070	*rplV*	1.00	-174.93	1.07E-10
0	0	683	t4071	*rpsC*	1.00	-174.93	2.06E-22
0	0	392	t4072	*rplP*	1.00	-174.93	3.57E-13
0	0	173	t4073	*rpmC*	1.00	-174.93	0.00000321
0	0	236	t4074	*rpsQ*	1.00	-174.93	3.21E-08
0	0	353	t4075	*rplN*	1.00	-174.93	6.19E-12
0	0	296	t4076	*rplX*	1.00	-174.93	3.99E-10
0	0	521	t4077	*rplE*	1.00	-174.93	2.87E-17
0	0	287	t4078	*rpsN*	1.00	-174.93	7.71E-10
0	0	374	t4079	*rpsH*	1.00	-174.93	1.33E-12
0	0	515	t4080	*rplF*	1.00	-174.93	4.44E-17
0	0	335	t4081	*rplR*	1.00	-174.93	2.31E-11
0	0	485	t4082	*rpsE*	1.00	-174.93	3.98E-16
0	0	161	t4083	*rpmD*	1.00	-174.93	0.00000773
0	0	416	t4084	*rplO*	1.00	-174.93	6.18E-14
0	0	338	t4087	*rpsM*	1.00	-174.93	1.85E-11
0	0	371	t4088	*rpsK*	1.00	-174.93	1.66E-12
0	0	602	t4089	*rpsD*	1.00	-174.93	7.68E-20
0	0	971	t4090	*rpoA*	1.00	-174.93	1.47E-31
0	0	512	t4161	*ssb*	1.00	-174.93	5.53E-17

Total inserts	Total reads	Gene length	Sys.ID	Name	Pvalue essential	Log likelihood ratio	Pvalue of no hit
0	0	614	t4337	*cI*	1.00	-174.93	3.19E-20
0	0	275	t4381	*groES*	1.00	-174.93	1.85E-09
0	0	527	t4402	*orn*	1.00	-174.93	1.85E-17
0	0	377	t4442	*rpsF*	1.00	-174.93	1.07E-12
0	0	296	t4443	*priB*	1.00	-174.93	3.99E-10
0	0	512	t4468	*ppa*	1.00	-174.93	5.53E-17
0	0	2837	t4510	*valS*	1.00	-174.93	8.3E-91
0	0	1082	t4513	-	1.00	-174.93	4.41E-35
0	0	719	t4586	*dnaC*	1.00	-174.93	1.48E-23
0	0	521	t4587	*dnaT*	1.00	-174.93	2.87E-17
1	2	2396	t3684	*gyrB*	0.83	-44.08	8.35E-77
1	9	2240	t3095	*parC*	0.82	-43.35	7.5E-72
1	1	2180	t3523	*priA*	0.81	-43.06	6.03E-70
1	2	1844	t2441	*dxs*	0.78	-41.23	2.81E-59
1	7	1628	t3678	*yidC*	0.76	-39.87	2.03E-52
1	4	1628	t4382	*groEL*	0.76	-39.87	2.03E-52
1	2	1499	t2958	*lysS*	0.74	-38.97	2.53E-48
3	6	4448	t1940	*mukB*	0.74	-38.85	5.84E-142
1	1	1409	t0160	*lpdA*	0.72	-38.29	1.82E-45
1	1	1397	t4152	*dnaB*	0.72	-38.20	4.39E-45
1	4	1352	t3657	*glmU*	0.71	-37.84	1.18E-43
1	4	1340	t0128	*murF*	0.71	-37.74	2.83E-43
3	6	4010	t3474	*rpoB*	0.71	-37.71	4.72E-128
1	1	1334	t0224	*yaeL*	0.71	-37.69	4.39E-43
1	6	1331	t3295	*accC*	0.71	-37.67	5.47E-43
1	5	1313	t4085	*prlA*	0.71	-37.52	2.04E-42
2	4	2618	t0592	*gyrA*	0.71	-37.48	7.46E-84
1	1	1274	t1971	*serS*	0.70	-37.18	3.53E-41
1	6	1241	t3224	*murA*	0.69	-36.90	3.94E-40
1	1	1241	t3380	*rho*	0.69	-36.90	3.94E-40
1	1	1226	t1701	-	0.69	-36.76	1.18E-39
2	10	2336	t0096	*imp*	0.68	-36.23	6.71E-75
1	1	1133	t0137	*ftsZ*	0.67	-35.89	1.06E-36
1	29	1130	t0230	*lpxB*	0.67	-35.86	1.32E-36
1	2	1112	t0586	*nrdB*	0.66	-35.68	4.92E-36
1	1	1082	t3682	*dnaN*	0.66	-35.38	4.41E-35
1	1	1061	t2987	*fba*	0.65	-35.16	2.05E-34
1	1	1049	t0132	*murG*	0.65	-35.04	4.92E-34
2	6	1964	t3131	*rpoD*	0.63	-34.31	4.35E-63
1	3	977	t1169	*gapA*	0.63	-34.25	9.51E-32
1	1	968	t0429	-	0.62	-34.15	1.84E-31
1	3	959	t1949	*lpxK*	0.62	-34.04	3.55E-31
1	1	956	t2492	*hemB*	0.62	-34.01	4.42E-31
1	2	944	t2371	*hemH*	0.62	-33.87	1.06E-30

Total inserts	Total reads	Gene length	Sys.ID	Name	Pvalue essential	Log likelihood ratio	Pvalue of no hit
1	5	911	t1726	*fabD*	0.61	-33.47	1.19E-29
3	9	2660	t3204	*infB*	0.60	-33.17	3.46E-85
1	11	881	t2440	*ispA*	0.60	-33.10	1.06E-28
1	6	860	t0370	*dapA*	0.59	-32.83	4.93E-28
1	4	860	t2637	*yfjB*	0.59	-32.83	4.93E-28
2	10	1700	t0245	*proS*	0.58	-32.70	1.05E-54
1	30	836	t1104	*kdsA*	0.58	-32.51	2.85E-27
1	10	836	t3954	*rpoH*	0.58	-32.51	2.85E-27
1	1	803	t0066	*dapB*	0.57	-32.06	3.18E-26
3	13	2402	t4141	*plsB*	0.57	-32.03	5.39E-77
3	6	2393	t0225	*yaeT*	0.56	-31.99	1.04E-76
2	12	1520	t2207	*lnt*	0.55	-31.44	5.45E-49
1	9	740	t0222	-	0.54	-31.14	3.19E-24
2	24	1457	t3951	*ftsY*	0.53	-30.97	5.46E-47
1	13	722	t3360	*hemD*	0.53	-30.87	1.19E-23
1	11	719	t2619	-	0.53	-30.82	1.48E-23
1	3	704	t2326	*ybbF*	0.52	-30.58	4.43E-23
1	1	644	t0176	*yadF*	0.49	-29.57	3.56E-21
1	1	638	t1316	*pdxH*	0.49	-29.47	5.52E-21
2	22	1256	t0334	*hisS*	0.48	-29.29	1.32E-40
3	19	1883	t2227	*pbpA*	0.48	-29.28	1.62E-60
1	1	623	t1294	*ribE*	0.48	-29.20	1.65E-20
1	7	614	t3620	-	0.48	-29.03	3.19E-20
1	1	611	t4065	*rplC*	0.47	-28.98	3.98E-20
1	1	602	t0146	*yacE*	0.47	-28.81	7.68E-20
2	15	1196	t0486	*fabB*	0.47	-28.73	1.06E-38
2	8	1190	t2139	*sucB*	0.46	-28.67	1.64E-38
1	2	587	t4066	*rplD*	0.46	-28.52	2.3E-19
2	2	1145	t2988	*pgk*	0.45	-28.23	4.4E-37
1	5	572	t1624	*ribA*	0.45	-28.22	6.88E-19
1	6	572	t2220	*rlpB*	0.45	-28.22	6.88E-19
2	8	1136	t3002	*metK*	0.45	-28.14	8.5E-37
1	12	557	t0270	*rpoE*	0.44	-27.92	2.06E-18
1	3	554	t4102	*yrdC*	0.44	-27.86	2.57E-18
5	18	2768	t3621	*polA*	0.44	-27.85	1.29E-88
2	6	1088	t1686	*trmU*	0.43	-27.65	2.84E-35
3	11	1619	t2854	*pyrG*	0.43	-27.56	3.92E-52
1	1	539	t0220	*frr*	0.43	-27.54	7.68E-18
2	12	1064	t4514	-	0.42	-27.39	1.64E-34
1	6	527	t3479	*nusG*	0.42	-27.29	1.85E-17
4	13	2096	t4059	*fusA*	0.42	-27.22	2.8E-67
4	10	2051	t3864	*glyS*	0.41	-26.97	7.52E-66
1	1	500	t1853	*fabA*	0.40	-26.68	1.33E-16
3	9	1475	t1750	*mviN*	0.40	-26.49	1.46E-47

Total inserts	Total reads	Gene length	Sys.ID	Name	Pvalue essential	Log likelihood ratio	Pvalue of no hit
1	1	491	t4098	*fms*	0.40	-26.47	2.57E-16
2	2	971	t2277	*fepG*	0.39	-26.34	1.47E-31
3	9	1415	t3120	*rfaE*	0.38	-26.01	1.18E-45
4	8	1871	t0159	*aceF*	0.38	-25.91	3.91E-60
2	2	935	t1727	*fabH*	0.38	-25.90	2.05E-30
2	30	929	t4097	*fmt*	0.37	-25.83	3.18E-30
2	6	923	t3359	*hemC*	0.37	-25.75	4.93E-30
4	11	1829	t2456	*secD*	0.37	-25.65	8.42E-59
2	3	914	t3809	*rfaD*	0.37	-25.64	9.52E-30
3	6	1352	t1688	*purB*	0.36	-25.48	1.18E-43
2	7	893	t3863	*glyQ*	0.36	-25.37	4.42E-29
1	1	443	t4408	*yjeE*	0.36	-25.28	8.59E-15
2	3	854	t4140	*ubiA*	0.34	-24.85	7.65E-28
3	117	1259	t3794	*kdtA*	0.34	-24.65	1.06E-40
2	2	839	t0223	*cdsA*	0.34	-24.64	2.29E-27
3	5	1250	t0499	*folC*	0.33	-24.56	2.04E-40
1	9	410	t3478	*rplK*	0.33	-24.37	9.59E-14
2	12	812	t0135	*ftsQ*	0.33	-24.26	1.65E-26
2	10	806	t3350	*dapF*	0.32	-24.17	2.56E-26
1	2	386	t1920	-	0.31	-23.66	5.54E-13
2	7	770	t1613	*fabI*	0.31	-23.63	3.56E-25
1	4	383	t3203	*rbfA*	0.30	-23.57	6.9E-13
5	11	1910	t2376	*dnaX*	0.30	-23.54	2.26E-61
5	10	1898	t0012	*dnaK*	0.30	-23.47	5.43E-61
7	11	2645	t0158	*aceE*	0.30	-23.41	1.04E-84
2	10	749	t2629	*trmD*	0.30	-23.31	1.65E-24
1	1	371	t0205	*yadR*	0.29	-23.20	1.66E-12
1	5	365	t3480	*secE*	0.29	-23.00	2.57E-12
1	2	365	t4091	*rplQ*	0.29	-23.00	2.57E-12
3	26	1085	t2447	*ribD*	0.28	-22.89	3.54E-35
2	21	710	t0588	*ubiG*	0.28	-22.67	2.86E-23
2	16	707	t0217	*rpsB*	0.28	-22.62	3.56E-23
1	2	347	t0125	*ftsL*	0.27	-22.40	9.59E-12
1	6	341	t3679	*rnpA*	0.26	-22.19	1.49E-11
2	20	677	t1057	-	0.26	-22.10	3.19E-22
5	9	1622	t3325	*aarF*	0.25	-21.60	3.15E-52
2	11	635	t3115	*ribB*	0.24	-21.34	6.88E-21
2	26	623	t2222	-	0.23	-21.11	1.65E-20
3	32	932	t0051	*lytB*	0.23	-21.07	2.55E-30
4	24	1223	t3124	*cca*	0.23	-20.88	1.47E-39
3	8	899	t0138	*lpxC*	0.22	-20.64	2.85E-29
3	11	896	t0498	*accD*	0.22	-20.60	3.55E-29
2	6	590	t4143	*lexA*	0.21	-20.45	1.85E-19
1	2	293	t2832	-	0.21	-20.37	4.97E-10

Total inserts	Total reads	Gene length	Sys.ID	Name	Pvalue essential	Log likelihood ratio	Pvalue of no hit
4	27	1157	t2202	*yleB*	0.21	-20.21	1.83E-37
5	14	1397	t0442	*gltX*	0.20	-19.79	4.39E-45
6	80	1649	t2189	*glnS*	0.19	-19.58	4.37E-53
7	60	1916	t3213	*ftsH*	0.19	-19.53	1.46E-61
10	30	2687	t0140	*secA*	0.18	-19.31	4.81E-86
5	35	1343	t2632	*ffh*	0.18	-19.30	2.27E-43
2	9	533	t0289	*yfhC*	0.18	-19.21	1.19E-17
1	1	266	t0630	*rplY*	0.18	-19.18	3.58E-09
4	16	1037	t3953	*ftsX*	0.17	-18.87	1.18E-33
3	11	776	t0216	*map*	0.17	-18.84	2.29E-25
3	5	776	t2910	*thyA*	0.17	-18.84	2.29E-25
2	24	512	t2181	*fldA*	0.17	-18.71	5.53E-17
5	16	1274	t0238	*mesJ*	0.17	-18.65	3.53E-41
4	21	1013	t2221	*holA*	0.17	-18.58	6.84E-33
4	10	1007	t0227	*lpxD*	0.16	-18.50	1.06E-32
1	3	251	t3201	*rpsO*	0.16	-18.47	1.07E-08
4	29	986	t1719	*holB*	0.16	-18.24	4.93E-32
5	12	1226	t0131	*ftsW*	0.16	-18.18	1.18E-39
3	45	722	t3415	-	0.15	-17.94	1.19E-23
4	10	959	t2444	-	0.15	-17.90	3.55E-31
4	27	956	t0275	*lepB*	0.15	-17.86	4.42E-31
4	21	950	t4400	*psd*	0.15	-17.78	6.85E-31
4	23	935	t3807	*waaC*	0.14	-17.58	2.05E-30
4	102	932	t3487	*coaA*	0.14	-17.54	2.55E-30
3	16	692	t2831	*ygbP*	0.14	-17.41	1.07E-22
4	11	887	t0276	*era*	0.13	-16.92	6.85E-29
3	20	659	t1941	*mukE*	0.13	-16.80	1.19E-21
1	8	218	t1724	*acpP*	0.12	-16.70	0.00000012
6	53	1292	t1703	-	0.12	-16.55	9.47E-42
4	24	848	t2321	*folD*	0.12	-16.35	1.19E-27
5	25	1034	t3288	*mreC*	0.11	-16.03	1.47E-33
3	7	608	t3214	*ftsJ*	0.11	-15.77	4.95E-20
2	16	401	t2445	*nusB*	0.10	-15.63	1.85E-13
2	7	398	t3009	-	0.10	-15.54	2.3E-13
14	38	2783	t2140	*sucA*	0.10	-15.52	4.3E-89
1	1	197	t3129	*rpsU*	0.10	-15.41	0.000000556
4	27	785	t0310	*suhB*	0.10	-15.36	1.19E-25
3	10	587	t3326	*yigP*	0.10	-15.32	2.3E-19
12	116	2336	t2410	*lon*	0.10	-15.25	6.71E-75
4	4	776	t2278	*fepC*	0.10	-15.21	2.29E-25
6	14	1160	t2975	*visB*	0.09	-15.17	1.47E-37
4	36	764	t3494	*murI*	0.09	-15.01	5.51E-25
2	4	374	t3260	*rpsI*	0.09	-14.73	1.33E-12
9	28	1655	t1953	*rpsA*	0.08	-14.52	2.82E-53

Total inserts	Total reads	Gene length	Sys.ID	Name	Pvalue essential	Log likelihood ratio	Pvalue of no hit
3	70	551	t3402	*cI*	0.08	-14.50	3.2E-18
6	50	1052	t3265	*degS*	0.07	-13.89	3.95E-34
2	6	338	t4055	*yheM*	0.07	-13.41	1.85E-11
10	16	1622	t2177	*pgm*	0.06	-12.86	3.15E-52
6	55	962	t2966	-	0.06	-12.71	2.85E-31
4	11	641	t2981	*rpiA*	0.06	-12.70	4.43E-21
8	64	1280	t2853	*eno*	0.06	-12.68	2.28E-41
3	12	473	t3287	*mreD*	0.06	-12.48	9.58E-16
2	3	311	t1835	-	0.05	-12.30	1.33E-10
13	29	1973	t2996	*tktA*	0.05	-11.97	2.25E-63
4	6	605	t3778	*gmk*	0.05	-11.92	6.17E-20
7	16	1043	t2730	*recA*	0.05	-11.72	7.63E-34
9	47	1340	t3369	*rffT*	0.05	-11.71	2.83E-43
3	22	440	t0123	*yabB*	0.04	-11.50	1.07E-14
1	7	146	t3157	-	0.04	-11.44	0.0000231
4	23	572	t2636	*grpE*	0.04	-11.16	6.88E-19
7	14	992	t0983	*ruvB*	0.04	-11.03	3.18E-32
5	73	707	t3670	*phoU*	0.04	-11.00	3.56E-23
7	104	953	t2455	*secF*	0.04	-10.48	5.5E-31
1	6	131	t4531	-	0.03	-9.94	0.0000693
16	34	2006	t3384	*rep*	0.03	-9.33	2.02E-64
4	13	500	t3235	-	0.03	-9.28	1.33E-16
8	13	989	t2276	*fepD*	0.02	-9.13	3.96E-32
4	12	491	t0424	*crr*	0.02	-9.03	2.57E-16
35	903	4205	t3473	*rpoC*	0.02	-8.72	3.04E-134
4	16	479	t3476	*rplJ*	0.02	-8.67	6.18E-16
17	85	2015	t0701	*metG*	0.02	-8.53	1.05E-64
8	29	932	t4411	*miaA*	0.02	-8.28	2.55E-30
10	63	1154	t3218	-	0.02	-8.14	2.28E-37
11	32	1262	t0203	*hemL*	0.02	-8.06	8.49E-41
9	13	1028	t3808	*waaF*	0.02	-7.99	2.28E-33
4	26	455	t2205	-	0.02	-7.93	3.57E-15
2	3	227	t3944	*yhhP*	0.02	-7.90	0.000000062
3	12	329	t2628	*rplS*	0.02	-7.40	3.58E-11
11	58	1196	t0313	-	0.02	-7.28	1.06E-38
10	48	1058	t4401	*yjeQ*	0.01	-6.87	2.55E-34
4	23	419	t4600	*holD*	0.01	-6.73	4.96E-14
2	23	209	t4444	*rpsR*	0.01	-6.69	0.000000231
12	54	1235	t0301	*glyA*	0.01	-6.46	6.11E-40
9	70	923	t0124	*yabC*	0.01	-6.41	4.93E-30
4	21	404	t3206	-	0.01	-6.18	1.49E-13
9	37	902	t1765	*htrB*	0.01	-6.07	2.29E-29
1	2	98	t4086	*rpmJ*	0.01	-5.73	0.0007732
3	13	290	t1622	-	0.01	-5.52	6.19E-10

Total inserts	Total reads	Gene length	Sys.ID	Name	Pvalue essential	Log likelihood ratio	Pvalue of no hit
10	36	938	t2274	*fepB*	0.01	-5.06	1.65E-30
5	8	458	t2641	-	0.01	-4.70	2.87E-15
1	2	89	t0694	-	0.01	-4.25	0.001493005
14	37	1232	t1682	*icdA*	0.01	-4.08	7.61E-40
17	75	1484	t3205	*nusA*	0.01	-3.95	7.58E-48
38	176	3185	t1734	*rne*	0.00	-3.31	7.4E-102
8	15	665	t1954	*cmk*	0.00	-3.18	7.67E-22
13	100	1067	t0480	*aroC*	0.00	-2.98	1.32E-34
21	53	1703	t1978	*cydC*	0.00	-2.78	8.44E-55
3	15	239	t2724	-	0.00	-2.50	2.58E-08
7	101	557	t3234	-	0.00	-2.48	2.06E-18
4	34	314	t2457	*yajC*	0.00	-2.26	1.07E-10
13	274	1010	t3489	*murB*	0.00	-2.09	8.52E-33

8.2 List of chromosomal genes comparing the differences between non-CmP6 and CmP6 passages

This list is the extension of Table 3-1. See folder Appendix 2 on CD.

8.3 List of plasmid stability genes

This list is the extension of Table 3-2. See folder Appendix 3 on CD.

8.4 TraDIS analysis scripts

See folder Appendix 4 on CD.

8.5 Isolates used for SNP typing

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
Ty2	1916	Russia	H10	H52	no	Kathryn Holt
M223	1939	unk	H8	H50b	no	Kathryn Holt
12-58	1958	Cameroon	H46	H81	no	François-Xavier Weill
27-58	1958	Morocco	H69	H52	no	François-Xavier Weill
14-58	1958	Cameroon	H77	H50a	no	François-Xavier Weill
05-59	1959	Viet Nam	H50	H50a	no	François-Xavier Weill
68-61	1961	Tunisia	H52	H52	no	François-Xavier Weill

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
66-61	1961	Tunisia	H50	H50b	no	François-Xavier Weill
69-61	1961	Tunisia	H56	H29	no	François-Xavier Weill
07-62	1962	Senegal	H52	H52	no	François-Xavier Weill
12-62	1962	Senegal	H52	H52	no	François-Xavier Weill
28-62	1962	Chad	H50	H50a	no	François-Xavier Weill
06-62	1962	Senegal	H50	H50a	no	François-Xavier Weill
13-62	1962	Senegal	H42	H50a	no	François-Xavier Weill
41-63	1963	Chad	H52	H52	no	François-Xavier Weill
63-63	1963	Chad	H52	H52	no	François-Xavier Weill
64-63	1963	Chad	H52	H50a	no	François-Xavier Weill
67-63	1963	Chad	H50	H50a	no	François-Xavier Weill
68-63	1963	Chad	H50	H50a	no	François-Xavier Weill
43-64	1964	Chad	H50	H58	no	François-Xavier Weill
06-64	1964	Chad	H52	H52	no	François-Xavier Weill
08-64	1964	Chad	H52	H52	no	François-Xavier Weill
10-64	1964	Chad	H52	H52	no	François-Xavier Weill
27-64	1964	Chad	H52	H52	no	François-Xavier Weill
62-65	1965	Ivory Coast	H83	H81	no	François-Xavier Weill
48-65	1965	Madagascar	H15	H52	no	François-Xavier Weill
49-65	1965	Madagascar	H15	H52	no	François-Xavier Weill
129-66	1966	Congo	H46	H81	no	François-Xavier Weill
29-66	1966	Algeria	H52	H52	no	François-Xavier Weill
19-66	1966	Congo	H52	H52	no	François-Xavier Weill
12-66	1966	Madagascar	H15	H52	no	François-Xavier Weill
171-66	1966	Morocco	H52	H52	no	François-Xavier Weill
84-66	1966	Tunisia	H52	H52	no	François-Xavier Weill
31-66	1966	Algeria	H36	H50b	no	François-Xavier Weill
49-66	1966	Algeria	H50	H50b	no	François-Xavier Weill
32-66	1966	Cameroon	H50	H50b	no	François-Xavier Weill
162-66	1966	Algeria	H50	H50a	no	François-Xavier Weill
37-66	1966	Cameroon	H50	H50a	no	François-Xavier Weill
102-66	1966	Senegal	H39	H50a	no	François-Xavier Weill
133-67	1967	Congo	H46	H81	no	François-Xavier Weill
42-67	1967	Ivory Coast	H81	H81	no	François-Xavier Weill
05-67	1967	Congo	H52	H52	no	François-Xavier Weill
31-67	1967	Congo	H42	H52	no	François-Xavier Weill
75-67	1967	Morocco	H42	H52	no	François-Xavier Weill
27-67	1967	Senegal	H52	H52	no	François-Xavier Weill
66-67	1967	Viet Nam	H1	H52	no	François-Xavier Weill
104-67	1967	Ivory Coast	H50	H50a	no	François-Xavier Weill
106-67	1967	Ivory Coast	H39	H50a	no	François-Xavier Weill
131-67	1967	Ivory Coast	H39	H50a	no	François-Xavier Weill
39-67	1967	Ivory Coast	H39	H50a	no	François-Xavier Weill

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
40-67	1967	Madagascar	H50	H50a	no	François-Xavier Weill
134-67	1967	Senegal	H39	H50a	no	François-Xavier Weill
49-67	1967	Viet Nam	H50	H50a	no	François-Xavier Weill
50-67	1967	Congo	H42	H42	no	François-Xavier Weill
67-67	1967	Viet Nam	H1	H1	no	François-Xavier Weill
69-67	1967	Viet Nam	H1	H1	no	François-Xavier Weill
72-1907	1972	Viet Nam	H68	H52	ST2	François-Xavier Weill
72-1910	1972	Viet Nam	H87	H52	no	François-Xavier Weill
72-1258	1972	Mexico	H11	H50a	ST3	François-Xavier Weill
ST3-40R181	1972	Mexico	none		ST3	Henry Smith
ST4-42R917	1972	Viet Nam	none		ST4	Henry Smith
73-43	1973	France	H52	H52	no	François-Xavier Weill
73-1102	1973	Viet Nam	H87	H52	ST4	François-Xavier Weill
73-114	1973	Viet Nam	H50	H50a	no	François-Xavier Weill
73-99	1973	Viet Nam	H50	H50a	no	François-Xavier Weill
75-2507	1975	India	H55	H50a	ST2	François-Xavier Weill
76-1292	1976	Zaire	H6	H81	no	François-Xavier Weill
76-54	1976	Chile	H50	H50b	7654	François-Xavier Weill
76-1261	1976	Zaire	H50	H50a	no	François-Xavier Weill
76-1406	1976	Indonesia	H42	H1	ST2	François-Xavier Weill
77-302	1977	India	H55	H50a	ST2	François-Xavier Weill
77-303	1977	India	H55	H50a	ST2	François-Xavier Weill
78-851	1978	Tunisia	H9	H52	ST2	François-Xavier Weill
80-2002	1980	Madagascar	H50	H50a	no	François-Xavier Weill
81-424	1981	Peru	H77	H50b	ST8	François-Xavier Weill
81-918	1981	Peru	H77	H50b	ST8	François-Xavier Weill
81-863	1981	Peru	H50	H50a	ST8	François-Xavier Weill
83-531	1983	Peru	H86	H50a	no	François-Xavier Weill
404ty	1983	Indonesia	H59	H29	no	Kathryn Holt
J-185SM	1985	Indonesia	H85	H29	no	Kathryn Holt
CT18	1993	Viet Nam	H1	H1	ST1	Kathryn Holt
ST1-pHCM1	1993	Viet Nam	none		ST1	John Wain
2T7466	1995	Ealing, UK	H58	H58	ST6	Fiona Cooke
230/95	1995	Viet Nam	H58	H58	no	François-Xavier Weill
162/95	1995	Viet Nam	H75	H50a	no	François-Xavier Weill
14/96	1996	Viet Nam	H58	H58	ST6	François-Xavier Weill
ST6-pSTY7	1996	Viet Nam	none		ST6	John Wain
205(97)S	1997	Viet Nam	H58	H58	ST6	François-Xavier Weill
209(97)S	1997	Viet Nam	H58	H58	ST6	François-Xavier Weill
226(97)S	1997	Viet Nam	H61	H58	ST6	François-Xavier Weill
43(97)S	1997	Viet Nam	H63	H58	ST6	François-Xavier Weill

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
E97-3246	1997	Madagascar	H17	H50a	no	François-Xavier Weill
E97-9141	1997	Turkey	H18	H50a	no	François-Xavier Weill
E98-8120	1998	Cameroon	H23	H81	no	François-Xavier Weill
E98-8119	1998	Peru	H22	H81	no	François-Xavier Weill
150(98)S	1998	Viet Nam	H63	H58	no	Kathryn Holt
30(98)S	1998	Viet Nam	H58	H58	no	François-Xavier Weill
31(98)S	1998	Viet Nam	H58	H58	ST6	François-Xavier Weill
318(98)N	1998	Viet Nam	H58	H58	no	François-Xavier Weill
339/98	1998	Viet Nam	H58	H58	ST6	François-Xavier Weill
358/98	1998	Viet Nam	H58	H58	ST6	François-Xavier Weill
38(98)S	1998	Viet Nam	H58	H58	ST6	François-Xavier Weill
39(98)S	1998	Viet Nam	H58	H58	no	François-Xavier Weill
43(98)S	1998	Viet Nam	H58	H58	ST6	François-Xavier Weill
49(98)S	1998	Viet Nam	H58	H58	ST6	François-Xavier Weill
E98-6926	1998	Mauritania	H21	H52	no	François-Xavier Weill
E98-3139	1998	Mexico	H50	H50b	no	Kathryn Holt
E98-0664	1998	Kenya	H55	H50a	no	Kathryn Holt
E98-4364	1998	Mexico	H50	H50a	no	François-Xavier Weill
E98-2107	1998	Senegal	H19	H50a	no	François-Xavier Weill
E98-2068	1998	Bangladesh	H42	H42	no	Kathryn Holt
E99-8067	1999	Algeria	H48	H81	no	François-Xavier Weill
E99-6785	1999	Morocco	H30	H81	no	François-Xavier Weill
E99-1028	1999	Senegal	H24	H81	no	François-Xavier Weill
E99-8635	1999	Nepal	H34	H58	no	François-Xavier Weill
192(99)S	1999	Viet Nam	H63	H58	no	François-Xavier Weill
197(99)S	1999	Viet Nam	H63	H58	no	François-Xavier Weill
219(99)S	1999	Viet Nam	H58	H58	ST6	François-Xavier Weill
E99-8095	1999	Algeria	H33	H52	no	François-Xavier Weill
E99-9794	1999	Comoros	H35	H52	no	François-Xavier Weill
E99-5920	1999	Tunisia	H26	H50b	no	François-Xavier Weill
E99-6359	1999	Mali	H27	H50a	no	François-Xavier Weill
E99-4879	1999	Morocco	H25	H50a	no	François-Xavier Weill
E99-8013	1999	Morocco	H32	H50a	no	François-Xavier Weill
E99-6646	1999	Algeria	H29	H29	no	François-Xavier Weill
E99-6478	1999	Guinea	H28	H29	no	François-Xavier Weill
E99-7012	1999	Morocco	H31	H29	no	François-Xavier Weill
E00-9821	2000	Congo	H46	H81	no	François-Xavier Weill
E00-7866	2000	Morocco	H46	H81	no	Kathryn Holt
E00-6111	2000	India	H58	H58	no	François-Xavier Weill
E00-9345	2000	India	H58	H58	no	François-Xavier Weill
CT111	2000	Viet Nam	H58	H58	no	OUCRU
CT115	2000	Viet Nam	H58	H58	ST6	OUCRU
E00-6172	2000	Indonesia	H52	H52	no	François-Xavier Weill

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
E00-6924	2000	Algeria	H36	H50b	no	François-Xavier Weill
E00-4624	2000	China	H8	H50b	no	François-Xavier Weill
E00-2756	2000	India	H37	H50b	no	François-Xavier Weill
E00-5869	2000	Bangladesh	H40	H50a	no	François-Xavier Weill
E00-6599	2000	Cape Verde	H41	H50a	no	François-Xavier Weill
E00-3201	2000	Mali	H38	H50a	no	François-Xavier Weill
E00-6999	2000	Peru	H43	H42	no	François-Xavier Weill
E00-7463	2000	Morocco	H44	H29	no	François-Xavier Weill
30laos	2000	Laos	H1	H1	no	OUCRU
57Laos	2000	Laos	H1	H1	ST1a	OUCRU
CT114	2000	Viet Nam	H1	H1	no	OUCRU
CT117	2000	Viet Nam	H1	H1	no	OUCRU
E01-5741	2001	Angola	H6	H81	no	François-Xavier Weill
E01-1747	2001	Cameroon	H47	H81	no	François-Xavier Weill
44Laos	2001	Laos	H58	H58	ST6	OUCRU
CT118	2001	Viet Nam	H58	H58	ST6	OUCRU
CT119	2001	Viet Nam	H58	H58	ST6	OUCRU
CT142	2001	Viet Nam	H58	H58	ST6	OUCRU
CT143	2001	Viet Nam	H58	H58	no	OUCRU
CT146	2001	Viet Nam	H58	H58	no	OUCRU
CT147	2001	Viet Nam	H58	H58	ST6	OUCRU
CT148	2001	Viet Nam	H58	H58	ST6	OUCRU
E01-7006	2001	Lebanon	H51	H52	no	François-Xavier Weill
E01-6750	2001	Senegal	H52	H52	no	Kathryn Holt
E01-8716	2001	Sri lanka	H53	H52	no	François-Xavier Weill
E01-5612	2001	Indonesia	H8	H50b	no	François-Xavier Weill
E01-7923	2001	Ivory Coast	H39	H50a	no	François-Xavier Weill
E01-1811	2001	Mali	H49	H50a	no	François-Xavier Weill
E01-9538	2001	Senegal	H80	H50a	no	François-Xavier Weill
E01-7101	2001	Togo	H39	H50a	no	François-Xavier Weill
CT120	2001	Viet Nam	H1	H1	no	OUCRU
E02-1180	2002	India	H45	H81	no	Kathryn Holt
E02-1497	2002	Ivory Coast	H81	H81	no	François-Xavier Weill
E02-2759	2002	India	H58	H58	no	Kathryn Holt
E02-1963	2002	Laos	H57	H58	no	François-Xavier Weill
E02-2159	2002	Sri lanka	H58	H58	ST6	François-Xavier Weill
DT003	2002	Viet Nam	H58	H58	ST6	OUCRU
DT006	2002	Viet Nam	H58	H58	ST6	OUCRU
DT007	2002	Viet Nam	H58	H58	ST6	OUCRU
DT008	2002	Viet Nam	H58	H58	ST6	OUCRU
DT009	2002	Viet Nam	H58	H58	ST6	OUCRU
DT014	2002	Viet Nam	H58	H58	ST6	OUCRU
DT015	2002	Viet Nam	H58	H58	ST6	OUCRU

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
DT019	2002	Viet Nam	H58	H58	ST6	OUCRU
DT021	2002	Viet Nam	H58	H58	ST6	OUCRU
DT023	2002	Viet Nam	H58	H58	ST6	OUCRU
DT025	2002	Viet Nam	H58	H58	ST6	OUCRU
DT026	2002	Viet Nam	H58	H58	ST6	OUCRU
DT027	2002	Viet Nam	H58	H58	ST6	OUCRU
DT029	2002	Viet Nam	H58	H58	ST6	OUCRU
DT030	2002	Viet Nam	H58	H58	ST6	OUCRU
DT031	2002	Viet Nam	H58	H58	ST6	OUCRU
DT032	2002	Viet Nam	H58	H58	ST6	OUCRU
DT033	2002	Viet Nam	H58	H58	ST6	OUCRU
DT034	2002	Viet Nam	H58	H58	ST6	OUCRU
DT035	2002	Viet Nam	H58	H58	ST6	OUCRU
DT036	2002	Viet Nam	H58	H58	ST6	OUCRU
DT037	2002	Viet Nam	H58	H58	ST6	OUCRU
DT038	2002	Viet Nam	H58	H58	ST6	OUCRU
DT039	2002	Viet Nam	H58	H58	ST6	OUCRU
DT040	2002	Viet Nam	H58	H58	ST6	OUCRU
DT041	2002	Viet Nam	H58	H58	no	OUCRU
DT042	2002	Viet Nam	H58	H58	ST6	OUCRU
DT043	2002	Viet Nam	H58	H58	ST6	OUCRU
DT044	2002	Viet Nam	H58	H58	ST6	OUCRU
DT045	2002	Viet Nam	H58	H58	ST6	OUCRU
DT046	2002	Viet Nam	H58	H58	ST6	OUCRU
DT047	2002	Viet Nam	H58	H58	ST6	OUCRU
DT048	2002	Viet Nam	H58	H58	ST6	OUCRU
DT049	2002	Viet Nam	H58	H58	ST6	OUCRU
DT050	2002	Viet Nam	H58	H58	ST6	OUCRU
DT051	2002	Viet Nam	H58	H58	ST6	OUCRU
DT052	2002	Viet Nam	H58	H58	ST6	OUCRU
DT053	2002	Viet Nam	H58	H58	ST6	OUCRU
DT054	2002	Viet Nam	H58	H58	no	OUCRU
DT055	2002	Viet Nam	H58	H58	ST6	OUCRU
DT057	2002	Viet Nam	H58	H58	ST6	OUCRU
DT058	2002	Viet Nam	H58	H58	ST6	OUCRU
DT059	2002	Viet Nam	H58	H58	ST6	OUCRU
DT061	2002	Viet Nam	H58	H58	ST6	OUCRU
DT062	2002	Viet Nam	H58	H58	ST6	OUCRU
DT063	2002	Viet Nam	H58	H58	ST6	OUCRU
DT064	2002	Viet Nam	H58	H58	ST6	OUCRU
DT065	2002	Viet Nam	H58	H58	ST6	OUCRU
DT066	2002	Viet Nam	H58	H58	ST6	OUCRU
DT068	2002	Viet Nam	H58	H58	ST6	OUCRU

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
DT069	2002	Viet Nam	H58	H58	ST6	OUCRU
DT070	2002	Viet Nam	H58	H58	ST6	OUCRU
DT071	2002	Viet Nam	H58	H58	ST6	OUCRU
DT073	2002	Viet Nam	H58	H58	ST6	OUCRU
DT074	2002	Viet Nam	H58	H58	ST6	OUCRU
DT075	2002	Viet Nam	H58	H58	ST6	OUCRU
DT076	2002	Viet Nam	H58	H58	ST6	OUCRU
DT078	2002	Viet Nam	H58	H58	ST6	OUCRU
DT080	2002	Viet Nam	H58	H58	ST6	OUCRU
DT081	2002	Viet Nam	H58	H58	ST6	OUCRU
DT082	2002	Viet Nam	H58	H58	ST6	OUCRU
DT083	2002	Viet Nam	H58	H58	ST6	OUCRU
DT084	2002	Viet Nam	H58	H58	ST6	OUCRU
DT085	2002	Viet Nam	H58	H58	ST6	OUCRU
DT086	2002	Viet Nam	H58	H58	ST6	OUCRU
DT087	2002	Viet Nam	H58	H58	ST6	OUCRU
DT088	2002	Viet Nam	H58	H58	ST6	OUCRU
DT089	2002	Viet Nam	H58	H58	ST6	OUCRU
DT090	2002	Viet Nam	H58	H58	no	OUCRU
DT091	2002	Viet Nam	H58	H58	ST6	OUCRU
DT092	2002	Viet Nam	H58	H58	no	OUCRU
DT093	2002	Viet Nam	H58	H58	ST6	OUCRU
DT094	2002	Viet Nam	H58	H58	ST6	OUCRU
DT095	2002	Viet Nam	H58	H58	ST6	OUCRU
DT096	2002	Viet Nam	H58	H58	no	OUCRU
DT097	2002	Viet Nam	H58	H58	ST6	OUCRU
DT098	2002	Viet Nam	H58	H58	no	OUCRU
DT099	2002	Viet Nam	H58	H58	ST6	OUCRU
1(02)C	2002	Viet Nam	H58	H58	no	François-Xavier Weill
11(02)S	2002	Viet Nam	H58	H58	ST6	François-Xavier Weill
12(02)S	2002	Viet Nam	H58	H58	no	François-Xavier Weill
14(02)S	2002	Viet Nam	H60	H58	no	François-Xavier Weill
17(02)S	2002	Viet Nam	H58	H58	ST6	François-Xavier Weill
19(02)S	2002	Viet Nam	H58	H58	ST6	François-Xavier Weill
2(02)N	2002	Viet Nam	H58	H58	ST6	François-Xavier Weill
2(02)S	2002	Viet Nam	H58	H58	no	François-Xavier Weill
20(02)N	2002	Viet Nam	H58	H58	no	François-Xavier Weill
3(02)C	2002	Viet Nam	H58	H58	no	François-Xavier Weill
4(02)C	2002	Viet Nam	H58	H58	ST6	François-Xavier Weill
4(02)N	2002	Viet Nam	H58	H58	ST6	François-Xavier Weill
4(02)S	2002	Viet Nam	H58	H58	no	François-Xavier Weill
7(02)N	2002	Viet Nam	H58	H58	no	François-Xavier Weill
8(02)C	2002	Viet Nam	H58	H58	ST6	François-Xavier Weill

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
8(02)S	2002	Viet Nam	H58	H58	no	François-Xavier Weill
SonLa-1	2002	Viet Nam	H68	H52	no	François-Xavier Weill
E02-0232	2002	French Guiana	H54	H50a	no	François-Xavier Weill
E02-0530	2002	Nigeria	H4	H50a	no	François-Xavier Weill
E02-1687	2002	Thailand	H79	H50a	no	François-Xavier Weill
E02-1687	2002	Thailand	H79	H50a	no	François-Xavier Weill
E02-5919	2002	China	H28	H29	no	François-Xavier Weill
E03-0658	2003	Philippines	H45	H81	no	François-Xavier Weill
E03-6418	2003	Bangladesh	H58	H58	no	François-Xavier Weill
E03-5712	2003	Cambodia	H58	H58	ST6	François-Xavier Weill
SD_A 102	2003	India	H64	H58	no	Shanta Dutta
SD_B0031	2003	India	H58	H58	ST6	Shanta Dutta
SD_B0111	2003	India	H58	H58	ST6	Shanta Dutta
SD_C 777	2003	India	H58	H58	no	Shanta Dutta
SD_C0294	2003	India	H58	H58	ST6	Shanta Dutta
SD_D 641	2003	India	H58	H58	ST6	Shanta Dutta
SD_D 878	2003	India	H58	H58	ST6	Shanta Dutta
SD_D0964	2003	India	H58	H58	no	Shanta Dutta
SD_E0176	2003	India	H58	H58	no	Shanta Dutta
ISP-03-07467	2003	Morocco	H58	H58	ST6	Kathryn Holt
E03-9804	2003	Nepal	H58	H58	ST6	Kathryn Holt
AG020	2003	Viet Nam	H58	H58	no	OUCRU
AG021	2003	Viet Nam	H58	H58	no	OUCRU
AG024	2003	Viet Nam	H58	H58	ST6	OUCRU
AG026	2003	Viet Nam	H58	H58	no	OUCRU
AG027	2003	Viet Nam	H58	H58	no	OUCRU
AG030	2003	Viet Nam	H58	H58	ST6	OUCRU
AG033	2003	Viet Nam	H58	H58	no	OUCRU
AG034	2003	Viet Nam	H58	H58	no	OUCRU
AG035	2003	Viet Nam	H58	H58	no	OUCRU
AG036	2003	Viet Nam	H58	H58	no	OUCRU
AG038	2003	Viet Nam	H58	H58	ST6	OUCRU
AG039	2003	Viet Nam	H58	H58	no	OUCRU
E03-6643	2003	India	H50	H50a	no	François-Xavier Weill
SD_B 765	2003	India	H42	H42	no	Shanta Dutta
E03-4983	2003	Indonesia	H59	H29	no	Kathryn Holt
03-4747	2003	Togo	H42	H29	ST2	François-Xavier Weill
ST8-SPA568	2003	Karachi, Pakistan	ParaA		ST8	Rumina Hasan
ISP-04-06979	2004	Central Africa	H58	H58	ST6	Kathryn Holt
04-2176	2004	India	H58	H58	no	François-Xavier Weill
SD_A 1672	2004	India	H58	H58	no	Shanta Dutta

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
SD_B 3235	2004	India	H58	H58	no	Shanta Dutta
SD_B1868	2004	India	H58	H58	ST6	Shanta Dutta
SD_B3274	2004	India	H58	H58	no	Shanta Dutta
SD_C 2114	2004	India	H58	H58	no	Shanta Dutta
SD_C 2115	2004	India	H58	H58	no	Shanta Dutta
SD_D 2190	2004	India	H58	H58	no	Shanta Dutta
SD_D1604	2004	India	H64	H58	no	Shanta Dutta
SD_G0677	2004	India	H64	H58	no	Shanta Dutta
ST148	2004	Jordan	H58	H58	no	Bianca Paglietti
ST298	2004	Jordan	H58	H58	no	Bianca Paglietti
ST661	2004	Jordan	H58	H58	ST6	Bianca Paglietti
ST721	2004	Jordan	H58	H58	ST6	Bianca Paglietti
SK_5982	2004	Kenya	H58	H58	ST6	Sam Kariuki
SK_5990	2004	Kenya	H58	H58	ST6	Sam Kariuki
8(04)N	2004	Viet Nam	H58	H58	no	Kathryn Holt
AG052	2004	Viet Nam	H58	H58	ST6	OUCRU
AG053	2004	Viet Nam	H58	H58	no	OUCRU
AG054	2004	Viet Nam	H58	H58	no	OUCRU
AG059	2004	Viet Nam	H58	H58	no	OUCRU
AG066	2004	Viet Nam	H58	H58	ST6	OUCRU
AG068	2004	Viet Nam	H58	H58	ST6	OUCRU
AG069	2004	Viet Nam	H58	H58	ST6	OUCRU
AG075	2004	Viet Nam	H58	H58	no	OUCRU
AG076	2004	Viet Nam	H58	H58	ST6	OUCRU
AG077	2004	Viet Nam	H58	H58	ST6	OUCRU
AG078	2004	Viet Nam	H58	H58	no	OUCRU
AG080	2004	Viet Nam	H58	H58	no	OUCRU
AG081	2004	Viet Nam	H58	H58	ST6	OUCRU
AG082	2004	Viet Nam	H58	H58	ST6	OUCRU
AG085	2004	Viet Nam	H58	H58	no	OUCRU
AG086	2004	Viet Nam	H58	H58	ST6	OUCRU
AG092	2004	Viet Nam	H58	H58	ST6	OUCRU
AG105	2004	Viet Nam	H58	H58	ST6	OUCRU
AG108	2004	Viet Nam	H58	H58	ST6	OUCRU
AG183	2004	Viet Nam	H58	H58	no	OUCRU
AG211	2004	Viet Nam	H58	H58	no	OUCRU
AG224	2004	Viet Nam	H58	H58	no	OUCRU
AG242	2004	Viet Nam	H58	H58	no	OUCRU
AG3	2004	Viet Nam	H58	H58	no	Kathryn Holt
16(04)S	2004	Viet Nam	H58	H58	ST6	François-Xavier Weill
2(04)S	2004	Viet Nam	H63	H58	no	François-Xavier Weill
21(04)S	2004	Viet Nam	H58	H58	no	François-Xavier Weill
3(04)C	2004	Viet Nam	H58	H58	ST6	François-Xavier Weill

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
4(04)C	2004	Viet Nam	H58	H58	no	François-Xavier Weill
8(04)S	2004	Viet Nam	H58	H58	no	François-Xavier Weill
SD_B1714	2004	India	H42	H42	no	Shanta Dutta
SD_C1449	2004	India	H42	H42	no	Shanta Dutta
SD_E728	2004	India	H42	H42	no	Shanta Dutta
K1071	2004	United States	H42	H42	no	CDC
K1124	2004	United States	H42	H42	no	CDC
04 6845	2004	Benin	H42	H29	ST2	François-Xavier Weill
1(04)C	2004	Viet Nam	H1	H1	no	François-Xavier Weill
05-3275	2005	Morocco	H6	H81	no	François-Xavier Weill
3T512	2005	Ealing, UK	H58	H58	no	Fiona Cooke
230	2005	India	H58	H58	no	Rajni Gaind
238	2005	India	H58	H58	no	Rajni Gaind
SD_A 1919	2005	India	H58	H58	no	Shanta Dutta
SD_A1921	2005	India	H64	H58	no	Shanta Dutta
SD_A2467	2005	India	H64	H58	no	Shanta Dutta
SD_B4716	2005	India	H58	H58	no	Shanta Dutta
SD_C 3495	2005	India	H64	H58	no	Shanta Dutta
SD_C 3634	2005	India	H64	H58	no	Shanta Dutta
SD_C 4277	2005	India	H58	H58	no	Shanta Dutta
SD_C 4932	2005	India	H58	H58	ST6	Shanta Dutta
SD_C4404	2005	India	H58	H58	no	Shanta Dutta
SD_C4529	2005	India	H58	H58	no	Shanta Dutta
SD_C4826	2005	India	H58	H58	no	Shanta Dutta
SD_C4903	2005	India	H58	H58	ST6	Shanta Dutta
SD_D 5140	2005	India	H58	H58	no	Shanta Dutta
SD_D4019	2005	India	H58	H58	no	Shanta Dutta
SD_E0856	2005	India	H58	H58	no	Shanta Dutta
SK_6010	2005	Kenya	H58	H58	ST6	Sam Kariuki
SK_6031	2005	Kenya	H58	H58	ST6	Sam Kariuki
SK_6066	2005	Kenya	H58	H58	ST6	Sam Kariuki
SK_6067	2005	Kenya	H58	H58	no	Sam Kariuki
SK_6073	2005	Kenya	H58	H58	ST6	Sam Kariuki
SK_6076	2005	Kenya	H58	H58	ST6	Sam Kariuki
SK_6130	2005	Kenya	H58	H58	no	Sam Kariuki
SK_6151	2005	Kenya	H58	H58	ST6	Sam Kariuki
SK_6152	2005	Kenya	H58	H58	ST6	Sam Kariuki
SK_6153	2005	Kenya	H58	H58	ST6	Sam Kariuki
SK_6155	2005	Kenya	H58	H58	ST6	Sam Kariuki
SK_6160	2005	Kenya	H58	H58	ST6	Sam Kariuki
AG118	2005	Viet Nam	H58	H58	ST6	OUCRU
AG119	2005	Viet Nam	H58	H58	ST6	OUCRU
AG121	2005	Viet Nam	H58	H58	ST6	OUCRU

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
AG123	2005	Viet Nam	H58	H58	ST6	OUCRU
AG126	2005	Viet Nam	H58	H58	ST6	OUCRU
AG128	2005	Viet Nam	H58	H58	ST6	OUCRU
AG129	2005	Viet Nam	H58	H58	ST6	OUCRU
AG130	2005	Viet Nam	H58	H58	ST6	OUCRU
AG142	2005	Viet Nam	H58	H58	ST6	OUCRU
AG144	2005	Viet Nam	H58	H58	ST6	OUCRU
AG148	2005	Viet Nam	H58	H58	ST6	OUCRU
AG152	2005	Viet Nam	H58	H58	ST6	OUCRU
AG153	2005	Viet Nam	H58	H58	ST6	OUCRU
AG165	2005	Viet Nam	H58	H58	ST6	OUCRU
AG167	2005	Viet Nam	H58	H58	ST6	OUCRU
AG169	2005	Viet Nam	H58	H58	no	OUCRU
AG172	2005	Viet Nam	H58	H58	ST6	OUCRU
AG174	2005	Viet Nam	H58	H58	ST6	OUCRU
AG175	2005	Viet Nam	H58	H58	ST6	OUCRU
AG270	2005	Viet Nam	H58	H58	ST6	OUCRU
AG282	2005	Viet Nam	H58	H58	ST6	OUCRU
AG284	2005	Viet Nam	H58	H58	ST6	OUCRU
SD_D5499	2005	India	H50	H50a	no	Shanta Dutta
SD_C 4190	2005	India	H42	H42	no	Shanta Dutta
SD_C 4809	2005	India	H42	H42	no	Shanta Dutta
SD_C4881	2005	India	H42	H42	no	Shanta Dutta
SD_C4892	2005	India	H42	H42	no	Shanta Dutta
SD_C5029	2005	India	H42	H42	no	Shanta Dutta
SD_C5146	2005	India	H42	H42	no	Shanta Dutta
SD_G 1197	2005	India	H42	H42	ST6	Shanta Dutta
SD_G1138	2005	India	H42	H42	no	Shanta Dutta
7	2006	India	H58	H58	no	Rajni Gaind
SK_6091	2006	Kenya	H58	H58	ST6	Sam Kariuki
SK_6301	2006	Kenya	H58	H58	ST6	Sam Kariuki
SK_6320	2006	Kenya	H58	H58	no	Sam Kariuki
SK_6322	2006	Kenya	H58	H58	ST6	Sam Kariuki
SK_6335	2006	Kenya	H58	H58	ST6	Sam Kariuki
SK_6335	2006	Kenya	H58	H58	ST6	Sam Kariuki
SK_6338	2006	Kenya	H58	H58	ST6	Sam Kariuki
JA_1031	2006	Kuwait	H58	H58	no	John Albert
JA_1050	2006	Kuwait	H58	H58	no	John Albert
JA_2076	2006	Kuwait	H62	H58	no	John Albert
JA_2342	2006	Kuwait	H62	H58	no	John Albert
JA_2553	2006	Kuwait	H58	H58	no	John Albert
JA_2610	2006	Kuwait	H58	H58	ST6	John Albert
JA_2739	2006	Kuwait	H58	H58	no	John Albert

Strain name	Year	Country	Haplotype	Haplo-group	PST	Source
JA_2756	2006	Kuwait	H58	H58	ST6	John Albert
JA_2827	2006	Kuwait	H58	H58	no	John Albert
JA_2849	2006	Kuwait	H58	H58	no	John Albert
JA_2895	2006	Kuwait	H62	H58	no	John Albert
JA_2941	2006	Kuwait	H58	H58	no	John Albert
JA_3059	2006	Kuwait	H62	H58	no	John Albert
JA_310	2006	Kuwait	H58	H58	no	John Albert
JA_3107	2006	Kuwait	H62	H58	no	John Albert
JA_3164	2006	Kuwait	H58	H58	no	John Albert
JA_3189	2006	Kuwait	H64	H58	no	John Albert
JA_3213	2006	Kuwait	H58	H58	ST6	John Albert
JA_3216	2006	Kuwait	H58	H58	ST6	John Albert
JA_3230	2006	Kuwait	*H58 precursor	H58	ST6	John Albert
JA_892	2006	Kuwait	H58	H58	no	John Albert
JA_182	2006	Kuwait	H52	H52	no	John Albert
JA_2815	2006	Kuwait	H25	H50b	no	John Albert
GG_871	2006	Ethiopia	H50	H50a	no	Getenet Bevene
JA_1251	2006	Kuwait	H42	H42	no	John Albert
JA_1254	2006	Kuwait	H42	H42	no	John Albert
JA_2316	2006	Kuwait	H42	H42	no	John Albert
GG_6	2006	Ethiopia	H31	H29	no	Getenet Bevene
84	2007	India	H58	H58	no	Rajni Gaind
90	2007	India	H58	H58	no	Rajni Gaind
SK_6665	2007	Kenya	H58	H58	ST6	Sam Kariuki
4968	2007	India	H42	H42	no	Rajni Gaind
4969	2007	India	H42	H42	no	Rajni Gaind
133	2007	India	H1	H1	no	Rajni Gaind
V.03052	?	Viet Nam	H58	H58	ST6	OUCRU
V.03818	?	Viet Nam	H58	H58	ST6	OUCRU
ST2-44R311	1970s	Thailand	none		ST2	Henry Smith
57	2005-2007	India	H58	H58	no	Rajni Gaind
32-TFS	unk	Bangladesh	H58	H58	no	OUCRU
40-TFS	unk	Bangladesh	H62	H58	no	OUCRU
CIS9662/06	unk	unk	H52	H52	no	François-Xavier Weill
M453	unk	unk	H52	H52	no	Derek Pikard
M464	unk	unk	H53	H52	no	Derek Pikard
CIS9661/06	unk	unk	H50	H50a	no	François-Xavier Weill
M513	unk	unk	H50	H50a	no	Derek Pikard
SD_C 1468	unk	India	ParaA		no	Shanta Dutta
SD_D 3581	unk	India	ParaA		no	Shanta Dutta

8.6 SNP data in phylip format

See folder Appendix 6 on CD

8.7 List of SNP used in the GoldenGate assay

See folder Appendix 7 on CD

8.8 Mobile elements and resistant genes detected by GoldenGate assay

See folder Appendix 8 on CD