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.