JOURNAL OF BACTERIOLOGY, Dec. 1993, p. 7863–7868
0021-9193/93/247863-06$02.00/0

Vol. 175, No. 24

The Ancestry of Insertion Sequences Common to *Escherichia coli* and *Salmonella typhimurium*

MARINA BISERČIĆ AND HOWARD OCHMAN*

Department of Biology, University of Rochester, Rochester, New York 14627

Received 29 April 1993/Accepted 19 October 1993

Despite very restricted gene exchange between *Escherichia coli* and *Salmonella typhimurium*, both species harbor several of the same classes of insertion sequences. To determine whether the present-day distribution of these transposable elements is due to common ancestry or to horizontal transfer, we determined the sequences of IS*1* and IS*200* from natural isolates of *S. typhimurium* and *E. coli*. One strain of *S. typhimurium* harbored an IS*1* element identical to that originally recovered from *E. coli*, suggesting that the element was recently transferred between these two species. The level of sequence divergence between copies of IS*200* from *E. coli* and *S. typhimurium* ranged from 9.5 to 10.7%, indicating that IS*200*, unlike IS*1*, has not been repeatedly transferred between these enteric species since *E. coli* and *S. typhimurium* diverged from a common ancestor. Levels of variability in IS*1* and IS*200* for strains of *E. coli* and *S. typhimurium* show that each class of insertion sequence has a characteristic pattern of transposition within and among host genomes.

The genomes of most bacteria contain many distinct classes of insertion sequences (IS; e.g., IS*1*, IS*2*, and IS*3*). These elements are small translocating segments of DNA that occur on the host chromosome and on plasmids and have the potential to move both within and among bacterial genomes. The ability of IS to mediate gene transfer and affect gene activity has led to speculation about their role in bacterial evolution (5), but until recently, the rate and extent of transposition of IS elements among natural isolates have remained unknown. On the basis of mutations generated upon insertion of IS elements, the frequency of transposition has been estimated to range from 10^{-3} to 10^{-7} per element per generation in *Escherichia coli* grown under laboratory conditions. In these experiments, transposition rates varied with the class of IS element, host genotype, target sequence, and physiological conditions (1, 4, 9, 27).

Most information on IS distribution and abundance stems from analysis of enteric bacteria, particularly *E. coli* (16, 21, 24, 29). Natural strains of *E. coli* recovered from diverse sources are highly variable in both the numbers and genomic locations of several classes of IS (22, 29), suggesting that elements transpose regularly within a host. By using pairs of identical isolates stored separately in stab cultures for over 50 years, Green et al. (10) were able to estimate an apparent rate of transposition of $\sim 10^{-2}$ per element per year for IS*5* within natural isolates of *E. coli*.

Recent studies have also examined the extent of horizontal transfer and mobility of IS elements in a phylogenetic context (18, 24). By analyzing IS distribution in strains with known genetic relationships, it has often been possible to reconstruct the evolutionary history, as well as the incidence of transfer of IS elements among extant lineages. For example, the sporadic distribution of IS*1* among enteric bacteria is usually cited as evidence of horizontal transfer after the divergence of these species, whereas the variation in IS copy numbers reflects proliferation within the host genome (15, 18, 21). Because of the potential for IS transfer within and among genomes, two problems arise when inferring the source and ancestry of particular elements. (i) The presence of the same IS class in related lineages could be a consequence of either common ancestry or independent acquisition. (ii) Both transposition within genomes and transfer between genomes result in an increase in IS copy numbers. In general, these issues can be resolved by assessing the genetic relationships of the IS elements themselves and comparing the extent of divergence among elements to that of chromosomally encoded genes from the host genomes.

In contrast to previous investigations on IS distribution among enteric bacteria, we recently established (3) that natural populations of *E. coli* and *Salmonella typhimurium* harbor at least three of the same classes of IS (IS*1*, IS*3*, and IS*200*). Preliminary evidence, based on comparative restriction mapping and selective amplification, indicated that *E. coli* and *S. typhimurium* harbor very distinct forms of IS*200* but very similar, if not identical, copies of IS*1*. Moreover, variation in the copy numbers and the extent of genetic diversity observed in these elements suggested different frequencies of transposition for each element in *E. coli* and *S. typhimurium*. To elucidate the ancestries of these elements, we obtained the nucleotide sequences of IS*1* and IS*200* from diverse isolates of *E. coli* and *S. typhimurium*. The IS of these closely related species of enteric bacteria exhibit characteristic patterns of evolution, and the differences in the mobility and dynamics of IS*1* and IS*200* can be attributed to features of each element rather than the host species.

MATERIALS AND METHODS

Bacterial strains. Strains of *E. coli* and *S. typhimurium* (subgroup I) were selected from the ECOR (23) and SARA (2) reference collections, respectively. Both collections contain isolates whose evolutionary relationships have been inferred on the basis of protein electrophoresis. The ECOR collection consists of 72 natural isolates and is representative of the range of genotypic diversity found within *E. coli*. We analyzed ECOR strains 8, 51, 63, and 72, which are distantly related and known to harbor single copies of IS*200* (3). The SARA collection comprises strains characteristic of the diversity found within subgroup I of the *S. typhimurium* complex and includes strains of the serovars *S. typhimurium*, *S. saintpaul*, *S. heidelberg*, *S. paratyphi* B, and *S. muenchen*. We examined 66 (SARA 1 to 4,

* Corresponding author.

7 to 13, 15 to 30, 33 to 43, and 45 to 72) of the 72 strains within the SARA collection. Species identification of strains used for sequence analysis of IS*1* or IS*200* was confirmed by ES MicroPlates (Biolog, Inc).

Southern blotting. Bacterial genomic DNAs were digested to completion with restriction enzyme *Eco*RI or *Hin*dIII, size fractionated through 0.9% agarose gels, and transferred to nylon membranes (Hybond-N; Amersham). Genomic DNAs cleaved with *Eco*RI were used in the analysis of IS*1*, since there were no recognition sites for this enzyme within the element. *Hin*dIII was used to analyze IS*200* copy numbers. There is a recognition site for *Hin*dIII at the 5′ end of IS*200* that cleaves 19 bp off the amplified product. In a previous study (3), no polymorphic *Hin*dIII sites have been detected within the region of IS*200* amplified from strains of *E. coli* and *S. typhimurium*.

A hybridization probe for detection of IS*1* was prepared from *E. coli* (ECOR strain 23) by polymerase chain reaction (PCR; 28) amplification of the element with the following primers: forward, 5′-GATTTAGTGTATGATGG-3′; reverse, 5′-GATAGTGTTTTATGTTC-3′. The IS*200* probe was prepared by amplification of the element from *S. typhimurium* LT2 DNA (SARA strain 2) with the following primer pair: forward, 5′-CTAGGCTGGGGGTTCCGGAA-3′; reverse, 5′-CAGATGCGCCTATAAGGCT-3′. (The direction of the IS*200* open reading frame [ORF] has been determined, and therefore, the orientation of the element and the direction of the primers are opposite to those reported previously.) Primer 5′-CTGTTCCTGTATTTGAAT-3′ and its complement, which span positions 436 to 453 of IS*200*, were used in conjunction with the IS*200* forward and reverse primers to assess the sequence heterogeneity of IS*200* in several strains of *S. typhimurium*. In addition, a primer four nucleotides shorter than the IS*200* reverse primer described above was used in combination with the IS*200* forward primer to detect elements in strains that did not yield products with the original pair of primers. Primers employed in the amplification of probe DNAs were based on available sequences of IS*1* (25) and IS*200* (12, 14). The IS*1* primers were designed to anneal internally to the inverted repeats. Each primer pair allowed amplification of almost the entire element, and the expected sizes of the PCR products were 729 bp for IS*1* and 660 bp for IS*200*. All amplification reactions proceeded for 30 cycles and were executed with 1 min of denaturation at 94°C, 30 s of annealing at 50°C, and 1 min of extension at 72°C. Amplification products were electrophoresed through 0.9% agarose gels, stained with ethidium bromide, and visualized under UV illumination to ensure amplification of IS*1* and IS*200*. PCR products were precipitated by addition of 2 volumes of ethanol, recovered by centrifugation, labeled to high specific activity by the method of Feinberg and Vogelstein (8), and hybridized to membranes containing digested bacterial DNAs for 16 h at 65°C in a high-phosphate buffer (0.5 M NaCl, 0.1 M NaH_2PO_4, 5 mM EDTA) supplemented with 0.2% Sarkosyl. Filters were washed at 20°C for a total of 30 min in five changes of 10 mM Tris–1 mM EDTA (pH 8.0) prior to autoradiography. Each hybridizing band can be taken to represent a single copy of an IS element, and the complement of hybridizing fragments for each strain has been designated an IS fingerprint (15, 34). In that comigrating and very large DNA fragments could represent multiple hybridizing fragments, IS copy numbers might be underestimated.

DNA sequencing. IS elements were amplified from genomic DNAs by the PCR, and single amplification products of the expected size, confirmed by electrophoresis, were directly purified with Prep-A-Gene (Bio-Rad) from the PCR mixture. The sequence of double-stranded DNA was determined by using Sequenase V.2.0 as specified by the manufacturer (U.S. Biochemicals), except that the primer-template mixture was denatured at 95°C for 5 min and cooled on ice prior to sequencing. For each IS element, the nucleotide sequence was determined for both strands over the entire amplified region. Sequencing was accomplished with the original primers used for PCR and with additional oligonucleotides whose sequences were complementary to the internal portions of IS*1* or IS*200*. The sequences of the IS*1* and IS*200* elements were obtained from SARA strains 17 (*S. typhimurium* Tm16) and 38 (*S. heidelberg* He5). These two *Salmonella* strains were selected for sequence analysis because they represent the most divergent lineages (on the basis of protein electrophoretic analysis) that contain both IS*1* and IS*200*. The *Salmonella* IS*200* sequence is identical to that obtained by Haack and Roth (12a) and differs from the original sequence reported by Gibert et al. (11, 12) at six positions over the region analyzed. In addition, we sequenced IS*200* from four strains of *E. coli* (ECOR 8, 51, 63, and 72).

Data analysis. The sequence data were analyzed by using programs from the University of Wisconsin Genetics Computer Group package (6). The BESTFIT program was used to calculate the degree of sequence identity with the algorithm of Smith and Waterman (33); divergence values were calculated by the DIVERGE program with the method of Perler et al. (26); and the FRAMES program was used to search for all possible ORFs in the sequence of each IS element. The degree of bias in synonymous codon usage within IS*1* and IS*200* was estimated by the codon adaptation index, which measures the extent of bias toward codons employed by highly expressed genes (32).

Nucleotide sequence accession number. The sequences reported here have been deposited in GenBank under accession numbers L25844 through L25848.

RESULTS

Distribution and abundance of IS*1* and IS*200*. Strains of *S. typhimurium* (subgroup I) from the SARA collection (2) were polymorphic with respect to their copy numbers and patterns of hybridization for both IS*1* and IS*200* (Fig. 1). Only 9 of the 66 *Salmonella* strains tested harbor IS*1*; compared with almost 90% of the *E. coli* strains from the ECOR collection. Moreover, those *Salmonella* strains containing the element have five or fewer copies whereas some *E. coli* strains are known to harbor over 20 copies of IS*1* (22, 29). None of the *Salmonella* strains displayed identical patterns of hybridization (IS*1* fingerprints), although some, such as SARA strains 33 and 34 (He1), were genetically indistinguishable by use of protein electrophoretic markers (2).

In contrast to the sporadic distribution of IS*1* among strains from the SARA collection, IS*200* is widespread and occurs in 65% of these strains, with some harboring as many as 19 copies of the element. In several strains containing hybridizing copies of IS*200* (SARA strains 35, 41 to 43, 45 to 49, 51 to 56, 61, and 62), this element could not be detected in this PCR (3), presumably because of variation in the sequence recognized by one or both of the original amplification primers. To determine which of these original amplification primers produced the negative results obtained with these strains, we performed a series of secondary amplification reactions by using the primers complementary to the center of IS*200* in combination with the original primers. Most strains (SARA strains 41, 42, 45, 46, 48, 49, 50, and 54) yielded a fragment of the proper length (432 bp) in these reactions containing the IS*200* forward primer,

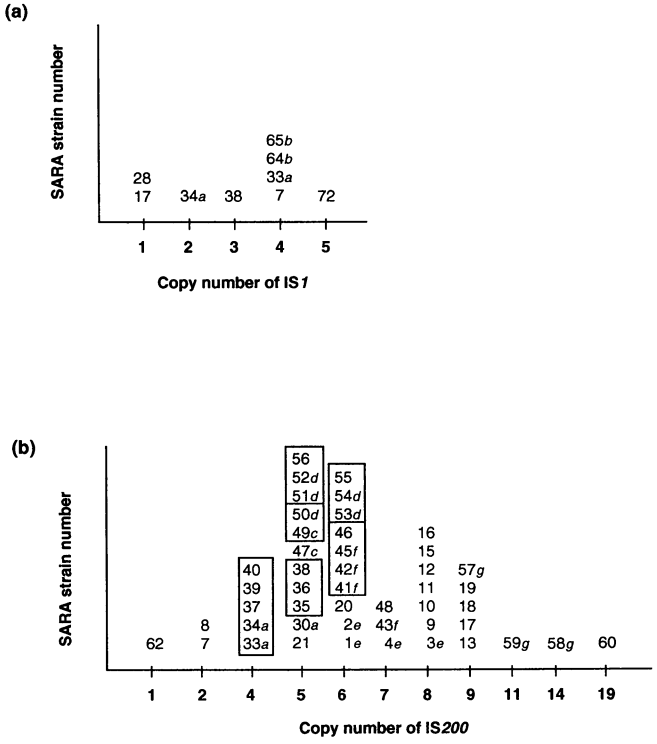

FIG. 1. Distribution and abundance of IS*1* (a) and IS*200* (b) among strains of the *S. typhimurium* subgroup I complex (2). The strains surveyed included SARA strains 1 to 4, 7 to 13, 15 to 30, 33 to 43, and 45 to 72. Only strains harboring one or more copies of IS*1* or IS*200* are shown. Strains having identical IS*200* fingerprints are boxed, and each box represents a unique IS fingerprint pattern. Strains with identical electrophoretic types are designated with the letters *a* to *g* as follows: *a*, *S. heidelberg* He1; *b*, *S. muechen* Mu1; *c*, *S. paratyphi* B Pb2b; *d*, *S. paratyphi* B Pb; *e*, *S. typhimurium* Tm1; *f*, *S. paratyphi* B Pb1; *g*, *S. paratyphi* B Pb5.

suggesting that the sequence variation corresponds to the 3′ end of the IS*200* reverse primer. To test this possibility, an additional primer, four nucleotides shorter than the original IS*200* reverse primer, was synthesized and used in amplification reactions with the original IS*200* forward primer. With the shortened reverse primer, we were able to obtain amplification products of the proper length and to confirm our expectation that IS*200* elements are heterogeneous near the 3′ end of the original primer. Therefore, by changing the target sequences recognized by the amplification primers, it is possible to achieve the same level of detection in both PCR and Southern hybridizations.

A majority of strains within the serovars *S. paratyphi* (group B) and *S. heidelberg* (group E) displayed identical, or very similar, IS*200* fingerprints, indicating that they share some or all of the sites at which IS*200* is present in the genome (Fig. 1). Strains within other phylogenetic clusters, such as *S. paratyphi* (group C), and the majority of the *S. typhimurium* strains (group F) had less similar patterns of hybridization, indicating some transposition or deletion of IS*200* within hosts. Electrophoretically indistinguishable strains of *S. typhimurium* Tm1 (SARA 1 to 3) vary in the number of copies of IS*200*.

Sequence divergence. To determine the extent of similarity between the IS*1* element detected in *S. typhimurium* and those previously isolated from other enteric bacteria (e.g., *E. coli*, *E. fergusonii*, *E. hermannii*, and *E. vulneris*), we obtained the nucleotide sequences of IS*1* elements from two *S. typhimurium* strains (SARA 17 and 38). Comparisons of the polymorphic sites in these elements with IS*1* from *E. coli* K-12 are presented in Fig. 2, and Table 1 summarizes the extent of sequence divergence between pairs of elements. SARA strains 17 and 38 harbor distinct forms of IS*1*, with nucleotide sequences that differ by 8.2%, and are more similar to the IS*1* elements present in other species of enteric bacteria than to one another. IS*1* from SARA strain 17 is identical to that from *E. coli* K-12, while the IS*1* from SARA strain 38 is very similar to the IS*1* recovered from *E. hermannii* (18), with a nucleotide sequence that differs by 1.4%.

```
                  11111111222222222223333333344444444444445555555555556666666
                  60012277901245556669003346690124456788999001223336788923566 77
                  50384758621373691258142653531145721167125081231484368710419 08

SARA 17,K12       GCGCCCAGTAGCTCGACCCCCTCCGCCAGGCCCATATACGGGAAACCGAGCAGAAGCCACG
SARA 38           AACTTTGACG.TCTCCGTA.A.AAATTTCATAAGG.GTTC.ACGGGGCGTAGATC.TACAA
E. hermannii      AACTTTGTCGATCTCCGTAAACAAATTTCATAAGGG.C.CAAC.GGGCG.A..TCATACAA
```

FIG. 2. Distribution of nucleotide sequence polymorphisms of IS*1* among enteric bacteria. K12, *E. coli* K-12 (23); S17 and S38, SARA strains 17 and 38 of the *S. typhimurium* subgroup I complex (2); Ehe, *E. hermannii* (18). Positions of polymorphic sites (numbered vertically) are listed above the sequence and are relative to the first position of the IS*1* inverted repeat. With the IS*1* element from *E. coli* K-12 as the reference sequence, dotted positions represent sequence identities, and positions that differ are marked with the substituted nucleotides.

The nucleotide sequences of IS*200* elements were the same for the two *Salmonella* strains (SARA strains 17 and 38) and identical to the sequence of IS*200* from *S. typhimurium* LT2

TABLE 1. Comparison of IS ORFs from enteric bacteria[a]

Sequences compared[b]	% DNA identity	% Divergence[c]		% Protein identity
		Silent	Replacement	
IS*1* ORF*a*, 56 (ATG)–331 (TAA)				
K12-S17	100.0	0	0	100.0
K12-S38	93.1	31.6	1.6	96.7
S17-S38	93.1	31.6	1.6	96.7
S17-Ehe	92.4	37.6	2.2	95.6
S38-Ehe	98.6	2.2	0	100.0
IS*1* ORF*h*, 250 (ATG)–729 (TAT)				
K12-S17	100.0	0	0	100.0
K12-S38	91.0	23.9	5.3	88.8
S17-S38	91.0	23.9	5.3	88.8
S17-Ehe	90.8	22.6	5.3	86.2
S38-Ehe	96.7	6.5	1.5	96.3
IS*200* ORF[d] 100 (ATG)–558 (TAA)				
S. typhimurium-*E. coli*	89.7	68.7	2.5	94.1
E. coli-*E. coli*	97.8	6.8	1.2	96.8

[a] IS*1* ORF designations (*a* and *h*) are those of Galas and Chandler (9). ORF*h* is not complete and extends to position 741.

[b] Sequence designations: K12, *E. coli* K-12; S17 and S38, *S. typhimurium* SARA strains 17 and 38; Ehe, *E. hermannii*.

[c] Identity values were calculated by the method of Smith and Waterman (33). Divergence values were calculated by the method of Perler et al. (26). Initiation and termination codon positions are relative to the first base of IS*1*.

[d] Initiation and termination codon positions relative to the first base of IS*200* in Fig. 2. There is no variation in IS*200* sequences among *Salmonella* strains. Sequence divergence and identity values represent averages for all pairwise comparisons.

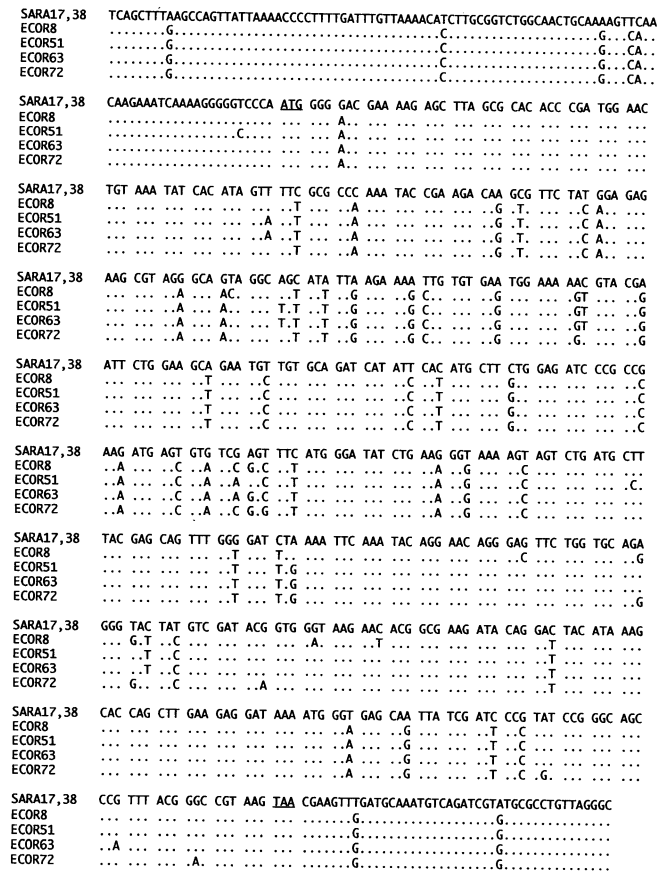

FIG. 3. Partial IS*200* sequences from *E. coli* and salmonellae. SARA strains 17 and 38 of the *S. typhimurium* subgroup I complex (2) and *E. coli* strains (ECOR strains 8, 51, 63, and 72) are from the ECOR collection (23). The sequence begins at position 48 from the actual start of IS*200*. Initiation and termination codons of the single ORF are underlined.

but polymorphic within *E. coli* (ECOR 8, 51, 63, and 72) (Fig. 3). Pairwise comparisons of these sequences revealed a maximum difference of 10.7% between the IS*200* sequences from *E. coli* and *S. typhimurium* and 2.7% difference between the IS*200* sequences from individual *E. coli* strains. While restriction fragment length polymorphism analysis of the IS*200* sequences from numerous *Salmonella* strains (3), as well as the identity of IS*200* sequences in SARA strains 17 and 38 and strain LT2, indicate that IS*200* is uniform within the *S. typhimurium* complex, the amplification reactions performed with additional sets of primers revealed a low level of sequence heterogeneity among strains.

Conservation of ORFs. IS*1* of *E. coli* contains two ORFs (*a* and *h*) which encode proteins essential for transposition (9), and IS*1* from *S. typhimurium* contained the same ORFs. To ascertain whether IS*1* elements from *Salmonella* strains are likely to be functional, we compared the nucleotide and corresponding amino acid sequences of ORFs *a* and *h* from SARA strain 38 to those of IS*1*, originally recovered from *E. coli* K-12 (Table 1). (IS*1* from SARA strain 17 is identical to that of *E. coli* K-12 and is presumably functional.) Several features of the IS*1* ORFs from SARA strain 38 suggest that the element has the potential to transpose: the excess number of synonymous-to-nonsynonymous substitutions and an absence of terminating substitutions within the ORFs. In addition, the preservation of a polyadenine (A_6C) signal implicated in the ribosomal frameshift control of transposition was conserved in IS*1* from SARA strain 38.

IS*200* sequences from both *E. coli* and *S. typhimurium* contain a single large ORF that extends almost the entire length of the element (from positions 100 to 558 in Fig. 2) and encodes a putative protein product of 152 amino acids (Table 2). This reading frame differs from those predicted by Gibert et al. (11), who detected three short ORFs for IS*200* from *S. typhimurium* LT2. The large ORFs within the IS*200* sequences from *E. coli* and *S. typhimurium* show 96% amino acid sequence identity, and most of the substitutions occur at synonymous sites. All other potential ORFs within IS*200* encoded products less than 50 amino acids long, and most lacked translation start and stop codons. This ORF has a low synonymous codon usage bias (codon adaptation index, 0.24), similar to that calculated for other IS elements from *E. coli* (16). The G+C content of IS*200* is, on average, 45.5%, lower than that of the *E. coli* and *S. typhimurium* genomes (51 to 53%), while the G+C content of IS*1* is 54.5% and is similar to that of other elements (16).

DISCUSSION

IS*200* was originally detected in salmonellae and thought to be confined to that genus (13, 14); however, copies of IS*200* are now known to reside within some *E. coli* (3) and *Shigella* (11) strains. Several features of IS*200* are atypical of transposable elements, including the lack of inverted terminal repeats and the absence of host target site duplications upon insertion, suggesting that transposition of the element might be restricted. To establish whether the present-day distribution of this element in enteric bacteria is due to horizontal transfer or common ancestry, we determined the IS*200* nucleotide sequences of *E. coli* and *S. typhimurium* isolates. In the absence of horizontal transfer, it might be expected that IS*200* in the two species would have diverged to the same extent as chromosomal genes and the phylogenies of strains based on the IS*200* sequences would be congruent with those inferred from chromosomal genes.

IS*200* elements from *E. coli* are genotypically distinct from those in *S. typhimurium*, differing by as much as 10% in the nucleotide sequence. This level of diversity is similar to the average extent of sequence divergence observed between pairs of homologous chromosomally encoded genes from the two species (15.6%) (31) and implies that (i) IS*200* was present in the common ancestor of *E. coli* and salmonellae, and (ii) copies of IS*200* have not been transferred between these species since they diverged. On the basis of the base composition (45 to 46% G+C) and phylogenetic distribution of IS*200*, the element probably arose in an organism with low G+C content and was acquired by the common ancestor of *E. coli* and salmonellae. From this, we might predict that elements containing the structural features of IS*200* would be present in low-G+C organisms. IS*200* has not been detected in *Citrobacter*, *Klebsiella*, *Serratia*, and *Proteus* spp. (14); however, these genera have not been surveyed extensively for the presence of this element.

On the basis of both experimental and comparative sequence data, the frequency of genetic exchange between *E. coli* and salmonellae is thought to be very low (17, 31). While many IS classes occur on extrachromosomal DNAs, which might aid in their dissemination, copies of IS*200* are rarely detected on plasmids (35, 36). In the absence of an association with such vectors, there may be little potential for horizontal transfer of

IS*200* between *E. coli* and *S. typhimurium*, which could account for the divergence observed in the elements from these two species. In addition, the limited amount of transfer of IS*200* among *E. coli* strains is also reflected in the phylogenetic distribution. Not only is the extent of sequence divergence of IS*200* within *E. coli* (2.7%) similar to that seen in other protein-coding regions within *E. coli* (7, 19, 20), but the relationships of the *E. coli* strains, based on the IS*200* sequences (with the *S. typhimurium* IS*200* sequence as the outgroup) are congruent with those based on chromosomal genes (19, 30). The overall similarity in the topologies of phylogenetic trees based on chromosomal genes and IS can be taken as evidence that neither character has been independently affected by horizontal processes.

In contrast to the situation in *E. coli*, the IS*200* nucleotide sequences determined for the two strains of *S. typhimurium* are identical. Restriction fragment length polymorphism analysis (3) has also indicated that IS*200* is very conserved among salmonellae. Despite the uniformity detected in sequenced and restriction mapped elements, there is some evidence that IS*200* is polymorphic among salmonellae, as seen by our inability to amplify IS*200* from some strains shown to contain the element by Southern hybridizations. The distribution of the variant form of this element is limited to strains within the *S. paratyphi* B and *S. muenchen* (designated groups A and B).

Initially, the homogeneity of IS*200* might imply that horizontal transfer of the element occurs to a greater extent in salmonellae than in *E. coli*. However, it is unlikely that this sequence uniformity was produced by horizontal transfer, for the following reasons. (i) The total extent of genetic diversity among the SARA strains based on protein electrophoretic loci is very low, less than one-third of that of the ECOR collection (2, 30). For example, the genetic distance between the *S. typhimurium* strains examined—SARA 17 and 38—is equivalent to that between the two closely related strains of *E. coli*, ECOR 51 and 63, whose IS*200* nucleotide sequences differ by only 0.6%. Therefore, the identity of IS*200* from SARA 17 and 38 might be expected even if the elements were ancestral to the two strains. (ii) The variant form of IS*200*, as revealed by our sequential PCR amplifications, is confined to a group of phylogenetically related lineages and displays no evidence of lateral transfer. Unlike those of *E. coli* strains, the IS*200* fingerprints of phylogenetically related *Salmonella* strains are very similar, which also reflects the close genetic relationships among strains. For instance, strains within some serovars (e.g., *S. paratyphi* and *S. heidelberg*) share most or all of the sites at which IS*200* sequences are present in the genomes, indicating that most of the elements have been transmitted vertically and are identical by descent.

It has previously been proposed that transposition of IS*200* may be subject to regulation by host strains (3, 11, 13). Although the copy numbers of IS*200* are similar for the two species, with as many as 11 and 19 copies harbored by strains of *E. coli* (3) and *S. typhimurium*, respectively, it is not known whether these species have different rates of IS*200* transposition. The ORF encoded by IS*200* is highly conserved, and any differences in the transfer rates would presumably not be attributable to changes in this protein-coding region. The possibility, however, that even minor changes in the protein sequence could affect rates of transposition cannot be excluded without additional experimental analysis.

On the basis of comparisons of IS fingerprints, there is some evidence that IS*200* transposes at very low rates within the genomes of individual strains. This is best illustrated by electrophoretically indistinguishable strains, such as *S. typhimurium* Tm1 (SARA strains 1, 2, and 3), that display similar but not identical patterns of hybridization. For example, SARA strain 2 contains six copies of IS*200*, and the element is present at the same six sites in the genome of SARA strain 1 and at only five of the sites in SARA strain 3. Moreover, SARA strains 1 and 3 share an additional hybridizing band not seen in SARA strains 2, and SARA strain 3 has two unique bands not present in SARA strains 1 and 2. Therefore, five copies of IS*200* have been vertically inherited by SARA strains 1, 2, and 3 and the unshared bands can be attributed to transposition of the element after the lineage separated. (A change in IS copy numbers is caused by transposition rather than mutations in restriction enzyme recognition sites, which generally alter the position but not the number of hybridizing fragments.) Some classes of elements, such as IS*1*, show variation in the length of target site duplications. This is attributed to mutations in the genes whose products are involved in the cleavage reactions, which in turn may affect the frequency of transposition. It not known whether the termini of IS*200* differ between *E. coli* and *S. typhimurium* or whether the IS*200* sequences in *E. coli* and *S. typhimurium* recognize different target sites. However, these issues could be elucidated by sequence analysis of the 3′- and 5′-flanking regions of IS*200* elements from several strains of both species.

Unlike that of IS*200*, the evolution of IS*1* among *S. typhimurium* strains has included several horizontal-transfer events. Although the nucleotide sequences of IS*200* elements from *S. typhimurium* SARA strains 17 and 38 are identical by descent, IS*1* sequences from these two strains are distinct. IS*1* from SARA strain 17 is identical to that recovered from *E. coli* K-12, while IS*1* from SARA strain 38 is very similar to that of *E. hermannii*. The *Escherichia* species have diverged substantially—*E. hermannii* is more distantly related to *E. coli* than are salmonellae—and the existence of very similar or identical IS*1* elements among *S. typhimurium* strains can only be the result of horizontal transfer. Our findings are consistent with previous studies investigating the level of sequence divergence of IS*1* elements among different species of enteric bacteria (18). In addition, these previous studies have also demonstrated that the phylogenies inferred from comparisons of IS*1* sequences were not congruent with the relationships inferred from chromosomal genes, as might be expected if the acquisition of the element was subsequent to the divergence of the species.

In contrast to IS*200*, IS*1* elements are often associated with plasmids. In the ECOR collection of *E. coli* strains, over 90% of the strains harbor IS*1* and over 10% of all copies reside on plasmids (29). It is not known whether any *S. typhimurium* strains contain plasmid-borne copies of IS*1*, but an association of IS*1* with extrachromosomal DNAs may have facilitated its transfer both within and among distantly related species of enteric bacteria (18). The sporadic distribution of IS*1* among salmonellae probably reflects recent introduction and dissemination of the element within the genus. This is also reflected in the relatively low copy number of IS*1* among *S. typhimurium* strains compared with that seen in natural populations of *E. coli*. The heterogeneity of IS*1* fingerprints among *S. typhimurium* strains suggests that IS*1*, unlike IS*200*, undergoes a fairly high rate of transposition. For example, *S. heidelberg* SARA strains 33 and 34 (He1) have indistinguishable protein electrophoretic profiles and IS*200* fingerprints but completely distinct patterns of IS*1* hybridization. The differing fingerprints and copy numbers suggest that acquisition of IS*1* by SARA strains 33 and 34 was the result of at least two independent transfer events.

By analyzing the present distribution and nucleotide sequences of several IS elements, we have determined that different IS classes, such as IS*1* and IS*200*, can exhibit distinc-

tive and characteristic patterns of evolution in natural populations of enteric bacteria. Moreover, the dynamics of these elements do not seem to be affected by their bacterial hosts: IS*1* is readily transferred both within and among strains of *E. coli* and *S. typhimurium*, whereas copies of IS*200* appear to be ancestral to lineages harboring the element. These data, along with evidence that enteric species share several other IS classes, can help identify the role of these elements in the evolution and organization of bacterial genomes.

ACKNOWLEDGMENTS

We thank Eric Mayer for technical support, Ken Haack for sharing unpublished data, and Ulfar Bergthorsson for reviewing the manuscript.

This work was supported by a grant from the National Institutes of Health (GM40995).

REFERENCES

1. **Arber, W., M. Hümbelin, P. Caspers, H. J. Reif, S. Iida, and J. Meyer.** 1981. Spontaneous mutations in the *Escherichia coli* prophage P1 and IS-mediated processes. Cold Spring Harbor Symp. Quant. Biol. **45:**107–113.
2. **Beltran, P., S. A. Plock, N. H. Smith, T. S. Whittam, D. C. Old, and R. K. Selander.** 1991. Reference collection of strains of the *Salmonella typhimurium* complex from natural populations. J. Gen. Microbiol. **137:**601–606.
3. **Bisercic, M., and H. Ochman.** 1993. Natural populations of *Escherichia coli* and *Salmonella typhimurium* harbor the same classes of insertion sequences. Genetics **133:**449–454.
4. **Calos, M., and J. H. Miller.** 1980. Transposable elements. Cell **20:**579–595.
5. **Campbell, A.** 1981. The evolutionary significance of accessory DNA elements in bacteria. Annu. Rev. Microbiol. **35:**55–83.
6. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
7. **DuBose, R. F., D. E. Dykhuizen, and D. L. Hartl.** 1988. Genetic exchange among natural isolates of bacteria: recombination within the *phoA* locus of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85:**7036–7040.
8. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
9. **Galas, D. J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 109–162. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
10. **Green, L., R. D. Miller, D. E. Dykhuizen, and D. L. Hartl.** 1984. Distribution of DNA insertion element IS*5* in natural isolates of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81:**4500–4504.
11. **Gibert, I., J. Barbe, and J. Casadesus.** 1990. Distribution of insertion sequence IS*200* in *Salmonella* and *Shigella*. J. Gen. Microbiol. **136:**2555–2560.
12. **Gibert, I., K. Carroll, D. R. Hillyard, J. Barbe, and J. Casadesus.** 1991. IS*200* is not a member of the IS*600* family of insertion sequences. Nucleic Acids Res. **19:**1343.

12a. **Haack, K., and J. R. Roth.** Personal communication.

13. **Lam, S., and J. R. Roth.** 1983. Genetic mapping of IS*200* copies in *Salmonella typhimurium* LT2. Genetics **105:**801–811.
14. **Lam, S., and J. R. Roth.** 1983. A *Salmonella*-specific insertion sequence. Cell **34:**951–960.
15. **Lawrence, J. G., D. E. Dykhuizen, R. F. DuBose, and D. L. Hartl.** 1989. Phylogenetic analysis of *Escherichia coli* based on insertion sequence fingerprinting. Mol. Biol. Evol. **33:**241–250.
16. **Lawrence, J. G., and D. L. Hartl.** 1991. Unusual codon bias occurring within insertion sequences in *Escherichia coli*. Genetica **84:**23–29.
17. **Lawrence, J. G., D. L. Hartl, and H. Ochman.** 1991. Molecular and evolutionary relationships among enteric bacteria. J. Gen. Microbiol. **137:**1911–1921.
18. **Lawrence, J. G., H. Ochman, and D. L. Hartl.** 1992. The evolution of insertion sequences within enteric bacteria. Genetics **131:**9–20.
19. **Milkman, R., and M. M. Bridges.** 1993. Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons. Genetics **133:**455–468.
20. **Milkman, R., and I. P. Crawford.** 1983. Clustered third-base substitutions among wild strains of *Escherichia coli*. Science **221:** 378–380.
21. **Nyman, K., K. Nakamura, H. Ohtsubo, and E. Ohtsubo.** 1981. Distribution of the insertion elements IS*1* in gram-negative bacteria. Nature (London) **289:**609–612.
22. **Nyman, K., H. Ohtsubo, D. Davison, and E. Ohtsubo.** 1983. Distribution of insertion elements IS*1* in natural isolates of *Escherichia coli*. Mol. Gen. Genet. **198:**516–518.
23. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:**690–693.
24. **Ohtsubo, E., H. Ohtsubo, W. Doroszkiewicz, K. Nyman, D. Allen, and D. Davison.** 1984. An evolutionary analysis of ISO-IS*1* elements from *Escherichia coli* and *Shigella* strains. J. Gen. Appl. Microbiol. **30:**359–376.
25. **Ohtsubo, H., and E. Ohtsubo.** 1978. Nucleotide sequence of an insertion element IS*1*. Proc. Natl. Acad. Sci. USA **75:**615–619.
26. **Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodgson.** 1980. The evolution of genes: the chicken preproinsulin gene. Cell **20:**555–566.
27. **Read, H., and S. R. Jaskunas.** 1980. Isolation of *E. coli* mutants containing multiple transpositions of IS sequences. Mol. Gen. Genet. **188:**157–164.
28. **Saiki, S., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487–491.
29. **Sawyer, S., D. Dykhuizen, R. DuBose, L. Green, T. Mutangadura-Mhlanga, D. Wolczyk, and D. Hartl.** 1987. Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. Genetics **115:**51–63.
30. **Selander, R. K., D. A. Caugant, and T. S. Whittam.** 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
31. **Sharp, P. M.** 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position and concerted evolution. J. Mol. Evol. **33:**23–33.
32. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:**1281–1295.
33. **Smith, T. F., and M. S. Waterman.** 1981. Identification of the common molecular sequences. J. Mol. Biol. **147:**195–197.
34. **Stanley, J., A. Burnens, N. Powell, N. Chowdry, and C. Jones.** 1992. The insertion sequence IS*200* fingerprints chromosomal genotypes and epidemiological relationships in *Salmonella heidelberg*. J. Gen. Microbiol. **138:**2329–2336.
35. **Stanley, J., M. Goldsworthy, and E. J. Threlfall.** 1992. Molecular phylogenetic typing of pandemic isolates of *Salmonella enteritidis*. FEMS Microbiol. Lett. **90:**153–160.
36. **Stanley, J., C. S. Jones, and E. J. Threlfall.** 1991. Evolutionary lines among *Salmonella enteritidis* phage types are identified by insertion sequence IS*200* distribution. FEMS Microbiol. Lett. **82:**83–90.