Chapter 6

Investigating Typhi populations using high throughput SNP typing

6.1 Introduction

Molecular typing of Typhi isolates is important in a variety of research and surveillance contexts. The aim is generally not just to distinguish genetically distinct groups of isolates, but to determine the phylogenetic relationships between those groups. It is therefore important that the genetic differences (mutations) uncovered by molecular typing are phylogenetically conserved, that is inherited directly during bacterial replication and not horizontally transferred or likely to revert to wildtype (e.g. temperate phage, genomic rearrangements). They should also be interpretable in a phylogenetic context, and reproducible in different sample populations and in different laboratories. Techniques currently in common use for typing of Typhi isolates (PFGE, ribotyping, phage typing) do not meet these criteria (see 1.3.1.2). Sequence-based analysis, however, provides the ultimate typing tool. Multi-locus sequence typing (MLST) has been introduced with great success in a wide variety of bacteria (458, 460). However Typhi is a relatively young serovar with very little sequence variation and current MLST schemes have virtually no power to discriminate within the Typhi population (1, 464).

Having compared whole-genome sequences of 19 Typhi isolates, we now have a set of nearly 2,000 SNPs with which to type large populations of Typhi isolates. By definition, typing these SNPs in Typhi populations will not lead to the identification of novel SNP loci, as MLST can within populations of sufficient nucleotide variation. However the comparative genome analysis revealed no regions of the Typhi genome that were sufficiently variable to construct a new Typhi-specific MLST scheme with sufficient resolution. A SNP typing scheme, based on 88 SNP loci identified in 66 Typhi gene fragments (2), has been used successfully to type Typhi isolates using Sequenom assays (256). However the scheme covers just a tiny fraction of the genome and provides limited resolution, particularly among the more common strain types, for instance H58 in South East Asia (2) or H59 in Indonesia (256). Furthermore, the Sequenom assay is difficult to scale up to large numbers of SNP loci, which will be important for large population-based studies of Typhi. To allow higher resolution SNP typing for Typhi, a genome-wide SNP typing scheme was developed using the GoldenGate platform (Illumina) (733, 734). This platform allows simultaneous typing of up to 1,536 SNP loci in 96 samples, and multiple sets of 96 samples can be assayed together with very small increases in handling time. In this chapter, the design and validation of GoldenGate arrays for SNP typing in Typhi is described, including not only chromosomal SNPs but also loci on the IncHI1 multidrug resistance plasmid, z66-encoding linear plasmid and cryptic plasmid pHCM2. This is followed by studies of a global collection of Typhi isolates, including many with MDR IncHI1 plasmids, and Typhi isolates from four endemic areas, demonstrating how high throughput SNP typing can contribute to studies of evolution and disease epidemiology in Typhi and potentially other bacterial pathogens.

Previous studies of Typhi populations have relied upon PFGE, phage typing or ribotyping to identify subpopulations of clones and study changes over time or space. For example, a study of 142 Typhi isolates from Northern, Central and Southern Vietnam from 1995-2002 using PFGE, Vi phage typing and ribotyping found that three quarters of the isolates from this period were attributable to one or two clones (the authors were unsure whether or not two distinct phage types represented two distinct clones) (735). Another study reported similar findings for Vietnamese Typhi isolates, but much more variability for isolates from Hong Kong (736). A study of isolates from four typhoid

outbreaks in Southern Vietnam in 1993-1997 used PFGE and phage typing to show that each outbreak was caused by a single clone, although different outbreaks were caused by different clones (380). Several other studies of typhoid outbreaks have attributed the outbreaks to a clonal source (357, 359, 474, 737, 738, 739), although there are also studies reporting diverse clones during outbreaks (740, 741, 742, 743, 744). These studies appear to reflect two different kinds of outbreaks: those associated with infection from a single source of Typhi, often an asymptomatic carrier (359) or a contaminated water source in a non-endemic area (357, 737); and those associated with a generalised increase in exposure to Typhi bacteria through a contaminated water supply in an endemic area (740, 742, 743).

The highest incidence of typhoid fever, among both inhabitants and travellers, occurs in Asia and in particular southern Asia (see Figure 6.1) (10, 11). In these high incidence regions, typhoid affects children more than adults, with the vast majority of patients aged under 20 (10). This chapter involves analysis of Typhi isolates collected from three sites where typhoid fever is particularly common - the Mekong Delta region of Vietnam, the city of Kathmandu in Nepal and an urban slum in Kolkata (Calcutta) in the east of India (see Figure 6.1). The isolates were collected during distinct studies in each site, the details of which are given in Table 6.1. The Mekong Delta study was a two-year hospital-based treatment study involving children and adults with typhoid fever (745), led by Christiane Dolecek of the Oxford University Clinical Research Unit (OUCRU) at the Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam. The Kathmandu study was a two-year hospital-based study of the burden of vaccine-preventable bacterial disease in children under 13 (746), led by Andrew Pollard of the Department of Pediatrics, University of Oxford, UK. The Kolkata study was a population-based cohort study designed to assess the burden of typhoid fever and the efficacy of a novel Vi conjugate vaccine (343, 421, 747), led by Shanta Dutta at the National Institute for Cholera and Enteric Disease (NICED) in Kolkata, India and John Clemens of the International Vaccine Institute (IVI), Korea. Typhi isolates from a fourth site, Nairobi in Kenya, were also analysed. Kenya, like most of Africa, is also considered an endemic area for typhoid fever but has a medium level of incidence (10-100/100,000 annually (10), see Figure 6.1). DNA from 96 Typhi isolates, collected in Nairobi as part of surveillance studies over a 21 year period, was provided by Sam Kariuki of the Kenya

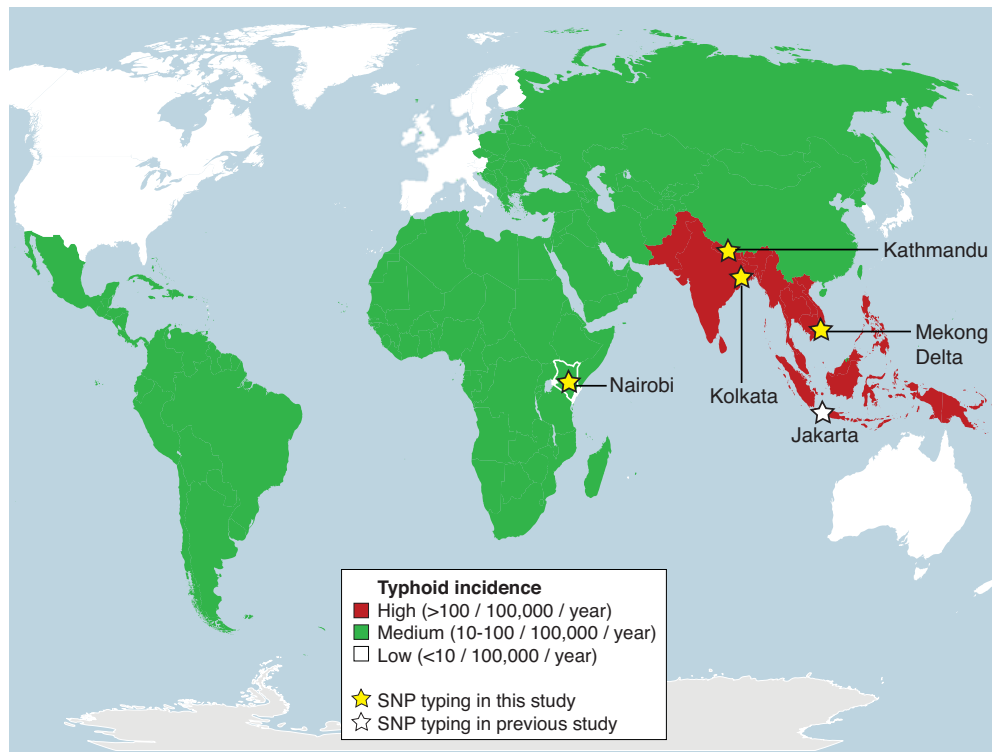Medical Research Institute (KEMRI), Nairobi, Kenya.

Figure 6.1: Typhoid incidence around the world and SNP typing study sites - Incidence levels are taken from a meta-analysis of typhoid incidence data and regional extrapolatations by Crump *et al.*, published in the Bulletin of the WHO, 2004 (10). A previous SNP typing study of Typhi in Jakarta was published by Baker *et al.* in 2008 (256).

	Mekong Delta, Vietnam	Kathmandu, Nepal	Kolkata, India
Site	Rural	Urban	Urban
Wet season	Apr-Nov	Jun-Aug	Jun-Sep
Typhoid incidence	78-198/100,000 (328)		214/100,000 (327)
Study design	Hospital-based	Hospital-based	Population-based
Typhoid patients	Adults and children	Children <13 years	Adults and children
Time period	Jun 2004 - Feb 2006	Apr 2005 - Dec 2006	May 2003 - Jan 2007
Collaborator	Christiane Dolecek	Andrew Pollard	Shanta Dutta
	OUCRU	Oxford	NICED & IVI
Reference	(745)	(746)	(343, 421, 747)
Typhi isolates	358	46	188

Table 6.1: Study sites for SNP typing of localised Typhi populations - Details of studies during which Typhi isolates were collected and later made available for SNP typing with the GoldenGate assay. Note that 96 isolates collected during surveillance studies in Nairobi, Kenya between 1988-2008 were also SNP typed.

6.1.1 Aims

The aim of the work presented in this chapter was to design and validate a novel high throughput SNP typing assay for Typhi and to apply the assay to study the population structure of Typhi from both global and regional perspectives. Specific aims were to:

- design a GoldenGate assay to type Typhi chromosomal and plasmid SNPs identified in Chapters 2 and 5;

- develop analysis methods and validate the assay using data from control isolates with known alleles;

- compare the assay to previously published SNP typing data for a global collection of isolates;

- study the movement of IncHI1 MDR plasmids in the Typhi population by analysing both IncHI1 plasmid types and Typhi chromosomal haplotypes of their host strains; and

- investigate the structures of Typhi populations circulating in endemic areas, including looking for correlations between clinical or epidemiological features of typhoid and particular Typhi haplotypes.

6.2 Methods

6.2.1 DNA preparation and quantitation

DNA was provided for SNP typing analysis by the collaborators listed in Appendix E. DNA was extracted from the 19 sequenced isolates by Stephen Baker, Robert Kingsley and Derek Pickard at the Sanger Institute.

Accurate analysis of the data generated by the GoldenGate assay depends on DNA from each sample being present in equal amounts, thus DNA quantitation is an important step in the preparation of DNA for SNP typing. Quantitation was performed using the Quant-iT PicoGreen dsDNA reagent (Invitrogen) which fluoresces when bound to double stranded DNA (dsDNA) but not when bound to nucleotides, single stranded DNA, RNA or contaminants. Quantitation was performed in black flat-bottomed 96-well plates (the PicoGreen reagent is light sensitive). To the first two columns were added 90μL of buffer (provided in the Quant-iT kit) and 10μL of DNA standards at concentrations of 0, 5, 10, 20, 40, 60, 80 or 100 ng/μL (Invitrogen). One μL of sample DNA and 99μL buffer was added to each of the remaining 80 wells. A 1:100 mix of Quant-iT PicoGreen dsDNA reagent was mixed with buffer and 100μL added to each well, to reach a total volume of 200μL per well. Plates were covered for three minutes, then the fluorescence read using a FluoSTAR (Omega) fluorescence microplate reader and standard fluorescein wavelengths (excitation 480 nm, emission 520 nm). A standard curve was calculated from fluorescence of the DNA standards and used to determine the concentration of each sample (using Omega STAR software). DNA samples were arrayed in 96-well plates for genotyping, with each plate including one blank well (water) and duplicate wells of at least one sequenced isolate as a control. DNA concentrations were adjusted so that each well containined 30 μL of DNA solution at a concentration of 10 ng/μL. DNA quantitation and preparation was performed by myself (sequenced isolates, Pasteur isolates, Kathmandu isolates), Christiane Dolecek (Mekong Delta isolates), Minh Duy Phan (IncHI1 plasmid control isolates) and Derek Pickard (Kolkata and Kenya isolates).

6.2.2   Illumina GoldenGate assay

The GoldenGate assay (Illumina) allows simultaneous genotyping of up to 1,536 loci in a single DNA sample. For each SNP locus, three oligonucleotides are designed: a locus-specific primer joined to an addressing sequence and a universal PCR primer sequence; and two allele-specific primers, each joined to a different universal PCR primer sequence. Up to 1,536 of these oligonucleotide sets are combined to form an oligonucleotide pool. The oligonucleotide pool is hybridised to the DNA sample, followed by extension and ligation to generate a single allele-specific product for each SNP locus, which incorporates both allele-specific and locus-specific sequences. These products are amplified using universal PCR primers; those that bind the allele-specific PCR primer sequences are Cy3- and Cy5-labelled. The resulting amplicons are bound to their complement bead types via the locus-specific addressing sequences and the presence of specific alleles on each bead type is measured via detection of Cy3 or Cy5 fluorescent signal. All steps in the GoldenGate protocol were performed by members of the dedicated genotyping lab at the Sanger Institute, under the guidance of Ranganath Bangalore Venkatesh, Rhian Gwilliam and Panos Deloukas.

SNP alleles and 100 bp sequences flanking each SNP locus were submitted to Illumina for design of oligonucleotides for use in the GoldenGate assay. The design process takes into account interference between oligonucleotides targeted to different loci in the same pool and the potential for non-specific binding in the genome. In particular, two SNPs that are less than 60 bp apart can not be targeted in the same oligonucleotide pool, as it is not possible to design non-overlapping oligonucleotides on either side of both SNPs. In this study two oligonucleotide pools were used, each containing 1,536 SNPs, so SNPs less than 60 bp apart could be targeted by assigning each SNP to a different pool. However if more than two SNPs were pesent within 60 bp, they could not all be typed using just two GoldenGate arrays. Furthermore, if two SNPs occur within 10 bp of each other, the variation can interfere with the binding of allele-specific oligonucleotides and so such SNPs should not be typed using GoldenGate. During the design of Typhi oligonucleotides, 1.8% of known SNP loci could not be targeted due to these issues.

6.2.3   Genotype calling from raw data

The raw data provided by the GoldenGate assay is, for each SNP in each sample, a fluorescence signal corresponding to each allele-specific fluorescent-labelled probe. Raw data was normalised using the proprietary Illumina BeadStudio software, but the mean normalised signal intensities and signal:noise ratios vary among SNPs. Thus turning allele-specific signals into genotype calls requires each SNP to be analysed individually, across a range of samples. This process is known as genotype clustering, and is essentially a two-dimensional clustering problem, where each allele-specific probe contributes one dimension. The problem is best illustrated by two-dimensional cluster plots, like those in Figure 6.2. DNA samples in which one allele is present will cluster at a point along the axis corresponding to the fluorescent marker attached to amplicons containing that allele. For example in Figure 6.2 samples lying in the x-axis cluster contain allele A, while samples lying in the y-axis cluster contain allele B. In a haploid bacterial sample this would imply the x-axis cluster allele or haplotype is A, and the y-axis cluster allele or haplotype is B (Figure 6.2a); in a diploid human sample this would imply the x-axis cluster genotype is AA and the y-axis cluster genotype is BB (Figure 6.2b). Among diploid samples like human DNA, heterozygotes are common, whereby both alleles are present in a single sample. For example in Figure 6.2b a number of samples cluster around a third point which has positive signal for both alleles, this cluster indicates the heterozygous genotype AB. In haploid bacterial samples we would not expect to see heterozygous clusters, unless the SNP locus is present in multiple copies within the bacterial genome.

In the present study care was taken not to include such SNP loci (see 2.3.1.5), so heterozygous clusters are not expected. However several loci were included that are expected to be absent from some samples, for example plasmid SNPs. These SNPs should generate a cluster around the origin of the graph, as in Figure 6.2c-d, corresponding to a lack of signal for either allele-specific probe at the SNP locus. The same lack of signal would be observed if the allele-specific probes failed to bind for some other reason, e.g. (a) a secondary SNP or indel mutation was present within one of the primer-binding sites or (b) a third allele was present at the SNP locus which would fail to bind to either of the two allele-specific primers. Thus while these 'no signal' clusters
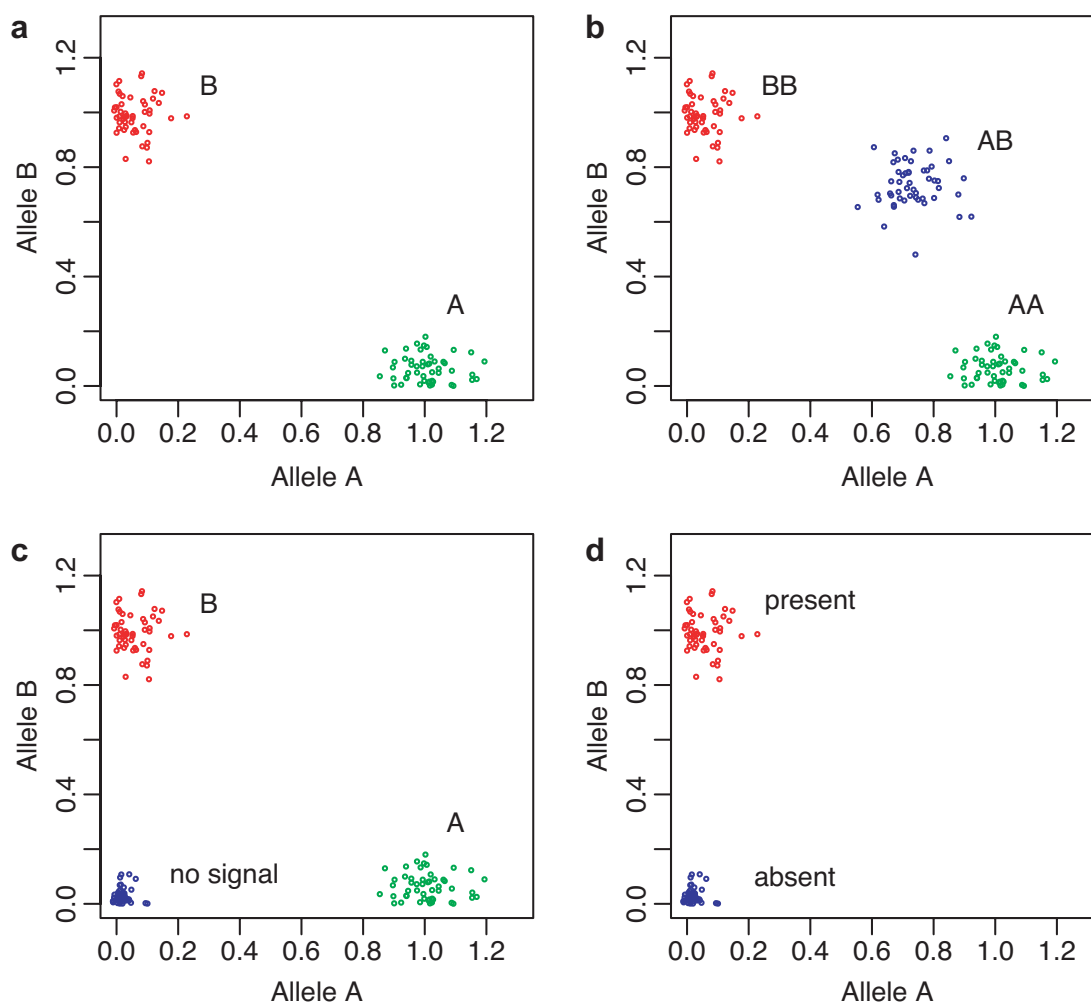
Figure 6.2: Example cluster plots for genotyping assays - Genotyping with the Illumina GoldenGate and similar array-based assays generate allele-specific signals for each SNP in each sample. By plotting these values for all samples at a given SNP, clusters can be identified which represent different genotypes at that SNP locus. (a) Haploid clusters at a SNP locus. (b) Diploid clusters at a SNP locus, with heterozygotes in blue. (c) Haploid clusters at a SNP locus with some samples displaying no signal for either allele; e.g. for plasmid SNPs, this would represent strains with either allele (A, B) or no plasmid present (no signal). (d) Haploid clusters at a nonpolymorphic locus with some samples displaying no signal; e.g. for a nonpolymorphic plasmid locus that is absent from some strains.

most likely indicate absence of known plasmid or chromosomal deletion loci, more care must be taken in the interpretation of 'no signal' clusters for chromosomal SNP loci.

6.2.3.1   Genotype calling with Illuminus-P

Since each SNP needs to be clustered individually, manual genotype clustering is extremely time consuming and can introduce bias into genotyping results. An automated clustering algorithm is implemented in the Illumina software (BeadStudio), but has been optimised for clustering genotypes in diploid samples, i.e. where three clusters are expected for each locus, corresponding to two homozygous and one heterozygous genotype (AA, BB or AB, as in Figure 6.2b). However for the present study involving haploid bacterial genotyping, we expect no heterozygous genotype clusters but some legitimate 'no signal' genotype clusters (Figure 6.2c-d). Illumina BeadStudio is proprietary software and unable to be modified, but third-party opensource genotype clustering algorithms are available. Among the best performing for the Illumina genotyping platform is Illuminus (748). On request the software's author, Yik Y Teo (Wellcome Trust Centre for Human Genetics, University of Oxford, UK), modified Illuminus to fit a third 'no signal' cluster centred at the origin rather than a heterozygous cluster. This version, referred to hereafter as Illuminus-P, was applied to genotype clustering of all GoldenGate data presented in this chapter.

6.2.3.2   Heuristic to identify 'no signal' cluster

Despite the modification, Illuminus-P occasionally failed to differentiate the cluster around the origin from the other clusters. To remedy this, a heuristic was applied following genotype clustering with Illuminus-P, such that for each SNP locus, samples with both allele signals below 15% of the maximum observed signal at that locus were assigned to the 'no signal' cluster. This heuristic improved genotype clustering accuracy for plasmid but not chromosomal loci and was applied to all plasmid loci in the analysis presented in this chapter.

6.2.3.3   Clustering across plates

Since clustering requires each SNP to be considered individually across a range of samples, the accuracy of genotype clustering depends on the number and uniformity of the

samples. Theoretically, increasing the number of data points increases the accuracy of cluster definition, so the more samples the better the performance. However clustering can be affected by the presence of outlier samples or subsets of samples. DNA samples are prepared for genotyping in 96-well plates, with all steps in the GoldenGate assay (PCR, array hybridisation and scanning) performed in 96-well format. We therefore expect to see little variation within 96-sample sets due to technical factors, but variation between 96-sample sets is sometimes observed. For example, one 96-sample set may display lower mean signal intensity for some SNPs compared to other sample sets, presumably due to run-specific technical factors such as reagent concentration or hybridisation conditions. In such cases including the plate with lower mean signal will diminish the clustering performance across all samples at loci that are most sensitive to technical variation.

For consistency in the present study, all sample sets were clustered together except where there was evidence of poor performance. Specifically, each 96-well plate was clustered individually in addition to clustering across all available sample sets. Each 96-well plate contained at least one control sample (a Typhi isolate whose genome has been sequenced and therefore all alleles were known), usually in duplicate. The accuracy of genotype calling in these control samples using individual-plate clustering was compared to that for all-plate clustering. Only four plates, all from the second oligonucleotide pool, gave better results on individual clustering. Each of these four plates displayed low mean signal intensity compared to the other plates and gave better results when clustered together as a set (compared to individual clustering or clustering with all other plates).

6.2.4 Phylogenetic analysis of genotyping data

Alleles determined by genotype clustering with Illuminus-P were analysed as follows. A Perl script was written to extract allele data from subsets of high-quality SNPs (determined by comparison with alleles for control isolates, see below 6.3.1.1) and samples from the clustered data set, see Figure 6.3. For example, chromosomal SNP loci were extracted for analysis separately from plasmid SNPs; samples from different data sets were analysed separately (e.g. global collection, individual local studies, etc). The
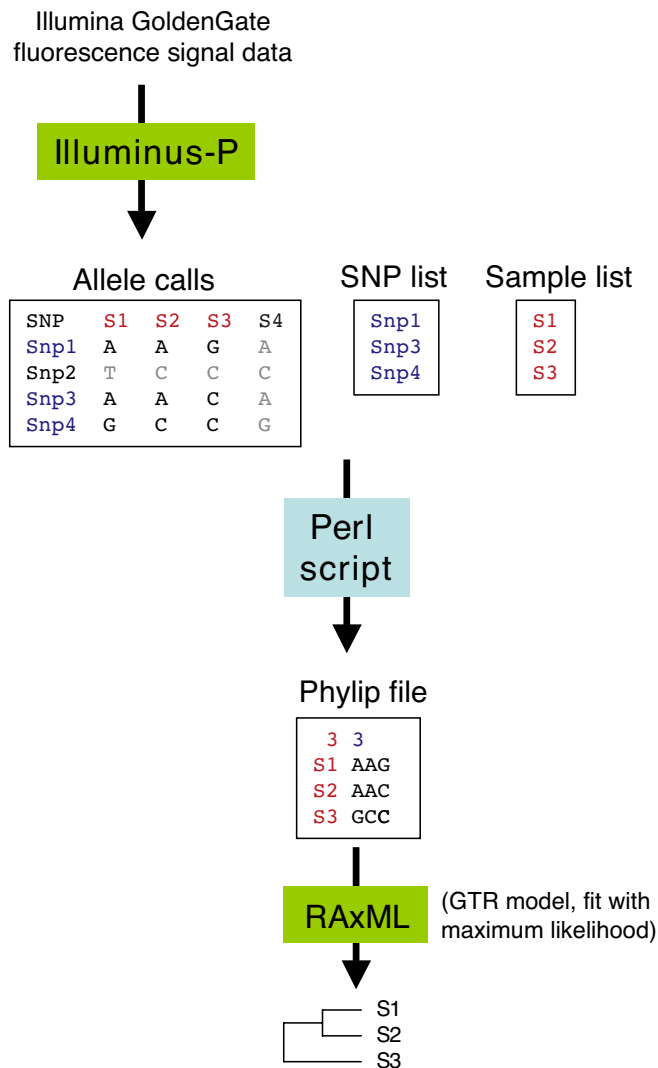
Figure 6.3: Phylogenetic analysis workflow for SNP typing studies - Illuminus-P was used to cluster fluorescence data across samples (including control and experimental samples from multiple studies) and call SNP alleles. A Perl script was written to extract allele data for subsets of SNPs (in this case Snp1, Snp3, Snp4) and subsets of samples (in this case S1, S2, S3) and convert it into a form suitable for phylogenetic analysis by RAxML.

script outputs alignments of concatenated strings of SNP alleles in phylip format, suitable for analysis using a range of phylogenetics software. Chromosomal alleles from a global collection of 180 isolates plus 19 sequenced isolates was analysed in ModelTest (645) (implemented in FindModel (646)), which suggested a general time reversible (GTR) model was most appropriate for analysis of this data. All phylogenetic analysis presented in this chapter was performed using maximum likelihood approaches to fit a GTR model, implemented in the software program RAxML (644). SNP typing with the GoldenGate assay only provides genetic information at the specific assayed loci; in the case of chromosomal SNPs, these are mostly loci determined by whole genome comparison of 19 Typhi strains (Chapter 2). Thus the SNP typing analysis essentially places each Typhi isolate at the appropriate position along branches defining the phylogenetic tree of these 19 strains (Figure 2.6) and branch lengths reflect genetic divergence only at the assayed loci. Short branches separating very closely related clusters (e.g. within the H58 group) were verified by manually inspecting cluster plots for SNPs that differentiated within those clusters. Similarly, plasmid SNPs were originally determined from comparison of a few plasmid sequences (Chapter 5) and thus genotyping at these loci essentially assigns each plasmid to a position in the phylogenetic tree defined by these SNPs (Figure 5.6).

6.2.5 Visualisation of temporal and spatial data

Data analysis was performed using the open-source statistical programming package R (749). For the Kathmandu data set, typhoid cases were so infrequent as to be easily visualisable using monthly counts; for the Kenya dataset years but not precise dates of isolation were available, so annual counts were used. For the Mekong Delta and Kolkata data sets typhoid cases were more frequent, so a smoothing function was used to visualise the incidence rate using the recorded day of isolation. Temporal distributions of Typhi isolates were plotted using R's `density` function for kernel density estimation, using days since study began to indicate the time of each isolation and bandwidth of 10 days.

Spatial clustering within the Kolkata data set was performed using Openshaw's Geographical Analysis Machine (GAM) (750), implemented in the R package `DCluster` (751). Note this is intended to highlight clusters of unusually high density but does not

provide a formal statistical test of spatial clustering, hence additional data and statistics is provided for each cluster discussed in the text. The centre of each geographical cluster (i.e. the clusters defined by the study team for distribution of vaccine) was determined using a spatial grid in which the smallest unit corresponded to the size of the smallest cluster (cluster 44). The observed and expected number of cases for each cluster i were defined as follows:

$$observed_i = n_i, \tag{6.1}$$

$$expected_i = \frac{p_i}{P} * N, \tag{6.2}$$

where n_i = number of cases observed in cluster i, N = $\sum_{i=1}^{k} n_i$, p_i = population of cluster i, P = $\sum_{i=1}^{k} p_i$. The `opgam` method of `DCluster` (751) was used to assess whether or not each point in the spatial grid represented a cluster of cases, based on a random Poisson distribution of cases among the study population; statistically significant clusters (p<0.005) were plotted.

6.2.6 Simpson's diversity index

Simpson's diversity index, 1-D was calculated as follows:

$$1 - D = 1 - \sum_{i=1}^{k} \frac{n_i * (n_i - 1)}{N * (N - 1)}, \tag{6.3}$$

where n_i number of Typhi isolates of haplotype i, N = total number of Typhi isolates.

Diversity indices were compared for two sets of Typhi isolates using Student's T-test, as follows:

$$s_x^2 = \frac{4}{N} * (\sum_{i=1}^{k_x} p_{i,x}^3 - \sum_{i=1}^{k_x} p_{i,x}^2)^2, \tag{6.4}$$

$$T = \frac{D_2 - D_1}{\sqrt{s_1^2 + s_2^2}}, \tag{6.5}$$

$$df = k_1 + k_2 - 2, \tag{6.6}$$

where $p_{i,x} = \frac{n_i}{N}$ for sample x, k_x = number of taxa in sample x.

For the Kolkata Typhi data, s^2 prior to December 2004 was 0.0005, s^2 from January 2005 was 0.0015, thus T = 82.85 with 14+10-2 = 22 degrees of freedom (df), which has a p-value of $<1 \times 10^{-6}$.

6.3 Results

6.3.1 Validation of GoldenGate assay for target loci in Typhi

The accuracy of the GoldenGate assay and genotype clustering for the two Typhi oligonucleotide pools was assessed using data from the first ten 96-well plates run on the arrays. These samples include all sequenced isolates, 180 isolates previously genotyped at 88 loci (isolates from the Pasteur Institute, see Appendix E (2)) and several hundred isolates from diverse regions across Asia and some from Africa (including Mekong Delta and Kathmandu collections, part of Kenya collection). Each of the 19 sequenced isolates was assigned alleles at each of these SNP loci based on sequencing data (Chapter 2). Each plate contained at least one of the sequenced isolates, often in duplicate, and comparison of the expected alleles with those assigned following genotype clustering of GoldenGate data was used to assess the accuracy of the GoldenGate SNP typing results.

6.3.1.1 Chromosomal loci

The GoldenGate assay utilises mega-plex PCR (up to 1,536 oligonucleotide nucleotides per pool) followed by hybridisation to custom bead arrays (see 6.2.2). Due to the PCR step, it is not possible to uniquely target two SNP loci separated by less than 60 bp in a single oligonucleotide pool. It is also not possible to assay any SNP locus that lies within 10 bp of another SNP, insertion or deletion, as these variants can interfere with primer binding. For these reasons, 35 (1.8%) of the 1,964 SNPs identified in Chapter 2 were not suitable for SNP typing with GoldenGate. Oligonucleotides were designed to target the remaining 1,929 SNPs, as well as 72 additional SNP loci ("BiPs") identified by Roumagnac *et al* (2). As explained above, the GoldenGate assay generates fluorescence signals for each SNP locus in each sample, which must be converted into genotype calls using a two-dimensional clustering approach (see 6.2.2). Illuminus (748), originally designed for clustering of diploid human genotypes, was used to assign alleles to the Typhi data. This resulted in perfect allele calls in control samples for just 1,104 SNPs (57%). A modified version of Illuminus, Illuminus-P (see 6.2.3.1) gave better results, with perfect allele calls in control samples for 1,436 SNPs (74%). Thus Illuminus-P was used for the remainder of this study.

Analysis of GoldenGate data for sequenced isolates was used to determine whether each SNP was (a) assayed successfully, (b) clustered accurately and (c) truly polymorphic as expected from the sequence data. Each SNP assay was considered successful if it generated signals of reasonable strength that were able to be clustered. For 1,402 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 detected no evidence of the derived allele (confirmed by manual inspection of the signal plots in addition to Illuminus-P clustering). These SNPs were originally detected in sequence data from just one isolate each and are likely to be genuinely nonpolymorphic sites. These SNPs were concentrated in isolates E01-6750 (5 SNPs), E03-9804 (3 SNPs), J185SM (3 SNPs) and M223 (3 SNPs), and were not included in downstream analysis. Note that none of these were nonsense SNPs, so would not affect analysis of pseudogenes or selection in Chapter 2. For a further 12 SNPs, GoldenGate analysis found evidence of the derived allele, but allele assignments did not match exactly those from sequence data. Signal plots for these SNPs were manually inspected and assessed to be good quality signals and accurately clustered. Therefore these loci are truly polymorphic, but the earlier sequence-based allele assignments likely contained some errors which can now be corrected by the GoldenGate analysis. These errors were concentrated in J185SM (4 SNPs), M223 (3 SNPs) and E01-6750 (3 SNPs), and resolve two apparent homoplasies (reported in Table 2.9) in genes STY0347 (*tsaC*) and STY1689 (*ydhD*).

Thus 1,448 SNP loci (75% of those targeted) were considered successful GoldenGate assays of polymorphic loci identified previously from sequence analysis. While the high rate (25%) of failed assays means a reduction in resolution, the SNP loci concerned were distributed evenly within the phylogenetic tree defined by the complete SNP set (Figure 2.6). Figure 6.4 shows the length of each branch as defined by the complete SNP set (x-axis) vs that defined by the successfully assayed SNPs (y-axis), demonstrating that these lengths are highly correlated (Pearson $R^2 = 0.994$, $p<2 \times 10^{-16}$). Thus the loss of resolution does not lead to significant change in relative branch lengths of the resulting phylogenetic tree.
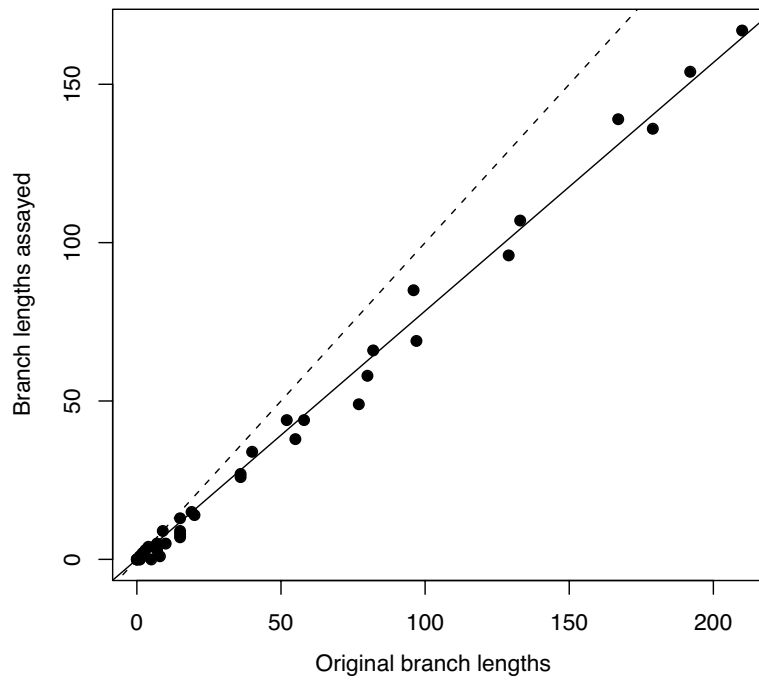
Figure 6.4: Effect of assay failure on relative branch lengths for Typhi chromosomal SNPs - Solid line indicates linear model fit (Pearson $R^2 = 0.994$), dashed line indicates y=x.
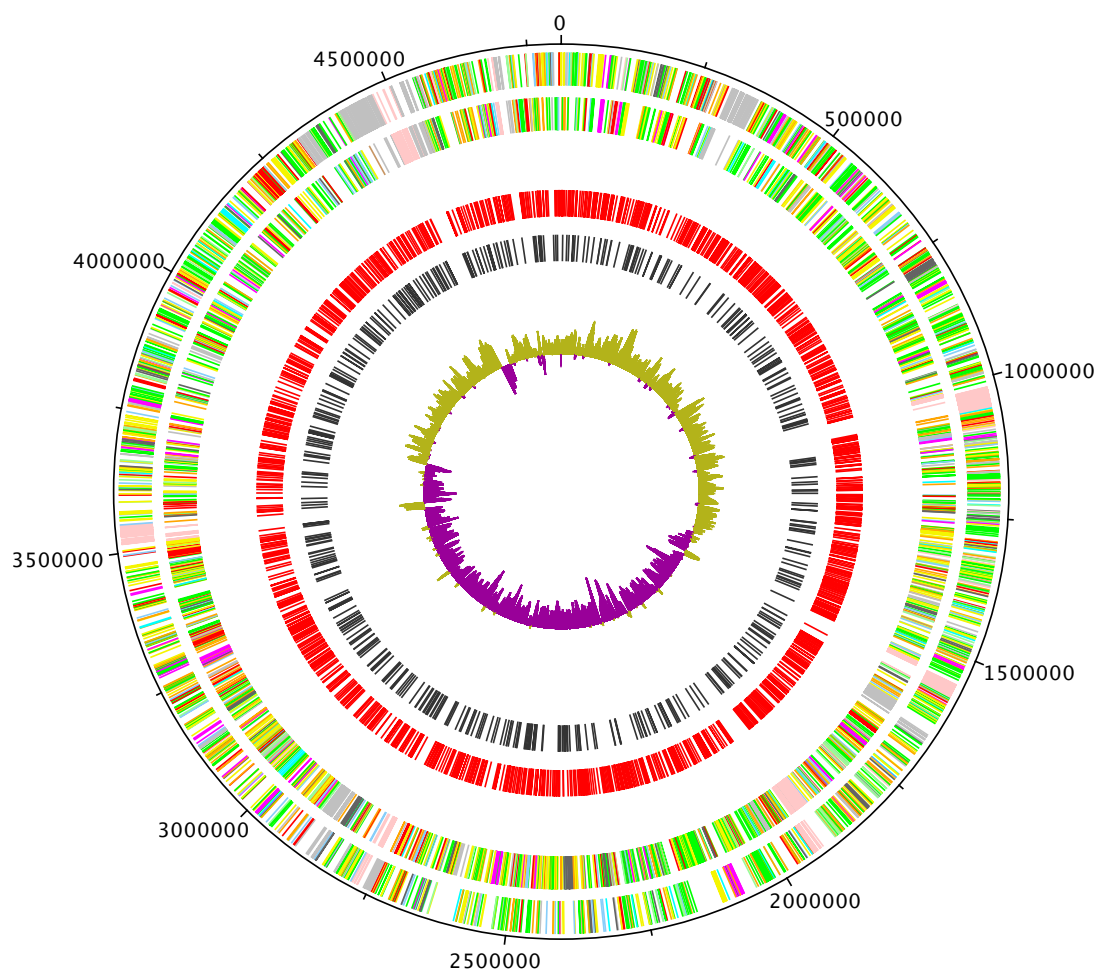
Figure 6.5: Distribution of assayed SNPs in the Typhi CT18 chromosome - Coordinates are bp in the Typhi CT18 genome. Genes annotated in the CT18 genome are shown in the two outer rings: outer-most ring = forward strand, second outer-most ring = reverse strand. Genes are coloured according to broad functional groups, note phage genes are coloured pale pink. Red and black rings show the location of all SNPs detected so far in the Typhi population (red) and those successfully assayed with GoldenGate (black). Inner-most rings indicate GC deviation in the CT18 genome ((G-C)/(G+C)), i.e. the difference in G content between the forward and reverse strands).

Signal plots were manually inspected to assess the GoldenGate assays of 72 SNPs defined by Roumagnac *et al* (2), of which 60 (83%) were of high quality (signals of reasonable strength, clustered accurately). Thus the phylogenetic analysis of experimental Typhi isolates presented in this study is based on 1,508 chromosomal SNP loci. These loci are distributed randomly in the Typhi chromosome (Figure 6.5), indicating that failures in the design, signal generation or clustering do not bias the distribution of SNPs that were assayed successfully with GoldenGate.

6.3.1.2   IncHI1 plasmid loci

A total of 345 SNPs were identified in the IncHI1 plasmid backbone in Chapter 5, 294 (85.2%) of these were included in the GoldenGate assay, and their genotypes called using Illuminus-P followed by the heuristic. IncHI1 plasmid SNPs were validated by comparing alleles from SNP typing and sequence data for the plasmids previously sequenced in Typhi isolates CT18, E03-9804, ISP-03-07467 and ISP-04-06979, the Paratyphi A plasmid pAKU_1 and the Typhimurium plasmid R27, as well as plasmids used to develop the pMLST scheme (Chapter 5, (725)). Alleles from genotyping and sequencing matched perfectly for 200 IncHI1 plasmid SNPs (68.0%) and these loci were used for the remainder of the study. The distribution of these SNPs is shown in Figure 6.6. Positive allele calls for IncHI1 plasmid SNPs (i.e. fluorescence signals clustered outside the 'no signal' cluster) correlated strongly with presence of the plasmid. The four control isolates known to contain IncHI1 MDR plasmids each gave positive allele calls for 193-197 of the 200 IncHI1 plasmid SNP loci, whereas the IncHI1 plasmid-free control isolates each gave positive signals for 0-5 plasmid SNPs. To assess whether excluding the poorly-assayed SNPs would bias phylogenetic analysis of the IncHI1 plasmid, the number of SNPs lying on each branch of the phylogenetic tree was compared for (a) the complete set of 345 SNPs identified in Chapter 5 and (b) 200 SNPs with 100% genotype calling accuracy. The branch lengths represented by the two SNP sets were highly correlated (R^2=0.964) (see Figure 6.7). This demonstrates that the exclusion of low quality SNPs does not significantly alter the relative branch lengths of phylogenetic trees determined from analysis of the SNP typing data and therefore will not bias estimates of relative genetic distance between samples. Note however that since 32% of SNP loci are being excluded, genetic distance will be underestimated by phylogenetic analysis of the SNP data.
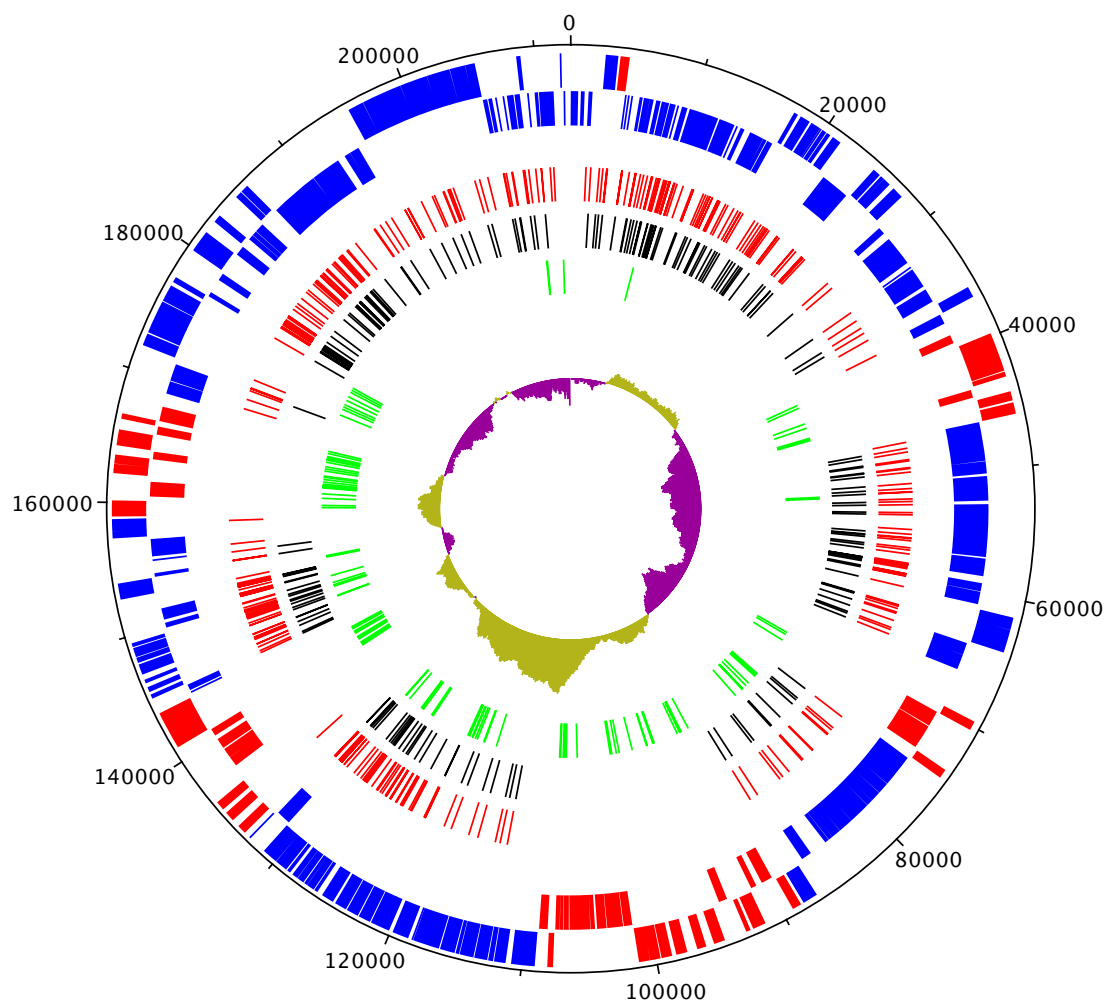
Figure 6.6: Distribution of assayed SNPs in the IncHI1 plasmid - Coordinates are bp in the IncHI1 plasmid pAKU_1. Genes annotated in pAKU_1 are shown in the two outer rings: outer-most ring = forward strand, second outer-most ring = reverse strand. Genes are coloured to indicate the IncHI1 conserved background (blue) or insertion sequences (red). Rings 3 and 4 show the location of all SNPs detected so far in the IncHI1 plasmid population (red) and those successfully assayed with GoldenGate (black). Ring 5 shows loci targeting resistance genes and deletions in the plasmid backbone (green). Innermost rings indicate GC deviation in pAKU_1 ((G-C)/(G+C)), i.e. the difference in G content between the forward and reverse strands).
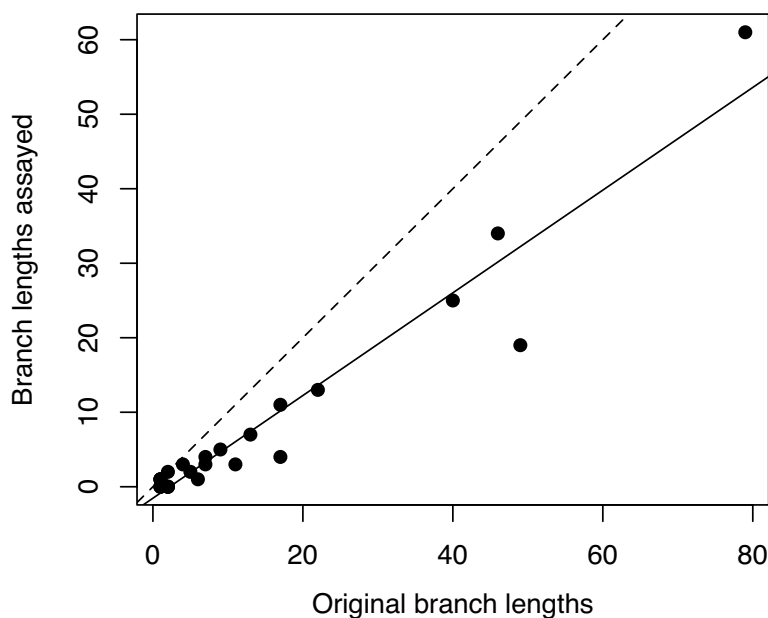
Figure 6.7: Effect of assay failure on relative branch lengths for IncHI1 plasmid SNPs - Solid line indicates linear model fit (Pearson $R^2 = 0.964$), dashed line indicates y=x.

A total of 218 SNPs designed to assess the presence or absence of resistance genes and specific IncHI1 sequences were included on the 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 (Chapter 5), shown in Table 6.2.

Gene(s)	Sequence type	Number of targets
*cat**	*Tn*9 - chloramphenicol res	3
*tetACDR**	*Tn*10 - tetracycline res	3, 3, 2, 2
*merAPRT**	*Tn*20 - mercury res	2, 1, 2, 2
*dfhR7**	integron cassette - trimethoprim res	3
*dhfR14**	integron cassette - trimethoprim res	3
*sul1**	integron cassette - sulfonamide res	1
bla$_{TEM-1}$*	bla/sul/str - beta-lactam res	2
*sul2**	bla/sul/str - sulfonamide res	1
*strAB**	bla/sul/str - streptomycin res	2, 2
IntI1	integrase of class I integron	1
tnpAB	*Tn*21 transposase	2, 2
*IS*10-L,-R	*Tn*10 transposase	2, 2
*IS*26	transposase	2
*IS*30	transposase	2
*IS*6100	transposase	2
repC	replication initiation gene	1
betU	*Tn*6062 - betaine transporter	1
SPAP0105	*Tn*6062 - hypothetical protein	2
citAB	citrate transporters	3, 2
nac	transcriptional regulator	1
SPAP0281	hypothetical protein	3
R0140	backbone in/del	1
SPAP0266-9	backbone in/del	2, 3, 3, 2
SPAP0329	backbone in/del	2
HCM25-6	backbone in/del	1
HCM102-106	backbone in/del	3, 2, 3, 0, 1
HCM160	backbone in/del	2
HCM163	backbone in/del	2
HCM170	backbone in/del	2
HCM174	backbone in/del	3
HCM177-8	backbone in/del	1, 2
HCM189-193	backbone in/del	3, 3, 2, 1, 3
HCM195-199	backbone in/del	2, 3, 1, 0 1
HCM203	backbone in/del	2
HCM278	backbone in/del	1

Table 6.2: SNPs for detection of resistance genes and IncHI1 plasmid deletions
- *Antimicrobial resistance genes, res resistance, *Tn* transposon, *IS* insertion sequence.

235

6.3.1.3   Other target loci

Two loci provided accurate detection of the cryptic plasmid pHCM2, which was successfully detected in the two sequenced isolates CT18 and E02-2759. A further two SNPs provided accurate detection of plasmid pBSSB1 which carries the z66 flagella antigen, present in the two sequenced H59 isolates 404ty and E03-4983. Replication (*rep*) genes from another 17 plasmids, detailed in (752), were also included on the array. Control plasmids were not available to test these assays, but allele signals were only detected in two isolates, both multidrug resistant Typhi containing plasmids of different incompatibility groups (see below 6.3.2.2).

Two SNPs specific to Paratyphi A were included in the assay to facilitate detection of 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 (see Figure 6.8b).

6.3.2   Validation of GoldenGate SNP typing in a global collection of Typhi isolates, previously typed at 88 loci

DNA from a global collection of 180 Typhi isolates was provided by Francois-Xavier Weill of the Pasteur Insitute, Paris, France. The isolates, listed in Appendix E, were collected between 1958 and 2004 from travellers returning to France with typhoid fever. Their geographical origin is considered to be the country in which the patient initially became ill, and includes Africa (89 isolates), Asia (77 isolates) and South America (10 isolates). This collection was previously used to identify SNPs in fragments of the Typhi genome (2), generating the tree reproduced in Figure 6.8a.

6.3.2.1   Phylogenetic analysis of chromosomal SNPs

As expected, phylogenetic analysis of chromosomal SNPs in this collection placed each isolate at different points on the phylogenetic tree defined by the 19 sequenced isolates (Figure 6.8b). The location of each isolate according to GoldenGate SNP typing was consistent with their previously defined haplotypes based on 88 SNPs (2) (Figure 6.8a),
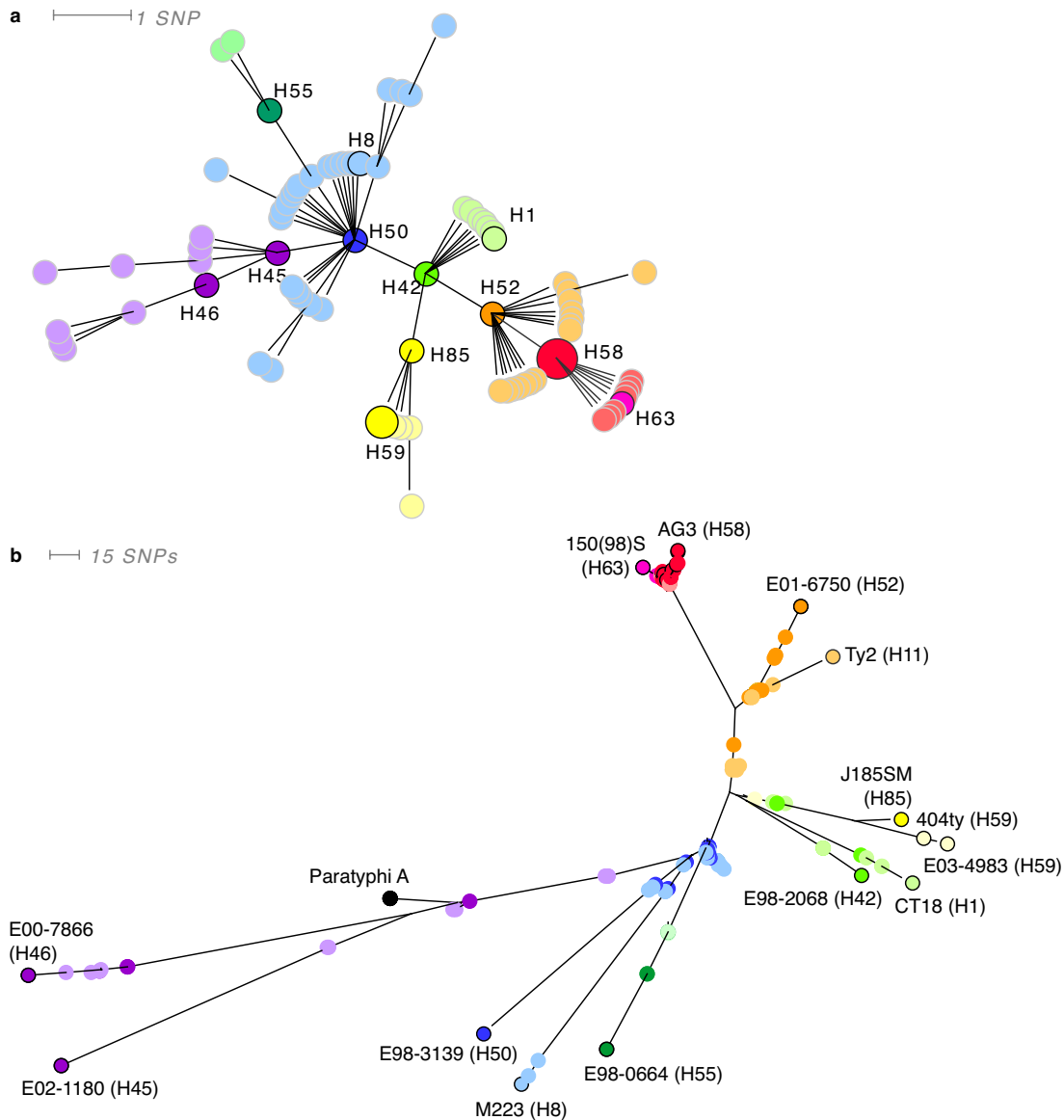
Figure 6.8: Phylogenetic trees for a global collection of 180 Typhi isolates (1958-2005) - (a) Minimum spanning tree based on 88 SNPs, reproduced from (2). Each circle (node) represents a distinct haplotype group. Haplotype groups from which isolates were sequenced are shown in bold colours with black outline, haplotypes not represented in sequencing are show in paler colours. Non-sequenced haplotypes are considered the result of clonal expansion of the sequenced haplotypes, and are coloured in a paler shade of the same colour as the haplotype from which they are assumed to have descended. (b) Maximum likelihood tree based on 1,508 chromosomal SNPs typed with the GoldenGate assay. Sequenced strains have black outlines; black circle is Paratyphi A strains, indicating the position of the root of the Typhi tree. All genotyped isolates lay along branches defined by the phylogenetic tree of sequenced strains. Colours indicate the previously determined haplotype of each isolate as given in (a); the same colour scheme is used in Figure 6.9.

demonstrating the ability of the Typhi GoldenGate assay to correctly cluster Typhi isolates into known haplotype groups.

In addition, the assay was able to differentiate within some haplotypes, even when only one member of the haplotype group had been included in SNP detection via sequencing. For example, E98-0664 was sequenced as a representative isolate of the H55 haplotype. As Figure 6.9 shows, the SNPs identified in this isolate were able to differentiate the four SNP typed H55 strains (77-303, 77-302, 75-2507 and E98-0664, coloured dark green) into three clusters lying along the branch that ends at E98-0664. Similarly, SNPs identified in M223 (H8) differentiate all three SNP typed H8 isolates (M223, E00-4626, E01-5612, coloured pale blue) and cluster H77 isolates into two groups (1458, 81424, 81918, pale blue). Differentiation between H50 isolates is also evident in Figure 6.9.
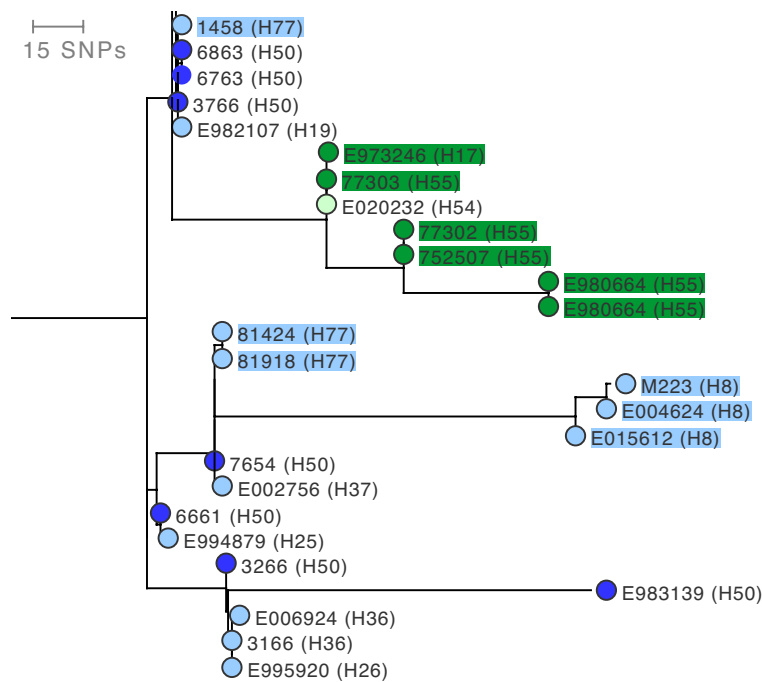

Figure 6.9: Discrimination within known Typhi haplotypes - Part of the phylogenetic tree based on chromosomal SNPs in the global collection of Typhi isolates. Isolates are coloured according to haplotype groups to which they were previously assigned (2) as shown in Figure 6.8a.

6.3.2.2 IncHI1 plasmids and multidrug resistance

A total of 40 isolates were recorded in the Pasteur laboratory as being multidrug resistant (defined as resistant to ampicillin, chloramphenicol and co-trimoxazole). The distribution of the number of IncHI1 plasmid SNPs in these and drug sensitive isolates is given in Figure 6.10. There was a clear distinction between isolates that gave positive signals for IncHI1 loci (>180 out of 200 SNPs) and those that did not (≤20 out of 200 SNPs, see Figure 6.10).
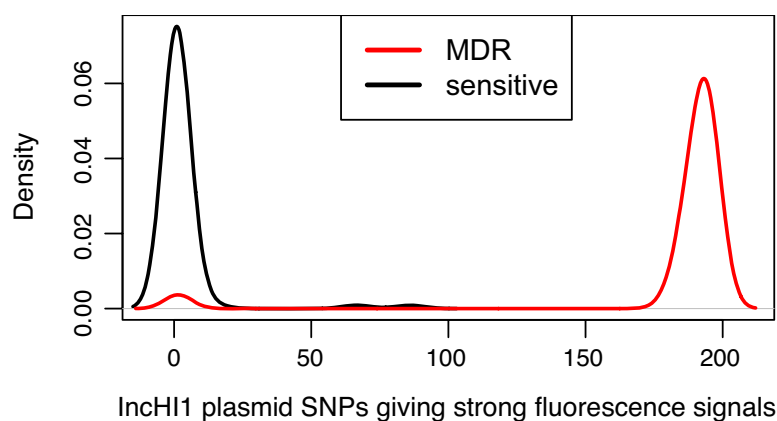

Figure 6.10: Distribution of IncHI1 plasmid SNPs detected in MDR and drug sensitive isolates - 'Strong fluorescence signal' is defined as clustering within an allele cluster as opposed to the 'no signal' cluster for a given SNP, this implies a fluorescence signal >15% that of the maximum signal detected for either allele of that SNP across all Typhi samples.

Thirty-eight of the isolates recorded as MDR had allele signals for >180 IncHI1 plasmid SNPs, consistent with the presence of an IncHI1 MDR plasmid. The remaining two MDR isolates had positive fluorescence signals for *rep* genes of other plasmid types. Isolate 76-1292 had positive signals for both IncI target loci, one of two IncK loci and one of three IncC loci. Plasmid conjugation and incompatibility group typing experiments performed by Francois-Xavier Weill (Pasteur Institute) confirmed that the isolate contained a 100 kbp plasmid of the IncI1 type, which transferred multidrug resistance to *E. coli*. Isolate 80-2002 had positive signals for two of three IncA/C target loci, one of two IncN loci and one of four IncW loci. Experiments by Francois-Xavier Weill confirmed that the isolate contained a 130 kbp plasmid of the IncA/C type, which transferred

multidrug resistance to *E. coli*. No evidence of IncHI1 plasmids was found in DNA from isolates recorded as drug sensitive. Phylogenetic analysis of the IncHI1 plasmid SNPs clustered the isolates into eight groups, shown in Figure 6.11.

The majority of plasmids (23) clustered into a single group of the ST6 IncHI1 plasmid type, defined by SNPs found in the sequenced Typhi H58 strains E98-0664, ISP-03-07467 and ISP-04-06979. All of these plasmids were found in H58 or H58-derived Typhi strains (see Figure 6.11), consistent with a single acquisition of the ST6 IncHI1 plasmid by a common ancestor of the H58 lineage. (Note that it was not possible to confirm this for the ST6 control plasmid pSTY7, which was transferred into *E. coli* for laboratory storage and experiments and the original host strain is not available.) Nearly all resistance gene target loci gave the same signals for all ST6 plasmids, matching the GoldenGate assay profile of pAKU_1. A notable exception was the isolate 38(98)S which appeared to be missing the composite transposon insertion (lacking *IS26*, *bla*, *sul2*, *strAB* (*bla/sul/str*); *sul1*, *dhfR7*, *mer* genes, *tnpA* (*Tn21*); and *cat* (*Tn9*); but carrying *tet* genes (*Tn10*)). The isolate, from which DNA had been prepared for SNP typing in July 2008, was re-tested in December 2008 by Francois-Xavier Weill at the Pasteur Institute, who confirmed that it was sensitive to all drugs at that time, including tetracycline. It is likely therefore that this isolate harboured the MDR IncHI1 plasmid when it was initially isolated in 1998, but the resistance genes have been lost from the plasmid during its 10 years in the laboratory.

Three of the Typhi IncHI1 plasmids clustered with ST8, originally defined by SNPs found in plasmid pAKU_1 sequenced from Paratyphi A (orange in Figure 6.11). The Typhi plasmids differed from the ST8 Paratyphi A plasmids (isolated from Karachi in 2003-2004 (725)) at just one SNP locus and were all isolated in Peru in 1981. The Typhi strains themselves were of haplotypes H77 (two isolates) and H50. This is the first clear evidence that very closely related IncHI1 plasmids have successfully transferred into both the Paratyphi A and Typhi populations. However, the resistance gene profiles of these three Typhi plasmids differed from that of pAKU_1. They gave positive signals for the *cat*, *strAB*, *sul1* and *tet* genes, consistent with their resistance phenotypes (chloramphenicol, streptomycin, sulfonamide and tetracycline resistant). However there was no evidence of *bla*, *dhfR7*, *dhfR14*, *sul2* or *IS26*. The two H77 strains were found to
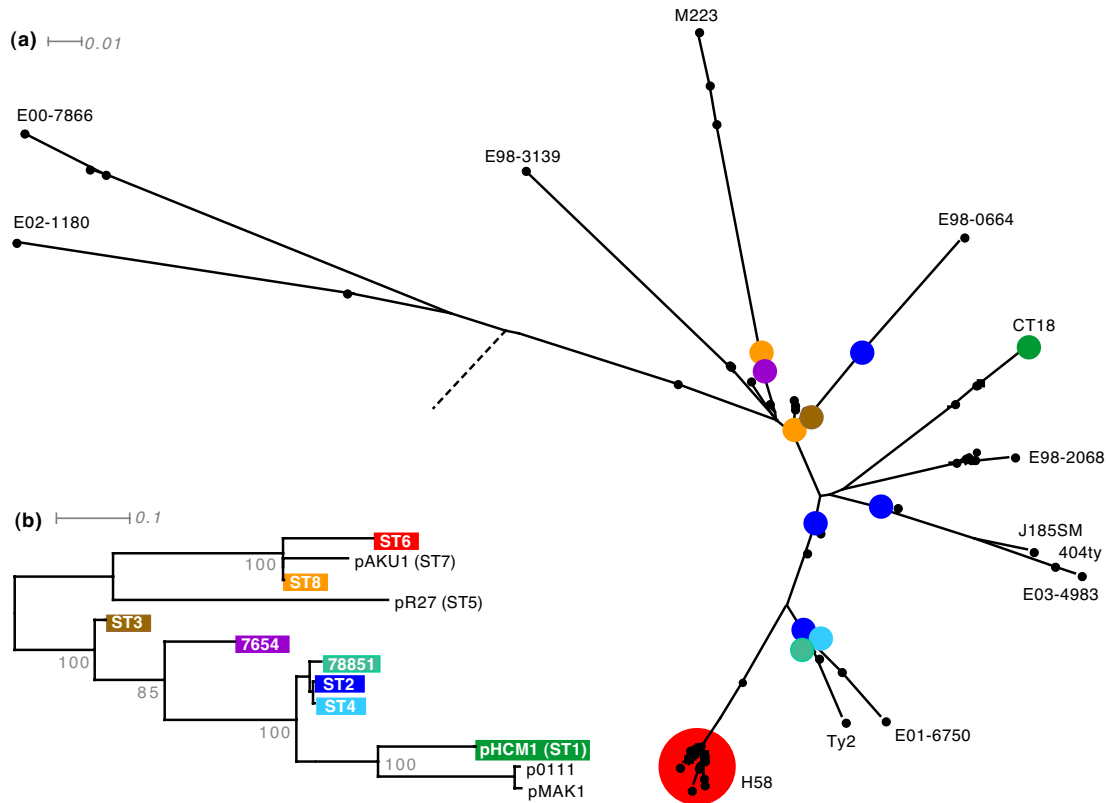
Figure 6.11: Phylogenetic trees of Typhi chromosomes and IncHI1 plasmids in a global collection of Typhi isolates - Maximum likelihood phylogenetic trees, scale bar is in divergence per assayed SNP (1,508 SNPs for a, 200 SNPs for b). (a) Tree of Typhi based on typing of chromosomal SNPs in a global collection of 180 isolates plus 19 control isolates (labelled). Nodes including isolates containing IncHI1 plasmids are highlighted with circles, coloured according to IncHI1 plasmid subgroup as shown in (b). (b) Tree of IncHI1 plasmids based on typing of SNPs in the conserved IncHI1 plasmid backbone in a global collection of 180 isolates, including 38 containing MDR IncHI1 plasmids, and control plasmids. Nodes are labelled by the plasmid ST type (as defined by plasmid MLST in (725)) and/or by the name of a representative plasmid. Each node including an IncHI1 plasmid that has been found in Typhi is assigned a unique colour. Uncoloured nodes represent plasmids that have so far only been found in other bacteria: pAKU_1 in Paratyphi A, R27 in Typhimurium, pMAK1 in Choleraesuis, p0111 in enterohemorrhagic *E. coli*. Bipartitions are labelled with bootstrap values from 1,000 bootstraps.

be trimethoprim resistant, so likely carry a different trimethoprim resistance gene from those sequenced to date from Typhi IncHI1 plasmids (*dhfR7*, *dhfR14*). The lack of resistance to ampicillin is consistent with the lack of signal detected for the beta-lactamase gene (*bla*). The SNP typing data for the Peruvian ST8 plasmids is consistent with the presence of *Tn*9, *Tn*21, *Tn*10 and *strAB* insertions in the same plasmid backbone as pAKU_1; the same elements are present in pAKU_1, with *Tn*21 inserted within *Tn*9.

To test whether the insertion sites were the same as in pAKU_1, PCR was performed using the same primer sets used to assay IncHI1 plasmids in earlier studies (see 5.2.4 and Table 5.3). Primers were designed by myself and PCR performed by Minh Duy Phan at the Sanger Institute. For all three plasmids, amplicons were generated across the left and right boundaries of the insertion site of *Tn*10 into pAKU_1 (PCR loci O, P in 5.2.4), and the insertion of the second copy of *strAB* into the pAKU_1 backbone (PCR locus Q in 5.2.4). For one plasmid, no further PCR amplicons could be generated. For the other two plasmids, 81863 and 81424, amplicons were generated across the insertion boundaries of *Tn*21 into *Tn*9 (PCR loci G, H in 5.2.4), and the insertion site of *Tn*9 into the pAKU_1 backbone (PCR loci J, K in 5.2.4). No amplicons were generated for the insertion sites of *Tn*9 or *Tn*10 observed in pHCM1 (PCR loci L, M, N in 5.2.4), or the insertion of *strAB* into *Tn*21. These results suggest that the Peruvian ST8 plasmids carry resistance insertions very similar to those in pAKU_1, with the exception of the *bla/sul/str* element. The failure of PCR to confirm the expected insertion sites of *Tn*9 and *Tn*21 in isolate 81918 suggests there may have been rearrangements within this particular plasmid, or even transposition of the *Tn*9-*Tn*21 composite transposon to a novel location. The Typhi strains carrying the ST8 plasmid type were all isolated in Peru in 1981, yet fall into distinct haplotypes H77 (isolates 81424 and 81918) and H50 (isolate 81863), separated by 28 target SNPs (Figure 6.9). This could be explained by a single acquisition of the plasmid by a common ancestor, followed by diversification of the strain types. However given the short time frame and genetic distance it is more likely that the ST8 plasmid was independently acquired by multiple Typhi strains circulating in Peru in 1981.

6.3.2.3 Distribution of other plasmids

Positive signals were detected in four experimental samples for a SNP in plasmid pHCM2 (coordinate 38,509). The SNP was originally detected between copies of the pHCM2 plasmid sequenced from genetically distant strains CT18 (H1) and E02-2759 (H58). The four novel strains harboured the E02-2759 allele at this SNP locus. Two of the strains were closely related (haplotype H39), but the other strains were genetically distant (H52 and H58-derived H34). To confirm the presence of pHCM2 in these four strains, PCR was performed as described in 2.2.2 using primers previously described in (278) (three targeting sites in pHCM2, one targeting chromosomal gene *aroC*). All three pHCM2 target amplicons, as well as the chromosomal target, were detected in the four novel isolates. In each case, amplicon sizes matched those amplified from control isolates CT18 and E02-2759. Thus, although rare in the Typhi population (278), pHCM2 can be found in distinct Typhi lineages, confirming multiple acquisition events as opposed to clonal spread within a single a lineage. Furthermore, we can expect that the array targets provide accurate detection of the pHCM2 plasmid.

The SNP typing assay included two targets for sequences within the linear plasmid pBSSB1, harbouring the z66 flagella antigen and previously found to be present only in H59 isolates (256). These targets gave positive signals for control isolates 404ty and E03-4983 (both H59) but not for any other isolates of the global collection (which did not include additional H59 isolates). This is consistent with PCR assays reported in (2), which failed to detect z66 sequences among these isolates.

6.3.3 Endemic typhoid in the Mekong Delta, Vietnam

Typhoid fever is endemic to the Mekong River Delta region in the south of Vietnam, shown in pink in Figure 6.12. With a mean of ∼80 cases per 100,000 people annually (328, 349, 422), the region accounts for over 75% of typhoid in Vietnam (354). The first MDR typhoid outbreak in Vietnam occurred in Kien Giang (Figure 6.12) in 1993 (380). MDR typhoid has continued to be a problem in the region, with rates peaking at 100% in 1996-1998 and declining to ∼50% in 2002-2004 (2, 16, 380). In Vietnam, MDR typhoid is usually treated with fluoroquinolones, however most Typhi isolated from the Mekong Delta (>95%) are now resistant to nalidixic acid (Nal) and show reduced
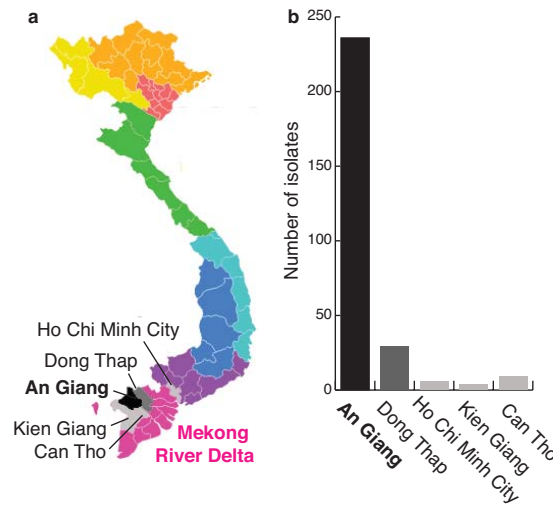
Figure 6.12: Geographical sources of isolates from the Mekong Delta - (a) Map of Vietnam, broken down into regions (shown in different colours) and provinces. Provinces from which Typhi was isolates for this study are highlighted. (b) Distribution of Typhi isolates from each province included in the study.

susceptibility to fluoroquinolones (2, 16). In order to compare alternative treatments for MDR Nal-resistant typhoid, a treatment study was conducted in 2004-2005 in the Mekong River Delta region, led by Christiane Dolecek at OUCRU, Ho Chi Minh City (745). Typhoid patients (adults and children) were recruited into the study from three hospitals: the Hospital for Tropical Diseases in Ho Chi Minh City, the Dong Thap Provincial hospital in Cao Lanh, Dong Thap province and the An Giang Provincial hospital in Long Xuyen, An Giang province (Figure 6.12). Patients were treated with gatifloxacin (a second generation fluoroquinolone) or azithromycin (745). A total of 282 patients had typhoid fever confirmed by blood culture during the two-year study period (745). The majority of these patients (82.2%) lived in An Giang province, with 10.1% of patients from Dong Thap, 1.7% from Ho Chi Minh City and the rest from neighbouring provinces (see Figure 6.12). In total, 96% of the Typhi isolates were Nal-resistant and 58% were MDR. Treatment failure with gatifloxacin or azithromycin occurred in 9% of patients, all of whom made a full recovery after treatment with ceftriaxone (745). A total of 264 of the 282 Typhi isolates were available for analysis and these were all SNP typed with the GoldenGate assay. DNA was prepared and quantified by Christiane Dolecek; data analysis was performed by myself as described above.

6.3.3.1 Phylogenetic analysis

The 264 SNP typed Typhi isolates are listed in Appendix E. A total of 258 isolates (97.7%) were of the H58 haplotype, the remaining isolates were of haplotypes H1 (N=3), H45, H50 and H52, see Figure 6.13a. H58 isolates displayed variation at 10 SNP loci, which differentiated seven distinct haplotypes, shown in Figure 6.13b. However 239 (92.6%) of these isolates belonged to just three closely related haplotypes, labelled C, E1 and E2 in Figure 6.13b. The sequenced isolate AG3, isolated in An Giang province during the study (March 2004), belongs to the H58-E2 haplotype. The SNPs separating E1 and E2 from C were originally identified from the AG3 sequence, i.e. the ability to differentiate within this cluster of 239 isolates is due to the inclusion of one of these isolates in the initial SNP detection study. Four isolates clustered with H58 based on chromosomal SNP alleles, but could not be fitted into the H58 phylogenetic tree due to conflicting SNP alleles. These isolates displayed unusual heterozygous signals for several chromosomal SNPs. They were also positive for IncHI1 plasmid loci previously identified in enterohemorrhagic *E. coli* (see 5.3.2.1), *qnr* (quinolone-resistance) gene loci which have not been reported in Typhi, and several additional plasmid *rep* genes. These samples are therefore likely to be contaminated with other bacterial DNA, possibly *E. coli*. There was no association between Typhi haplotype and patient age, fever clearance time or relapse. Diarrhea was more common in patients infected with H58-C compared to H58-E2 (74% vs 55%, p=0.006 with Pearson χ^2 test; 64% among other haplotypes), whereas constipation was more common in patients infected with H58-E2 (14% H58-E2, 6% H58-C, p=0.076 with Pearson χ^2 test; 6% among other haplotypes). Headache was also more common among patients infected with H58-C compared to H58-E2 (70% vs 55%, p=0.036 with Pearson χ^2 test; 64% among other haplotypes) and was associated with diarrhea (p<10 x 10^{-5}, Pearson χ^2 test). Only one instance of Typhi carriage was observed during 3 month follow-up, this was an MDR H58-C isolate.
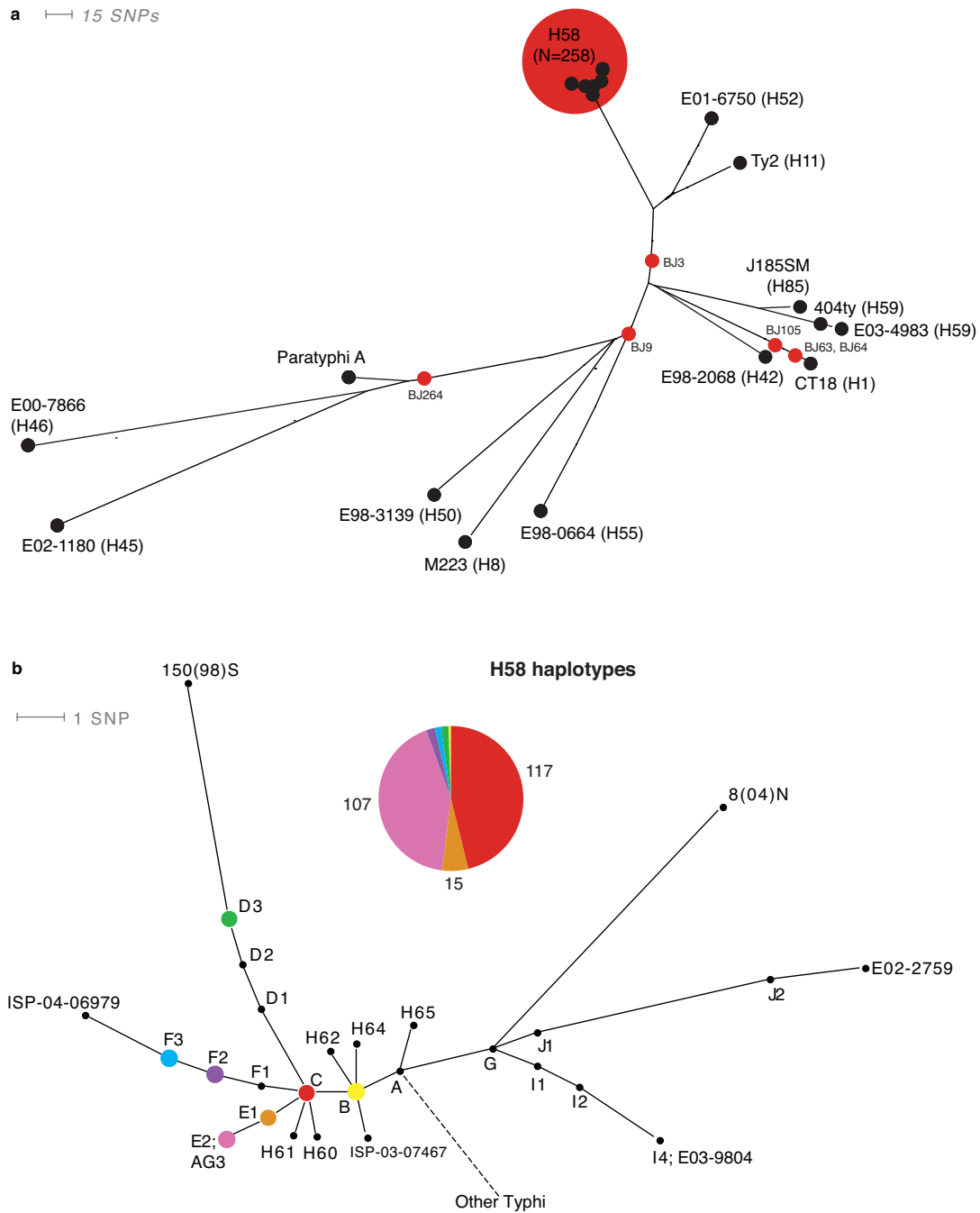
Figure 6.13: Phylogenetic distribution of Typhi isolates from the Mekong Delta - (a) Typhi phylogenetic tree. Black nodes indicate control isolates, red nodes show the position of 264 isolates from the Mekong Delta. Most isolates (258) were of the H58 haplotype, non-H58 isolates are named individually. (b) Zoom in on Phylogenetic tree of the H58 cluster. Coloured nodes indicate haplotypes that were detected among Typhi isolates from the Mekong Delta, inset pie chart indicates the frequency of each of these nodes.

6.3.3.2 Plasmids and drug resistance

The presence of fluorescence signals for IncHI1 SNP loci indicated that a total of 137 samples contained IncHI1 plasmids. All plasmids were of the ST6 type, and all host isolates were of the H58 haplotype. Of these isolates, 135 were classified as MDR by resistance testing performed by Christiane Dolecek at OUCRU (defined as MIC $\geq$ 32 μg/mL for chloramphenicol and ampicillin, and MIC $\geq$ 8/152 μg/mL for co-trimoxazole) (745). The other two isolates tested positive by GoldenGate SNP typing for *sul1*, *sul2*, *dfrA7*, *tetACDR*, *strAB*, *bla* and*cat*, just like the MDR isolates, despite very low recorded MICs for chloramphenicol, ampicillin and co-trimoxazole. An additional 18 isolates were recorded as MDR but did not test positive for IncHI1 plasmid SNPs. One of these, BJ5, did give positive signals for resistance genes *strA*, *bla*, *sul1*, *sul2*, *dfrA7*; the *repC* replication initiation gene of IncHI1; transposases and *merAPTR* from *Tn21*; and *IS*26 and *IS*10 transposases. The MDR IncHI1 plasmid was much more common among C and E1 isolates than E2 isolates (80% vs 15%, see Figure 6.14).
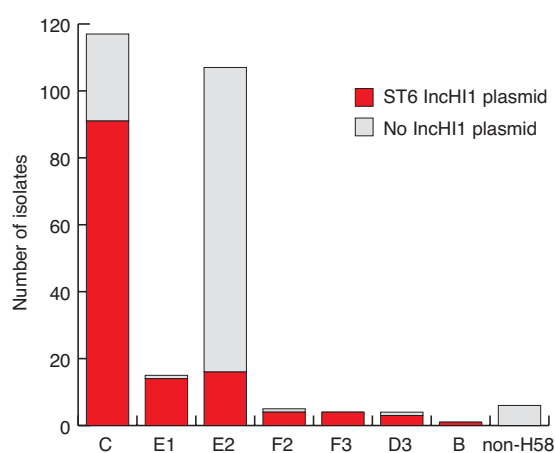

Figure 6.14: Distribution of IncHI1 plasmids among Typhi isolates from the Mekong Delta - B-F H58 haplotypes.

Nal resistance was tested by Christiane Dolecek at OUCRU. A total of 254 isolates were Nal resistant, defined as MIC $\geq$ 32 μg/mL (745). All of these were H58 isolates and all were susceptible to the fluoroquinolone drugs gatifloxacin, ciprofloxacin and ofloxacin. The only Nal susceptible H58 isolates were the singleton H58-B isolate (see Figure 6.13b), one H58-C isolate and the two isolates suspected of contamination.

6.3.3.3 Spatial and temporal distribution

Figure 6.15 shows the distribution of Typhi haplotypes among Mekong Delta provinces. All of the non-H58 isolates were from patients living outside An Giang province. Two H1 isolates BJ63 and BJ64 were identical at all assayed SNP loci and were taken from patients in Dong Thap on consecutive days. The third H1 isolate (BJ105 in Figure 6.13a) differed from these at 16 SNP loci and was collected in Dong Thap 14 months after BJ63 and BJ64. The H45 isolate BJ264 originated in Can Tho province, the H50 isolate BJ9 originated in Ho Chi Minh City and the H52 isolate BJ3 originated in Dong Nai province. The H58 C/E1/E2 cluster was dominant among isolates from each province (except Dong Nai from which just one isolate originated).
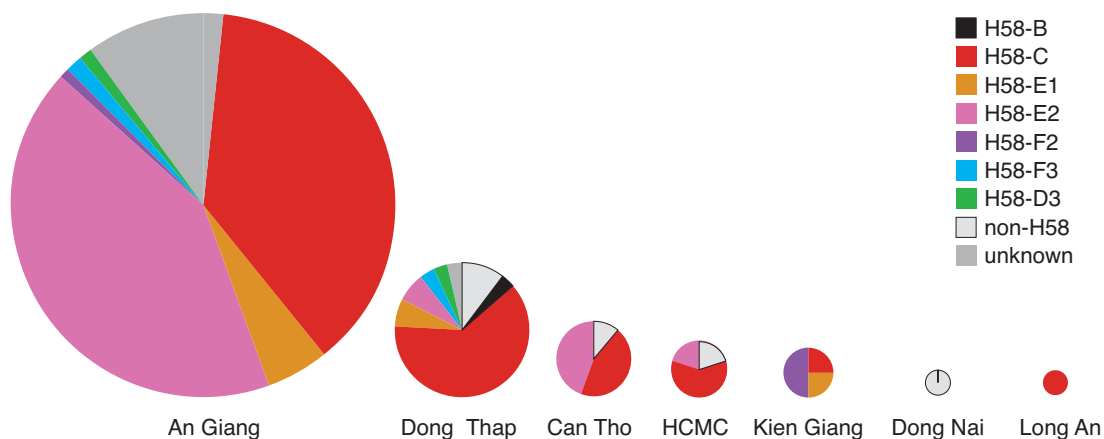

Figure 6.15: Distribution of haplotypes among provinces in the Mekong Delta
- Pie charts are scaled to represent the total number of isolates from each province.

The distribution of Typhi haplotypes over time is shown in Figure 6.16a. The majority of typhoid cases occurred in the wet season, between July and December each year. In the season of 2004 H58-E2 and H58-C were both prevalent, whereas very few isolates of H58-E2 Typhi were collected during the 2005 wet season. The decline of H58-E2 may be associated with the IncHI1 MDR plasmid (see Figure 6.14), which was much more common in H58-C. As Figure 6.16b highlights, the majority of isolates collected during the second season were MDR and carried the IncHI1 plasmid ST6.
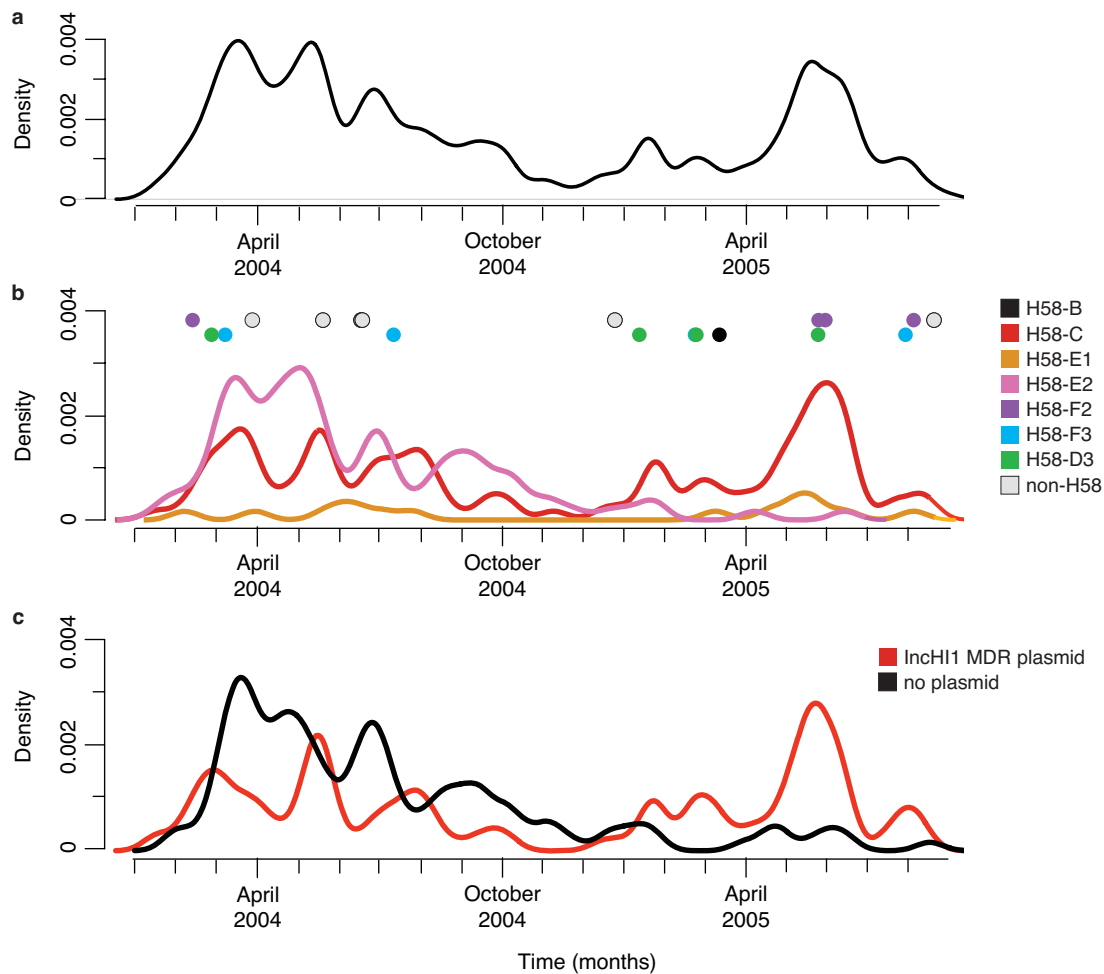
Figure 6.16: Distribution of typhoid fever cases over two years in the Mekong Delta - (a) All typhoid fever cases in the study. (b) Typhoid fever cases split by haplotype. Density distributions are shown for the three dominant H58 haplotypes C, E1 and E2. Other haplotypes are shown as circles. All lines and circles are coloured by haplotype according to the legend provided. (c) Typhoid fever cases split by presence of the MDR IncHI1-ST6 plasmid.

249

6.3.4 Pediatric typhoid in Kathmandu, Nepal

Although precise incidence data is not available, the prevalence of typhoid fever in Kathmandu, the capital city of Nepal, has been observed to be very high (753). A ten-year retrospective study of typhoid in 1993-2003 found the number of enteric fever cases (including both Typhi and Paratyphi A) more than doubled in 2001-2003 compared with the previous three years (336). MDR was not a significant problem during the study period, although there were increasing levels of reduced susceptibility to fluoroquinolones (336). In 2005-2006, researchers from Oxford University and Patan Hospital in Kathmandu conducted a study of the burden of disease caused by encapsulated bacteria among children in Patan, a central district of Kathmandu (746). Children under 13 years of age admitted to Patan Hospital with suspected bacteraemia, meningitis or pneumonia were recruited into the study (N=2,039), and blood cultures performed (N=141 positive cultures). Typhi (N=53 isolates) and Paratyphi A (N=6 isolates) were responsible for 49% of all bacteraemias and were the most frequently cultured pathogens in children older than 12 months. DNA samples from 46 of the 53 Typhi isolates was provided by Andrew Pollard of the Department of Paediatrics, University of Oxford, UK and were SNP typed at the Sanger Institute using the GoldenGate assay (following DNA quantitation by myself).

6.3.4.1 Phylogenetic analysis

The SNP typed Typhi isolates are listed in Appendix E. The majority of isolates were H58 (32, or 70%), with most of these belonging to the specific haplotype H58-G (30 isolates), see Figure 6.17. There was also a significant phylogenetic cluster of nine isolates (20%) of a subgroup of H42 (H42-A, blue in Figure 6.17), demonstrating that the dominance of H58 was not as complete in Kathmandu as it appeared to be in the Mekong Delta. There was also a cluster of three H50 isolates and two isolates from different subgroups of H42 (yellow and white in Figure 6.17). There was no association between Typhi haplotype and patient age or sex, see Table 6.3. The slightly higher mean age of children infected with Typhi of the H58-G haplotype (4.6 years vs 2.9 years for other haplotypes) is most likely due to the higher frequency of H58-G, as infections in older children were generally rarer and thus the chance to observe older children infected with other haplotypes was reduced compared to the more common H58-G

haplotype. In support of this, direct comparison of the age distributions of children infected with H58-G versus other haplotypes (see Figure 6.18) showed no evidence that the underlying age distributions differ between haplotypes (two-sample Kolmogorov-Smirnov test, p=0.45).
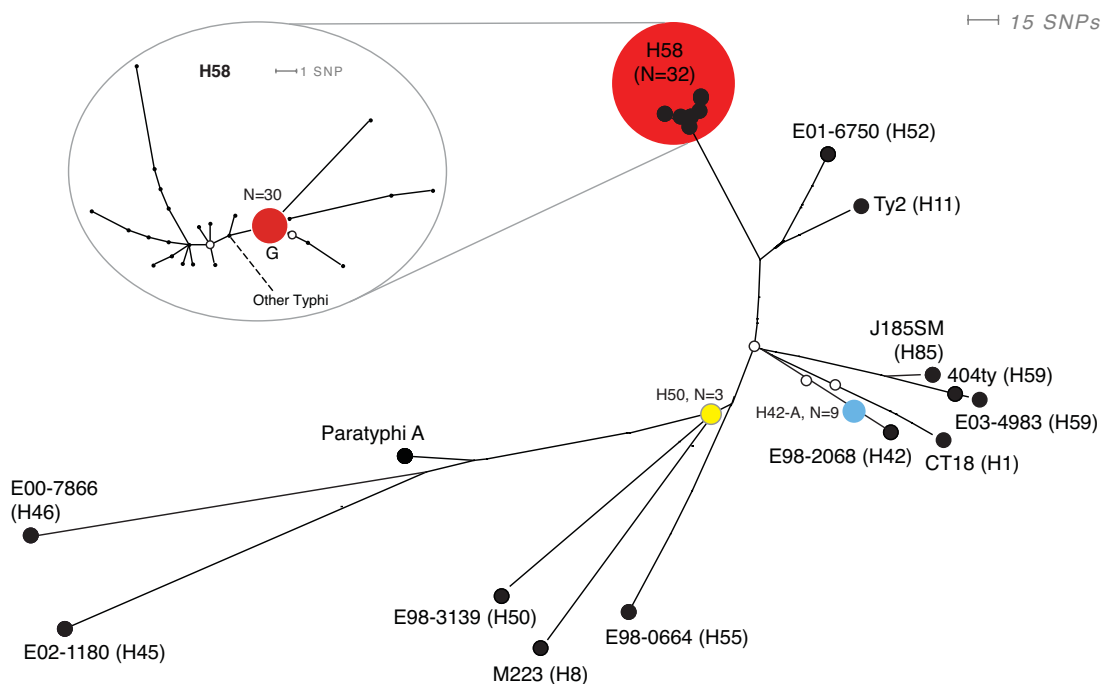

Figure 6.17: Phylogenetic distribution of Typhi isolates from Kathmandu - Black nodes indicate control isolates, coloured nodes show the position of 46 isolates from pediatric typhoid cases in Kathmandu. Most isolates were of H58 haplotypes, inset shows phylogenetic structure within the H58 group.

Haplotype	No. isolates	Mean age	Female	Nal-R
H58-G	30	4.6	47%	28
H42-A	9	3.8	44%	0
Other	7	1.9	57%	1
All	46	4.0	48%	29

Table 6.3: Typhoid case parameters by Typhi haplotype in Kathmandu - Nal-R indicates resistance to nalidixic acid.
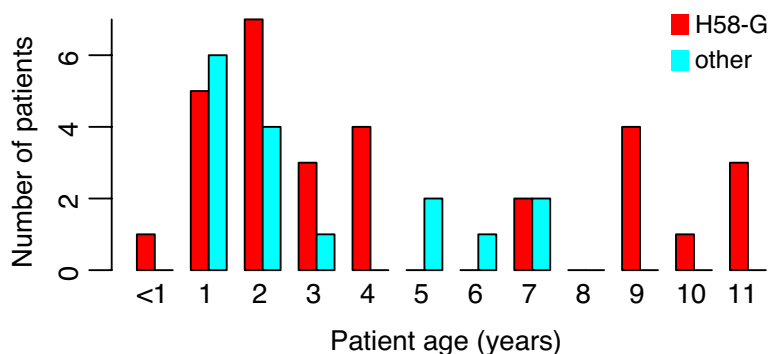
Figure 6.18: Distribution of patient ages for H58-G vs other haplotypes detected in Kathmandu - Cases are split according to Typhi haplotype, coloured as shown.

6.3.4.2 Drug resistance

Typhi isolates were tested for resistance to ampicillin, chloramphenicol, co-trimoxazole, gentamicin, ciprofloxacin, ceftriaxone and nalidixic acid (Nal). (Resistance testing was performed by David Murdoch at the University of Otago, Christchurch, New Zealand.) All isolates were susceptible to the former six drugs, while 29 isolates were resistant to Nal. MDR Typhi has been reported previously in Nepal, usually associated with the presence of plasmids (16, 598, 687, 754). However the GoldenGate assay detected no signals for resistance genes, IncHI1 plasmid SNPs or other plasmid loci in these isolates, consistent with the absence of resistance phenotypes. Nal resistance was restricted to H58 isolates, with 28 of 30 H58-G isolates and the closely related isolate 872 exhibiting MICs greater than 256 μg/mL. Quinolones target bacterial topoisomerase genes, in particular DNA gyrase (GyrA), and mutations at positions 83 and 87 of the GyrA protein have been shown to confer Nal resistance in Typhi (396). These SNPs could not be typed using the GoldenGate assay, which cannot target SNP loci that lie within 10 bp of each other. Instead, all H58 isolates were tested by Yajun Song at the Environmental Research Institute, Cork, Ireland for six known GyrA mutations (Pro83, Phe83, Tyr83, Asn87, Tyr87 and Gly87 (2)) using a Luminex 200 assay (755). All H58 isolates harboured the Phe83 mutation, with the exception of the H58-B isolate 959 which was Nal sensitive and carried wildtype alleles at GyrA positions 83 and 87.

6.3.4.3 Temporal distribution of haplotypes

The distribution of typhoid cases across the course of the study is shown in Figure 6.19. Surprisingly, the incidence of pediatric typhoid cases requiring hospitalisation was fairly constant throughout the study period (mean 2.5 cases per month), with no increase associated with the wet season (see Figure 6.19). Patterns of infection differed among Typhi haplotypes (Figure 6.19b-c). Hospitalisation of patients infected with H58-G Typhi occurred at a constant rate during the course of the study (mean 1.4 cases per month, Figure 6.19b). However H42-A Typhi was not seen until the second half of the study (Figure 6.19c, blue), after which point this haplotype was detected at a mean rate of 0.8 cases per month. Hospitalisations due to infection with other Typhi haplotypes occurred throughout the study, at a mean rate of 0.33 cases per month.
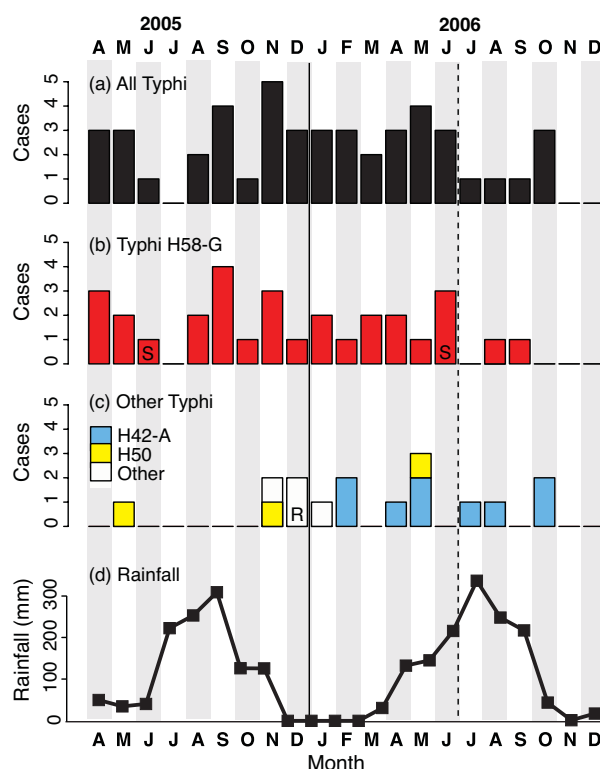

Figure 6.19: Distribution of typhoid cases in Kathmandu by month - (a) All cases. (b) Infections with haplotype H58-G, note all of these cases were nalidixic acid resistant except two marked 'S'. (c) Infections with Typhi of other haplotypes, colours indicate haplotype according to legend provided. All of these cases were nalidixic acid sensitive except one marked 'R'. (d) Total monthly rainfall at Kathmandu airport.

6.3.5 Endemic typhoid in an urban slum in Kolkata, India

The annual incidence of typhoid fever in India has been estimated at 662/100,000 among inhabitants and 42/100,000 among travellers, the highest rates in the world (10, 11). In Kolkata in the east of India, typhoid is most common in urban slum areas ("bustees"). The first MDR typhoid outbreak in Kolkata occurred in 1989-1990 (756) and MDR typhoid has persisted, declining from 100% in 1991-1992 to below 15% in 2003-2004 (417, 418, 757). In 2004, the first fluoroquinolone-resistant Typhi isolates were observed in Kolkata, with MICs >16 μg/mL to ciprofloxacin and ofloxacin (401). In 2003, a four year study of typhoid fever was begun in an urban slum in East Kolkata (Wards 29 and 30), led by researchers at NICED in Kolkata and IVI in Seoul, Korea. The 1 km^2 study site was home to nearly 60,000 people, who were encouraged to report all episodes of fever lasting at least three days to study centers, where they were offered free diagnosis and reimbursed transport costs as well as treatment costs for enteric fever and malaria (343, 421, 747). Surveillance began on May 1 2003 and continued until 31 January 2007, during which time 378 cases of typhoid fever were confirmed by blood culture (2.8% of fever cases reported; 50% of fever cases reported in children aged 5-15 (327)). During December 2004, residents were vaccinated with either a Vi conjugate vaccine (designed to protect against Typhi), a Hepatitis A vaccine, or no vaccine (421). Residents were assigned to one of 80 geographical clusters based on the location of their dwelling, geographic clusters were assigned to either the Vi conjugate or Hepatitis A vaccine (40 clusters each, see Figure 6.22 below), and 25-85% of individuals within each cluster received vaccine. DNA was available for 188 (50%) of the 378 cases of typhoid fever confirmed during the study (see Appendix E); these samples were SNP typed with the GoldenGate assay. DNA was prepared in Kolkata by Shanta Dutta at NICED and quantified by Derek Pickard at the Sanger Institute. Data analysis was performed by myself.

6.3.5.1 Phylogenetic analysis

The majority of isolates were H58 (139, or 74%), see Figure 6.20a. There was also a significant phylogenetic cluster of 28 isolates (20%) of H42-A (blue in Figure 6.20a), as well as nine H50 isolates and 12 other isolates scattered around the phylogenetic tree (white in Figure 6.20a). Although the basic composition of the population was similar to that observed among isolates from Kathmandu, Nepal (70% H58, 20% H42-A) the H58 subgroups were different. Among the Kolkata H58 isolates, the majority were of subgroups B or G (77 and 43 isolates respectively), which represent distinct sublineages of H58 (see Figure 6.20b-c). There were also small clusters of subgroups A and H64 (nine and eight isolates respectively), which are each separated from subgroup B by a single target SNP (see Figure 6.20b-c). The remaining two H58 isolates were of two different, more distantly related subgroups, separated from G by at least two target SNPs (see Figure 6.20b). There was no association between Typhi haplotype and patient age or sex, see Table 6.4. IncHI1 plasmids were detected in just six isolates, all of them H58 (see Table 6.4).

Haplotype	No. isolates	Mean age	Female	IncHI1
H58-B	77	10	43%	0
H58-G	43	12	42%	4
Other H58	19	10	47%	2
H42-A	28	10	43%	0
Other	21	10	44%	0
All	188	10	45%	6

Table 6.4: Typhoid case parameters by Typhi haplotype in Kolkata - IncHI1 indicates the presence of the IncHI1 plasmid as detected by SNP typing with GoldenGate.
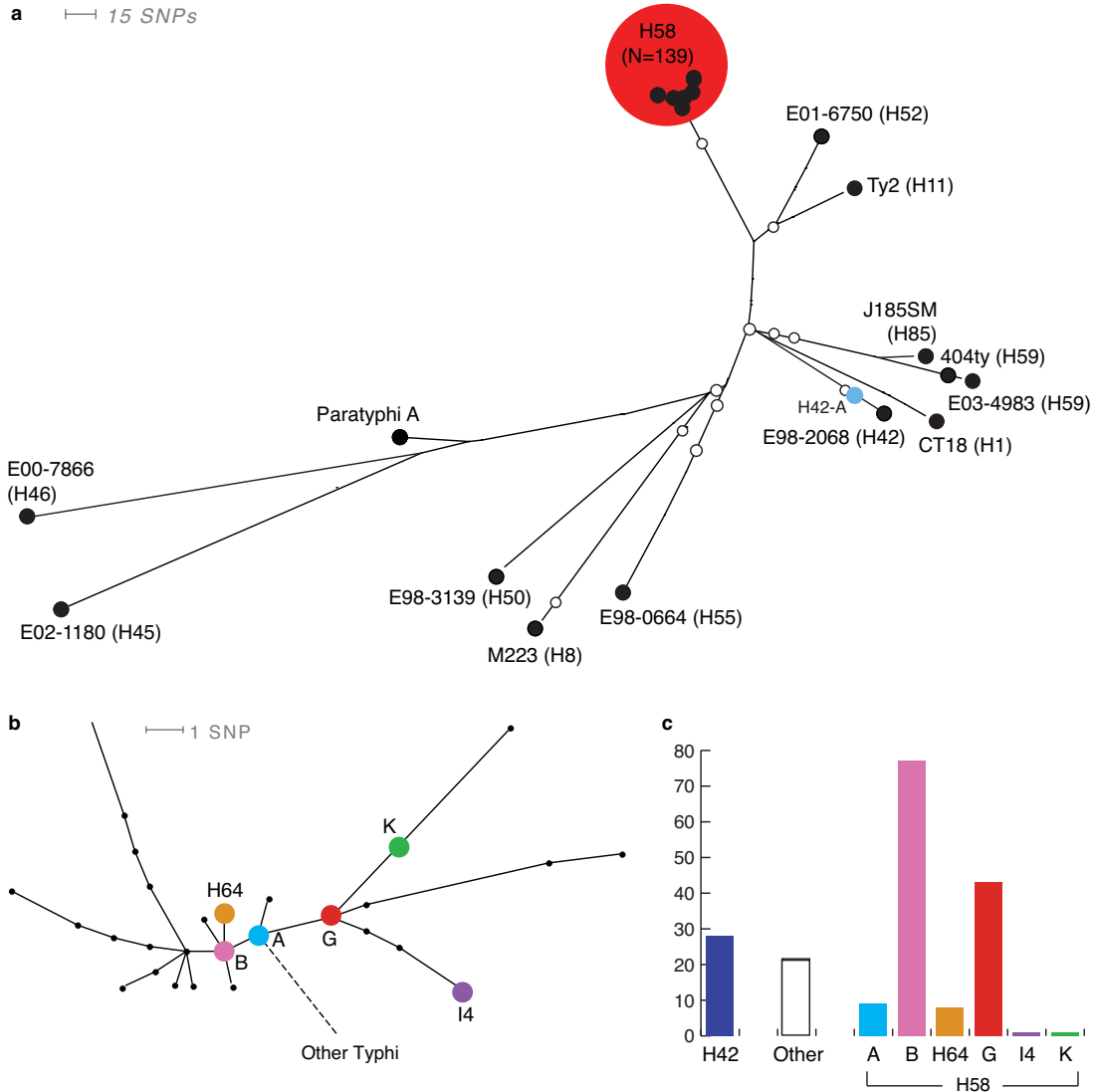
Figure 6.20: Phylogenetic distribution of Typhi isolates from Kolkata - (a) Typhi phylogenetic tree. Black nodes indicate control isolates, other nodes show the position of 188 isolates from Kolkata. Most isolates (139) were H58, although there were also 28 H42 isolates (pale blue). (b) Zoom in on phylogenetic tree of the H58 cluster. Coloured nodes indicate haplotypes that were detected among the Kolkata isolates. (c) Frequency of each haplotype observed. Nodes are coloured as in a-b.

6.3.5.2   Spatial and temporal distribution of haplotypes

Confirmed typhoid cases occurred at a rate of 0.16 per day (approximately one per week) during the study period, both before and after the introduction of the vaccine in December 2004. The distribution of typhoid cases across the course of the study is shown in Figure 6.21a, which shows a number of peaks in typhoid incidence. The distribution of haplotypes, ascertained for 188 (50%) of the Typhi isolates, is shown in Figure 6.21b. From this plot it is clear that peaks in typhoid cases in 2004 resulted from infections with a diverse range of Typhi haplotypes (Figure 6.21b, peaks 1 and 2). However later peaks in typhoid incidence were due almost entirely to infection with Typhi haplotype H58-B (peaks 4 and 5) or H58-G (peak 6, see Figure 6.21b).

The spatial distribution of typhoid cases was non-random, with a higher incidence of cases in certain geographical clusters including high population density clusters in the south-west of Ward 29 (Figure 6.22). There were also spatial areas in which specific haplotypes were overrepresented (see 6.2.5), shown in Figure 6.22. For example, 27 out of the 77 infections with H58-B occurred in a spatial area containing seven of the geographical clusters (red in Figure 6.22), at a rate of 546/100,000 people compared to 91/100,000 across the rest of the study site and making up 79% of isolates collected in these clusters compared to 40% of all SNP typed isolates. Seventeen of these patients lived in just six dwellings in the same street, with 2-4 cases per household. In geographical cluster 17 H58-G was overrepresented, with incidence 456/100,000 compared to 66/100,000 in other clusters, and 50% of Typhi infections in this cluster due to H58-G compared to 23% of all SNP typed Typhi isolates (blue in Figure 6.22). H42 isolates were overrepresented in geographical clusters 32 and 17 (incidence 230/100,000 vs 40/100,000; 50% in these clusters compared to 15% overall, green in Figure 6.22) and H64 isolates were overrepresented in geographical cluster 3 (incidence 412/100,000 vs 8/100,000; 60% compared to 4% overall, pink in Figure 6.22). Two of the five typhoid cases (40%) in neighbouring geographical clusters 9 and 10 were of haplotype H50 (the two patients lived in the same street), compared to 5% of all cases (incidence 134/100,000 vs 12/100,000, pale blue in Figure 6.22). The haplotype-specific spatial clusters did not show strong correlations with peaks in typhoid incidence highlighted in Figure 6.21, however most peaks had identifiable spatial foci, shown in Figure 6.23.
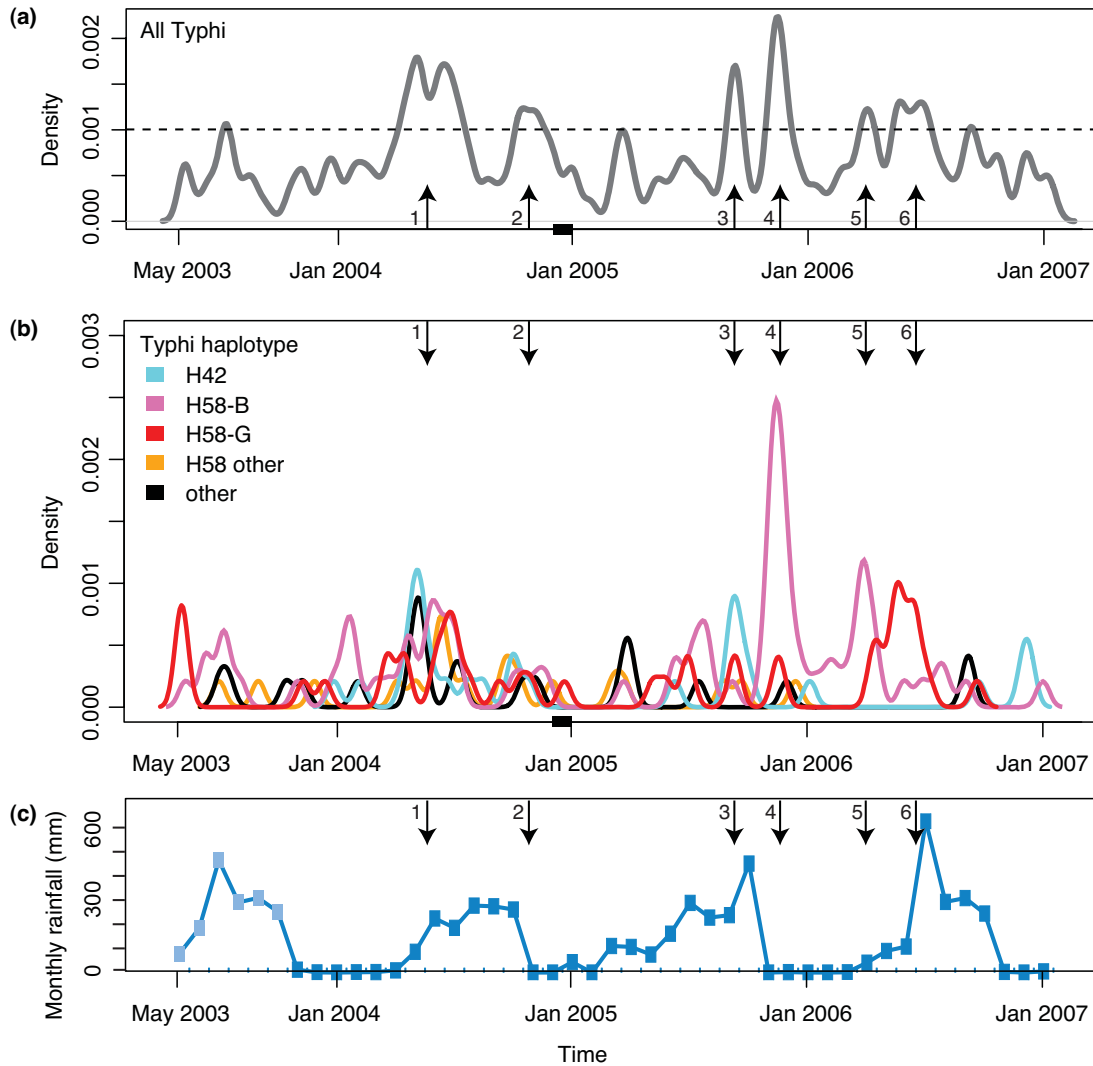
Figure 6.21: Distribution of typhoid cases during a four year study in Kolkata
- Black rectangle indicates December 2004, during which time vaccinations were given.
Outbreaks labelled 1-6 are highlighted with arrows. Arrows indicate outbreaks discussed
in the text. (a) All typhoid cases (N=378). (b) SNP typed typhoid cases (N=188) broken
down by Typhi haplotype, coloured as shown. (c) Monthly rainfall in the Howrah district
where the study site is located. Data was sourced from the India Meteorological Depart-
ment (http://www.imd.gov.in). Data was not available for 2003, so the mean values across
2004-2006 were plotted in its place (pale blue).

Figure 6.22: Spatial clustering of typhoid cases in Kolkata - Geographic clusters 1-80 are shown as orange and blue blocks. Countour plots (black lines) show the distribution of typhoid fever cases during the study period. Geographic clusters with an overrepresentation of typhoid cases of particular haplotypes (using Openshaw's Geographical Analysis Machine to test for spatial clustering within each haplotype) are coloured according to the legend.
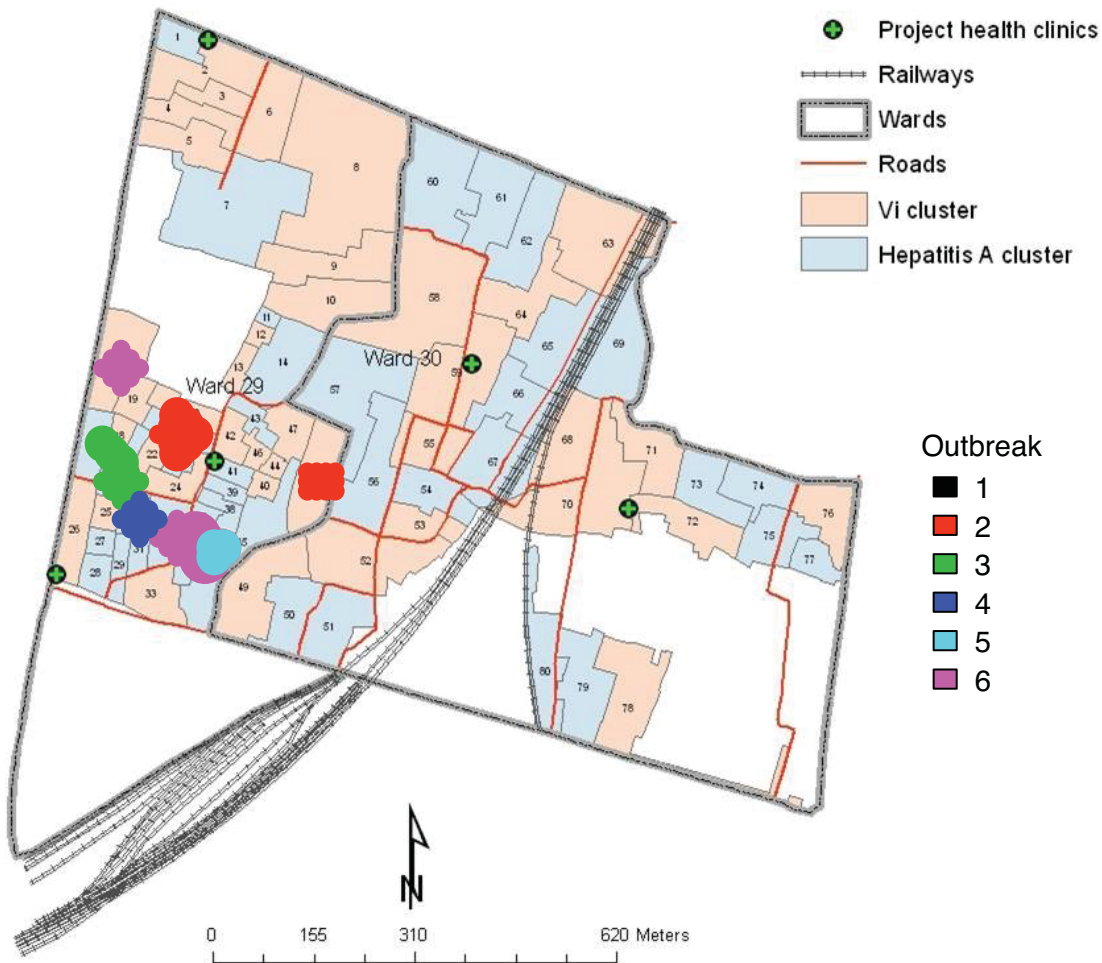
Figure 6.23: Spatial clustering during typhoid peak-incidence periods in Kolkata - Geographic clusters 1-80 are shown as orange and blue blocks. Geographic clusters with an overrepresentation of typhoid cases during six periods of high incidence (labelled 1-6 as in Figure 6.21) (using Openshaw's Geographical Analysis Machine to test for spatial clustering within each period) are coloured according to the legend.

There was evidence of typhoid fever cases clustering within househoulds, with 88 (47%) of the SNP-typed Typhi isolated from patients living in 34 (0.3%) of the 10,954 households in the study site (2-6 per household, median 2, see Figure 6.24a). Nineteen of these households had cases of infection with multiple Typhi haplotypes, see Figure 6.24a. The median time between cases in the same household with the same haplotype was 93 days (3 months; range 1-722 days), while the median time between any cases within a household was 154 days (5 months; range 1-1078 days). The downwards skew of time between infections in the same household (see Figure 6.24b) suggests these infections are not independent. Infections within a household may be caused by transmission between individuals (i.e. same haplotype), or multiple people sharing infections from a common external source (which could be of the same or different haplotypes). The distribution of times between infections with the same haplotype was not significantly different from those between infections with different haplotypes (p=0.3, one-sided Kolmogorov-Smirnov test). The high number of households reporting infections with multiple Typhi haplotypes (Figure 6.24) suggests that shared exposure to external sources of contaminated food or water may make an equal if not greater contribution to shared typhoid fever illness than direct transmission of Typhi between household members, at least in high-risk endemic areas.
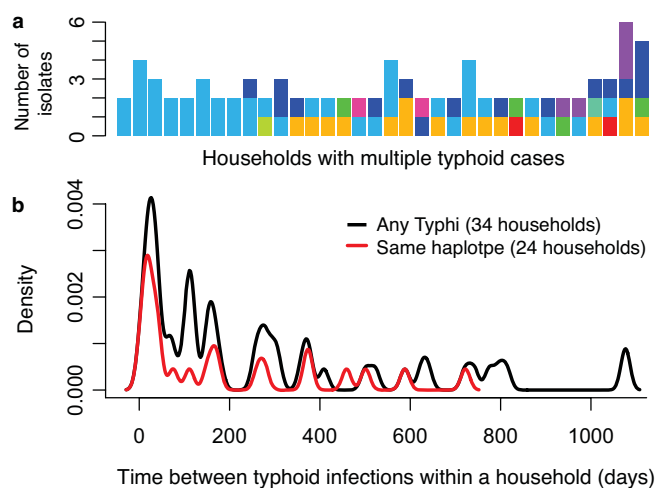

Figure 6.24: Distribution of Typhi cases among households in Kolkata - (a) Typhi diversity among households with multiple SNP-typed Typhi infections. Colours represent distinct Typhi haplotypes. (b) Distribution of time between Typhi cases within a household.

6.3.5.3 Association with the vaccination programme

The Vi conjugate vaccine was effective at reducing the incidence of typhoid fever among vaccinees, with overall effectiveness reported at 61% (421). Among the post-January 2005 cases for which isolates were SNP typed, the odds ratio (OR) for typhoid among Vi vaccinees compared to unvaccinated inhabitants of clusters assigned the Vi vaccine was 0.55 (95% confidence interval 0.28-1.08, p=0.04 (758, 759)). This is equivalent to vaccine effectiveness of (1-0.55) x 100% = 45%. The diversity of Typhi haplotypes causing disease appeared to have been reduced after the introduction of the vaccine, with a Simpson's diversity index (1-D) of 0.81 prior to December 2004 and 0.68 after January 2005 (p-value$<1\text{x}10^{-6}$, see 6.2.6; see also Figures 6.21b and 6.25). The proportion of Typhi isolates of the H58-B haplotype increased after the vaccinations, from 33% to 50%. In addition, the vaccine appeared to be ineffective against H58-B, with an OR of 1.56 (95% CI 0.44-7.16, p=0.27) for H58-B infection among vaccinees compared to unvaccinated inhabitants, in contrast to OR 0.33 (95% CI 0.14-0.77, p=0.005) for infection with other Typhi haplotypes. All isolates expressed Vi (421) and all of the H58-B isolates gave positive signals for all SNP loci targeted in SPI7, including seven SNPs within the *tviD* and *tviE* genes involved in Vi biosynthesis. It is possible that the H58-B isolates may have a modified Vi structure or a modified pattern of Vi expression which facilitates their escape from immunity conferred by the Vi vaccine, however further experiments will be needed to test these hypotheses.
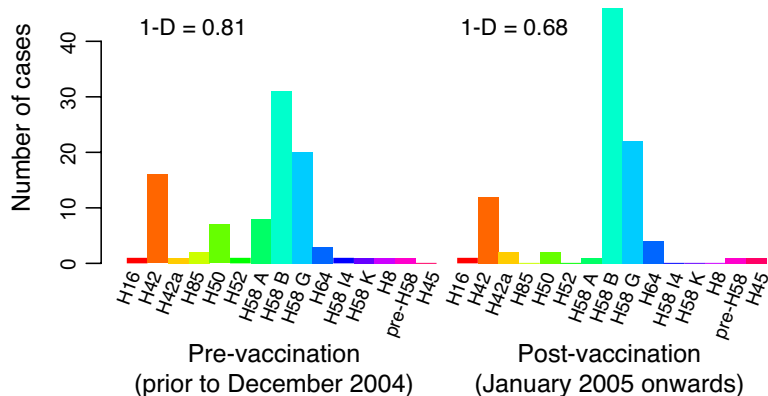

Figure 6.25: Frequency distribution of Typhi haplotypes before and after the introduction of a Vi conjugate vaccine in Kolkata - 1-D = Simpson's index.

6.3.6   The Typhi population in Nairobi, Kenya over a 21 year period

The incidence of typhoid fever in Kenya has not been studied, however the annual incidence of typhoid fever in Eastern Africa has been estimated at 39/100,000, and 50/100,000 for Africa as a whole (10) (see Figure 6.1). Diagnostic tests and surveillance systems for *Salmonella* are not as well developed in Kenya as they are in high-incidence endemic areas in South Asia (760, 761, 762) and as a result there are few studies addressing the population structure or drug resistance of Typhi. However available studies reveal a high rate of MDR typhoid, associated with the presence of a >100 kbp plasmid (383, 763). A collection of 96 Typhi isolates was assembled by Sam Kariuki at KEMRI from hospitals in Nairobi, Kenya (see Appendix E). The isolates were collected between 1988 and 2008 for a number of surveillance studies. The collection is biased towards more recent isolates, with 73 isolates (76%) collected between 2001-2008. DNA was provided by Sam Kariuki for SNP typing on the GoldenGate array. DNA quantitation was performed by Derek Pickard at the Sanger Institute, analysis was performed by myself.

Figure 6.26 shows the distribution of Typhi haplotypes over the 21 year period represented by the Nairobi collection. The majority of isolates (76%) were H58, with the remaining isolates distributed among seven distinct haplotypes (see Figure 6.26a). The H58 isolates increased in frequency over time, but at least one H58 strain was isolated in every year represented, including the earliest year 1988. The distribution of other haplotypes was more sporadic, with most detected in just one or two years (see Figure 6.26b). The distribution of H58 subgroups is shown in Figure 6.26c-d. The most common H58 subgroup was J1, represented by 47 isolates (64% of H58 isolates) spread across the entire period (see Figure 6.26d). Two isolates of H58-J2, derived from H58-J1 (see Figure 6.26c) were detected among those from 2001. H58-B was detected among isolates between 2004-2008. The H58-B subgroup is from a different sublineage to H58-J1 and H58-J2 (see Figure 6.26c), thus the appearance of H58-B in 2004 may represent the introduction and spread of a novel clone in this region.
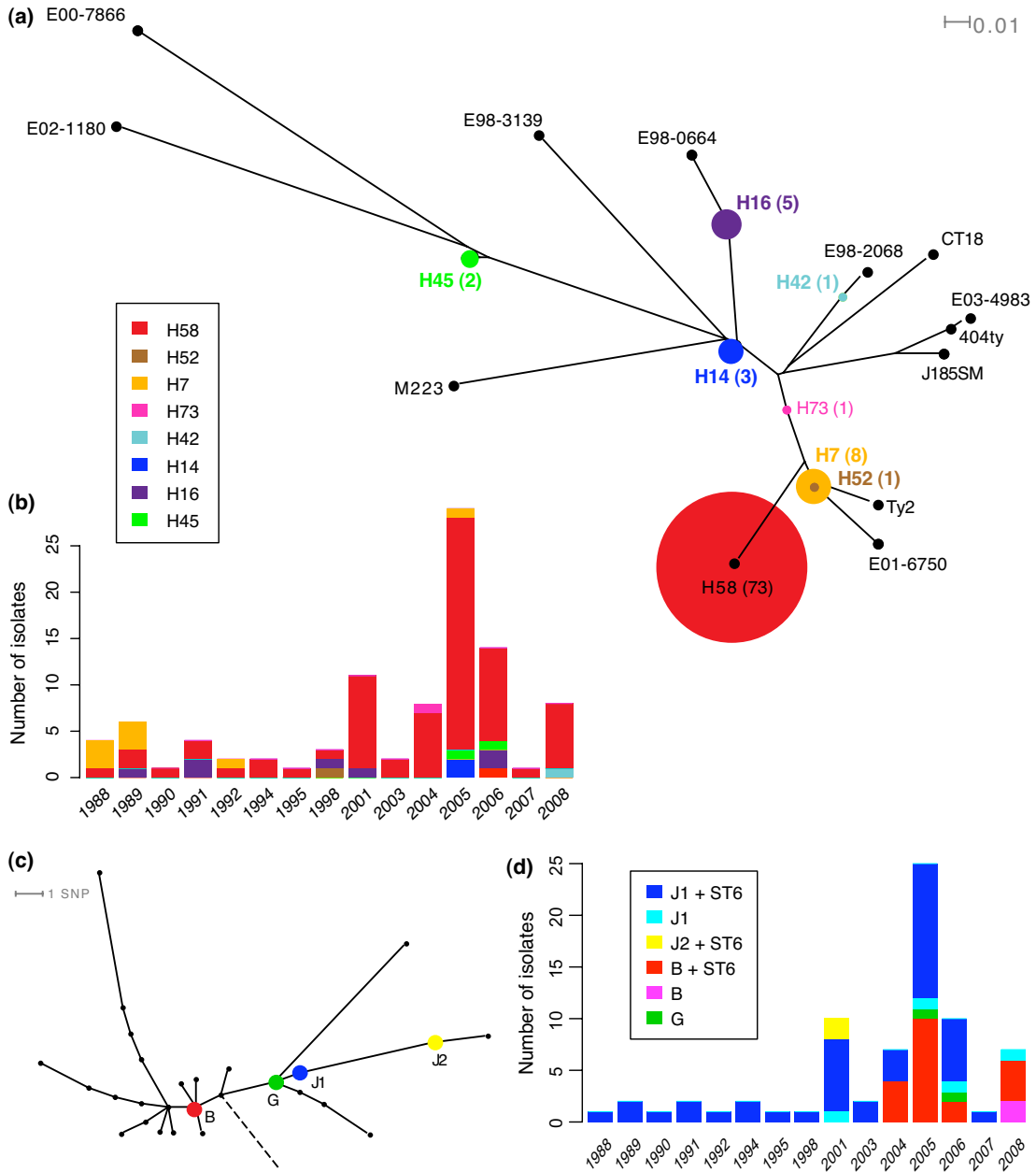
Figure 6.26: Distribution of Typhi haplotypes in Nairobi, Kenya - (a) Phylogenetic tree of Typhi, the position of Kenyan isolates is shown with coloured nodes, labelled by haplotype and the number of isolates (in brackets). The size of nodes are also scaled to represent the number of isolates. (b) Distribution of haplotypes over time, colours indicate haplotypes as shown in the legend and in (a). (c) Phylogenetic tree of Typhi H58, the position of Kenyan isolates is shown with coloured nodes, labelled by H58 subgroup. (d) Distribution of H58 subgroups over time, colours indicate subgroup and presence of the ST6 IncHI1 plasmid, as shown in the legend and in (c).

A total of 66 IncHI1 plasmids were detected, of which 65 were ST6 plasmids found in H58 strains. The exception was a single plasmid, similar to the ST6 plasmids but with different alleles at 18 SNP loci and no signal at 39 loci, found in an isolate of the H73 haplotype collected in 2004 (pink in Figure 6.26a). Among the H58 isolates, ST6 plasmids were found in the majority of J1 and B isolates (96% and 91%, respectively) and in both J2 isolates (see Figure 6.26d). The two H58-G isolates, collected in 2005 and 2006, were plasmid-free. The plasmid-free B and J1 isolates were observed in later years (see Figure 6.26d), consistent with a gradual loss of the MDR plasmid over time.

6.3.7 Typhi H58 and the IncHI1 ST6 plasmid

Figure 6.27 shows the distribution of Typhi H58 subtypes among the regional datasets analysed above. As those analyses showed, the distribution of H58 subtypes was different in the different regions studied. However Figure 6.27 highlights some broader trends in the geographic distribution of H58. In particular node C and derived lineages (D, E, F) was mainly restricted to Vietnam, while node G and derived lineages (I, J, K) were common among isolates from India but not Vietnam. The Kenyan isolates were mostly from the JI and B nodes, which were associated with Indian but not Vietnamese isolates (B was common in Kolkata; the J nodes are derived from the G node which was also common in Kolkata). This may reflect human patterns of travel and migration between India and Kenya, which has had an Indian community since the early 20th century, generating a constant flow of migrants and visitors between the two countries. The isolates found in Kathmandu were from the G node, which was also common in Kolkata, perhaps reflecting the close geographic proximity of India and Nepal.

The earliest H58 and derived haplotypes were first reported by Roumagnac *et al.* (2), including a 1958 H61 isolate from Morocco, a 1966 H62 isolate from Morocco and a 1968 H60 isolate from the Cote d'Ivoire (see Figure 6.27). The 1964 H58 isolate 43-64 from Chad was retyped in the present study and found to be H58-E2, indicating that diversification of the H58 B lineage had already occurred by the early 1960s. The earliest representative of the G lineage was among the earliest isolates from the Kenya collection (1988). Thus the earliest record we have of H58, including diverse sublineages, is from Africa. This is consistent with an African origin for H58, however there
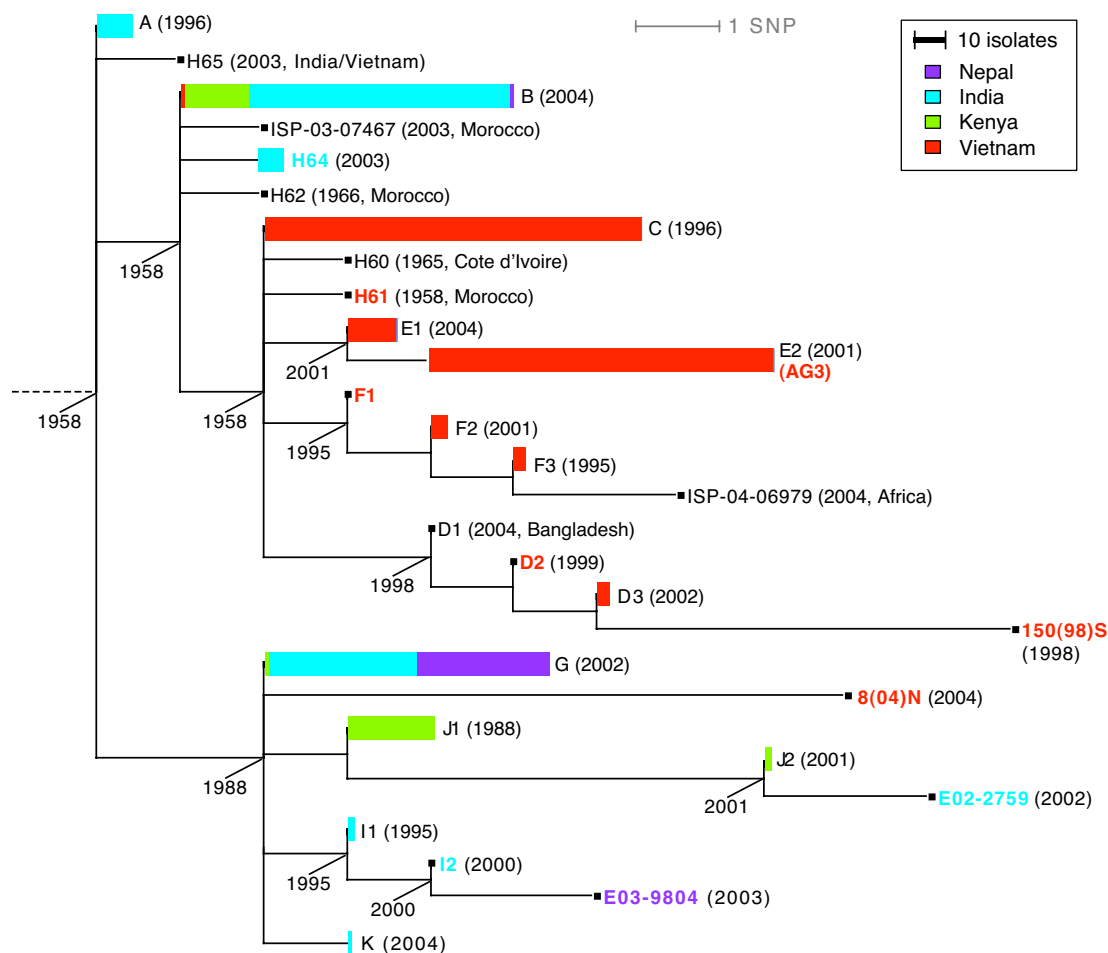
265

Figure 6.27: Distribution of H58 subtypes among Typhi isolates from four regional collections - The phylogenetic tree of H58, dashed line represents the root of this tree, i.e. where H58 joins the rest of the Typhi phylogeny. Nodes are labelled with their name as assigned in this study (A-K) or previously assigned H-group, or with the sequenced isolate which defines the node. Coloured bars indicate the number of isolates at that node from each of the four regional collections. Nodes that are not represented by isolates from these collections (i.e. defined by isolates from the Pasteur collection, (2) or sequenced isolates) are coloured to indicate their origin from Vietnam, India, Nepal or Kenya, or are labelled to indicate their origin elsewhere. Nodes are also labelled with the earliest year of isolation (among the isolates typed in this study or in (2)).

is not enough data available from early Asian isolates to confirm this. Only ten pre-1995 Asian isolates have been tested (sourced from the Pasteur collection and tested both here and in (2)), all of which were from Vietnam and belonged to the non-H58 associated haplotypes H1 (N=3), H50 (N=4), H87 (N=2) and H68 (N=1). The earliest appearance of H58 among Vietnamese isolates in the Pasteur collection was 1995, with an isolate from the H58 F lineage (reported in (2) and analysed in this study). The earliest isolate of the H58 D lineage was from 1999 and the earliest H58 E lineage isolate from Vietnam was from 2001. The earliest example of the J lineage was a 1988 isolate from the Kenyan collection, and the earliest example of the I lineage was an Indian 1995 isolate from the Pasteur collection. The non-random distribution of H58 subtypes shown in Figure 6.27 is in contrast to the global distribution of Typhi haplotypes first noted in (2). This is likely a function of the different time scales and levels of resolution involved - the diversification of H58 likely represents well under 100 years of evolution, with much of this occurring during an expansion over the last 10-20 years; the broader scale comparison of H-groups represents diversification over many thousands of years. Thus while it is broadly true that Typhi has a global distribution, high-resolution SNP typing highlights some geographic patterns. This suggests there are barriers to the spread and maintenance of novel Typhi clones, even between high-incidence endemic areas that are relatively close geographically.

In the data presented above, a clear trend emerged in which MDR IncHI1 plasmids detected in the last ten years have been almost exclusively of the ST6 type (see summary plot in Figure 6.28c) and have been found in H58 isolates. While the data is biased towards the four locations studied intensively (Mekong Delta, Kathmandu, Kolkata and Kenya), the Pasteur collection revealed Typhi H58/ST6 plasmid isolates originating from numerous locations in Asia, Africa and the Middle East (see Figure 6.28a), suggesting that this MDR clone has global reach. Figure 6.30 shows the distribution of IncHI1 plasmids within H58 subgroups, as well as the distribution of *IS*1. This IS element is encoded in the IncHI1 plasmid (as part of *Tn*9, *Tn*6062 and singleton insertions) and in the sequencing study presented in Chapter 2 was also found inserted in the chromosomes of IncHI1 plasmid-containing isolates (see 2.3.3). The association between plasmid and *IS*1 is further supported by the GoldenGate data, with *IS*1 detected far more frequently within haplotypes in which plasmids had also been detected. The

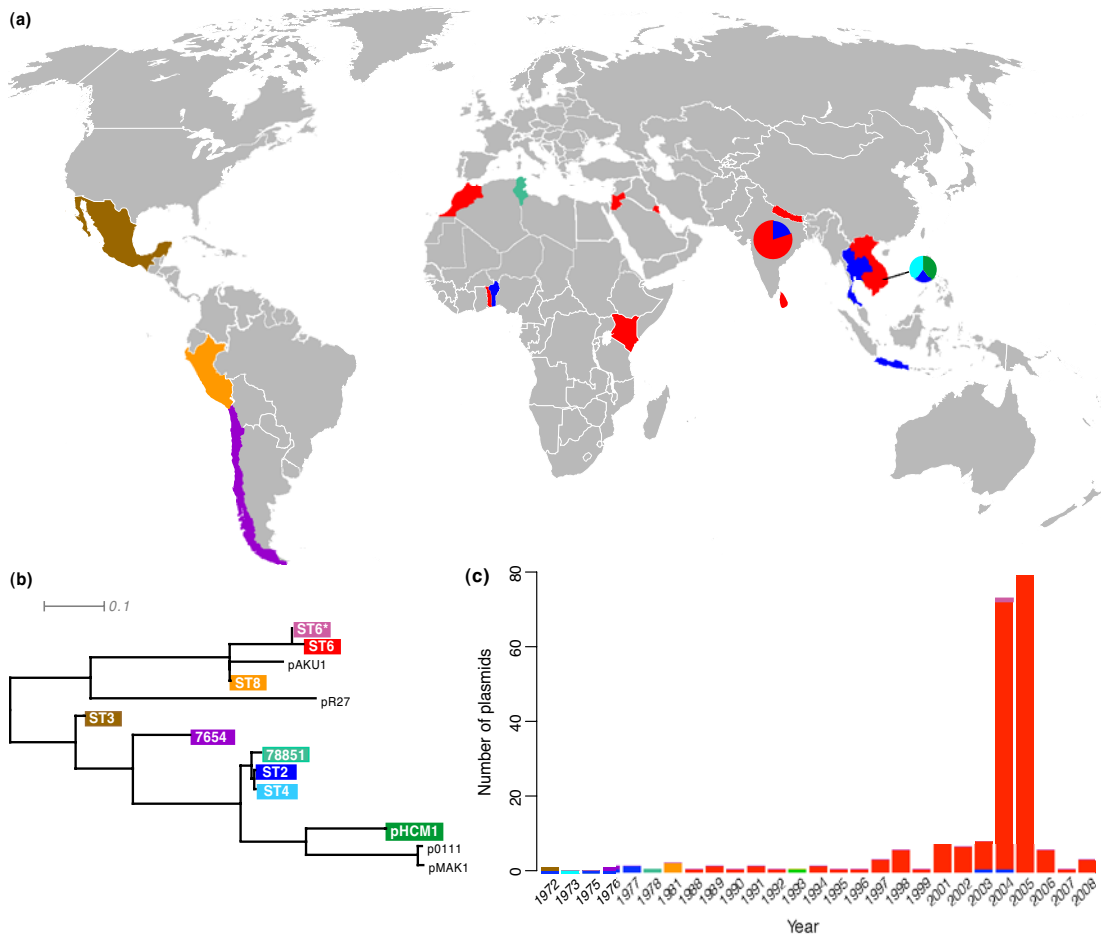**Figure 6.28: Distribution of IncHI1 plasmids in time and space** - Each colour indicates a unique IncHI1 subtype, as labelled in (b). (a) Countries from which IncHI1 plasmids were analysed, coloured to indicate the IncHI1 subtype detected. (b) Phylogenetic tree of IncHI1 plasmids detected in this study, scale bar indicates divergence among 200 SNPs. (c) Distribution of IncHI1 plasmid types over time.
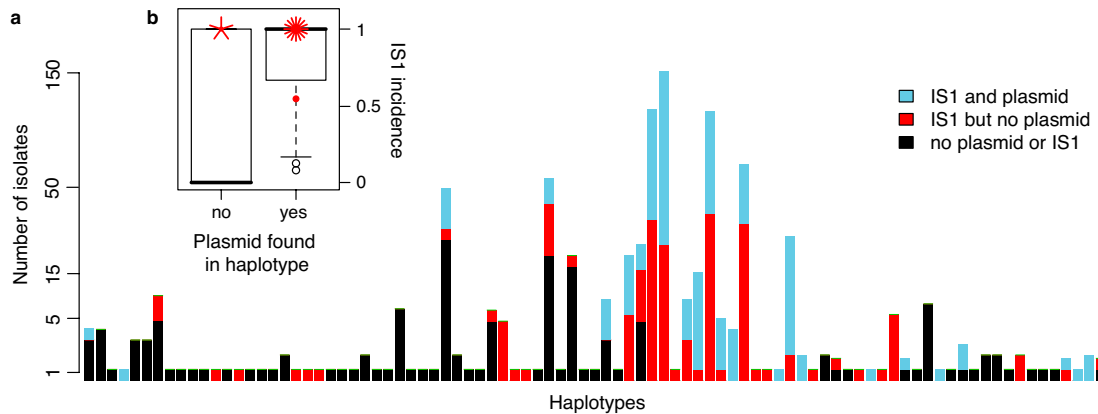
268

Figure 6.29: Distribution of IncHI1 plasmids and *IS*1 among Typhi haplotypes - (a) Presence of *IS*1 and IncHI1 plasmid among Typhi haplotypes. All plasmid-containing isolates also carried *IS*1. Note the y-axis is on the log scale. (b) Incidence rate of *IS*1 among haplotypes in which the presence of IncHI1 plasmids was demonstrated. Data from H58 haplotypes are highlighted using a sunflower plot, in which each radial line represents a different data point (i.e. haplotype) sharing the value represented by the central point.

*IS*1 probe gave positive signals for every isolate in which the IncHI1 plasmid was found (230 isolates), demonstrating its reliability in detecting the presence of *IS*1. It was detected in only 21 out of 61 haplotypes in which plasmids were not detected, which could be the result of plasmid loss after transposition of *IS*1 into the chromosome. Figure 6.29 shows the distribution of plasmids and *IS*1 elements in all haplotypes, as detected by GoldenGate assay and Figure 6.30 shows the distribution within the H58 phylogenetic tree. As the latter shows, the ST6 IncHI1 plasmid was detected in all lineages of the H58 phylogenetic tree, although many isolates did not have the plasmid. However, apart from those in the ancestral node A, all H58 isolates gave positive signals for two *IS*1 loci targeted in the GoldenGate assay. This is consistent with a single acquisition of the ST6 plasmid by the common ancestor of H58, perhaps belonging to node A itself, followed by transposition of *IS*1 from the plasmid into the chromosome and gradual loss of the plasmid from sublineages following the diversification of H58. However, given that H60 and H61 were present as early as 1965 and 1958, respectively (see Figure 6.27) this would imply that the ST6 plasmid had already been acquired by this time, which seems unlikely. It is not impossible, given that chloramphenicol resistant typhoid was described as early as 1950 (363), but would imply that further resistance genes accumulated in ST6 while it was resident in H58 Typhi.
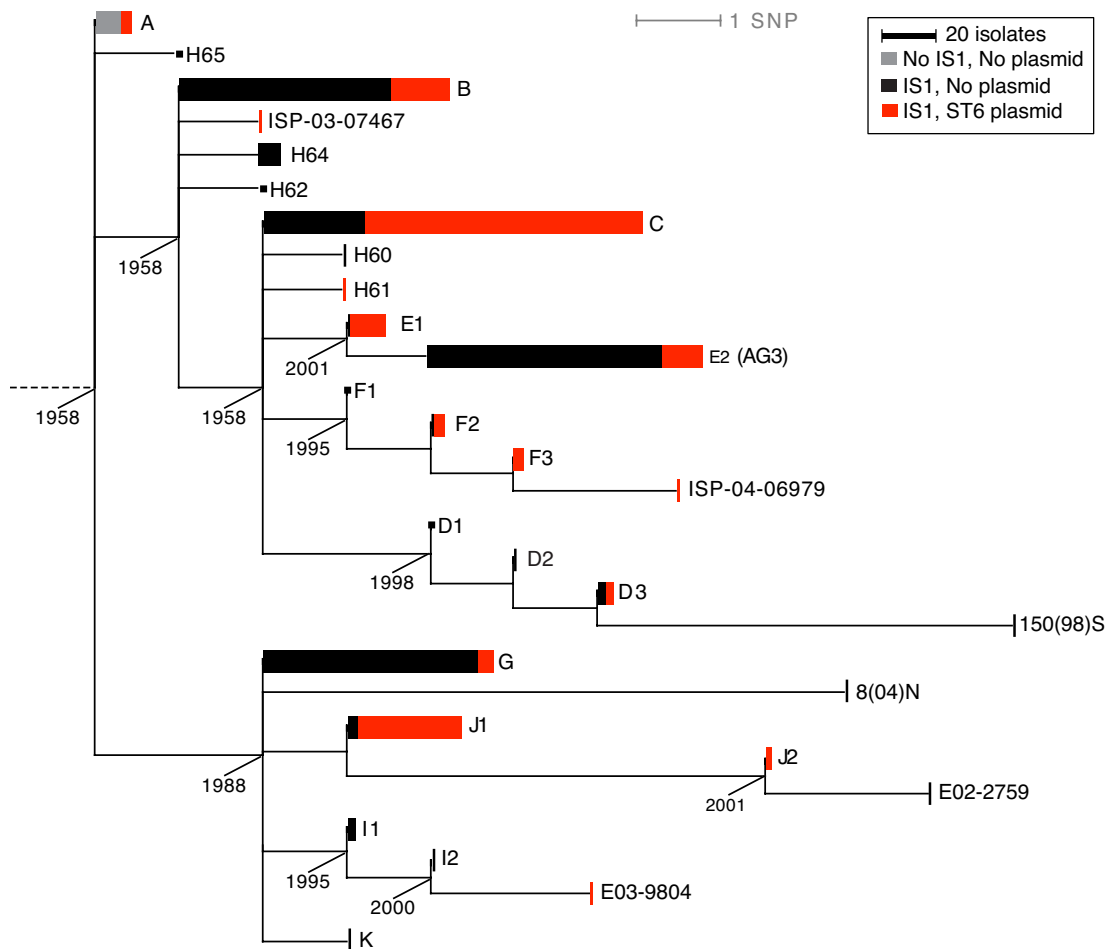
Figure 6.30: Distribution of IncHI1 plasmids and *IS*1 among Typhi H58 subtypes - The phylogenetic tree of H58, dashed line represents the root of this tree, i.e. where H58 joins the rest of the Typhi phylogeny. Nodes are labelled with their name as assigned in this study (A-K) or previously assigned H-group, or with the sequenced isolate which defines the node. Coloured bars indicate the number of isolates at that node (from the regional collections or the Pasteur collection) that contain *IS*1 and/or the IncHI1 plasmid ST6 as indicated in legend. Base nodes are labelled with the earliest year of isolation (among the isolates typed in this study or in (2)).

Figure 6.31: GyrA SNPs distributed among Typhi H58 subtypes - The phylogenetic tree of H58, dashed line represents the root of this tree, i.e. where H58 joins the rest of the Typhi phylogeny. Nodes are labelled with their name as assigned in this study (A-K) or previously assigned H-group, or with the sequenced isolate which defines the node. Coloured bars indicate the number of isolates at that node that have been SNP typed at GyrA positions 83 and 87 (total of 97 H58 isolates). Base nodes are labelled with the earliest year of isolation (among the isolates typed in this study or in (2)).

6.3.8 Typhi H58 and mutations in GyrA
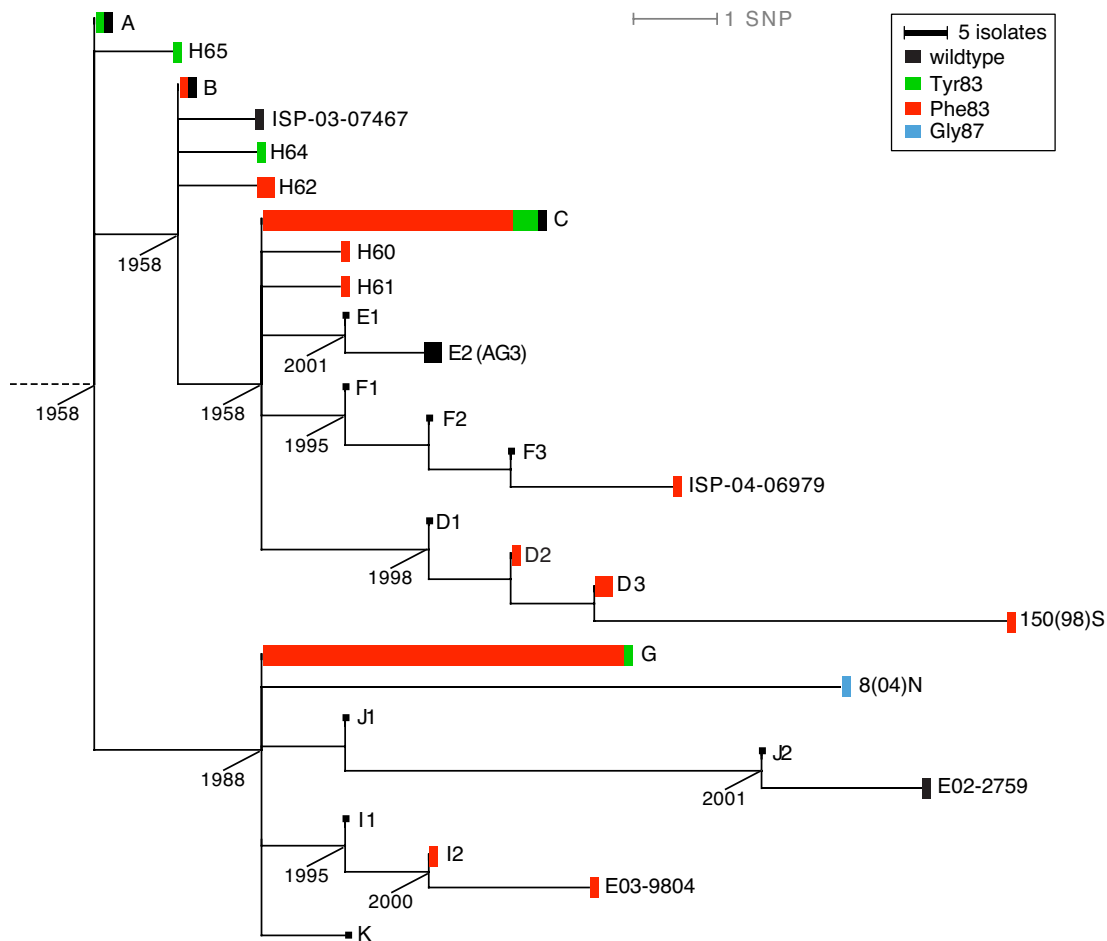
A total of 246 isolates (from the Pasteur collection (2), the Kathmandu collection above and the sequenced isolates) have been typed for SNPs in *gyrA* that are known to be associated with fluoroquinolone resistance. Three isolates had a SNP at position 87: Asn87 in a H6 isolate from Morocco in 2005 (Pasteur collection), and Gly87 in sequenced isolates E02-1180 (H45, from India, 2002) and 8(04)N (H58, from Vietnam, 2004). Of the remaining isolates, 97 were from the H58 cluster, of which 83 carried the Phe83 SNP and five carried the Tyr83 SNP. The distribution of these SNPs among H58 subtypes is shown in Figure 6.31, and supports the notion that mutations arose at GyrA position 83 not once but multiple times within the H58 cluster (2). The sequenced H50 isolate E98-3139 also carried the Phe83 SNP.

6.4 Discussion

6.4.1 Strengths and limitations of the study

The major strengths of the GoldenGate high throughput SNP typing assay for the study of Typhi populations are (a) the reproducibility and phylogenetic informativeness of the sequence-based approach, (b) the increased resolution offered over other sequence-based approaches that have been attempted previously (MLST (1, 725) and SNP typing at <100 loci (2, 256)), (c) the ability to simultaneously target hundreds of loci on both chromosome and plasmid, and (d) the ability to screen hundreds of isolates in a single assay. The increased resolution, particularly within the H58 cluster, has revealed dynamics of the Typhi population at a scale that was largely inaccessible until now. For example, while it has been observed previously that H58 is common in South Asia (2, 570) and Typhi lineages are globally distributed (2), high resolution SNP typing in this study revealed fluctuations within the H58 populations of endemic regions over time (see Figures 6.16, 6.19 and 6.21) as well as geographic differences in population structure (see Figure 6.22). These observations were only possible using high resolution analysis of large numbers of isolates and sequence-based technologies that facilitated direct comparison between data sets. By linking Typhi strain types with IncHI1 plasmid types for the first time, this study provided direct evidence of multiple independent acquisitions of distinct IncHI1 plasmids by distinct strain types,

spread of IncHI1 plasmids within specific Typhi strain types, and the presence of very closely related IncHI1 plasmids in Typhi and Paratyphi A (see Figure 6.28). It also highlighted the close association between H58 Typhi and ST6 plasmids in recent years (see Figures 6.11 and 6.30) and provided evidence for multiple independent occurrences of the same fluoroquinolone-resistance mutations within closely related H58 strains (see Figure 6.31).

The GoldenGate assay has some major limitations, including the inability to accurately type a quarter of known SNP loci, and in particular the inability to target SNPs that are close to other mutations, including the known fluoroquinolone resistance-associated SNPs in *gyrA* and other topoisomerase target genes. Thus to study the acquisition of Nal resistance within the Typhi population would require additional assays to screen for mutations in these genes, such as the Luminex assay used to type *gyrA* SNPs within the Nal-resistant isolates from Kathmandu (6.3.4.2) or the Sequenom SNP typing assay (not yet demonstrated for Nal-resistance Typhi SNPs). However, all SNP typing approaches suffer the much more general problem of being limited to known SNP loci. This limitation is acceptable for many purposes, for example the present study as well as other SNP-typing studies (2, 256) have yielded novel insights into the evolution and population structure of Typhi. However SNP typing by its nature cannot discover novel mutations, thus there will always be a limitation to the resolution possible with this technique. Since H58 was known to be the most prevalent haplotype (2), care was taken in this study to screen for SNPs that would discriminate within the H58 cluster (seven out of nineteen sequenced isolates were chosen from within this cluster, Chapter 2) and indeed 45 SNPs were successfully typed. Yet it is almost certain that there is much more variation within the larger nodes of the H58 tree than we were able to detect. This is most clearly illustrated in Figure 6.27, which shows that the vast majority of H58 isolates tested fell into internal nodes (B, C, G). In contrast the leaf nodes were rarely found, except for E2 which contained nearly half of the isolates from the Mekong Delta study and was defined by SNPs detected in an isolate taken directly from the study population for sequencing (AG3 from An Giang, 2004). On the other hand, within the rest of the Typhi phylogeny there was a lot of redundancy, since the Typhi isolates tested clustered to some extent into clonal groups as opposed to a continuum of haplotypes (see Figures 6.8 and 6.9). Although this may be unsurprising,

it could not have been known conclusively before the present study, and even now it is not entirely clear which SNPs along the longer internal branches of the phylogenetic tree would be the most informative for a lower-resolution set of target loci. Finally, it is difficult to interpret a lack of signal from the GoldenGate assay. For plasmid and resistance gene loci, a lack of signal was interpreted in this study as absence of the plasmid or target sequence from the isolate. This is reasonable as lack of signal at IncHI1 loci was strongly correlated between IncHI1 loci and with lack of MDR pheno-type (see 6.10). However it is difficult to interpret lack of signal at chromosomal loci, which may be due to deletion of chromosomal sequence or simply mutations within oligonucleotide binding sites. This should not affect phylogenetic inference though, as a lack of signal was represented in the allele alignments as the gap character '-', which does not contribute to phylogenetic inference using maximum likelihood methods.

The collections of Typhi isolates that were SNP typed in this study have been col-lected by different research groups over different time periods, for different purposes and using different sampling approaches (see Table 6.1 and the introduction to this chapter 6.1). As such the Typhi populations under study are not directly comparable, although each of the regional collections gives a snapshot of the Typhi population in a defined time and place. In order to directly compare Typhi populations across time and space, a more consistent sampling approach would need to be used, and ideally would include a strategy to collect isolates from asymptomatic carriers as well as typhoid fever patients. The availability of spatial data associated with the Kolkata study provided the opportunity to examine spatial clustering of typhoid cases and Typhi haplotypes. The detection of haplotype-specific spatial clusters in this study suggests that spatial data will be useful in future studies of Typhi populations and potentially genomic epi-demiology of typhoid; ideally high-resolution data would be collected using geographical positioning systems (GPS) which are increasingly cheap and easy to implement. The population-based sampling approach implemented in the Kolkata study should provide a fairly complete view of typhoid fever within the study site, although there still may be issues with residents reluctant to report illness to the health services involved in the study. In particular, clustering within households may be confounded by household preferences to report or not report febrile episodes. The analysis of Typhi haplotypes

presented above from the Kolkata study is currently underpowered, as only half of available isolates were typed. This was due to problems with obtaining adequate yields from DNA extractions at the NICED laboratories in Kolkata, but repeat extractions will be performed to allow all isolates to be SNP typed and provide a complete set of SNP data for this collection. This will provide additional power to detect haplotype-specific differences and may clarify the apparent difference in vaccine efficacy on the Typhi H58-B haplotype. The Nepal study is much smaller and more specific than the Mekong Delta and Kolkata studies, focussing only on pediatric cases of typhoid fever requiring hospitalisation at Patan Hospital. This should be considered the 'tip of the iceberg' of typhoid fever in Kathmandu, as the majority of enteric fever patients presenting at the hospital are treated as out-patients (279) and pediatric enteric fever is rarely reported in Kathmandu (279, 764). The studies of local populations presented here provide proof-of-principle that high-resolution SNP typing, including maximum resolution in H58 and inclusion of IncHI1 plasmid SNPs, is informative for studying Typhi populations, which have been relatively impenetrable to low-resolution typing methods that have achieved success in studying other more diverse bacterial populations. These studies also demonstrate that there is substructure within the H58 population, which can be used to distinguish clones with distinct spatio-temporal distributions.

6.4.2 Typhi populations in endemic areas

This study provides high-resolution data on the structures of Typhi populations within high- and medium-incidence endemic areas. The only such study published to date used the Sequenom platform to target 88 SNPs within 140 Typhi isolates from Jakarta, Indonesia, which is also a high-incidence typhoid endemic area (256) (note I was involved in the analysis and am a co-author on this study). This revealed the presence of nine distinct Typhi haplotypes co-circulating within Jakarta, and geo-positioning data for 54 (39%) of the isolates revealed some spatial clustering within the city. However in Indonesia, typhoid fever follows quite different patterns to those observed in mainland Asia, including a lack of drug resistance (16, 327, 765, 766), prevalence of z66 antigen and haplotype H59 (2, 248, 251, 252, 253, 256, 767) and lack of prevalence of H58 (2, 256).

The present study focused on three high-incidence typhoid endemic regions in Asia and one medium-incidence endemic region in Africa (see Figure 6.1). Although the isolate collections from each region span different time periods and were collected via different sampling methods (Table 6.1), some clear similarities emerge from the SNP typing data. In each region multiple Typhi haplotypes were co-circulating, with H58 by far the most common group in each case (70-98%), see Table 6.5. Some of the haplotypes were shared between regions, while others appeared to be geographically restricted. The clusters of H42 isolates making up the second most frequent groups in Kolkata, Kathmandu and Kenya (labelled H42-A) were identical at all target loci. While H58 was dominant in each study site, the structure of the H58 populations varied and provided the strongest examples of geographical restriction (see Figure 6.27). There was usually a single dominant subtype of H58, which differed in each region (see Table 6.5), with the exception of the Mekong Delta isolates which were fairly evenly split between the closely related subtypes C and E2. However there was clearly variation from year to year, with different haplotypes appearing and disappearing over the course of the studies. This was particularly evident among isolates from Nairobi, which cover the longest time period (21 years). In the late 1980s to early 1990s, H58 made up less than 50% of isolates, although sample sizes were low (see Figure 6.26b). There is a gap in the sampling, but isolate numbers increased from 2001 and from this point H58 was clearly dominant (Figure 6.26b). In Nairobi H58-J1 was the only H58 subtype detected between 1988 and 1998 and the J1-derived J2 node was detected only in 2001 (see Figure 6.26d). These were the only H58 subtypes detected until the appearance of H58-B in 2004, which co-circulated with J1 for a few more years and had all but replaced it by 2008 (see Figure 6.26d).

Patterns of drug resistance varied markedly between the regions studied (see Table 6.5). In Kenya and Vietnam, there were high rates of MDR, associated with IncHI1 ST6 plasmids in H58, and the presence of the plasmids appeared to contribute to the success of the dominant Typhi haplotypes. In Vietnam the H58-E2 subtype, which was generally plasmid free, was replaced in 2005 by the H58-C subtype which had a high rate of the MDR plasmid (see Figures 6.14 and 6.16b-c). In Kenya nearly all H58 isolates carried the MDR plasmid and only two cases of the (plasmid-free) H58-G subtype were detected, possibly due to lack of the plasmid. In Nepal and Kolkata,

	Mekong Delta	Kathmandu	Kolkata	Nairobi
H58	98%	70%	74%	76%
Dominant H58	C (45%)	G (94%)	B (55%)	J1 (64%)
subtype(s)	E2 (41%)	-	G (31%)	B (30%)
MDR (overall)	58%	0	3%	69%
MDR (H58)	59%	0	4%	89%
Nal-R (overall)	96%	63%	nd	nd
Nal-R (H58)	98%	91%	nd	nd

Table 6.5: Summary of Typhi populations from endemic areas - MDR = multidrug resistant, Nal-R = nalidixic acid resistant.

however, MDR was very rare (0% - 3%, see Table 6.5). Nalidixic acid resistance was studied only in the collections from Nepal and Vietnam, and high rates were found in both locations. In Nepal, 91% of H58 isolates were Nal resistant and 97% carried the Phe83 mutation in *gyrA*. The dominant H58-G isolates, all of which carried this mutation, occurred at twice the rate of other, Nal-sensitive haplotypes. In Vietnam, Nal resistance was only observed among H58 isolates, at a rate of 98%.

The peaks observed in these studies are considered to be random (or seasonal) fluctuations in an endemic area, as opposed to outbreak events. In the Mekong Delta, Kathmandu and Kolkata studies, typhoid fever cases showed varying degrees of seasonality (see Figures 6.16, 6.19 and 6.21). In areas that experience a wet season, peaks in typhoid fever and other water-borne bacterial disease is usually associated with onset of the wet season (343, 354, 768, 769, 770). It has been suggested that such peaks tend to occur just before the beginning of the wet season, when water scarcity leads to difficulties in sourcing drinking water and compromised hygiene practices (354). Flooding, too, may be associated with water-borne disease as it becomes hard to keep drinking water and sewerage systems separate (343, 771). In countries with a dry climate, seasonal peaks in typhoid incidence are usually associated with high-temperature (772, 773). The seasonality of typhoid fever and other enteric bacterial disease in the Mekong Delta has been studied over much longer time scales, which revealed weak correlations between high rainfall and periods of high-incidence of typhoid fever, peaking between April and July (354). This is broadly consistent with the present study, in which typhoid fever peaked between March-July in 2004 and May-June in 2005 (see

Figure 6.16). Rainfall data was not available, but this timing corresponds to the usual onset of the wet season (April-November).

In the Kolkata study, similar seasonal patterns were visible, although less pronounced. The first major peak (1 in Figure 6.21) corresponded with the onset of the wet season, a pattern which was also evident in 2006 (5, 6 in Figure 6.21). However other peaks occurred at the beginning of the dry season, or during the wet season (2-4 in Figure 6.21), consistent with previous reports of a lack of seasonality to typhoid outbreaks in India (774). It is interesting to note that greater seasonality was observed in the rural setting of the Mekong Delta than the urban slum setting of Kolkata. This may be due to climatic differences between the two regions, but could also be associated with the different pressures on the water supply. In the Mekong Delta, water scarcity at the end of the dry season is a very real problem, as nearly half the population relies on rivers, streams and ponds for their drinking water and over two-thirds of the population have their toilet facilities directly over water (349). In the Kolkata slums drinking water is taken from public taps, but water pipes and sewerage pipes lie close together, so although contamination is more likely to occur as a result of flooding during the wet season, it is a risk all year round (343).

No seasonality was observed in the Kathmandu study (see Figure 6.19), although this may be due to the low number of cases included in the study, which were limited to children under 12 hospitalised at Patan Hospital with blood-culture confirmed typhoid fever. The majority of enteric fever patients presenting at Patan Hospital are treated as out-patients (279), and pediatric enteric fever is rarely reported in Kathmandu (279, 764) so the present study is unlikely to detect seasonal patterns in typhoid incidence that may exist in Kathmandu.

6.4.3 The evolution of drug resistance in Typhi

This study is the first to link IncHI1 plasmid variation with Typhi strain variation in a phylogenetically informative way. Most studies have been unable to detect differences between plasmids in distinct strain backgrounds (277, 735) or have addressed only plasmid differences (275) or strain differences (276) but not both. One study reported two distinct PFGE patterns for 200 kbp plasmids from strains that had closely related but

distinct chromosomal PFGE patterns, and a third PFGE pattern for a 200 kbp plasmid in an unrelated strain type (736). This is consistent with either co-diversification of plasmid and strain type or independent acquisition of distinct plasmid types by distinct strain types, but these scenarios could not be differentiated due to the lack of phylogenetic informativeness of PFGE. By simultaneously targeting chromosomal and IncHI1 plasmid SNP loci, the GoldenGate assay enabled each IncHI1 and host strain to be positioned on phylogenetic trees, providing direct evidence of independent plasmid acquisition events (Figure 6.11). The plasmid-strain associations revealed instances of the same plasmid being present within distantly related strains (e.g. ST2, ST8, see Figure 6.11) which confirms multiple acquisition of highly similar plasmids by distinct Typhi lineages. In some cases these coincided in time and space, for example ST8 plasmids were found in distinct Typhi strains isolated from Peru in 1981, which although circumstantial does suggest direct transfer between Typhi cells or at least transfer from a single third-party bacteria into multiple Typhi cells at around the same time and place.

The most striking plasmid-strain association identified in this study was that between the ST6 IncHI1 plasmid and H58 Typhi. Although there were hundreds of examples of H58 isolates without plasmids, in particular in Nepal (100% plasmid-free) and Kolkata (96% plasmid-free), the presence of IS1 in all but the most ancestral node of the H58 phylogeny (see Figure 6.30) suggests that the plasmid was acquired early in the life of H58 and subsequently lost. The loss of the plasmid is presumably associated with reduced selective pressure following the switch to fluoroquinolones for the treatment of enteric fever. However the maintenance of the plasmid in the Mekong Delta (59% of H58 isolates) suggests it may still provide some selective advantage, at least in this region. The dominance of the plasmid-containing H58-C isolates in the second year of the study (see Figure 6.16) may be associated with such an advantage. Although inappropriate antibiotic prescribing and dispensing practices are common in Vietnam (775) and high rates of drug resistance have been reported among bacterial contaminants in the human food chain (776, 777), water systems in the Mekong Delta do not contain unusually high levels of antibiotics (778) and a survey of *Salmonella* isolates originating from food, animals and human diarrhea patients in the Mekong Delta found less than 10% of isolates were resistant to any of the drugs tested (including chloramphenicol, streptomycin, ampicillin and nalidixic acid among others) (779). The ST6 IncHI1

plasmid may provide some selective advantage unrelated to drug resistance, perhaps associated with osmoprotectant properties of *betU* encoded on *Tn*6062 (721) or mercury resistance encoded on *Tn*21 (although this was present in most IncHI1 plasmid types). *BetU* may be associated with urinary carriage (721). Successful culturing of Typhi from urine has been reported from typhoid fever patients and asymptomatic carriers, albeit at much lower rates than isolation from stool (295, 780, 781, 782), and urinary shedding has been linked to disease transmission (780).

Unfortunately, SNPs known to confer nalidixic adid resistance were unsuitable targets for the GoldenGate assay (see 6.2.2), and thus the present study offers few novel insights into the evolution of fluoroquinolone resistance in Typhi. This could be remedied in future studies by additional typing of Nal resistance loci using another method such as Luminex SNP typing (as was done for the Kathmandu study), Sequenom SNP typing or targeted resequencing. However, the study did provide increased resolution within the H58 cluster for isolates that had previously been typed at Nal resistance SNP loci in GyrA (the Pasteur collection), or sequenced (see Figure 6.31). The results add further support to earlier assertions that the prevalence of resistance-associated GyrA SNPs among H58 isolates is due to multiple independent mutations arising in different H58 subtypes and not the expansion of a single clone following acquisition of resistance mutations (2).