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.

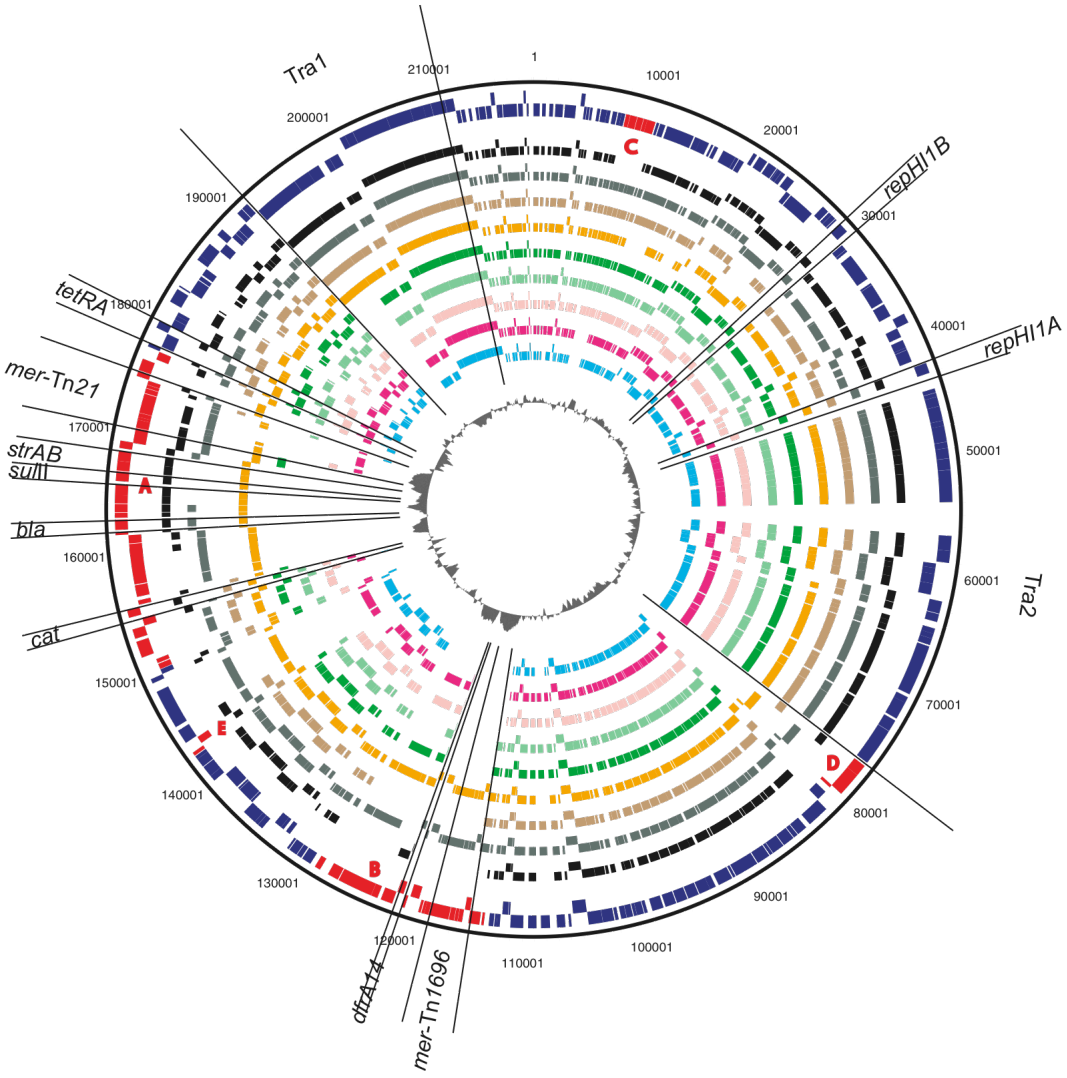


Figure 4-1 Genetic variation in IncHI1 plasmids from *S.* Typhi as determined by microarray analysis.
Predicted coding sequences (CDS) are defined in the sequence of pHCM1 in the outer ring (NC_003384). The inner concentric rings each represent the CDS predicted as present or absent in each of the different 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.

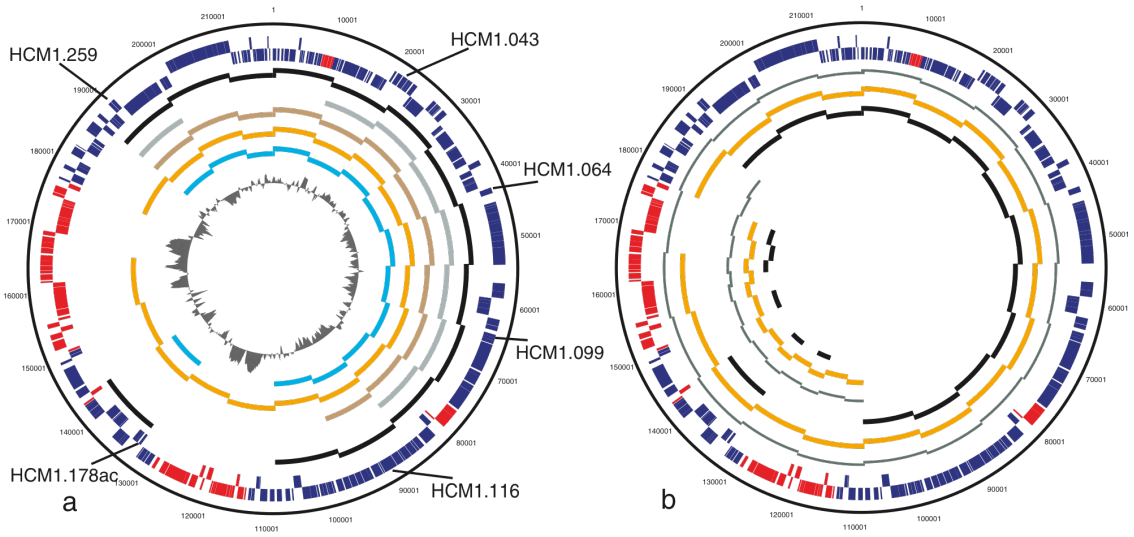


Figure 4-2 Syntenic regions present on eachIncHI1 plasmids.
2a: Long range PCR of ~10Kb amplicons around IncHI1 plasmids of *S.* Typhi. The products generated from R27 (inner grey ring) are predicted from the DNA sequence and the primer binding sites are represented at either end of the grey bars on 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.

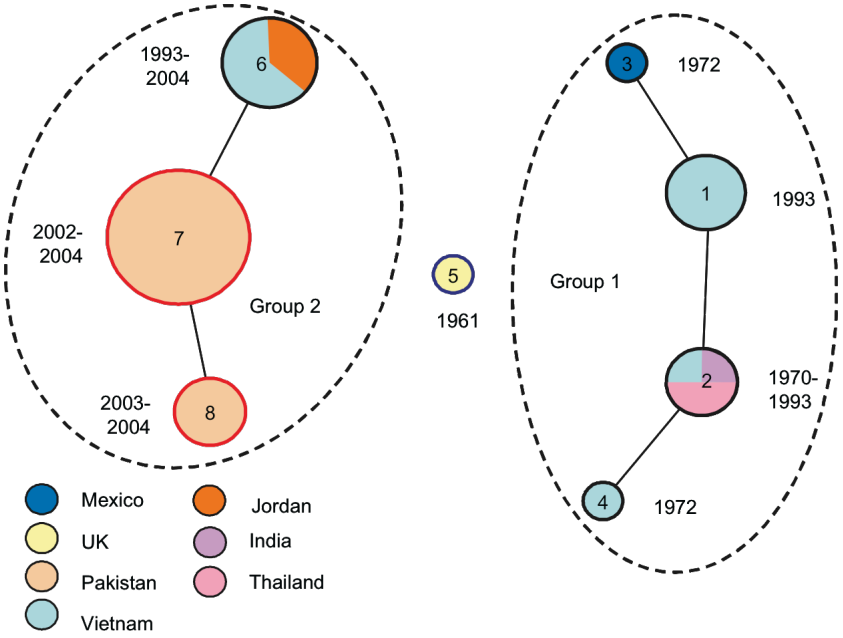


Figure 4-3 Modified eBURST diagram for plasmid multi-locus sequence typing.
The numbers inside the circles represent the plasmid sequence types (PSTs). Lines connecting circles represent single locus variants. The dotted circles represent groups of related plasmids. Plasmids in group 1 are all from before 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.

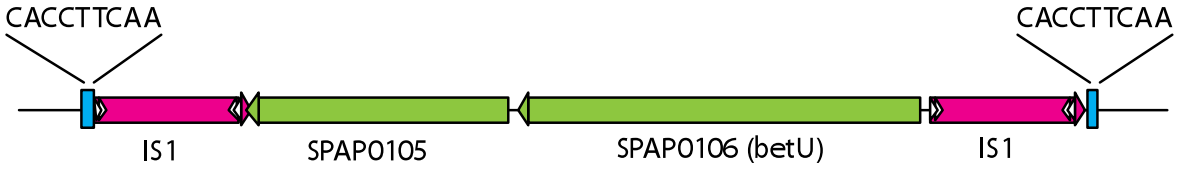


Figure 4-4 Schematic structure of Tn*6062*
Blue: 9-bp direct repeat; Pink: IS*1*; Green: coding sequence.

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.

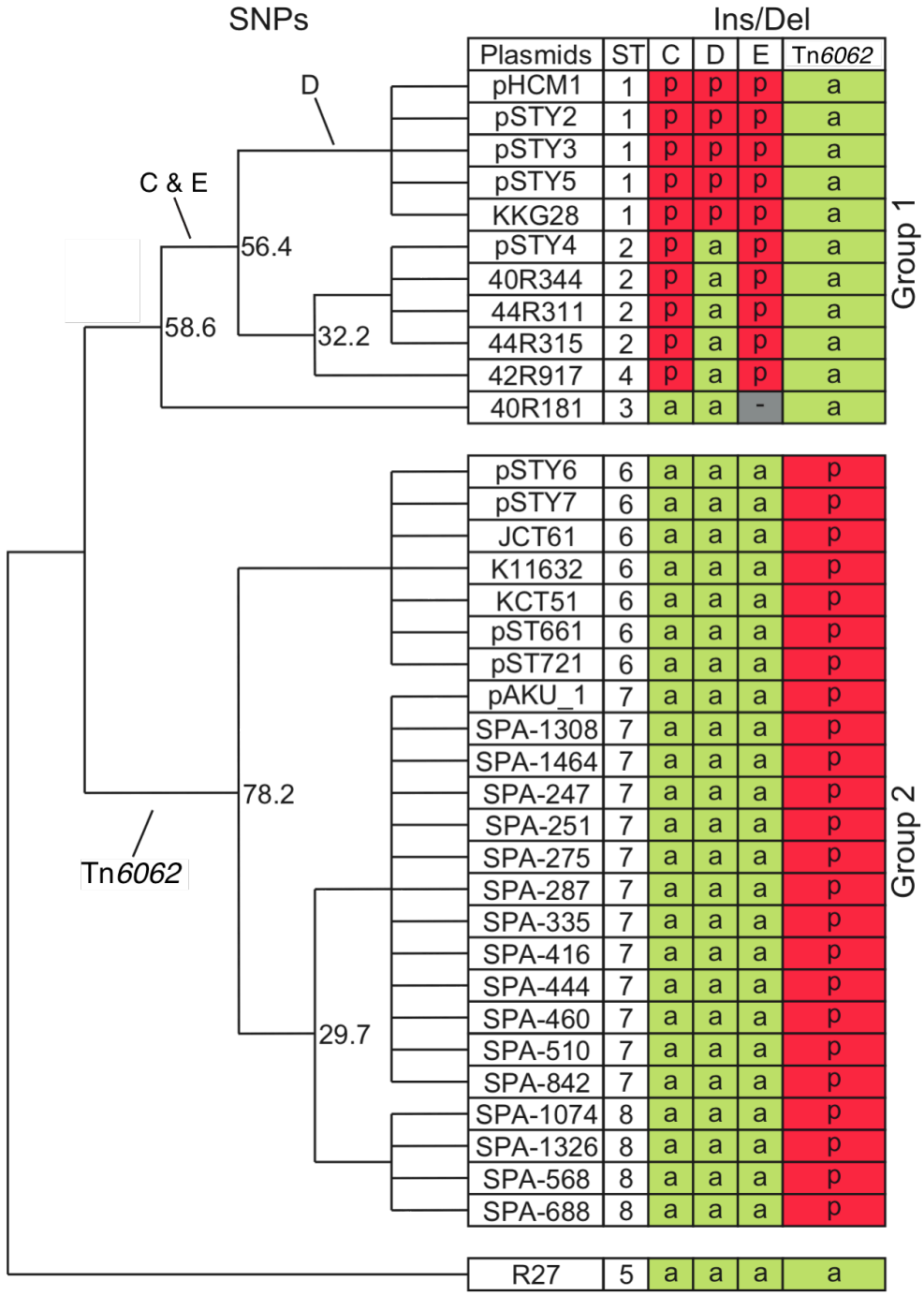


Figure 4-5 IncHI1 plasmids phylogenetic tree and observed insertion/deletion events
The phylogenetic tree was built based on the SNPs identified within PMLST sequences using maximum parsimony for discrete character data (Phylip software package), bootstrap values shown from 1000 re-sampling. The insertion/deletion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

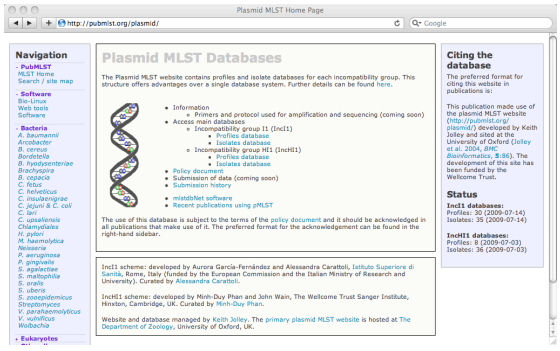
the phylogenetic tree and PCR results, we conclude that plasmids of group 2 (post-1993) did not evolve directly from group 1 (pre-1993) but belong to two different lineages.

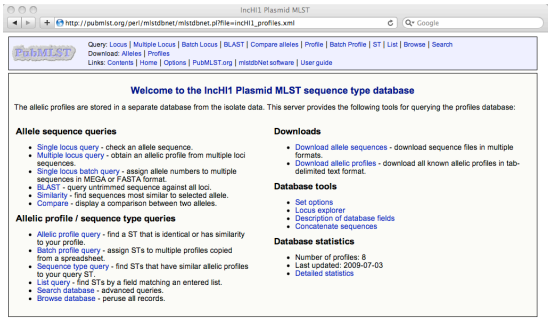
4.1.2.7 Online database for PMLST

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).

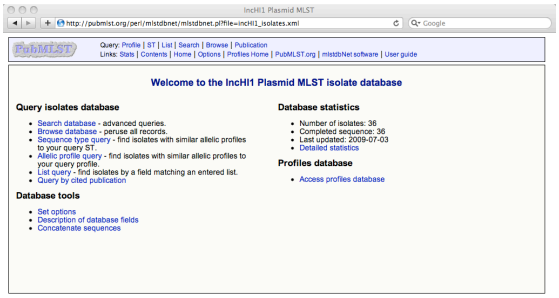
The PMLST information is organised into two interlinked databases: a profile database and an isolate database. The Profile database contains allelic sequences of each locus and plasmid sequence types. The Isolate database provides information on each isolate/plasmid including year of isolation, geographical 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.