Chapter 5

IncHI1 multidrug resistance plasmids in Paratyphi A and Typhi

5.1 Introduction

Antibiotic treatment is central to the control of enteric fever. Although vaccines against Typhi have been available for a long time, they are used mainly in travellers rather than inhabitants of endemic areas (15, 425, 429) and there is currently no vaccine that provides strong protection against Paratyphi A (280, 431). Chloramphenicol resistant Typhi was first reported in 1950 (363), just two years after the introduction of chloramphenicol for the treatment of typhoid fever (362). By the early 1970s, Typhi resistant to both chloramphenicol and ampicillin had been observed (364), and multidrug resistant (MDR) Typhi emerged soon after (268, 684), see Figure 1.7. MDR has been broadly defined as resistance to three or more first-line antibiotics, but in most studies of enteric fever refers to resistance to chloramphenicol, ampicillin and co-trimoxazole (e.g. (16, 327)). Since the 1990s the recommended treatment for enteric fever has changed to quinolones and more recently fluoroquinolones (297, 685), although susceptibility to these drugs has been declining among Typhi and Paratyphi A isolates since this time (16). Rates of MDR Typhi fluctuate over time and geographical locations (see for example (16, 327)), but MDR is still a problem in many areas despite the switch to fluoroquinolones (16). For example, in southern Vietnam, over 80% of Typhi isolates tested in 2005 were MDR (16). Unlike Typhi, Paratyphi A isolates have predominantly been susceptible to antibiotics (384, 385). However, in recent years the incidence of MDR Paratyphi A has increased, particularly in Pakistan and India where rates as high as 45% of Paratyphi A isolates have been reported (386, 387, 388). Higher rates of fluoroquinolone resistance among Paratyphi A isolates compared to Typhi isolates have been reported $(279, 336, 686)$, and a recent study in Nepal found MDR was more common among Paratyphi A than Typhi isolates (687). The situation is perhaps most extreme in China, where in some regions more than 50% of enteric fever is now caused by Paratyphi A and is largely drug resistant (285, 335). Among the 159 Paratyphi A isolates examined in pools in Chapter 3, the GyrA-Phe83 SNP which confers increased fluoroquinolone resistance (391, 393) was detected in 17 pools, with an estimated frequency of 55 isolates (35%) .

MDR in Typhi is almost exclusively plasmid-mediated, with the majority of plasmids analysed being of the HI1 incompatibility type (IncHI1) (268, 269, 270, 271, 272, 273, 274, 275, 276), although other plasmids have been reported (277). The few studies which have reported MDR in Paratyphi A have pointed to a key role for plasmids in mediating resistance although few molecular studies have been undertaken (688). A large transferable plasmid of 140 MDa (∼230 kb, similar in size to the 120-200 kbp IncHI1 MDR plasmids found in Typhi) was found in 73% of MDR strains in Bangladesh in 1992 to 1993 (284). A similarly sized plasmid was reported in recent Chinese Paratyphi A isolates (285). However, in Calcutta, India, a smaller plasmid (∼55 kb) was responsible for conferring MDR in Paratyphi A isolates (286). Recent work by Satheesh Nair at the Sanger Institute demonstrated that in Paratyphi A isolated from Pakistan between 2002-2004, MDR was associated with IncHI1 plasmids of approximately 220 kbp (283).

The prototype IncHI1 plasmid is R27, isolated from Typhimurium in 1961 (269) (note that some manuscripts have incorrectly reported that R27 was isolated from Typhi). R27 is self-transmissible and can transfer between Enterobacteriacae and other gramnegative bacteria (689). Conjugal transfer is encoded in two regions, $tra1$ and $tra2$ (690). Tra2 contains genes for pilus production and mating pair formation (691), while tra1 contains genes required for translocation of DNA across the bacterial cell membrane and initial replication upon entering the recipient cell (692). Conjugal transfer of IncHI1 plasmids is thermo-sensitive, with maximum transfer efficiency at ambient temperatures (14-27 \degree C) and highly impaired transfer rates at 37 \degree (689, 692). The finished sequence of R27 was published in 2000 (692), followed one year later by that of pHCM1, the IncHI1 plasmid of the sequenced Typhi isolate CT18 isolated in 1993 (46). The two plasmids shared 168 kbp with 99% sequence identity, including the tra1 and tra2 regions, but differed in their resistance gene insertions. R27 contained just one resistance locus - Tn10, encoding resistance to tetracycline. The Typhi plasmid pHCM1 contained the same Tn10 transposon, along with several other mobile elements including integrons and the transposons $Tn21$ and $Tn9$, carrying several drug resistance genes: dhfR14 (trimethoprim resistance), sul2 (sulfonamide resistance), catI (chloramphenicol resistance), bla (TEM-1; ampicillin resistance) and $strAB$ (streptomycin resistance). These transposons coincide with those found in IncHI1 Typhi plasmids in the early 1970s, including $Tn3$ (of which $Tn21$ is a subtype) encoding ampicillin resistance, T_{n9} encoding chloramphenicol resistance and T_{n10} encoding tetracycline resistance (269).

The Paratyphi A isolate AKU 12601, isolated in Pakistan in 2002 and presented in Chapter 3, was multidrug resistant and contained a 212 kbp IncHI1 plasmid. The plasmid, pAKU 1, was also sequenced and finished at the Sanger Institute. This Chapter begins with the annotation of the pAKU 1 sequence and comparison of this plasmid sequence to those of R27 and pHCM1. The remainder of the chapter focuses on phylogenetic relationships between these and more recently sequenced IncHI1 plasmids.

5.1.1 Aims

The aims of this chapter were to annotate the recently completed sequence of the Paratyphi A MDR IncHI1 plasmid pAKU₁ and compare the sequence to that of other available IncHI1 plasmids. Specific aims of the analysis were to:

- identify the conserved backbone of the IncHI1 plasmid and determine the phylogenetic relationships between available plasmid sequences;
- determine the mobile genetic elements encoding MDR in pAKU₁ and other IncHI1 plasmids, and examine their distribution among the population of IncHI1 plasmids; and

• gain insight into the spread of drug resistance via IncHI1 plasmids in Typhi and Paratyphi A.

5.2 Methods

5.2.1 Annotation

The 212,711 bp plasmid pAKU₁ was sequenced and assembled by members of the Pathogen Sequencing Unit at the Sanger Institute using capillary-based Sanger sequencing. Gene prediction was performed by Nick Thomson at the Sanger Institute using Glimmer. Predicted genes or CDSs were assigned systematic identifiers with the prefix 'SPAP'. The nucleotide sequences and translated protein sequences of CDSs were used to batch query several databases and motif recognition programs to assist with annotation: BLASTN search of EMBL database (nucleotide similarity), fasta3 search of EMBL (protein similarity), Pfam search (protein domains (693)), TMHMM analysis (transmembrane hidden Markov model to identify protein transmembrane domains (657)), SignalP (to identify signal peptides, indicative of protein export (694, 695)), HTH (to identify helix-turn-helix DNA-binding motifs within proteins (577, 696)) and tandem repeats (to identify tandem repeats within coding sequences). Each CDS was annotated manually in EMBL format, using Artemis (697) to coordinate the display of results from the aforementioned database searches for consideration during annotation. Start and stop codons were determined on the basis of (a) similarity with known proteins, (b) presence of Shine-Dalgarno sequence (AGGAGG) (698) and (c) GC frame plots across the CDS and flanking sequence. Annotation included assignment to a functional category listed in Table 5.1, using the '/colour' identifer in the EMBL format.

These functional assignments took into account evidence from the database searches described above (e.g. genes with transmembrane domains identified using TMHMM would be assigned to category 3). Annotation of IS elements was done using the IS finder database (699) to identify the coding sequence for the transposase gene and flanking inverted repeat sequences. Transposons were annotated with reference to previously described transposon sequences and structures identified using sequence homology and literature searches. Each CDS was assigned a confidence value (identifier '/confidence level') for the annotation: 1=experimental evidence of function in this

Table 5.1: Functional categories for genome annotation - Each open reading frame is assigned to a category using the /colour identifier in the EMBL-style annotation.

species, 2=experimental evidence of function in related species, 3=presence of functional domains, 4=no real evidence of function.

5.2.2 Sequence comparison and SNP detection

Plasmid sequences were compared using BLASTN and the comparisons viewed in the Artemis Comparison Tool (ACT) (604), which aids visualisation of synteny and similarity between pairs of sequences. SNPs between finished plasmid sequences were identified using MUMmer 3.1 (576). MUMmer's nucmer algorithm was used to align the nucleotide sequences of R27 and $pHCM1$ to $pAKU_1$, and its show-snps algorithm was used to detect SNPs based on this alignment. SNPs were identified from Solexasequenced Typhi IncHI1 plasmids by mapping reads generated from the three Typhi isolates to the finished pAKU 1 sequence, using Maq (564) and quality filters described in 2.3.1.3.

5.2.3 Phylogenetic analysis

SNPs called in repetitive regions or inserted sequences were excluded from phylogenetic analysis, so that phylogenetic trees were based only on the conserved IncHI1 backbone sequence. SNP alleles were concatenated to generate a multiple alignment of SNP alleles. A maximimum likelihood phylogenetic tree was fit using RAxML (644) with a GTR+Γ model and 1,000 bootstraps. A phylogenetic network was also constructed, using the parsimony splits method implemented in SplitsTree4 (603) as described earlier (3.2.3).

5.2.4 PCR

PCR primers were designed using Primer3 (700) according to the following criteria: melting temperature 56° C, no hairpins or dimers affecting 3' ends, no cross-dimers between forward and reverse primers. Primer sequences are given in Table 5.2. Primers were designed by myself, PCR assays were performed by Minh Duy Phan (Sanger Institute).

5.2.5 Accession codes

The annotated plasmid sequence was submitted to the EMBL database with accession number AM412236. Plasmid seqeunces used for comparative analysis were: R27 - AF250878; pHCM1 - AL513383; R478 - BX66401; pRSB107 - AJ851089); pU302L - AY333434; DT193 - AY524415; IncI plasmid of Enteritidis - AJ628353; pMAK1 - AB366440.

		Length	Length	Target
Label	Forward and reverse primer	pAKU ₋₁	pHCM1	feature
IncHI1	CGAAATCGGTCCAACCCATTG	110		repHIA
	CGACAACTCATCAGAAGCGTCAAC			
\mathbf{A}	GAAAGGAATCATCCACCTTCA	419	990	del on pAKU ₋₁
	AACTGTCGCTACGCCTGACT			
$\mathbf B$	ATCCAGCGTGCAAAGATTTC	407	2589	del on pAKU ₋₁
	TGGGGGAGAACACCACTTTA			
$\mathbf C$	AAAGATGCAATGGGAGGAGA	289	4399	insert on pHCM1
	GCGCAGCTGCTTCAATTA			
D	TAGGGTTTGTGCGGCTTC	3138	none	insert on pAKU ₋₁
	CCTTCTTGTCGCCTTTGC			
$\mathbf E$	TCAAGGCAGATGGCATTCCC	156	none	sul1
	CGACGAGTTTGGCAGATGATTTC			
$\mathbf F$	GTGTCGAGGAAAGGAATTTCAAGCTC	191	none	dhfR7
	TCACCTTCAACCTCAACGTGAACAG			
G	GATGGAGAAGAGGAGCAACG	989	989	$bla/sul/str$ -Tn21
	TTCGTTCCTGGTCGATTTTC			$(strB/tniA\Delta)$
H	GTGCTGTGGAACACGGTCTA	271	1598	$Tn21-Tn9$
	TCATCAACGCTTCCTGAATG			$(Tn21 \; tmpA/Tn9 \; acetyltransferase)$
\bf{I}	ACGAAAGGGGAATGTTTCCT	163	1490	$Tn21-Tn9$
	CGAGTGGGAATCCATGGTAG			(merR/Tn9)
$_{\rm J}$	CAAAATGTTCTTTACGATGCC	2219	none	$Tn9-tra2$
	CCAGACAGGAAAACGCTCA			(cat/trhN)
$_{\rm K}$	CTGTGCCGAGCTAATCAACA	1314	none	$Tn21-Tn9-tra2$
	ACGAAAGGGGAATGTTTCCT			(merR/trhI)
L	TTTTAAATGGCGGAAAATCG	none	1872	$Tn9-Tn10$
	GCCAGTCTTGCCAACGTTAT			(insA/tetA)
$_{\rm M}$	GGGCGAAGAAGTTGTCCATA	none	2195	$Tn9$ -backbone (pHCM1 site)
	ATTCGAGCAAAACCATGGAA			cat/HCM1.203
$\mathbf N$	CGGGATGAAAAATGATGCTT	none	2180	$Tn10$ -backbone (pHCM1 site)
	GGTCGGTGCCTTTATTGTTG			$Tn10$ /HCM1.247
\circ	GCGTACAAAAGGCAGGTTTG	1823	none	$Tn10/backbone (pAKU_1 site)$
	GCTTGATGATGTGGCGAATA			$tetD$ /SPAP0276
P	TGGTCGGTGCCTTTATTGTT	1899	none	$Tn10/backbone (pAKU_1 site)$
	GGGCGTCAGAGACTTTGTTC			SPAP0261/Tn10
Q	TTCGCCCGATATAGTGAAGG	1923	none	$strAB/backbone (pAKU_1 2nd copy)$
	CTAACGCCGAAGAGAACTGG			strA/SPAP0228

Table 5.2: PCR primers for analysis of IncHI1 plasmids - Used to detect features of the IncHI1 backbone and resistance gene insertions. The locations of A-Q in pAKU 1 and/or pHCM1 are shown in Figures 5.2 and 5.4.

5.3 Results

5.3.1 Characterisation of IncHI1 plasmid backbone and resistance gene insertions

Coding sequences in the pAKU₋₁ plasmid were annotated as described in 5.2.1. The 212,711 bp plasmid contained 237 coding sequences, the majority of which are of unknown function. Figure 5.1 shows the distribution of annotated functional groups, as well as a more specific breakdown of plasmid and resistance functions.

Figure 5.1: Functions of genes annotated in the IncHI1 plasmid pAKU 1 from Paratyphi A - A total of 237 coding sequences were annotated and assigned to functional groups 0-13 as defined in Table 5.1; these are labelled with the corresponding group numbers on the x-axis. A subset of these genes (66) have known functions associated with multidrug resistance plasmids, these are further divided into specific plasmid and resistance functions (labelled 'plasmid functions' on the x-axis).

5.3.1.1 The conserved IncHI1 backbone

Comparison with EMBL/GenBank sequence databases (January 2007) revealed high DNA sequence similarity between pAKU₋₁ and the IncHI1 plasmids R27 (692, 701) and pHCM1 (46). Detailed comparative analysis of the three plasmid sequences revealed a 164.4 kb shared IncHI1-associated backbone, with 99.7% nucleotide identity among the three plasmids. This shared backbone constitutes 83% of pAKU₋₁ sequence and includes the IncHI1 incompatibility locus, the tra1 and tra2 conjugative transfer regions (690, 691) and three potential replicon elements (RepHI1A, RepHI1B, RepFIA) characteristic of IncHI1 plasmids (702), the locations of which are indicated in Figure 5.2a. The rest of the shared backbone harbours genes involved in the core plasmid functions of replication, maintenance and conjugative transfer, as well as many hypothetical genes with no database matches to sequences outside IncHI1 and the Serratia marcescens IncHI2 plasmid R478 (703) (EMBL/GenBank, June 2009).

The shared IncHI1 backbone sequences of pAKU₋₁, R₂₇ and pHCM1 were aligned and nucleotide differences determined using MUMmer (5.2.2). This analysis found pAKU 1 shared 99.71% nucleotide identity with pHCM1 and 99.89% with R27. Figure 5.2b shows an unrooted phylogenetic tree based on the number of single nucleotide changes found between the three IncHI1 backbones. As the tree shows, the plasmid backbone of pAKU 1 was closer to that of R27 than pHCM1. This is supported by the presence of shared variations in the backbones of pAKU₋₁ and R₂₇ relative to pHCM₁. These are marked $(**)$ in Figure 5.2a and include a small inversion near the $5'$ end, two deletions downstream of this inversion and a gene (annotated as R0107 in R27 and SPAP0320 in $pAKU_1$) inserted at the 3' end of the shared backbone. A large region was inverted on pAKU 1 relative to R27 and pHCM1, however this occurred on pAKU 1 or a similar precursor plasmid rather than a common ancestor of R27 and pHCM1 (see Figure 5.4).

The 19 kbp $tra1$ region contains nine genes essential for transfer, the $oriT$ (origin of transfer) site and a further 4-5 CDSs of unknown function (692). Tra1 was conserved as a single sequence block with $>99.9\%$ identity across pAKU₋₁, pHCM1 and R27, differing from each other by only 14-16 bp. Few of these changes occurred in coding regions: nearly all the encoded proteins were 100% identical at the amino acid

Figure 5.2: Comparison of three complete IncHI1 plasmid sequences from Salmonella - (a) Representative alignment of the 164kb IncHI1 backbone sequences of pAKU 1, R27 and pHCM1, with the sites of major insertions, deletions and inversions indicated. Note that the plasmids are actually circular and are shown as linear here merely for ease of comparison. Red boxes show the sites of IncHI1 replicons, green box represents the incompatibility region. Blue boxes represent resistance gene insertions, scaled to indicate relative size compared to backbone, and are labeled as in Figure 5.3 (a=Tn9, b=Tn21, c=Class I integron, $d=bla/sul/str, e=Tn10, e^*$ =truncated $Tn10$). Black bars indicate PCR target amplicons, labeled as in Table 5.2 and Figure 5.4. Transposon insertion sites are labelled in red, corresponding to labels given in the text and Figure 5.6. Other insertions are shown in white boxes (ins), note that the insertion targeted by PCR D is $Tn6062$. Inversions are shown as graded black/grey boxes, gradient indicates direction. (b) Tree showing the relationship between 164kb IncHI1 backbone sequences of the three plasmids, based on SNPs. Branch lengths are proportional to the number of SNPs, indicated next to branches. The position of four major differences (** in a) is indicated; the position of the root is imprecise due to lack of suitable plasmid sequences for use as outgroups.

level, with the exception of the pAKU 1-encoded $trhG$ which differed from the R27 and pHCM1 orthologs at two amino acid residues (1253 and 1321) and from the pHCM1 ortholog at a third residue (670). The *oriT* site was completely conserved across all three plasmids (although not precisely defined, the site lies in the region between traH and $trhR$, which was 100% identical). The 23 kbp $tra2$ region contains genes required for synthesis of the H-pilus, mating pair stabilisation and DNA transfer, genes required for plasmid partitioning and stability and the plasmid incompatibility region (inc). Apart from some insertions and rearrangements within $tra2$ in pAKU₋₁ and pHCM1, the sequences first characterised in R27 (690, 691, 692) were highly conserved among pAKU₋₁, pHCM1 and R27 ($>99.85\%$ nucleotide identity). Most of the tra2-encoded proteins were 100% identical at the amino acid level, with single amino acid differences in four pHCM1 orthologs (trhC, parA, orf16, trhP), one R27 ortholog (trhW) and three residue changes in the pAKU₋₁ ortholog of *orf9*.

5.3.1.2 Comparison of drug resistance genes in pAKU 1 and pHCM1

The pAKU 1 plasmid sequence contained multiple antibiotic resistance gene elements inserted into the IncHI1 backbone. These insertions were highly clustered relative to the conserved IncHI1 backbone and were related to antibiotic resistance genes found on pHCM1 but not R27 as shown in Figure 5.2a. These resistance genes can be attributed to the insertion of previously described transposable elements (see Figure 5.3) into different positions in the plasmid backbones.

Tn10, carrying genes for tetracycline resistance (tet) $(704, 705, 706)$ (Figure 5.3e), was present on pAKU₋₁ as well as R27 (706) and pHCM1 (46), although part of the transposon was missing from pHCM1. The insertion of $Tn10$ is mediated by flanking elements IS10-left and -right (704), and insertion of the transposon generates 9 bp direct repeats of the target sequence (target site duplications) (707). The site of $Tn10$ insertion into the backbone was different in each plasmid (see Figure 5.2a), generating distinct flanking repeats in each case and indicating that the transposon was independently acquired in each plasmid rather than by a common ancestor. No further resistance insertions were present on R27 (692).

Figure 5.3: Transposons identified in pAKU 1 - The gene order shown here agrees with the consensus from other sequences; note that $Tn9$ and $Tn21$ have been disrupted by rearrangements in pAKU 1 and pHCM1 (see Figure 5.4a,d). Upper and lower bands represent forward and reverse strands; open triangles represent inverted repeats, filled triangles represent direct repeats (target site duplications). The insertion sites shown for (d) into (c), (c) into (b) and (b) into (a) are conserved in $pAKU_1$, $pHCM1$ and $pRSB107$.

Figure 5.4: Rearrangements of composite transposons inserted in IncHI1 plasmids - The inferred ancestral version of the composite transposon (b) and its rearrangements in three plasmids. $*=$ genes inserted relative to the composite transposon; $\hat{ }=$ genes in the variable integron gene cassette. Colours and genes are the same as in Figure 5.3; upper and lower bands represent forward and reverse strands; open triangles represent inverted repeats. Filled triangles represent target site duplications: green="TTGCGCCG"; orange="AAAAAAAG". Regions of identical sequence are joined by coloured boxes (tan when in direct orientation, green when inverted). Black lines indicate PCR amplicons, labels correspond to those in Table 5.2 and Figure 5.2.

Tn9, with identical copies of the chloramphenicol resistance gene cat (368) (Figure 5.3a) was present on pHCM1 (46) and pAKU 1. The transposition of $Tn9$ is accompanied by 9 bp target site duplications (708). Distinct insertion sites, accompanied by distinct target site duplications, were evident in the two plasmids, suggesting that $Tn9$ was inserted independently into the backbones of pAKU₋₁ and pHCM1 (sites shown in Figure 5.2a).

 $Tn21$, harbouring a class I integron and mercury resistance (*mer*) operon (709) (Figure 5.3b-c) was also identified on both pHCM1 (46) and pAKU 1, although in each case the primary transposable element has been disrupted by IS26 insertions and subsequent sequence rearrangements (Figure 5.4). Tn21 was inserted at the same site within $Tn9$ in pAKU₋₁ and pHCM1; pHCM1 also harbours a second, divergent copy of $Tn21$ elsewhere on the plasmid (a in Figure 5.2a). The resistance gene cassettes associated with the class I integrons (Figure 5.3c) differ in the two plasmids: $dh\hat{R}$ in pAKU₋₁ and $dhfR14$ in pHCM1 (both encoding trimethoprim resistance (710, 711)). The sull gene, encoding resistance to sulfonamides (712), was adjacent to $dh\hat{r}\hat{\gamma}$ in pAKU₋₁. Sulf is frequently associated with class I integrons, however it is not believed to be part of the integron cassette (709).

An identical ∼9 kb sequence, incorporating bla_{TEM-1} (a beta-lactamase), sul2 (sulfonamide resistance (713)) and $strAB$ (streptomycin resistance (714)) genes flanked by IS26 elements (Figure 5.3d), was present on both pHCM1 (46) and pAKU 1. BLAST searching (most recently in June 2009) revealed that this is a promiscuous sequence, referred to hereafter as $bla/sul/str$, that is also present in the 120 kb IncF plasmid pRSB107 (unknown host, Germany, 2005) (715) and the F-like plasmid pU302L of Typhimurium strain G8430 (Centre for Disease Control, USA) (716). The sequence has also been identified in the chromosome of Typhimurium strain DT193 (Ireland, 1998) and (in part) in an IncI plasmid of Enteritidis (Italy, 1997) (717). As has been suggested previously (716) , it is likely that this $bla/sul/str$ sequence has moved as a single unit among enteric bacteria. Two IS26 in direct orientation can mediate tranposition of intervening sequences, generating 8 bp direct repeats on either side of the conjugate transposon (718). However no such repeats could be identified in the sequences under study. Thus the $bla/sul/str$ sequence may have originally been transferred into

Tn21 via recombination between its flanking IS26 sequences and IS26 elements already inserted within Tn21.

5.3.1.3 A composite resistance transposon

Although the transposons of pAKU 1 and pHCM1 have been disrupted by several IS element-mediated insertions and sequence rearrangements, sequence identity at the boundaries of Tn21 and the $bla/sul/str$ insertion suggests that Tn9, Tn21 and $bla/sul/str$ may have been transferred as a single unit between plasmids. Specifically, it is hypothesised that some plasmid first acquired $Tn9$, followed by the transposition of $Tn21$ into Tn9, 3' of the cat gene (see Figure 5.3a,b). A sequence described as $Tn2670$, identified in the *Shigella* IncFII plasmid NR1 (later called R100) (709), contains $Tn21$ inserted at the same locus in T_n9 and is therefore a likely precursor. At some point $bla/sul/str$ was inserted into the integron in Tn21, adjacent to $\text{tni} \Delta \Delta$ (Figure 5.3c,d). The resulting 24 kbp composite transposon has since been transferred between plasmids, at the very least between distinct IncHI1 backbones. The transposition mechanism is presumably by the IS1 ends of Tn9, as direct repeats are evident at opposite ends of the IS1 elements in pAKU₋₁. The same composite transposon is evident in plasmid pRSB107, sequenced from an unknown bacterial host from a waste water-treatment plant, albeit with additional resistance gene insertions (Figure 5.4c).

Once inserted into the ancestors of pAKU 1 and pHCM1, the composite transposon sequence has been disrupted by rearrangements mediated by IS elements (see Figure 5.4). In pAKU₋₁, two inversions in the $5'$ end of the composite transposon appear to have been mediated by IS26 elements integrated into the plasmid in direct orientation (green regions in Figure 5.4a). A large inversion appears to have occurred between IS26 elements inserted in the backbone and into the $Tn21$ tnpR gene (marked *IS26) in Figure 5.4), separating the 5' ends of $Tn9$ (IS1, cat) and $Tn21$ (tnpA, tnpR-3' fragment) from the rest of the composite transposon. This is supported by the present arrangement of 8 bp target site duplications adjacent to IS26 elements (green and orange arrows in Figure 5.4). A smaller inversion appears to have occurred between the IS26 inserted in $tmpR$ and the 5' IS26 of $bla/sul/str$, disrupting the class I integron (see Figure 5.4a). This is supported by analysis of the configuration of $IS26$ target site duplications, which were inverted along with the rest of the sequence between IS26 elements (green and orange arrows in Figure 5.4). These inversions have presumably deactivated the composite transposon in this plasmid, as the IS1 genes are now in opposite orientation and separated by 62 kb, thus disrupting $Tn9$. $Tn21$ and the integron are similarly disrupted, although $bla/sul/str$ may still be capable of transfer via IS26-mediated transposition or homologous recombination.

In pHCM1, recombination between an $IS26$ element inserted between $tnpA$ and $tnpR$ and the 5' IS26 element of $bla/sul/str$ resulted in deletion of $tmpR$, $tmpM$, $int1$ and the integron gene cassette (Figure 5.4d). IS4321 elements (purple in Figure 5.4d) were also inserted within the $Tn21$ inverted flanking repeats, demonstrated to be a preferred target site for this IS element (719). In pRSB107, there were two additional resistance gene insertions within the composite transposon (Figure 5.4c). The $Tn4352B$ kanamycin/neomycin-resistance transposon was inserted at the 3' end of $bla/sul/str$. This transposon comprised the aph gene (aminoglycosid 3'-phosphotransferase, conferring kanamycin resistance (720)) flanked by IS26 elements, so may have been inserted at this position via recombination with the $5'$ -end IS26 element of $bla/sul/str$. A macrolide resistance module was also inserted between the integron gene cassette and bla/sul/str.

5.3.1.4 Other insertions in pAKU 1

Plasmid pAKU₁ contained an additional transposon not present in pHCM1, designated Tn6062, inserted within $tra2$ (D in Figure 5.2a). The transposon is made up of four genes, including $betU$ and a conserved hypothetical protein (SPAP0105) flanked by IS1 elements in direct orientation. A 9 bp target site duplication was evident on either side of $Tn6062$ (as with $Tn9$, which is also composed of genes flanked by IS1 elements), confirming that the four genes were inserted as a single unit. BetU contains a betaine-choline-carnitine transporter family domain and encodes a betaine uptake system, capable of transporting glycine betaine and proline betaine (721). It was first described in E. coli strains causing polynephritis (ascending urinary tract infection) and is believed to be an osmoregulator, allowing E. coli to survive the high osmolality and urea content in urine (721). However the gene is distributed among E. coli with a range of pathogenic phenotypes, so its osmoprotectant properties may be useful in other environmental contexts (722). The pAKU₋₁ betU sequence shares 99% identity with

the amino acid sequences found in E. coli, and 99% and 98% identity respectively with protein sequences found in Shigella boydii and Klebsiella pneumoniae. The conserved gene SPAP0105 contains a signal peptide sequence and four probable transmembrane helices, suggesting it may be an outer membrane protein, however it contains no protein domains of known function. BLASTN searching of the EMBL database revealed the pair of genes, $SPAP0105$ and $betU$, are present adjacent to each other in the chromosomes of several $E.$ coli strains and the plasmid pKPN3 of $K.$ pneumoniae. However, the precise structure of $Tn6062$, with two flanking IS1 elements, was not detected outside of pAKU₁ (BLASTN search of EMBL database, June 2009).

Plasmid pAKU₋₁ also contained a second region, encoding four genes SPAP0280-83, that was not present in pHCM1. The genes include citrate transporters $citA, citB, a$ hypothetical protein (SPAP0281) and lysR-family transcriptional regulator nac, and were inserted 3' of Tn10 in pAKU₋₁. The entire sequence was also present in R27, as well as the chromosomes of Typhi, Paratyphi A, Paratyphi C and Choleraesuis (23% divergence between chromosomal and plasmid sequences at the nucleotide level). Early studies of IncHI1 plasmids found that most IncHI1 plasmids, but not other plasmids, were able to confer citrate utilisation Cit+) upon transfer to otherwise Cit- E. coli and Shigella strains (723, 724). However the Typhi strains from which the Cit+ IncHI1 plasmids were isolated did not appear able to utilise citrate in culture (723), despite the presence of the $citA$ and $citB$ transporters and $citAB$ two-component system that has subsequently been found in all sequenced Typhi genomes to date (see 2.3.4.1).

5.3.2 Evolution of IncHI1 plasmids and MDR

5.3.2.1 Phylogenetic analysis of the IncHI1 plasmid backbone

A further two finished IncHI1 plasmid sequences have recently become available, plasmid pMAK1 from S. enterica serovar Choleraesuis (EMBL: AB366440) and p0111 from an enterohemorrhagic E. coli (EHEC) strain (sequence provided by Tetsuya Hayashi, University of Miyazaki, Japan; May 2008). Both plasmids were similar to pHCM1 in their conserved backbone sequence and resistance gene insertions, and were highly similar to each other. These plasmids, as well as pHCM1 and R27, were compared to pAKU₋₁ using MUMmer to identify SNPs within the conserved IncHI1 backbone (5.2.2). SNPs were also identified within three IncHI1 plasmids sequenced with Solexa from H58 Typhi isolates E03-9804, ISP-03-07467 and ISP-04-06979 (2.3.3.3, see methods in 5.2.2). These plasmids were similar to $pAKU_1$, with 100% coverage of the pAKU 1 sequence, and were indistinguishable from each other.

A total of 345 SNPs were identified among the backbone sequences of all available IncHI1 plasmids, shown in Figure 5.5. Phylogenetic analysis of the SNP data confirmed that the plasmids from Typhi H58 isolates were much more closely related to the Paratyphi A plasmid pAKU 1 than the Typhi plasmid pHCM1 (Figure 5.6). It also confirmed that p0111 and pMAK1 from E. coli and Choleraesuis were closest to pHCM1 (Figure 5.6). An additional 16 SNPs were identified among plasmid sequences present in the Paratyphi A pools, as described in 3.3.3.6. These formed a tight cluster with pAKU₋₁, each differing from pAKU₋₁ and one another at $\langle 10 \rangle$ loci (see Figure 3.22). An MLST approach was recently developed by Minh Duy Phan at the Sanger Institute to study IncHI1 plasmids by comparing sequences from six gene fragments. The analysis of 40 plasmids (including pHCM1, pAKU₋₁, R₂₇ and E03-9804) identified eight SNPs, defining eight sequence types (STs) (725). The relationships between STs were consistent with the phylogenetic relationships described here, with pAKU₁ and E03-9804 belonging to distinct STs separated by a single SNP (ST7 and ST6 respectively), and pHCM1 and R27 being much more distantly related (ST1 and ST5) (725).

Figure 5.5: Distribution of SNPs in the pAKU 1 IncHI1 plasmid - Outer two rings = coding sequences on forward and reverse strands; blue = conserved backbone, red $=$ insertions, labelled in red text. Third ring (black) $=$ SNP loci included in phylogenetic analysis. Central plot shows GC deviation $((G-C)/(G+C))$, i.e. the difference in G content between the forward and reverse strands); inwards $=$ negative deviation (low G), outwards $=$ positive deviation (high G). Outer labels show pAKU₋₁ sequence coordinates (bp).

Figure 5.6: Phylogenetic trees of IncHI1 plasmids based on sequence data - (a) Maximum likelihood phylogenetic tree; each bipartition has 100% support from 1,000 bootstraps; scale bar is in substitutions per site estimated by maximum likelihood; position of the root (inferred from separate phylogenetic analysis of sequences shared between IncHI1 plasmids and IncHI2 plasmid R478) indicated with open circle. Inferred independent acquisitions of transposons are shown in grey; $Tn9$ has distinct insertion sites in two lineages, $Tn10$ has different insertion sites in three lineages, labelled A-C as in the text; positions of these sites in the IncHI1 backbone are indicated in Figure 5.2a. (b) Split network; scale bar is number of SNPs.

5.3.2.2 Drug resistance insertions in IncHI1 plasmids

The H58 Typhi plasmids contained the $Tn9-Tn21-bla/sul/str$ composite transposon inserted within the *tra* conjugal transfer region, the same insertion site as in pAKU₋₁ (labelled Tn9-A in Figure 5.6a, site indicated in Figure 5.2a). They also contained Tn10, inserted in the same site as in pAKU₋₁ (Tn10-A in Figures 5.2a and 5.6a). This was confirmed by the presence of reads mapping across each insertion boundary in each isolate, and the successful amplification of PCR products across these boundaries (see 5.2.4 and Table 5.3). PCR results agreed with insertion sites determined from sequence data wherever both data types were available (see Table 5.3). Comparison of the finished sequences showed that plasmids p0111 and pMAK1 also contained the composite transposon, inserted within $Tn10$ at the same insertion site as in pHCM1 (Tn9-B in Figures 5.2a and 5.6a). The insertion site for $Tn10$ itself also matched that of pHCM1 $(Tn10-B)$ in Figures 5.2a and 5.6a).

A study of MDR IncHI1 plasmids collected in Vietnam between 1993 and 1996 found that pHCM1-like plasmids, common until 1996, were replaced by a novel IncHI1 plasmid with a distinct RFLP pattern (275). A plasmid representative of the novel type, pSTY7, was found by plasmid MLST to be of the ST6 type, confirming its distinction from pHCM1 (ST1) (725). PCR analysis of resistance genes in the Vietnam study showed that pSTY7 contained *sul1*, dfrA7 and the full tet operon tetRACDD, similar to pAKU₋₁ but not pHCM1 (275). To check how closely related pSTY7 (ST6) was to pAKU 1 (ST7) and the Typhi H58 plasmids (ST6), PCR was performed as outlined above (5.2.4). The results indicated that pSTY7 was highly similar to the ST6 plasmids from Typhi, with the same insertion sites for all resistance genes tested (see Table 5.3). All ST6 plasmids contained the same resistance insertions as pAKU₁ (ST7), with the exception of an additional copy of $strAB$ that was detected only in pAKU₋₁. These results are consistent with one acquisition each of $Tn10$ and the composite transposon in a common ancestor of all ST6 and ST7 plasmids, as shown in Figure 5.6a.

Table 5.3: Resistance gene insertions in IncHI1 plasmids determined from sequence data and PCR - PCR products are labelled as in Table 5.2 (which gives primer sequences) and Figures 5.2a and 5.4 (which illustrate the positions of target amplicons); yes=successful amplification, no=no amplification with these primers; *=amplicon larger than that from pAKU₋₁. Insertion sites for $Tn9$ and $Tn10$ are labelled as in Figure 5.2a.

5.4 Discussion

5.4.1 IncHI1 plasmids in Paratyphi A and Typhi

The IncHI1 plasmid pAKU₋₁ was the first MDR plasmid from Paratyphi A to be sequenced and analysed in detail. Plasmids of this type were responsible for MDR in the majority of clinical isolates analysed from Pakistan in 2004 (283) and plasmids of a similar size have also been associated with Paratyphi A in Bangladesh and China (284, 285). Like IncHI1 plasmids isolated from MDR Typhi, the plasmid DNA sequence was composed of an IncHI1 backbone with numerous insertions of mobile elements, encoding resistance to chloramphenicol, streptomycin, beta-lactams, trimethoprim, sulfonamides and tetracycline. The comparative analysis presented here showed that pAKU₁ shares an IncHI1 backbone with plasmids that have been sequenced from other organisms: R27 from Typhimurium, pHCM1 from Typhi CT18, three plasmids from H58 Typhi, pMAK1 from Choleraesuis and p0111 from enterohemorrhagic E. coli. This shared backbone must have been inherited vertically from a common ancestral plasmid and was therefore analysed separately from the mobile elements contained in the plasmid sequences, which encode drug resistance genes and can be readily transferred horizontally into distinct DNA backbones.

Phylogenetic analysis of the IncHI1 backbone sequences confirmed the presence of distinct lineages of IncHI1 MDR plasmids within distinct lineages of Typhi. Specifically, pHCM1 (plasmid type ST1) was found within Typhi CT18 (Typhi haplotype H1), while plasmids of the ST6 type were found within Typhi isolates of the H58 haplotype (Figure 5.6a). The presence of distinct plasmid types in Typhi was detected previously by MLST (725), but this study links plasmid type to Typhi strain type for the first time, providing direct evidence for independent acquisitions of distinct MDR IncHI1 plasmids by distinct Typhi lineages. The spread of MDR Typhi may therefore be attributed not just to the spread of a particular plasmid within the Typhi population, or the expansion of a particular Typhi clone following the acquisition of resistance, but to the spread of multiple MDR plasmids within the Typhi population. A corollary of this is that distinct plasmid types may compete for maintenance in the Typhi population, driving selection for plasmid maintenance unrelated to drug resistance, and thereby contributing to competition between Typhi lineages. For example, the replacement of ST1 plasmids (pSTY1-3,5) with ST6 plasmids (pSTY6-7) observed in Vietnam in the mid-1990s (275) may be related to the success of the H58 Typhi lineage (see 2.4.2.1, (2, 570)), as pSTY1 was present in H1 Typhi (CT18) and ST6 plasmids have been confirmed only in H58 Typhi isolates (see 5.3.2.1, Figure 5.6a). Unfortunately the haplotypes of the Typhi strains hosting pSTY6-7 cannot be determined as the strains themselves have not been maintained (the plasmid was transferred to E. coli for study in the laboratory), so this question can never be settled directly.

Phylogenetic analysis of the IncHI1 backbone sequences showed that the Paratyphi A plasmid pAKU 1 was distinct from plasmids analysed so far from Typhi and other organisms, although closely related to the plasmids found in H58 Typhi. So far there is no evidence of multiple independent acquisitions of MDR IncHI1 plasmids by Paratyphi A. Several Paratyphi A plasmids have been analysed by MLST and found to cluster into two closely related sequence types (ST7 and ST8), distinguished by a single deletion (725). In Chapter 3, just 16 SNPs were identified between pAKU 1 and sequencing reads from Paratyphi A pools. This small degree of variation could be due to mutations arising within the plasmid following a single acquisition event, or even sequencing errors in the case of SNP analysis. The question of multiple acquisitions may be resolved in the future by typing of individual plasmids and their host strains, which was not possible in this study since isolates were sequenced in pools (see 3.3.3.6).

5.4.2 Acquisiton of MDR by IncHI1 plasmids

Plasmids pAKU₁ and pHCM1 share very similar resistance gene complements, while R27 has only one resistance gene element $(Tn10, \text{ encoding tetracyclic}$ resistance). However comparative sequence analysis found that while pAKU₁ and pHCM1 shared near-identical mobile elements encoding drug resistance, they were inserted into different loci within the IncHI1 backbones, indicating that they were acquired via independent insertion events. The accumulation of resistance in IncHI1 since the 1960s when R27 was first isolated, is therefore the result of independent acquisition of resistance genes by distinct IncHI1 plasmid lineages, rather than accumulation of resistance genes in the R27 lineage as first supposed (46) . For example, R27, pAKU₋₁ and pHCM1 each encode a tetracycline resistance transposon $Tn10$, but the insertion site is at different positions in the IncHI1 backbone $(Tn10-A,B,C)$ in Figure 5.2a), suggesting that it has been independently acquired by each plasmid since their divergence. The insertion sites in pMAK1 and p0111 match the insertion site in pHCM1, consistent with a single acquisition of $Tn10$ in a common ancestor of these three plasmids $(Tn10-B)$ in Figure 5.2a). Similarly, the insertion sites in the Typhi H58 plasmids match that of pAKU 1, consistent with a single acquisition of $Tn10$ in a common ancestor of these plasmids $(Tn10-A)$ in Figure 5.2a).

In addition to $Tn10$, pHCM1 and pAKU₋₁ carry a highly similar set of mobile elements encoding resistance to chloramphenicol, streptomycin, beta-lactams, trimethoprim and sulfonamides. However this is not due to common ancestry of the plasmids, but to the independent acquisition of a single composite transposon by both plasmids (Figures 5.3, 5.4), which has since been subject to different rearrangements in each (Figure 5.4). The proposed composite transposon includes T_{n9} , T_{n21} and a stretch of sequence including the bla_{TEM-1} , sul2 and strAB resistance genes that may itself be mobile $(bla/sul/str,$ Figure 5.3d). The insertion site of the composite transposon is different in pAKU 1 and pHCM1, supporting the hypothesis that the plasmids acquired their similar resistance genes independently by horizontal transfer rather than vertical inheritance from a common ancestral plasmid. BLAST searching the proposed composite transposon sequence (Figure 5.4b) against the EMBL database revealed its presence, without rearrangements, in a plasmid from an unknown source, pRSB107 (Figure 5.4c). This plasmid has a distinct IncF backbone, thus the composite transposon appears capable of insertion into a variety of genetic contexts. The strongest evidence for the transfer of the composite transposon as a single unit is the 100% sequence identity in pAKU₋₁, pHCM1 and pRSB107 across the boundaries of insertion of (a) $Tn21$ into $Tn9$, and (b) $bla/sul/str$ into Tn21. If Tn9, Tn21 and $bla/sul/str$ were acquired independently in each plasmid, it is unlikely that the insertion sites of $Tn21$ and $bla/sul/str$ would be identical at the nucleotide level as they are in these three sequences. It is possible that multiple independent acquisitions of $bla/sul/str$ via homologous recombination could result in identical sequences, however this would require identical recombinations between IS26 elements to occur at least three times. Thus the most parsimonious explanation is that the insertions occurred once to form a composite transposon, which was then able to move between distinct plasmid backbones as a single unit using the IS1 ends of Tn9.

As with Tn10, the insertion sites of the composite transposon in pMAK1 and p0111 match pHCM1, while those in the Typhi H58 plasmids and pSTY7 match pAKU₋₁ (Table 5.3). This suggests that the plasmids sequenced so far belong to just three lineages, represented by pHCM1, pAKU₋₁ and R27 (see Figure 5.6). This is compatible with the three major 'groups' proposed by MLST analysis of a global collection of IncHI1 plasmids: group 1 (including pHCM1), group 2 (including pAKU₋₁) and group 3 (R27) (725). The sequence analysis presented here suggests that group 1 and group 2 lineages each acquired, independently, $Tn10$ and the composite transposon before diversifying into the subtypes that are evident today; group 1 diversity including ST1 (pHCM1, pMAK1, p0111), ST2, ST3, ST4 and group 2 diversity including ST6 (pSTY6-7, Typhi H58 plasmids), ST7 (pAKU 1) and ST8 (725). Since the earliest reports of tetracycline resistant pathogens $(Tn10)$ date back only 50 years (726) and multidrug resistant pathogens (Tn9, Tn21, $bla/sul/str$) less than 40 years (268, 684, 727), it follows that this diversification must represent recent evolution.

5.4.3 The spread of MDR via IncHI1 plasmids

The close relationships between IncHI1 backbones and resistance insertions presented above demonstrates that p0111, pMAK1, pHCM1, pAKU 1 and R27 share a recent common ancestry, indicative of spread between bacterial populations. The IncHI1 plasmid is self-transmissible, due to a conjugal transfer system which enables the plasmid to construct a pilus and transfer directly between bacterial cells (692). Each of the IncHI1 plasmids in this study (which includes all of those sequenced to date) was discovered in a different human enteric pathogen - enterohemorrhagic E. coli, and S. enterica serovars Choleraesuis, Typhi, Paratyphi A and Typhimurium, suggesting that the plasmids are able to spread between distinct pathogens occupying the human enteric niche. However it is unlikely that conjugal transfer occurs during coinfection of humans or any other mammalian host, since transfer is temperature-sensitive in IncHI1 plasmids, occurring at much higher efficiency at ambient temperatures (14-27◦C) than in vivo temperature $(37°)$ $(689, 692)$. This is consistent with selection for plasmid exchange outside the animal host, perhaps in water which is a common transmission route for enteric and other bacteria.

Whatever conditions are required for conjugal transfer to take place, it is clear that MDR can be transferred between enteric pathogens via IncHI1 plasmids. There is also evidence that MDR can be transferred via phage (728) and integrated into the chromosome of host bacteria (e.g. $Tn9-Tn21$ composite transposon in the Typhimurium DT193 chromosome (717) . This mobility suggests that selection for resistance in one pathogen (via antibiotic treatment) can impact the development of resistance in another. It also suggests that selection for resistance to one antibiotic may lead to the proliferation of resistance to many, as complex composite transposons encoding resistance to multiple drugs can move together as a single unit. This is an important consideration in a clinical environment as it highlights that treatment choices for infection with one pathogen can impact the development of resistance in other pathogens, leading to a narrowing of options for the treatment of other, perhaps more serious infections. The phenomenon of MDR has already driven a switch to fluoroquinolone-based drugs for the treatment of many bacterial diseases including enteric fever (3, 297, 685), see Figure 1.7. Resistance to fluoroquinolones is on the rise in Typhi (16, 401), Paratyphi A (279) and other pathogens (729), however the most common mechanism of resistance to fluoroquinolones is mutations in the topisomerase targets $gyrA$, $gyrB$, $parE$ and $parC$ (396, 397, 398), which are encoded on the chromosome and therefore unlikely to spread between bacteria via plasmids or bacteriophage. However genes encoding resistance to fluoroqinolones (qnr genes) have been discovered in MDR plasmids present in many pathogens (730, 731, 732) including Typhimurium (407, 408, 409) and other S. enterica serovars (410, 411, 412, 413, 414, 415), including plasmids of different incompatibility groups (409). Thus it may simply be a matter of time before plasmids encoding MDR and fluoroquinolone resistance appear in Typhi and Paratyphi A, leaving very limited options for the treatment of enteric fever.