The general characteristics and genetics of insertion sequence (IS) elements are well-established. For *Escherichia coli* IS elements, mechanisms of transposition and mutation are known and their recombinogenic role in the bacterial genome has been investigated. Population models relate the distribution of these IS elements to autoregulation of their transposition. IS200, the smallest known element, is confined to the salmonellae and several lineages of *E. coli*. It exhibits atypical molecular features. The population dynamics of IS200 make it a particularly effective marker of chromosomal genotype in many *Salmonella* serovars. Molecular epidemiological typing with IS200 has been developed for important serovars in groups D1, C1, C2 and B. Findings for *S. Enteritidis*, *S. Panama*, *S. Infantis*, *S. Typhimurium*, *S. Heidelberg*, *S. Paratyphi B* and *S. Java* are reviewed. Of the 12 IS elements found in mycobacteria, IS6110, found in *Mycobacterium tuberculosis* and *M. bovis*, exhibits the greatest potential for molecular epidemiological applications. Although *M. tuberculosis* is a single serogroup, and its genome is otherwise highly homogeneous, strains are highly polymorphic with respect to copy number and location of IS6110. A standard IS6110 typing method has been established, together with novel PCR-based approaches to IS6110 fingerprinting.

### Introduction

Bacterial insertion sequences (IS elements) are mobile DNA elements that are capable of transposition into diverse sites within the genome. They vary in genetic organisation and behaviour, are sized between c. 0.7 and 2.5 kb, and are found in copy numbers of one to several hundred per genome in many species of bacteria. They characteristically terminate in perfect or nearly perfect inverted repeat sequences that flank a unique central region. This central region contains an open reading frame (ORF) encoding a transposase that recognises the ends of the element, breaks DNA at those ends and joins them to target sequences. In some cases, genes for proteins or RNA transcripts that autoregulate transposition are also found [1], but little or no other genetic information is carried. Rarely, but often at well above the usual mutation rates, IS elements may transpose (jump) to a new locus by a replicative mechanism. They are ‘selfish’ (parasitic) DNA elements which are maintained in bacterial populations even under adverse natural selection because they can replicate and transpose independently of chromosome replication. Thus IS insertion patterns are subject to continuous variation during the history of a bacterial strain. Appropriate interpretation of restriction fragment length polymorphisms (RFLPs) around IS insertion sites (‘IS-profiling’, ‘IS typing’ or ‘IS fingerprinting’) may reveal much about the epidemiology and evolutionary genetics of bacterial pathogens.

This review focuses on the use of IS elements in studying the epidemiology of bacteria pathogenic for man. IS elements have now been characterised from several important human pathogens including *Burkholderia cepacia* [2,3], *Brucella* spp. [4], *Bordetella* spp. [5], *Enterococcus* spp. [6], Neisseria meningitidis [7], *Pseudomonas aeruginosa* [8], *Staphylococcus aureus* [9,10], *Streptococcus pyogenes* [11] and *S. pneumoniae* [12] and *Yersinia pestis* [13]. IS200 and IS6110 (Fig. 1) are then considered in detail. These elements have been used extensively for genotyping natural isolates of *Salmonella* and *Mycobacterium tuberculosis*, respectively. First, those aspects of IS element genetics that have a bearing on their use as tools for molecular epidemiology are considered.
Overview

Genetics of IS elements

IS elements were discovered as a result of the investigation of unstable polar mutations in *Escherichia coli* [14,15]. Hybridisation and heteroduplex analysis demonstrated that these mutations were caused by insertions of the same DNA sequence in different positions and in different orientations [16]. It became clear that transposition of these mobile sequences to new genetic loci led to insertion mutations, the majority of which were highly polar with respect to the expression of genes downstream. Like the maize transposable elements described by McClintock [17], bacterial IS elements are naturally resident in their host genome. As in the case of maize, their discovery undermined the concept of the bacterial genome as an inherently stable entity. Subsequent investigation has shown that the Mutator transposable element of maize encodes two proteins, one of which shares amino acid similarity to bacterial IS elements [18].

It is characteristic of IS elements that they generate small, directly repeated, duplications of host DNA at the point of insertion, probably as a result of staggered cleavage of the target by the transposase. The length of these duplications, 2–13 bp, is specific to an individual IS element. Excision of the IS element restores the function of an insertionally inactivated gene, or relieves the polar mutation of the downstream genes. Besides disruption of gene sequences, gene functions may also be affected by terminators within IS elements [19], and IS element insertion may activate cryptic genes or cause over-expression of contiguous genes from the element’s internal promoters or from newly formed junctions that act as ‘hybrid’ promoters [20,21]. IS elements are recombinogenic—it they can bring about various internal re-arrangements in bacterial genomes by initiating deletions, inversions and replicon fusions for intervening regions of the host genome. IS element activities generate a join between non-homologous DNA segments independently of the programmed genetic recombination (recA-dependent) system of the cell—a form of ‘illegitimate’ recombination.

The genome of a bacterial cell can be modified by point mutations or by re-arrangements, and illegitimate recombination contributes to the overall frequency of mutation in bacteria. The frequency of IS-mediated mutation is influenced by the location of target DNA (e.g., plasmid versus chromosome), by the site-specificity of a particular IS, by its copy number in the host cell, and by the growth conditions of the culture [22,23]. The first bacterial IS elements to be characterised were from *E. coli* and were termed IS1–IS5. Their copy numbers were deduced from genomic Southern blots. This method remains the basis for studying IS distribution in bacterial pathogens, but it should be noted that it does not yield information about the sequence identity or activity of the IS copies so detected.

Many IS elements are an integral part of composite transposons, in which two IS elements orientated in directly repeated or opposing directions flank a
sequence which is itself non-transposable. Most of these interstitial sequences are determinants of antibiotic resistance. Pre-antibiotic era enterobacterial plasmids did not contain composite transposons encoding antibiotic resistance [24]. The association of IS elements with these genes, resulting in the phenomena of mobile and transmissible antibiotic resistance, appears to have been selected by the widespread use of antibiotics in human and animal medicine over the last four decades. Certain well-characterised genes for pathogenicity determinants also exhibit genetic mobility mediated by IS elements. One example is the afa-3 gene cluster encoding the afimbrial adhesins expressed by uropathogenic strains of E. coli, which is flanked by two directly oriented IS1 elements that mediate its translocation [25].

Distribution of IS elements

Naturally occurring strains of E. coli are polymorphic for the presence or absence of individual IS elements. For example, while E. coli strain W carries no copies of IS1 and strain E. coli K12 carries up to 10 copies, some strains of Shigella dysenteriae carry over 200 copies [26]. In a reference collection of wild-type strains shown by multilocus enzyme electrophoresis (MLEE) to represent the evolutionary genetic diversity of E. coli [27], 15% of strains lacked IS1 in the chromosome, but carried it on plasmids [28]. This finding is generally applicable to the Enterobacteriaceae; typical IS elements occur more frequently, per unit length of DNA, on plasmids than in the chromosome. It means that plasmids play a key role in the dissemination of most IS elements between strains and species of gram-negative enteric bacteria. This must be borne in mind when interpreting data for bacterial strains carrying transmissible plasmids.

The overall abundance of IS elements in E. coli is determined by processes which can be analysed by population genetic models. The most applicable models for transposition of IS1–IS5 assume nearly clonal reproduction of E. coli [29] and include no factor for excision, as there is evidence that this occurs at a much lower rate than transposition [30]. These branching-process models define the transposition rate of an element (T) and the reduction of fitness of strains containing it (D) as functions of the number of copies of the element (n). Individual cells with n occupied loci are assumed to undergo single transposition events to the state (n + 1), with a transposition rate T(n) which is influenced by autoregulation. The distribution and abundance of IS1–IS5 in the ECOR collection indicates that they exhibit different degrees of autoregulation of transposition. IS1 and IS5 copy number distributions fit populations models with little or no autoregulation, while IS2 and IS4 distributions fit models with moderate autoregulation, and IS3 distribution fits models with strong autoregulation [28].

The potential of IS profiling for epidemiological studies was first recognised by Sawyer et al. [31], who analysed the number and distribution of six IS elements in defined MLEE clones of E. coli. Strains that did not differ in any enzyme mobilities differed in 57% of IS bands, while strains differing in a single enzyme mobility differed in up to 72% of IS bands. Thus the number of DNA restriction fragments containing IS elements (‘IS bands’) could be used to differentiate even between closely related strains.

Molecular genetics of IS200

IS200 was first characterised by Lam and Roth [32] in S. Typhimurium strain LT2. The mutation hisD984, isolated after chemical mutagenesis in the second structural gene of the histidine operon, was found to result from a DNA insertion of c. 700 bp. The inserted sequence, named IS200, was found to be homologous to sequences present elsewhere in the chromosome, with a copy number of six in the strain LT2 genome as deduced from Southern blots with several enzymes. Other strain LT2 derivatives were shown to contain one or more extra copies, and single transposition events were found to occur occasionally in stab cultures [32].

The IS200 insertion loci in strain LT2 were genetically mapped by screening strains carrying random insertions of the 10-kb tetracycline resistance transposon Tn10 by Southern blotting with a probe for IS200. Genomic DNA was digested with PvuII, an enzyme which does not cut within the sequence of either IS200 or Tn10. Tn10 insertions in PvuII fragments containing IS200 increased their size by 10 kb, and this was detectable by altered electrophoretic mobility. Correlation with genetic mapping then located six IS200 copies at evenly spaced sites around the strain LT2 chromosome [33]. The positions and orientations of the six IS200 insertions were located subsequently on the macrorestriction map of strain LT2 by pulsed-field gel electrophoresis, hybridisation and phage transduction [34]. Other than hisD984, none of these insertions resulted in detectable mutations. Only one mutation caused by insertion of IS200 has been identified; a mutant resistant to 8-azaguanine was shown to contain a novel IS200 insertion in the gpt gene which encodes guanine-xanthine phosphoribosyltransferase [35]. Thus, while spontaneous mutations in E. coli are frequently caused by IS insertions, this is only rarely the case in S. Typhimurium [36].

The nucleotide sequence of IS200 shows that it is the only known IS that contains neither inverted nor direct repeats. Phage Mu, which like IS200 does not revert spontaneously, is the only other transposable element without obvious inverted repeats. There was difficulty
in defining the ends of IS200 precisely, as it had inserted into a poly-A sequence, and a 2-bp target site duplication was estimated. The promoter is located some 100 bp from the left end of the element, and a terminator close to the right end provides a strong block to transcription across the insertion site [37]. The element contains a single large ORF that extends almost the whole IS length (nucleotides 100–558), encoding a putative transposase of 152 amino acids [38]. The G+C content of the element averages 45.5%, lower than that of the S. Typhimurium (53%) or E. coli (50%) genomes.

IS200 was originally thought to be limited to Salmonella, while the E. coli elements IS1–4 were not thought to occur in Salmonella. Up to 20 copies of IS200 occur in 15 Salmonella subspecies I serovars, including S. Typhimurium, S. Typhi, S. Dublin and S. Heidelberg. Two serotypes, S. Agona and S. Typhlisus appear to lack IS200, as does subspecies IIIa. The element was absent from laboratory strains of E. coli (K12, B, C), and no homology was found in other enteric genera except Shigella spp. [32], where partial homology was detected but shown subsequently to result only from cross-hybridisation with very short regions of the Shigella elements IS640 and IS600 [39, 40].

In contrast to the initial observations, Bisercic and Ochman [41] showed that certain MLEE clones of E. coli and S. Typhimurium contained polymorphic forms of IS200, but isogenic forms of IS1. IS1 sequences in S. Typhimurium and E. coli were identical, consistent with recent horizontal (plasmid-mediated) transfer between the two species. IS200 sequences in Salmonella were essentially homogeneous, but IS200 sequences from four MLEE clones of E. coli exhibited 9.5–10.7% sequence divergence from those in Salmonella. This resembles the average sequence divergence between pairs of homologous chromosomal genes of the two species, indicating that IS200, unlike IS1, must have been present in the common ancestor of E. coli and Salmonella, and that it has not been transferred between the species since their evolutionary divergence [38, 41]. These findings offer an explanation as to why IS200 is a particularly useful marker of chromosomal genotype in many Salmonella serovars. The element seems to have little if any affinity for extrachromosomal sites in a genus which is host to a large variety of transmissible and mobilisable plasmids. Thus it can act to fingerprint the vertically inherited chromosome independently of laterally transferable replications.

Context of IS200 profiling

Clinical and epidemiological studies aimed at determining routes and vehicles of infection depend on accurate classification of the organism and clear determination of its association with the disease or outbreak. Unique characteristics are required for specific strain identification, distinguishing it from the large background of non-pathogenic or non-outbreak strains for epidemiological determination of the source of an outbreak [42]. The utility of a particular characteristic for typing is related to its stability within a strain and its diversity within the species. The key features of an epidemiological typing scheme are typability (unambiguous positive results for each isolate tested), reproducibility, discriminatory power (ability to differentiate among unrelated strains) and ease of use [43].

The comprehensive population genetics framework for Salmonella provided by MLEE shows that subspecies I serovars are basically clonal, and that major serovars are associated with a relatively small number of MLEE clones [44]. However, MLEE is not easy to use and generally has inadequate discriminatory power for epidemiological typing. Therefore, in applying IS profiling to Salmonella serovars for molecular epidemiology, the most useful starting point is the demonstration that IS numbers and positions in the E. coli genome change much more rapidly in evolution than do electrophoretic mobilities of proteins [31]. Clonality is not an absolute concept, but can be defined with increasing certainty depending on the resolving power of the techniques used to characterise a culture. A clone has been defined [45] as any strain belonging to a set of microbial isolates that have been obtained ‘independently from different sources, in different locations, and perhaps at different times, but showing so many identical phenotypic and genotypic traits that the most likely explanation for this is a common origin’. A feature of IS200 profiling studies has been the identification of clonal lines within Salmonella serovars whose genotype can be delineated by possession of conserved insertion sites [46, 47].

The long established serotyping scheme for subspecies I salmonellae provides a uniquely well-differentiated context for genotyping studies, identifying about 2250 serovars by their possession of particular O (somatic) and H (flagellar) antigen profiles. Various phenotypic characteristics have been used to discriminate between strains within certain serovars. In particular, phage typing schemes have been developed for S. Typhi and several non-typhoidal serovars of particular public health importance such as S. Enteritidis [48] and S. Typhimurium [49]. Plasmids in Salmonella occur in a wide variety of sizes. Plasmid profiling, which examines the numbers and molecular sizes of intact extrachromosomal DNA elements, is a rapid and simple adjunct technique that increases the discriminatory power of phage typing [50]. However, not all serovars carry plasmids, and plasmids are laterally transferrable between strains. Therefore, identical plasmid profiles do not necessarily indicate genetic clonality or an epidemiological relationship. Evaluation of IS200 profiling must take into account the data
provided by these typing schemes, which continue to provide a foundation for *Salmonella* epidemiology.

**The group D1 serovars**

O-serogroup D1 contains many serovars of particular importance in medical microbiology, notably the agent of typhoid fever (*S. Typhi*) and the commonest agent of non-typhoidal salmonellosis (*S. Enteritidis*). The distribution of IS200 was investigated principally in genomic digests with *PstI*, an enzyme which does not cut within the element, and Southern blots probed with internal fragments of IS200. Type strains of 30 serovars and of the principal phage type reference strains of *S. Enteritidis* were examined for RFLPs at IS200 and 16S rRNA loci, and for plasmid profiles [5]. Simple conserved features (two IS200 bands in *PstI* or *BglII* digests) were found for serovars sharing the g phase-I flagellar antigen. A small number of bands is not useful for epidemiological purposes, but has significance in evolutionary studies [46]. There was an exact parallel between the IS200 and 16S rRNA gene profiles, dividing these serovars into two phylogenetic subgroups with *S. Enteritidis* in a pivotal position. Three distinct IS200 profiles occurred within *S. Enteritidis* and were characteristic of its most prevalent phage types, PT4, PT8 and PT11. Two 16S ribotypes were also found, in PT4 and PT8, and in PT11, respectively. *S. Blegdam* and *S. Moscow* share the IS200 and ribotypes of *S. Enteritidis* PT4, while *S. Dublin* and *S. Rostock* share those of *S. Enteritidis* PT11. The distribution of virulence plasmid size classes again follows these subgroupings: virulence plasmids of 36–38 MDa are associated only with the PT4 and PT8 groups, whereas virulence plasmids of 50–59 MDa are found only in the PT11 group. These data indicate that the *S. Enteritidis* PT4 and PT8 subgroups possess features of an ancestral type from which serovars *S. Antartica*, *S. Moscow* and *S. Blegdam* may have evolved. Similarly, the PT11 subgroup may have been the ancestor of serovars *S. Dublin*, *S. Rostock* and *S. Berta* [46]. Such an interpretation is consistent with findings from MLEE [52, 53] suggesting that *S. Enteritidis* occupies an ancestral position among this group of serovars. Further evidence for the ancestral position of *S. Enteritidis* among serovars sharing the g flagellar antigen comes from the demonstration that there is a single IS200 copy in *S. enteritidis*, into which IS1351, a novel insertion sequence belonging to the IS3 family, has itself inserted. This configuration of insertions is shared by *S. Dublin*, *S. Berta* and *S. Pullorum* (A. P. Burnens, personal communication).

*S. Enteritidis* is currently the commonest serovar isolated from cases of human salmonellosis in many countries [54]. Up to 20,000 human isolates are typed each year by the Public Health Laboratory Service – about four times as many as the next most common serovar, *S. Typhimurium*. Phage typing of *S. Enteritidis* distinguishes at least 27 phage types [48], but the current epidemic in the UK and several European countries is very largely caused by strains of a single phage type, PT4. PT4 is highly adapted to its poultry host, as is the most prevalent such phage type in North America, PT8 [55]. When the 27 phage type reference strains of *S. Enteritidis* were examined, all fell into one of the three IS200 profile subgroups described above. Phage types of major epidemiological significance occurred in each IS200-defined subgroup. The first and third most prevalent phage types in the UK (PT4 and PT6) were found in subgroup I, the second (PT8) in subgroup II, and the fourth (PT11) in subgroup III. The subgroups were termed evolutionary lines, Se Cl I (containing 13 PTs), Se Cl II (containing eight PTs) and Se Cl III (containing five PTs) [56]. PT8 isolates which predominate in the North American epidemic, and PT4 isolates which predominate in the epidemic in other European countries such as Switzerland, belong to the same evolutionary lines as UK strains [57, 58]. Further investigation of these evolutionary lines was made with a combination of genomic typing methods including *NorI* macrorestriction, pulsed-field gel electrophoresis (PFGE) and standard RFLP typing with random cloned chromosomal fragments. These methods confirmed the clonality of the Se Cl I and Se Cl II subgroups and showed that Se Cl III was too genetically diverse to be regarded as clonal [59]. A fourth evolutionary line with an additional IS200 band, termed Se Cl IV, was found for poultry strains from France [60].

Other serovars carrying the g flagellar antigen have also been analysed by IS200 profiling. Faecal and blood isolates of *S. Dublin* from England and Wales all shared the same profiles (Se Cl III), and strains could be distinguished only by 16S ribotyping [61]. Strains of *S. Berta* fell into one of two IS200 profiles which shared a common IS200 band in *PstI* digests and were distinguished by possession of one or two further IS200 bands [62]. Other IS200 profiles found among type strains of serogroup D1 were diverse, and many appeared unique. High IS200 copy numbers and unrelated RFLP characteristics were found for *S. Typhi*, *S. Lomalinda* and *S. Portland* – three serovars solely or predominantly host-adapted to man – which happen to exhibit characteristic metabolic deficiencies. Many Group D1 serovars carried no plasmids, but had more than five IS200 copies in the chromosome, at which point fingerprinting with the element becomes useful for epidemiological purposes [51]. For such serovars (*S. Typhi*, *S. Panama*, *S. Saarbrucken*, *S. Pensacola*, *S. Durban*, *S. Onarimon*) the technique appeared particularly appropriate. An assessment of the applicability of IS200 profiling to the epidemiology of *S. Typhi*, showed that all strains yielded 11 or more IS200 bands, while 14 profiles occurred among the Vi phage type reference strains [63]. Consistent with *S. Typhi* being genetically homogeneous, three
IS200 bands were unique for the serovar and conserved among the majority of phage type reference strains and recent clinical isolates (A. P. Burnens, personal communication). Although the discriminatory power of IS200 profiling with PstI alone was less than that of Vi phage typing for some lysosvars, it was still possible to subdivide other phage types [63].

IS200 profiling of *S. Panama* [64] provided several interesting points of contrast with *S. Enteritidis*. *S. Panama*, one of the most commonly isolated serovars from human gastro-enteritis, is also a significant cause of invasive disease. A phage-typing scheme distinguishes eight types, but leaves a substantial group of non-typable isolates—a common problem with phenotypic methods. Preliminary studies indicated that the reference strain contained at least six IS200 bands, in contrast to their relative paucity in *S. Enteritidis*. IS200 profiles were therefore generated from a collection of *S. Panama* strains with enzymes that do not cut within the element (*PvuII* or *PstI*) or that cut only once within the element (*EcoRI* or *EcoRV*). Up to nine bands or up to 18 bands, respectively, were found. When comparison was made with phage typing, 16S ribotyping and plasmid profiling, it was clear that IS200 profiling had superior typability and discriminatory power. The best enzymes in combination (*PvuII* and *EcoRI*; Fig. 2) yielded a discriminatory index power (D) of 0.78. When plasmid profiling was used as an adjunct technique, the D value (0.97) approached the theoretical maximum (1.0).

**Salmonella serovars of groups C1 and C2**

IS200 profiling and 16S rRNA gene typing were used to systematise molecular subtyping of *S. Infantis* (serogroup C1). Like *S. Panama*, this is one of the commonest non-typhoidal serovars, but lacked an effective subtyping scheme. All strains contained up to seven copies of IS200, with four conserved and serovar-specific IS200 bands. Among the 15 genotypes distinguished by a combination of IS200 and ribotype, there were three major IS200 types—profile A consisted of strains almost solely of animal origin, profile B of strains of both animal and human origin, and profile F of strains of human origin [65].

---

**Fig. 2.** IS200 profile types of *S. Panama*. Genomic DNA from 30 strains of the serovar was digested with *PvuII* (A) or *EcoRI* (B) and hybridised with a 692-bp PCR-generated internal probe for IS200 [65]. The 12 IS200 types found with each enzyme [64] are shown.
At least two examples of incomplete distribution of IS200 within a serovar have been characterized. In S. Bovismorbificans (serogroup C2) – a veterinary pathogen which also causes sporadic human infections – three clonal lines were characterised, but not all strains carried IS200. The latter strains had unique ribotypes and protein electrophoretotypes, and their IS200 bands were not conserved – they were minor clones within the serovar [66]. In S. Virchow (serogroup C1), the third most common serovar isolated from man in the UK, IS200 was found in only 10% of the 57 phage type strains and occurred only in low copy number (1–4) with no conserved bands [67]. Thus for these two serovars, IS200 profiling alone is of limited use for epidemiological typing. However, there is an interesting evolutionary genetic question as to whether those strains that have the element acquired it before or after acquisition of the serovar determinants.

**S. Typhimurium and other group B serovars**

Strains of *S. Typhimurium* descended from the original Lilleeengen phage types (LTs) were used by Zinder and Lederberg [68] and Demerec et al. [69] in classical studies of genetic exchange. Most mutants used for genetic analysis of the *S. Typhimurium* chromosome are derivatives of strain LT2 in which IS200 was discovered. A phage typing scheme [49] that defines over 200 definitive phage types (DTs) for medical and public health microbiology purposes was developed independently of genetic analysis of LT strains. Thus the relationship between natural (pathogenic) isolates and laboratory (genetic) strains remained obscure, as did relationships between the DTs themselves. Substantial epidemiological surveillance information exists concerning the prevalence of different DTs in human and animal salmonellosis. In 1990–91 the most common strain of *S. Typhimurium* isolated from man in the UK was DT193, while in 1993–94 it was DT104. Other phage types of epidemiological significance in this period included DTs 208, 170, 12, 108, 49 and 24 [70].

IS200 profiling was used to compare genetic and reference strains of *S. Typhimurium* with phage typed clinical isolates [47]. All strains yielded at least six IS200 bands, three of which (PstI fragments of 1.9, 2.4 and 4.8 kb) were conserved in all strains. The latter were found to be serovar-specific; they did not occur in any serovar examined except *S. Typhimurium*. Additional evidence for the conserved nature of the 2.4-kb band and its exclusive occurrence among strains of *S. Typhimurium* was obtained by cloning and sequencing of the IS200 insertion locus contained therein. This particular copy of IS200 was shown to be inserted in the *flIB-flIA* intergenic region (A. P. Burnens, personal communication). The six IS200 bands characteristic of strain LT2 occurred in a number of genetic and reference strains of consider-

ably different origin [47]. For example, strain NCTC 3048, isolated in 1923, had the same IS200 profile as mutagenicity-testing strains [71] derived from strain LT2 60 years later, indicating that IS200 profiles can be remarkably stable. The highest numbers of IS200 bands were obtained with DT (type) strains, which are natural isolates. Thus, while studies of genetic strains indicated that IS200 transposition is rare in the laboratory, natural isolates growing under conditions of natural selection and zoonotic transmission have somewhat higher rates of IS200 transposition. Nevertheless, most additional IS200 bands were conserved in strains of a given phage type. All but two phage type strains tested had specific IS200 profiles. Thus, it appears that while some DTs correspond to genotypic clones, a few are phage receptor-site variants of the same clone.

It is significant that IS200, unlike the *E. coli* IS elements, was not found on plasmids in any *S. Typhimurium* strains. IS200 profiles of genetic strains and natural isolates were clearly related, and insertion profile (IP) groups were identified by possession of shared IS200 (PstI) bands in the chromosome. Over and above the three serovar-specific bands, the predominant group, termed IP1.0, shared additional conserved IS200 bands of 3.2 and 7.1 kb and included strain LT2 itself. A distinct profile shared by DT49, DT204 and some culture collection strains contained conserved IS200 bands of 9.4, 22 and 24 kb, and was termed IP2.0. These were the two major lines of descent within the serovar. Among clinical isolates, further clusters of closely related IS200 profiles associated with IP1.0 and IP2.0 were found. For example, phage types DT170 and DT108 (10 IS200 bands) and DT208 (11 IS200 bands) shared specific PstI bands of 3.5 and 5 kb, characteristic of a subgroup termed IP1.1. Taken as a whole, these data show that a low rate of IS200 transposition had acted as a marker of vertical descent and evolutionary radiation in *S. Typhimurium*. The evolutionary relationship of strains could be established by the proportion of conserved IS200 bands, *viz* IS200 loci and copy number resulting from a continuous time-branching process. Fig. 3 shows the proposed phylogenetic history of *S. Typhimurium* strains.

In subsequent studies it was confirmed by IS200 profiling that all strains of certain phage types were clonally related. This was the case for DT204C, a phage type belonging to IP2.0, which was implicated in the early spread of multiple antibiotic resistance to the human food chain [72]. In contrast, other phage types appear composite; for example, DT193 contains highly diverse IS200 profiles, consistent with a number of different clones sharing its configuration of phage receptor sites [73]. From the epidemiological point of view, these studies extended the scope of IS200 profiling to subtyping within phage types of *S. Typhimurium*. 
Fig. 3. Evolutionary genetic relationships among S. Typhimurium strains, modelled on the basis of copy number (scale at top) and conservation of IS200 bands [47]. Groups of IS200 profiles (IP groups) had a range of IS200 bands (boxed) or a discrete number of bands (vertical line).

Several studies have demonstrated the value of these stable IS200 profiles as genetic markers. IS200 profile and other molecular methods were used to demonstrate that multiple episodes of salmonellosis in an AIDS patient resulted from the persistence of a single S. Typhimurium strain rather than re-infection with new strains [74]. Similarly, IS200 profile, ribotype and plasmid profile were used to differentiate a vaccine strain of S. Typhimurium from the background of wild-type isolates of the serovar in cattle [75]. Stable IS200 profiles were characteristic of 15 clonal lines of S. Typhimurium identified from different poultry flocks in France [60].

The Group B serovar S. Heidelberg is found consistently among the ‘top 10’ agents of human bacterial gastro-enteritis, but there was no previous subtyping methodology other than plasmid profiling. All strains of S. Heidelberg were found to contain a common serovar-specific IS200 band, part of a profile of four of more bands. IS200 profiles within this serovar were related by a time-branching process model, defining seven clonal lines, two of which were associated with distinctive epidemiological contexts. In no case was IS200 found in plasmids [76]. The IS200-defined genetic lineages of S. Heidelberg and S. Typhimurium are distinct, in contrast to MLEE data [77], indicating that IS200 analysis may be better suited than MLEE to elucidate genetic relationships between Salmonella serovars.

Two Salmonella biovars in serogroup B that share the same O and flagellar antigens, yet cause quite different diseases, exemplify the discriminatory power of IS200 profiling. The first, S. Paratyphi B, cannot ferment D-tartrate (dTa-), produces a slime wall and causes paratyphoid fever in man. The second, S. Java, ferments D-tartrate (dTa+), forms no slime wall and causes gastro-enteritis in animals and man. The IS200 profiles of all strains of S. Paratyphi B and S. Java contained six or more IS200 bands and the core patterns of their profiles were distinct. Although MLEE [78] had not succeeded in defining distinct genotypes for S. Paratyphi B or S. Java, unique IS200
bands (PstI, 3.7 kb; PvuII, 8.4 and 8.8 kb) were specific for strains of S. Paratyphi B and absent from strains of S. Java. Thus, IS200 profiling is the first genetic method to discriminate between S. Paratyphi B and S. Java, with one copy of the element specific for the former and carried by all isolates. Compared to the human-adapted S. Paratyphi B, greater genetic diversity was found among profiles of S. Java strains, as might be expected for an organism with many animal reservoirs. One of the IS200 profiles found for S. Paratyphi B occurred among strains of widely different geographical origin, i.e., it represented a globally-distributed clone of this biovar [79].

S. Brandenburg is a generally uncommon Group B serovar which was the cause of a prolonged nationwide outbreak of human salmonellosis in Switzerland throughout 1992. All strains examined contained three serovar-specific IS200 bands (PstI) in the chromosome, but lacked plasmid DNA. When the outbreak was analysed by a combination of molecular methods, pulsed-field gel electrophoresis was found to be more discriminatory than IS200 profiling because of the low copy number of the element in this serovar. Nonetheless, both methods concurred in showing that the 1992 isolates belonged to a single genotypic clone, characterised by a fourth IS200 site and found in multiple food vehicles of infection. This clone was distinct from pre- or post-1992 sporadic isolates of the serovar in Switzerland, and from strains isolated in other European countries [80].

**Insertion sequences of mycobacteria**

Infections caused by pathogenic mycobacteria were, until recently, declining in number in developed countries. This trend has now been reversed and rising numbers of cases of tuberculosis have been reported. In addition, infections caused by other mycobacteria, such as *M. avium* are increasing. Factors that may contribute to the resurgence of tuberculosis include immigration from areas where tuberculosis is endemic and poverty among sections of the community. HIV infection is another predisposing factor and is the most important factor contributing to increases in infection with mycobacteria other than *M. tuberculosis*. These threats are compounded by the emergence of drug-resistant strains.

At least 12 IS elements have been found in mycobacteria [81–83]. Some – e.g., IS900 from *M. paratuberculosis* – have specific sites of insertion while others – e.g., IS6110 – appear to insert randomly. They vary in size from 880 bp (IS6100 from *M. fortuitum*) to 2260 bp (IS1096 from *M. smegmatis*). They are related phylogenetically to elements from both gram-positive (IS1245, IS1081) and gram-negative (IS6110) bacteria indicating that mycobacteria have acquired them as exogenous genetic material.

Mycobacterial insertion sequences can be used as species- or strain-specific markers. IS900 is specific for *M. paratuberculosis* within the *M. avium* complex [84], while IS6110 has been used extensively for strain typing in *M. tuberculosis*. Three of five elements that can transpose in *M. smegmatis* (IS6110, IS6120 and IS1096) have features that may make them useful for insertional mutagenesis [81].

IS1081 was cloned from *M. bovis* and characterised by Collins and Stephens [85]. This 1324-bp element has terminal 15-bp inverted repeats and is found in five-to-seven copies in all strains belonging to the *M. tuberculosis* complex. The putative transposase of IS1081 has high similarity to the products of the equivalent ORF of IS1245 (see below) and of IS256 from *Staph. aureus*. The copy number and chromosomal locations of IS1081 in *M. tuberculosis* strains are apparently well conserved. Of 99 *M. tuberculosis* and *M. bovis* strains, 61 shared all six sites as judged by *Pvu*II RFLP analysis with a 300-bp IS1081 probe. This IS element may be a useful subsidiary marker in detailed epidemiological studies of *M. tuberculosis* [86], but is unlikely to be of value on its own because of its low discriminatory power. However, it may serve to differentiate wild-type *M. bovis* isolates from BCG strain derivatives [87].

The *M. avium* elements IS1245 and IS1311 were identified by Guerrero and colleagues [82, 83]. Sequence analysis showed that they were distinct but related 1.3-kb elements that shared 85% base sequence homology. In addition to being found in typical *M. avium* strains, IS1245 was shown to be present in the *M. avium* subspecies *silvaticum* and *paratuberculosis*. It was absent from *M. intracellulare* and other more distantly related mycobacterial strains. The host range of IS1311 has not been reported. The *M. avium* strains analysed carried up to 27 copies of the elements [82, 83]. The putative transposase of IS1245 has high similarity (64% shared amino acids) with IS1081 and clusters in the *Staph. aureus* IS256 (55% amino acid homology) family. The inverted repeats are dissimilar to those found in IS1081, being comprised of 38 bp at the termini. Further inverted repeats of 15 bp are present internally on the element close to the 38-bp sequences. At one end the inverted repeats are adjacent, but an intervening sequence of 24 bp separates them at the 5′ terminus. Southern blotting of different *M. avium* strains shows great variability in IS1245 and IS1311 profiles, and they are clearly promising elements for use in epidemiological typing systems. Further investigations are required to establish the rates of transposition of these elements. If transposition is infrequent the discriminatory power of the system will be correspondingly low, but if it is very frequent it may not be possible to link even
closely related strains. A moderate rate of transposition is required if these IS elements are to provide useful epidemiological data. These two elements may also be used to identify *M. avium*.

**Molecular genetics of IS6110**

The insertion sequence IS6110 has been of particular interest because of its potential diagnostic and epidemiological applications. Three copies of the element (originally designated IS6110, IS986 and IS987) have been isolated from different *M. tuberculosis* or *M. bovis* strains. These have been sequenced and found to differ in only a few nucleotides [88–91]. IS6110 is present in the other species of the MTB-complex although not in the more distantly related mycobacteria [92].

IS6110 is a 1360-bp element with imperfect 28-bp inverted repeats at either end. Flanking the element are 3-bp direct repeats, probably resulting from target site duplication. The largest ORF encodes a putative transposase with pronounced (51%) amino acid similarity to the corresponding ORF of IS3411, one of the enterobacterial IS3 family elements [90]. In common with other IS3-like elements, IS6110 contains at least two ORFs. Analysis of the sites of insertion of IS6110 in a strain with multiple copies of the element [93] showed that all the sites were distinct, suggesting that random transposition events had occurred. However, studies of *M. tuberculosis* and *M. bovis* strains, which have only one copy of IS6110, show that it is inserted conservatively into one of an array of 36-bp directly repeated (DR) sequences [91]. Therefore, the extent of random versus site-specific transposition remains unclear. The mechanism of transposition of IS6110 was shown to be orthodox in *M. smegmatis* by use of artificial composite transposons (a kanamycin resistance gene cassette flanked by copies of IS986) [94].

**IS6110 typing of *M. tuberculosis***

Typing of tuberculosis strains is necessary for epidemiological control since it allows cases to be linked, outbreaks to be traced with precision, and changing patterns of disease to be monitored. Typing by classical methods has proved difficult because of the phenotypic and genotypic homogeneity of the species. Serotyping shows that *M. tuberculosis* strains belong to a single group [95], and both DNA-DNA hybridisation [96] and restriction digest analysis of total DNA preparations [97] demonstrate very little genetic variation. By the latter technique, all strains of the species share a large majority of restriction fragments, although some polymorphism is observed among large DNA fragments [98]. This genomic homogeneity hampered early efforts to develop genotyping of *M. tuberculosis*.

Different *M. tuberculosis* strains show great variability in the chromosomal copy number and location of IS6110 against a stable genetic background. Consequently, restriction fragments carrying the element are highly polymorphic. A standardised methodology exploiting this polymorphism for discriminatory DNA fingerprinting of *M. tuberculosis* strains was published by van Embden and colleagues [99]. This method involves genomic digestion with *Pvu*II, which cleaves the insertion sequence at a single site, followed by hybridisation with a 245-bp probe that hybridises only to restriction fragments carrying the 3’ end of the IS6110 sequence. Each band on the Southern blot corresponds to a chromosomal copy of IS6110. Fig. 4 shows typical typing results for 11 strains. The reproducibility of the method is excellent and the IS element copy number and integration site are stable when individual strains are passaged in guinea-pigs [100], subjected to serial subculture on laboratory media [101] or isolated from individual patients at intervals of up to 4.5 years [102]. However, evidence for replicative transposition or deletion of IS6110 within a single strain is also available [101–103] and in our laboratory we have noted a relatively high frequency of such events. Of 22 groups of related strains studied, 12 included profiles that differed in the presence or absence of at least one band, providing evidence that replicative transposition of IS6110 may occur relatively frequently among UK isolates and indicating that strains from a single source may exhibit different IS6110 band profiles. The apparent discrepancy between passage studies, which did not show transposition, and these findings may result from differences in the numbers of cells transferred. In the laboratory studies many cells were inoculated at each stage, allowing the majority IS6110-type to overgrow any mutants, but when the organism is transmitted from person to person, small numbers of bacteria may survive to establish the new infection. The overall discriminatory power of the standard method [99] is very high, but some evidence suggests that, among some human population groups, strains carrying one copy of the sequence are relatively common. In these strains the IS element appears to be located within the DR sequence [104]. While typability of strains is the rule, two reports describe strains in which IS6110 is absent [104, 105].

An important aspect of IS6110 typing is the evaluation of band profiles produced when it is not possible to compare all strains of interest in a single experiment. This represents a significant problem because the great inter-strain variability of IS6110 profiles is compounded by small differences in electrophoretic and blotting conditions. The best solution is to use a computerised database of profile types, corrected against a suitable standard. Computer programs such as GelCompar (Applied Maths, Belgium) have been used widely for storage and analysis of IS6110 typing profiles.
Fig. 4. Example of standard IS6110 typing of M. tuberculosis strains. DNA from 11 strains (lanes 1–11) was digested with PvuII, electrophoresed in agarose and vacuum blotted on to nylon membrane. Restriction fragments containing the 3' (downstream) end of IS6110 were detected colorimetrically by hybridisation with a digoxigenin-labelled probe. Lanes labelled λ contain PstI and EcoRI fragments of phage lambda.

As discussed above, it may be possible to determine clonality and to reconstruct the evolutionary histories of bacterial populations by careful analysis of IS element profiles. This has been difficult for M. tuberculosis because of the relatively frequent transposition of IS6110 and because no comparative phenotypic typing or genotyping data are available. However, there is evidence that strains with a common geographic origin may have similar IS6110 typing patterns [101, 104].

**Epidemiological studies based on IS6110 typing**

IS6110 typing has been used to study tuberculosis in defined human populations. These include surveys of strains isolated in San Francisco [106], Greenland [86], Denmark [86], different regions of Tunisia [103], strains from a New York hospital [107] and strains from HIV-seropositive and seronegative patients in Dar es Salaam [108]. In these studies, strains that gave IS6110 banding patterns identical to those of one or more of the other strains analysed were investigated further to determine whether there was a common source of infection. For a high proportion of strains that were in such clusters, an epidemiological link could be identified to other strains in the same cluster. An exception to this was that strains with only one or two copies of IS6110 could generally form clusters with no known epidemiological relationship.

In the UK, a study of strains isolated from patients at three London hospitals [109] identified eight clusters of strains with multi-copy patterns that were either identical or differed by a single band position. Careful review of the clustered cases was able to establish an epidemiological link in each. These studies do not exclude the possibility that two strains from unrelated cases will by chance have the same banding pattern. However, it seems that the rate of transposition (and possibly deletion) of IS6110, while not enough to obscure relationships between recently disseminated strains, ensures that strains diverge relatively rapidly into different types. Consequently, very many different types are circulating, or are latent, within infected individuals in each part of the human population at any one time.

As IS6110 typing exhibits excellent discriminatory power, the number of cases within clusters has been used to estimate the rate of new infection versus reactivation of tuberculosis within a community. Although IS6110 typing of strains with one or two copies of the element is not sufficient to establish a link between such strains in the absence of other evidence, it can provide good supplementary evidence when strains are suspected of being from a common source. For example, in the study of isolates from three London hospitals [109], the largest single cluster comprised nine strains that carried only one copy of
IS6110 (generating a single band of 1.4 kb on Southern blotting). It may be calculated that the chance of two unrelated strains being of this type was <1%. Therefore, IS6110 typing can be of value for outbreak investigation even when the strains involved carry few copies of the element.

**PCR methods for IS6110 typing**

Although the standard RFLP typing method [99], which employs Southern blotting, is highly discriminatory and reproducible, it has the drawback of requiring several micrograms of chromosomal DNA. Obtaining an adequate cell mass may take several weeks because of the slow growth rate of *M. tuberculosis*. To avoid the necessity for prolonged culture, several PCR-based approaches to IS6110 typing have been developed. The first approach amplifies DNA separating adjacent copies of IS6110 or DNA between IS6110 and another repetitive DNA [110–112]. The main drawback of this approach is that production of the PCR amplicons depends on priming sites provided by the repeated elements being close enough for efficient amplification. In practice few amplicons are produced, limiting the discrimination achieved. The second approach uses ligation-mediated PCR, in which a specific synthetic oligonucleotide priming site is ligated to a restriction site flanking IS6110 to allow amplification between the IS element and the restriction site [113, 114]. This relatively complex method requires careful optimisation. Although its discriminatory power is satisfactory, it seems that additional steps are required to improve reproducibility. These extra manipulations further diminish ease of use. The third approach is inverse PCR [115], based on sequences in the IS6110 element [116]. This technique results in amplification of part of IS6110 together with its upstream flanking sequence, as outlined in Fig. 5. The size and number of PCR amplicons produced depends on the number of copies and sites of integration of IS6110. The method is technically simple and has a discriminatory power that bears comparison with standard RFLP analysis. Typical IS6110 typing results for a group of *M. tuberculosis* isolates are shown in Fig. 6.

**Concluding remarks**

Changes in IS element copy numbers resulting from transposition change the number of fragments hybridising to an IS element probe. On the other hand, mutations in restriction enzyme recognition sites (conventional RFLPs) alter the length (position), but not the number of hybridising fragments. As a result, there is a unique form of variation in IS element numbers and loci, presenting a powerful tool for DNA fingerprinting of bacteria. Genotyping of salmonellae and of *M. tuberculosis* with IS200 and IS6110 exemplifies the application of this variation to epidemiological typing of bacterial pathogens. There are interesting points of comparison between these insertion sequences. IS200 is unusual among enterobacterial IS elements because its rare or negligible occurrence in

---

**Fig. 5.** Diagrammatic representation of the inverse PCR method for IS6110 typing. Genomic DNA is digested with the enzyme *Bsr*FI and ligated to form circular DNA molecules. PCR with hemi-nested primers (position and orientation indicated by arrows) results in the formation of a linear product. Sequence derived from IS6110 is shown in grey while the flanking DNA sequence amplified is shown in black.
Fig. 6. Example of typing of \textit{M. tuberculosis} strains by inverse PCR. DNA from 15 strains (lanes 1–15) was digested with \textit{Bsr} FI, ligated and subjected to hemi-nested PCR. First round amplification primers were 5'\textit{TGAACCACTGACCATGACCC}3' and 5'\textit{GTCTACTTGGTGTTGGCTGC}3'. In the second round PCR, 1 \textmu l of this product was amplified with primers 5'\textit{GACATGCCGGGGCGGTTC}3' and 5'\textit{GTCTACTTGGTGTTGGCTGC}3'. The final product was electrophoresed on Metaphor agarose (Flowgen) 2% w/v and stained with ethidium bromide. Lanes labelled \textPhi show \textit{Hae} III fragments of phage \textPhi X174.

plasmids, combined with a low transposition rate, has resulted in characteristic serovar-specific loci among the salmonellae. Within many serovars, the level of insertion site variation permits stable profiles and types to be found. These can be used to plot evolutionary genetic relationships between strains. In molecular epidemiological terms, greatest discriminatory power is obtained for serovars where the copy number is five or more. Finally, IS200 profiling yields DNA fingerprints that can be stored and compared in a band-matching database, and that integrate well with other typing methods to give coherent descriptions of strain genotypes. IS6110 generates highly discriminatory DNA fingerprints of \textit{M. tuberculosis} strains. Its rate of transposition appears to be dependent on subtle genetic differences – in some strains transposition is so infrequent that a single site of insertion is conserved, while in many others there are multiple copies of IS6110 and transposition is frequent. PCR can also be applied to rapidly reveal the genetic heterogeneity resulting from IS element transposition. Methods such as inverse-PCR fingerprinting of IS6110 loci point a way forward to strain genotyping methods that are both rapid and genetically coherent.

There is considerable scope for developing IS element genotyping methods for various bacterial pathogens, particularly in circumstances where traditional phenotypic methods cannot discriminate adequately between strains. At the time of writing there are several pathogenic genera that present increasing threats to public health – e.g., the enterococci and streptococci – for which IS element typing could offer significant insights into evolutionary genetics and improvements to epidemiological studies.

References

8. Kropinski AM, Farinha MA, Jansons I. Nucleotide sequence of the \textit{Pseudomonas aeruginosa} insertion sequence IS222: another member of the IS3 family. \textit{Plasmid} 1994; 31: 222–228.


