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

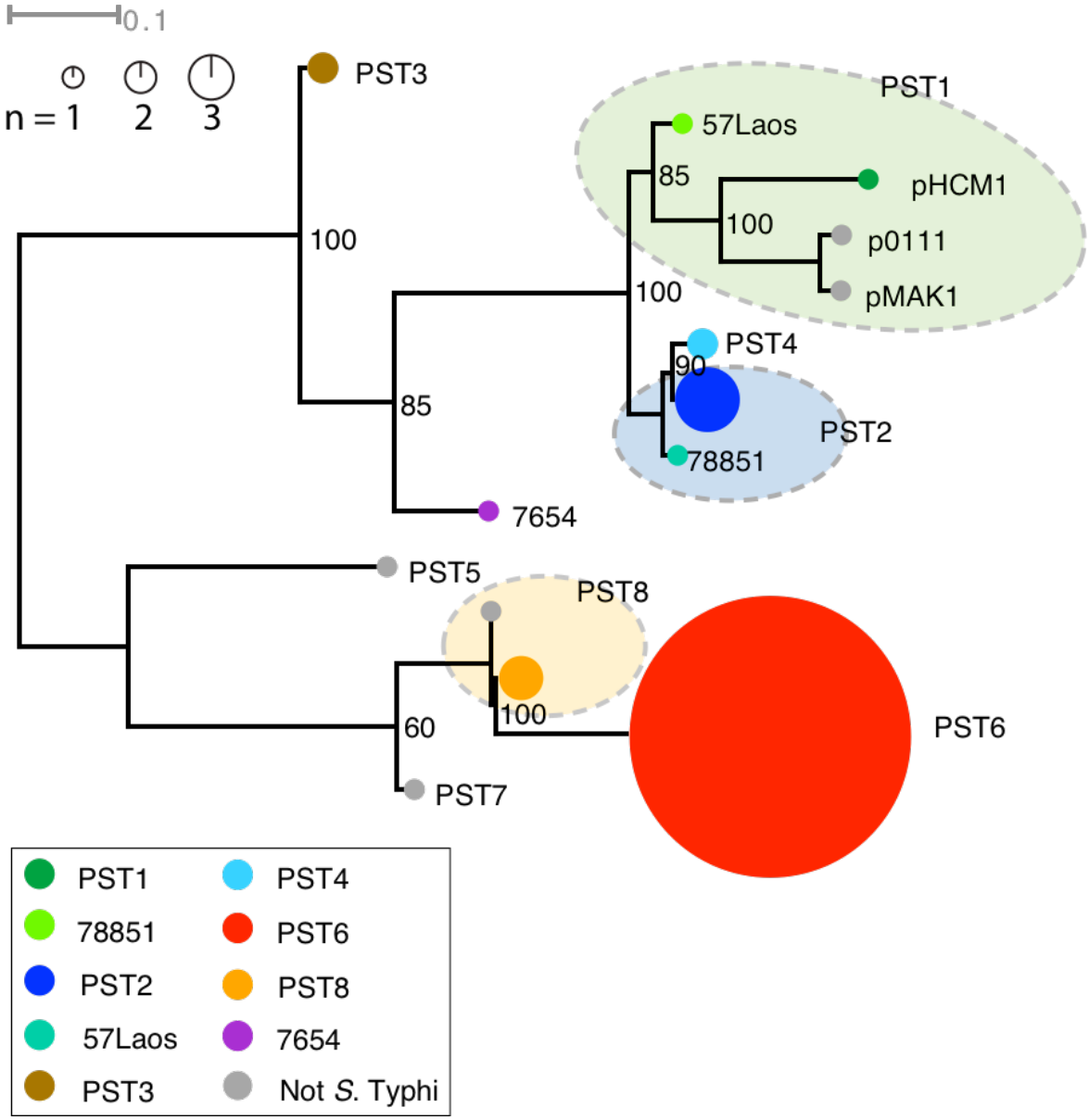


Figure 5-1 Relationship of plasmid sequence types
Demonstrated by a maximum likelihood phylogenetic tree drawn using SNP data from the Golden Gate platform. The PST numbers denotes the sequence types define by PMLST. Bootstrap values after 1000 iterations is shown.

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.

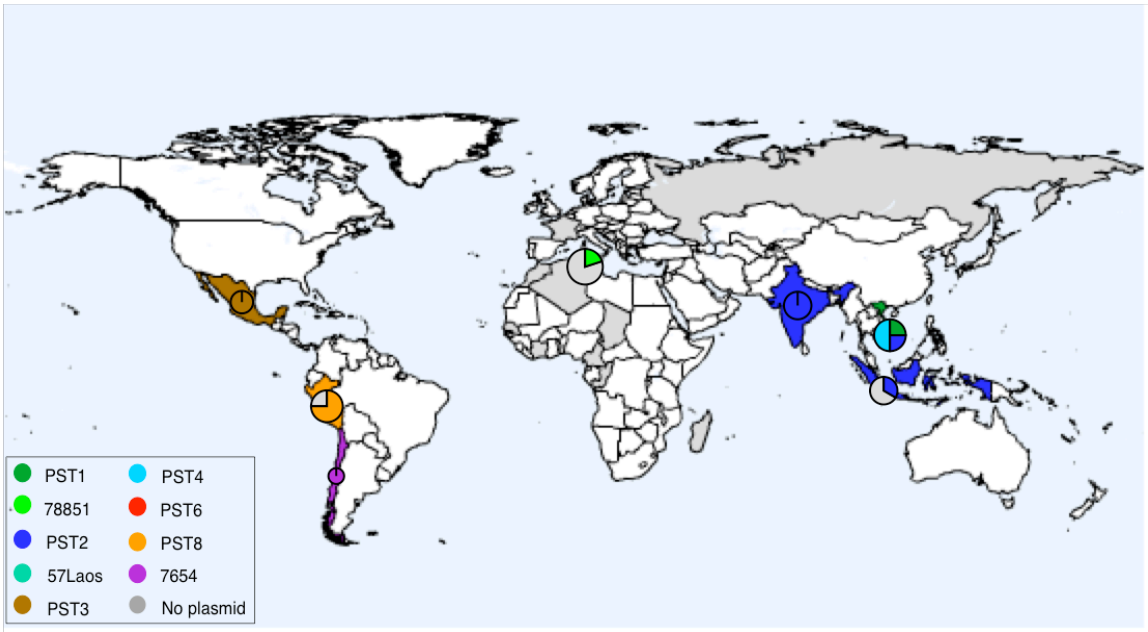
PST6 had the largest number of plasmids in our study. Although there were 180 isolates in this group, they are identical in all SNP loci investigated. All PST6 plasmids in this strain collection were isolated after 1995. This suggested a recent clonal expansion of PST6 plasmids, which supported our conclusion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

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

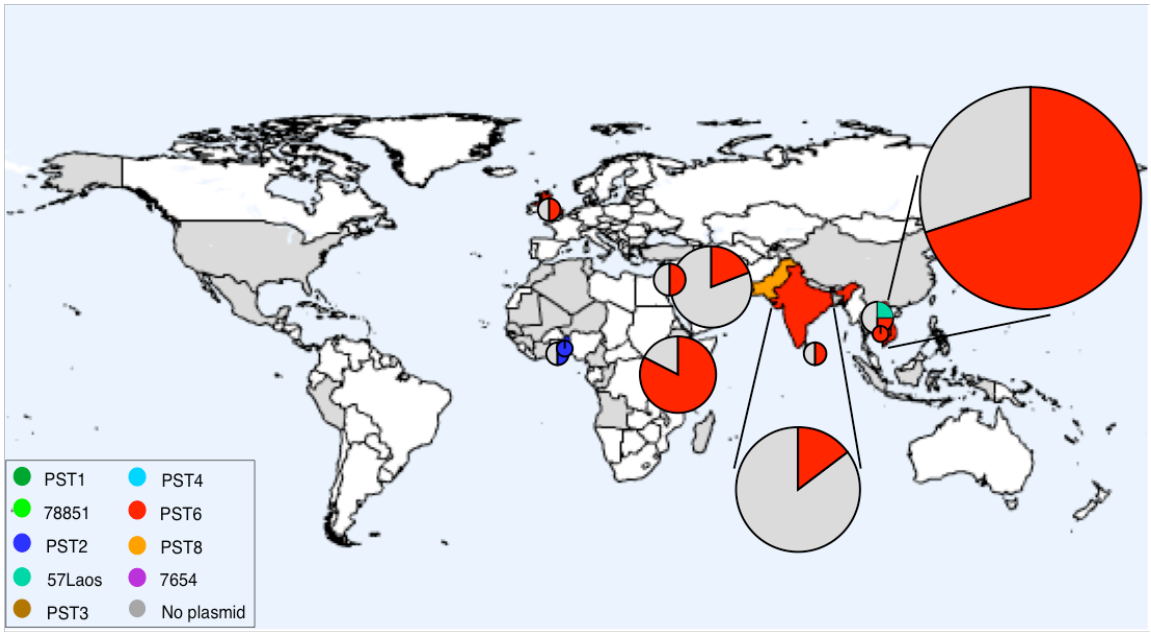
	Before 1993	After 1993	Total
7654	1	0	1
PST1	2	0	2
57Laos	0	1	1
PST2	5	3	8
78851	1	0	1
PST3	2	0	2
PST4	2	0	2
PST6	0	180	180
PST8	3	0	3
No plasmid	66	197	263
Total	82	381	463

- 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).

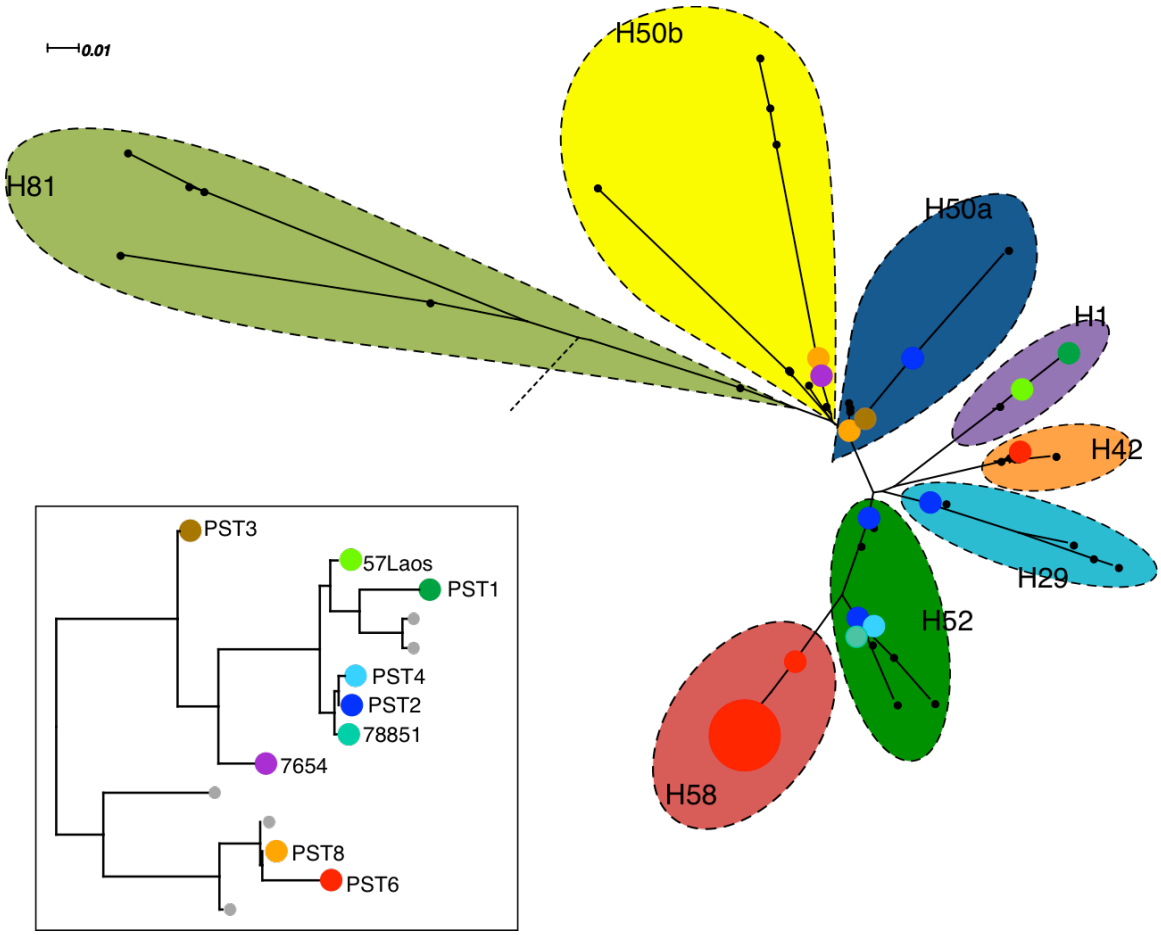
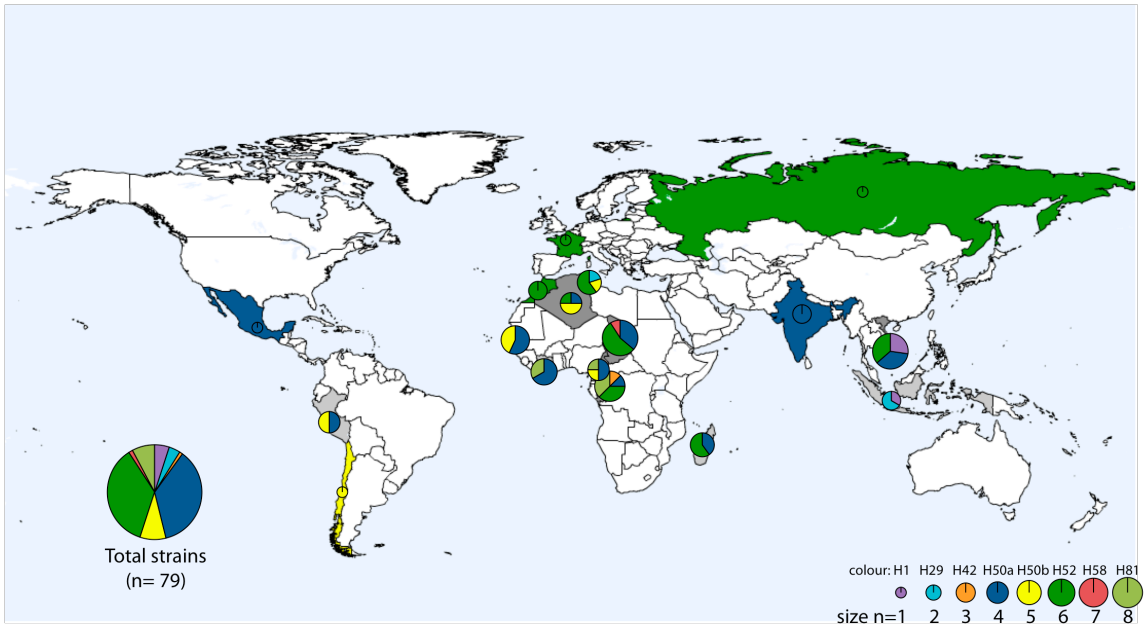
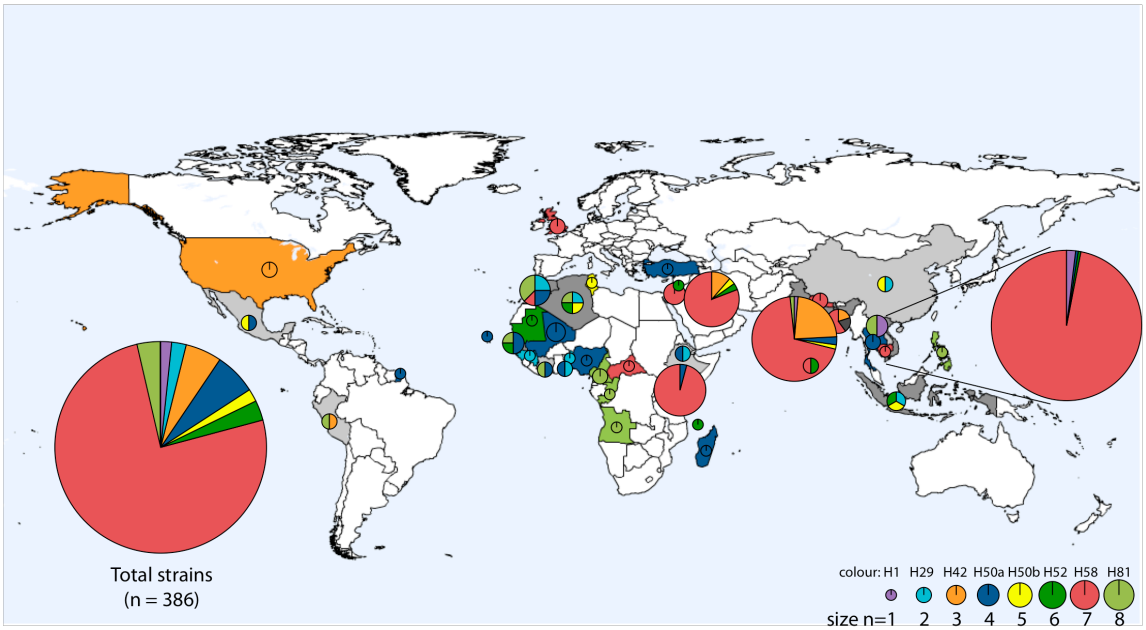


Figure 5-3 The association of plasmid type and strain background.
Tree drawn from chromosomal SNPs (colours represent the different haplogroups in all subsequent figures.). Internal circles represent the different plasmid STs (same colours as those in the plasmid tree inlet). All the branching shown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

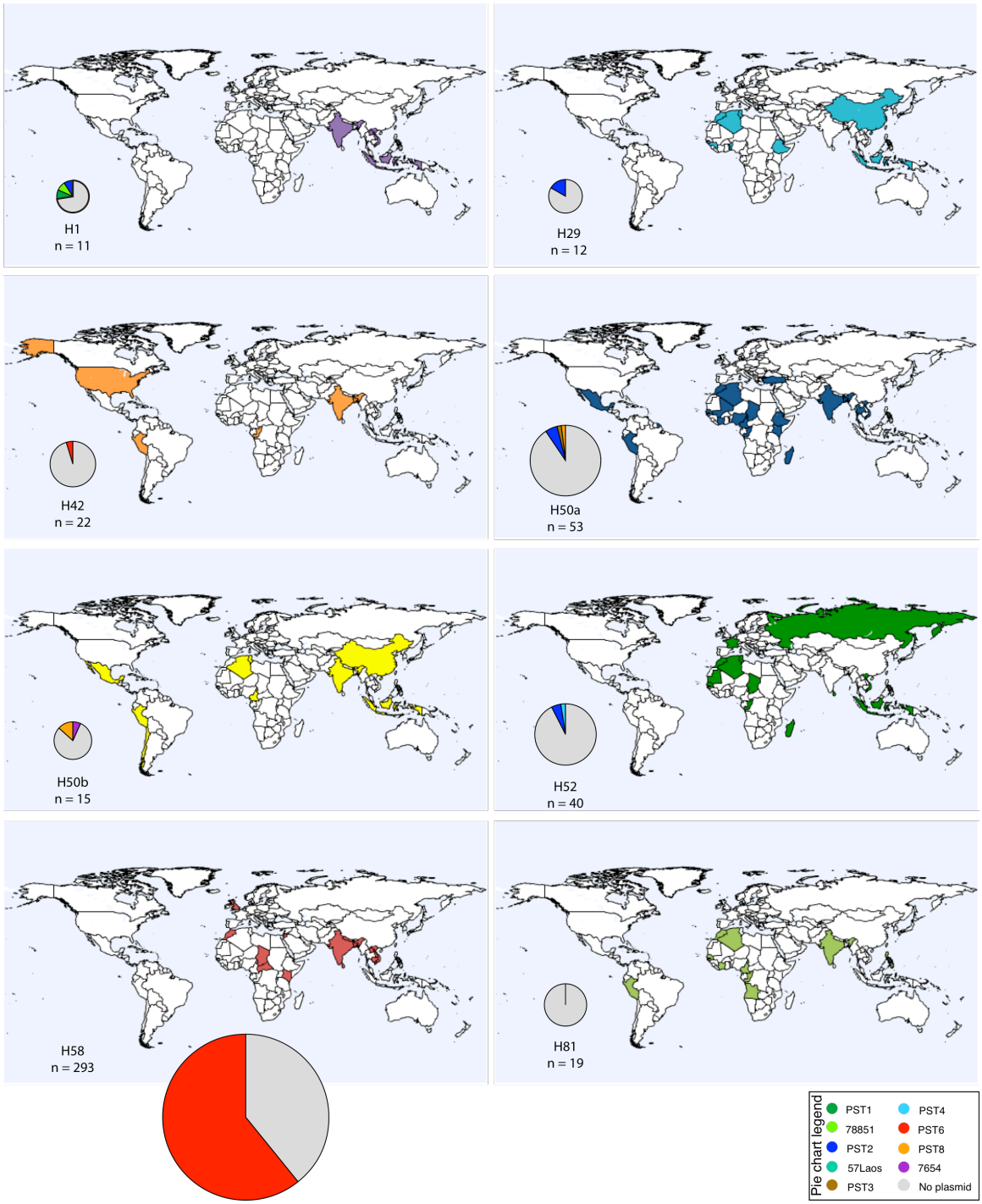


Figure 5-5 The distribution of individual haplogroups, and their plasmids, across the world. Colours within the map represent the distribution of the haplotype stated underneath the pie chart. The pie chart represents the number of isolates (size) and the colour the relative numbers of plasmid types.

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.

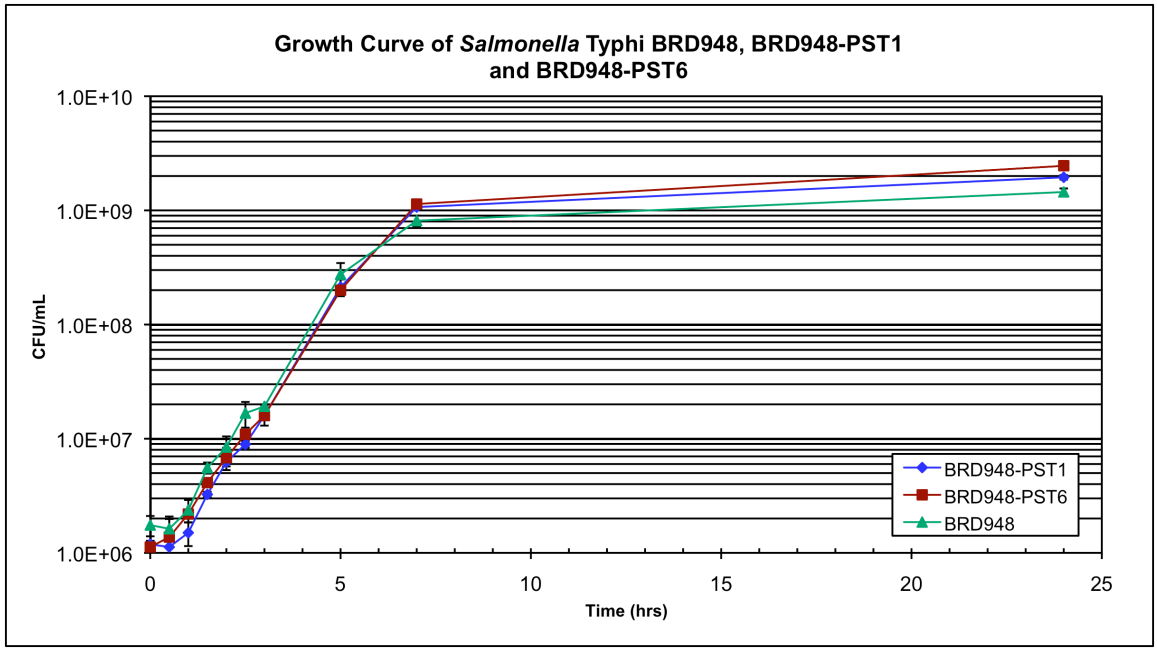


Figure 5-6 Comparing the growth of *S.* Typhi harbouring PST1 and PST6 plasmids.

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.