© 1997 Oxford University Press
*Nucleic Acids Research,* 1997, Vol. 25, No. 7 **1355–1361**

# Conserved structure of IS*200* elements in *Salmonella*

**Carmen R. Beuzón and Josep Casadesús***

Departamento de Genética, Facultad de Biología, Universidad de Sevilla, Apartado 1095, E-41080 Sevilla, Spain

Received December 12, 1996; Revised and Accepted February 7, 1997

DDBJ/EMBL/GenBank accession nos X56834, Y09564, Y08755

## ABSTRACT

**Sequence analysis of three IS*200* elements (two from *Salmonella typhimurium*, one from *Salmonella abortusovis*) reveals a highly conserved structure, with a length of 707–708 bp and absence of terminal repeats. IS*200* contains an open-reading-frame (ORF) which potentially encodes a peptide of 151 amino acids, with a putative ribosome-binding-site properly placed upstream of the ORF. A potential RNA stem–loop structure that might occlude the ribosome-binding-site of the ORF is also found. Another conserved trait is a potential RNA hairpin which resembles a Rho-independent transcription terminator, located near one end of IS*200*. The junctions between IS*200* and host DNA sequences are A+T-rich. Upon insertion, IS*200* duplicates 1–2 bp of host DNA sequences. The observation that IS*200* elements characterized as 'hops' are roughly identical to those residing in the *Salmonella* genome suggests that IS*200* transposition is unlikely to generate inactive copies. If such is the case and many or all IS*200* elements are active, the extremely low frequency of IS*200* transposition may reflect the normal behavior of the element.**

## INTRODUCTION

IS*200* is an insertion element abundant in the genus *Salmonella* (1,2) and sporadically found in strains of *Shigella* (2), *Escherichia coli* (3) and *Yersinia* (4). The chromosome of *Salmonella typhimurium* LT2 contains six copies of IS*200* (1,5,6); in other *S.typhimurium* strains the number of IS*200* copies ranges from 1 to >12 (2). One paradoxical trait of IS*200* behavior is its poor contribution to spontaneous mutagenesis (7,8), which can be correlated with the extremely low transposition frequency of the element (8–10). In fact, only two IS*200* insertion mutations have been reported in *S.typhimurium*: one in the *his* operon (1), another in the *gpt* gene (8). Hunts for IS*200*-induced mutants in *S.typhimurium*, sometimes involving positive selection strategies, have confirmed that transposition is rare (7–10). Furthermore, surveys carried out in field isolates have indicated that IS*200* transposition is also infrequent in natural populations of *Salmonella* (11).

One possible explanation for the low activity of IS*200* might be the generation at relatively high frequencies of defective copies of the element. Such a behavior has no precedents among prokaryotic insertion elements (12) but is relatively common in eukaryotic transposons (13). To investigate this possibility, we compared the structure of IS*200* 'hops' with that of 'genomic copies' of the element. The results were unequivocal: elements characterized as IS*200* hops are extremely similar, if not identical, to IS*200* elements residing in the *Salmonella* genome. Thus IS*200* transposition is unlikely to generate inactive or aberrant copies. A corollary is that the extremely low frequency of IS*200* transposition and its scarce contribution to spontaneous mutagenesis likely reflect the normal behavior of the element.

Cladograms constructed with the predicted peptides of IS*200* ORFs from *Salmonella*, *E.coli* and *Yersinia* match the phylogenetic tree of the Enterobacteriaceae (14,15). This observation gives support to the hypothesis that IS*200* is an ancestral element of enteric bacteria (16).

## MATERIALS AND METHODS

### Bacterial strains and plasmids

All strains and isolates of *S.typhimurium* used in this work derive from the standard wild-type LT2. Strain SS44 of *Salmonella abortusovis* (17) was provided by S. Rubino, Institute of Microbiology, Università degli Studi di Sassari, Sassari, Sardinia, Italy. The plasmid vectors used for cloning were pBluescript I KS(+) and pBluescript II SK(+) (Stratagene, La Jolla, CA). pIZ46 is a pUC19 derivative carrying a tail-to-tail dimer of the 0.3 kb *Eco*RI–*Hin*dIII fragment of IS*200* (2). pIZ47 is a pUC19 derivative carrying a tetramer of the internal *Taq*I fragment of IS*200,* flanked by *Eco*RI sites (I. Gibert and J. Casadesús, unpublished). Transformation with plasmid DNA followed the procedure of Inoue *et al*. (18). The recipient strain for transformation was *E.coli* DH5α (19).

### Media and chemicals

Luria–Bertani medium (LB) with added NaCl (5 g/l) was used as the standard rich broth. Solid medium contained Difco agar at a final concentration of 1.5%. EBU indicator plates were prepared as described elsewhere (20). Concentrations of antibiotics were as follows: kanamycin sulfate (Sigma, St Louis, MO), 50 mg/l; ampicillin (Sigma), 100 mg/l. The indicator 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside ('X-gal', United States Biological, Swampscott, MA), was dissolved in *N,N*-dimethylformamide (2 mg/ml) and used at the final concentration of 25 μg/ml. For gel electrophoresis, agarose (SeaKem, FMC, Rockland, ME) was prepared in either TAE or TBE buffers (21). The final agarose concentration (0.7–1.0%) depended on the range of DNA

*To whom correspondence should be addressed. Tel: +34 5 455 7105; Fax: +34 5 455 7104; Email: genbac@cica.es

fragments to be separated. E buffer contained 40 mM Tris-base and 2 mM EDTA; pH was adjusted to 7.9 with glacial acetic acid. The lysis solution contained 3% sodium dodecyl sulfate, 50 mM Trizma base and 72 mM NaOH.

### Bacteriophages and transduction

The transducing phage was P22 HT 105/1 *int201* (22), henceforth referred as P22 HT. Phage sensitivity was tested with the clear-plaque mutant P22 H5. To obtain phage-free isolates, transductants were purified on EBU plates. For transduction of antibiotic-resistance markers (e.g. $Km^r$), transducing mixtures were made on LB plates and incubated 4–5 h before replica-printing to selective plates with sterile velvets.

### Virulence plasmid curing

Curing of the virulence plasmid (pSLT) of *S.typhimurium* was achieved by destabilization of the *par* locus with a kanamycin-resistant cartridge (23). The $Km^r$ cartridge was introduced into the strain to be cured by P22 HT transduction. $Km^r$ transductants were streaked on EBU plates; individual phage-free colonies were then scored for loss of kanamycin resistance.

### DNA extraction and purification

Genomic DNA preparations were obtained according to Ausubel *et al.* (24), using 10 ml of an overnight culture. Plasmid DNA for both clone analysis and DNA sequencing was obtained with the alkaline lysis method, without phenol extraction (25). Plasmid DNA preparations for the generation of nested deletions were obtained with the 'boiling' method (26), followed by purification in Sephadex G-50 columns. Virulence plasmid DNA was extracted as described elsewhere (11).

### Digestion, end modification and ligation of DNA fragments

Restriction enzymes were purchased from Promega Biotech (Madison, WI), Boehringer Mannheim (Mannheim, Germany) and New England Biolabs (Beverly, MA). The buffers used were those provided by the supplier. For multiple digestions, we used the 'One-phor-all' buffer (Pharmacia Biotech, San Francisco, CA). Deoxyribonucleotides and Klenow DNA polymerase were purchased from Promega Biotech. T4 polynucleotide ligase was from Boehringer Mannheim; ligation was achieved by incubating >12 h at 16°C. For blunt end ligation, a low-ATP buffer was used (New England Biolabs, unpublished observations).

### DNA hybridization

Digestion of DNA with restriction enzymes, electrophoretic separation of restriction fragments, DNA denaturation, transfer of DNA from agarose gels to nylon filters, DNA labeling and DNA hybridization followed the procedures described by Southern (27) and Sambrook *et al.* (21). Recovery of DNA from agarose gels was achieved with the GeneClean system (Bio 101, La Jolla, CA). The standard IS*200* probes were the *Eco*RI fragments of pIZ46 and pIZ47. These fragments were recovered from agarose gels and purified by the GeneClean system (Bio 101), modified according to Boyle and Lew (28). DNA probes were labeled with the DIG DNA labeling kit from Boehringer Mannheim.

### Generation of nested deletions

Deletions were generated with *E.coli* exonuclease III, using the double-stranded nested deletion kit from Pharmacia Biotech. Protected, 3′ overhanging ends were generated by *Apa*I digestion, while the substrate for exonuclease III digestion was generated with *Cla*I. Deleted plasmids were treated with nuclease S1 to generate blunt ends; these were then ligated with T4 polynucleotide ligase.

### DNA sequencing

Sequencing reactions were performed with the dideoxy chain termination procedure (29), using the Sequenase kit version 2.0 (United States Biochemical Corporation, Cleveland, OH). The primers used for sequencing are shown in Table 1. Sequencing gels were prepared in TBE and contained 6% acrylamide and 500 g/l urea. Gels were run in a Poker Face SE1500 sequencer (Hoeffer Scientific Instruments, San Francisco, CA), dried in a Slab Gel Dryer, model SE1160 (Hoeffer) and developed by exposure to an X-ray film.

### Sequence analysis and phylogeny methods

We used the computer analysis package of the Genetics Computer Group, University of Wisconsin, Madison, WI (30). Free energies of RNA secondary structures were calculated with the program 'Fold RNA' (31). Evolutionary distances between peptide sequences were measured using the Jukes–Cantor algorithm (32). Phylogenetic trees were constructed with the GrowTree program, using the 'unweighted pair-group artihmetic average' (UPGMA) method (33).

## RESULTS

### Sequence analysis of the insertion *hisD984::*IS*200*

The original mutation *hisD984*::IS*200* (1) had been previously cloned on several plasmid vectors (2). One of these clones, pIZ44, is a pUC9 derivative carrying a ~2 kb insert which includes the insertion *hisD984*::IS*200* and flanking sequences of the *S.typhimurium his* operon (2). DNA sequencing of the *hisD984*::IS*200* insertion (henceforth called IS*200*-HIS) was carried out directly on pIZ44, using primers I, II and V (Table 1). These primers were designed using previous sequencing data (34). The EMBL accession number for IS*200*-HIS is X56834. The predicted IS*200* ORF (see below) runs opposite to *hisD* (for a map of the histidine operon; ref. 35).

**Table 1.** DNA primers used for sequencing IS*200* elements

| Primer | Sequence (5′–3′) | Origin |
|---|---|---|
| T3 | ATTAACCCTCACTAAAG | pBluescript |
| T7 | AATACGACTCACTATAG | pBluescript |
| I | GTGCAGAGGGTACTATC | IS*200* |
| II | GAGCTTAGCGCACACCC | IS*200* |
| III | CCCGCCGAAGATGAGTGTGT | IS*200* |
| IV | GTCTTCGGTATTTGGGCGC | IS*200* |
| V | CTGCCTACTGCCCTACGCTTCT | IS*200* |

### Characterization of an IS*200* element inserted in the *S.typhimurium* virulence plasmid

Increase in the number of IS*200* elements is observed at low but detectable frequencies among the survivors of stab cultures (1,9,10), a behavior previously reported for other insertion elements (36–38). However, in a given strain, IS*200* copy number can potentially increase by duplication of chromosomal regions (39). This caveat directed our search to the detection of IS*200* insertions in the *S.typhimurium* virulence plasmid; the rationale was that a 'hop' in *trans* was unlikely to be caused by DNA rearrangements other than transposition.

Because the virulence plasmid (pSLT) of *S.typhimurium* LT2 does not carry IS*200* (5), the presence of IS*200* insertions in pSLT can be scored by comparing the IS*200* profiles of isolates carrying extra IS*200* copies and those of derivatives cured of the virulence plasmid. Curing of pSLT was achieved with the method of Tinge and Curtiss III (23). Analysis of 54 LT2 derivatives carrying more than six IS*200* copies yielded three independent examples of plasmid-borne IS*200* elements. Southern hybridization analysis of pSLT DNA digested with *Pvu*II, *Hin*cII, *Pst*I and *Ava*I (and with combinations of these enzymes) indicated that *Pvu*II–*Hin*cII digestion of one of the candidates generated a band of only 3.7 kb, absent in the cured derivative. This 3.7 kb fragment was cloned onto the *Sma*I site of pBluescript II SK(+) to generate plasmid pIZ801. Nested deletions with exonuclease III reduced the size of the insert of pIZ801 to 1 kb; Southern hybridization analysis indicated that the resulting plasmid, pIZ802, still carried IS*200*. The insert of pIZ802 was sequenced using primers T7, T3, III and IV (Table 1). This IS*200* element inserted in the virulence plasmid (EMBL accession no. Y09564) will be henceforth called IS200-VP.

### Characterization of an IS*200* element from *S.abortusovis*

Strain SS44 of *S.abortusovis* carries an IS*200* element on a ~9 kb *Pst*I chromosomal fragment (11). *Pst*I fragments of ~9 kb from *S.abortusovis* SS44 were cloned onto pBluescript I KS(+). The presence of IS*200* among $Lac^-$, $Ap^r$ transformants of *E.coli* DH5α was screened by Southern hybridization, using the *Eco*RI fragment of pIZ46 as a probe (2). One positive isolate was the source of plasmid pIZ72. Restriction analysis proved that pIZ72 carried an insert of 9.5 kb. This insert was split in two fragments of 6.5 and 3 kb by *Stu*I digestion. The 3 kb fragment (but not the 6.5 kb fragment) hybridized against IS*200* probes, suggesting that the IS*200* element was located in the small *Stu*I fragment. The latter was cloned onto pBluescript I KS(+) to generate plasmid pIZ73.

To reduce the size of the 3 kb *Stu*I fragment of pIZ73, nested deletions were generated with exonuclease III. The presence of IS*200* among deletion derivatives of pIZ73 was detected by Southern hybridization, as above. One plasmid carrying a ~1 kb insert (pIZ74) was chosen for DNA sequencing with primers I, II, V, T3 and T7. This IS*200* element (EMBL accesion no. Y08755) will be henceforth called IS200-SAO.

### General features of IS*200* elements from *Salmonella*

The three IS*200* elements sequenced are very similar. IS*200*-HIS is 707 bp, while IS*200*-VP and IS*200*-SAO are both 708 bp. Some relevant features found in their sequences are as follows.

(i) All elements lack terminal repeats, as previously reported (8,34,40). The ends of the three elements are highly conserved: in the left-most 50 bp, only 1 bp difference is found between IS*200*-HIS and IS*200*-VP, while IS*200*-SAO shows a 2 bp difference with both IS*200*-HIS and IS*200*-VP. In turn, the right-most 50 bp are identical in IS*200*-HIS and IS*200*-VP and 1 bp difference (with both) is found in IS*200*-SAO.

(ii) The three elements contain a single ORF which can potentially yield a peptide of 151 amino acids. The predicted peptides encoded by IS*200*-HIS and IS*200*-VP are identical, while the predicted peptide of IS*200*-SAO shows four amino acid changes (Fig. 1). The number of changes in the nucleotide sequence of the ORFs are as follows: (a) only 1 bp difference (at a third-codon position) between IS*200*-HIS and IS*200*-VP; (b) 10 nt differences (six at third-codon positions) between IS*200*-SAO and IS*200*-HIS; (c) 11 nt differences (seven at third-codon positions) between IS*200*-SAO and IS*200*-VP. Thus the divergence is higher at the DNA level than at the peptide level, suggesting that the coding capacity of the ORF is selectively maintained.

A databank search indicated that the IS*200* ORF is related to the transposase of IS*1004*, an insertion element found in *Vibrio cholerae* (41). Additional evidence that the IS*200* ORF may encode the transposase of the element is provided by the observation that expression of the IS*200* ORF from an inducible promoter increases the frequency of various IS*200*-associated DNA rearrangements (42; C. R. Beuzón and J. Casadesús, unpublished). Unfortunately, the rarity of IS*200* transposition has so far hampered a direct survey of the function of ORF1 (e.g. mutations in ORF1 should render a non-transposing element).

(iii) A potential ribosome-binding site (5′-AGGGG-3′) is located between nucleotides –5 and –9 upstream of the start codon of the ORF. The sequence differs only 1 bp from the consensus and is located at a standard distance from the start codon (43,44). This putative ribosome-binding-site is 100% conserved in all elements characterized in this study (and in IS*200* elements from other sources: see below).

(iv) Two regions that can be expected to give rise to secondary structures in their transcripts are found upstream of the ORF. The left-most structure (nucleotides 9–32) can give rise to a hairpin ($\Delta G_0$: –14.1 kcal/mol) which reminds of a Rho-independent transcription terminator (Fig. 2). This region has been actually shown to cause the polarity of the insertion *hisD984*::IS*200* on the distal gene *hisC* (40). The potential hairpin is identical in IS*200*-HIS and IS*200*-VP and shows a 1 bp difference (located in the stem) in IS*200*-SAO. Transcriptional terminators, both Rho-dependent and Rho-independent, have been found near the ends of many insertion sequences and are usually regarded as control elements that terminate transcripts entering the IS (12).

The second structure is defined by two inverted repeats located between nucleotides 66–85 and 120–139. The second repeat is only 3 bp away from the start of the ORF. The inverted repeats can give rise to a stem–loop structure in the transcript ($\Delta G_0$: –25.8 kcal/mol). If formed, this structure might cause occlusion of the ribosome-binding-site (Fig. 2). The potential stem–loop region is identical in IS*200*-HIS and IS*200*-SAO and shows a 1 bp difference (in the stem) in IS*200*-VP.

### Restriction analysis of IS*200* elements from *S.typhimurium* LT2

Five IS*200* elements present in the chromosome of *S.typhimurium* LT2 (copies I, II, III, V and VI; 5,6) were cloned using 'locked-in'

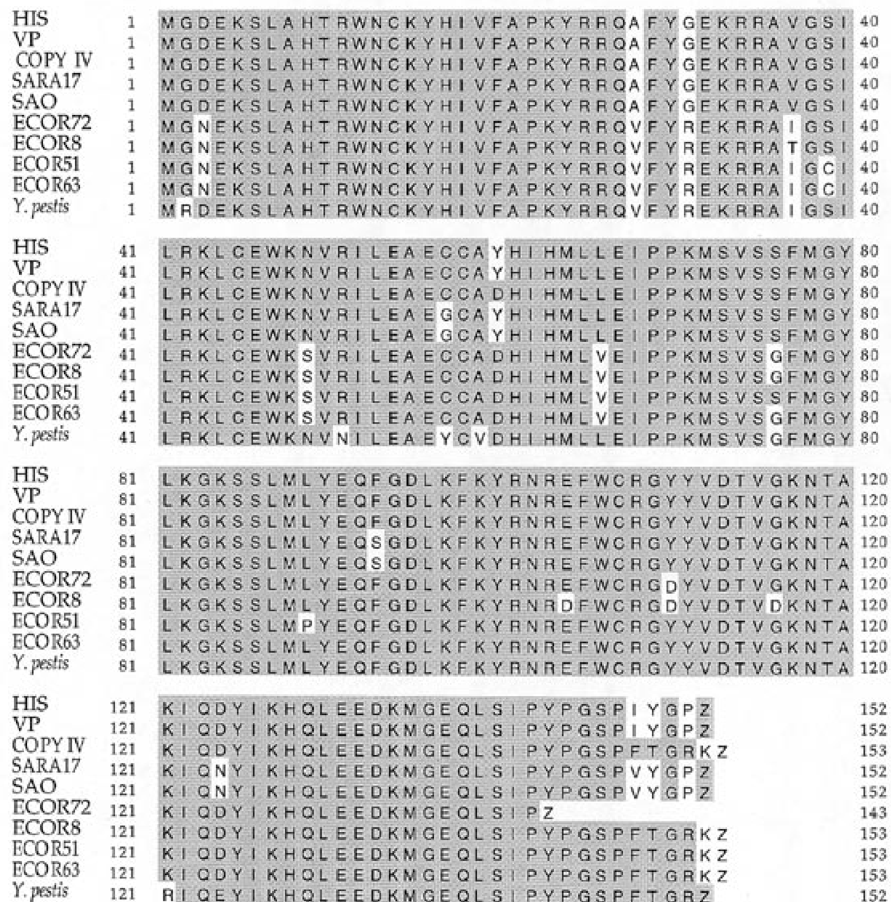


**Figure 1.** Alignment of the predicted peptides encoded by the ORF of IS*200* elements from various sources. HIS, VP and SAO are the three IS*200* elements sequenced in this work. Copy IV is one of the IS*200* elements of the *S.typhimurium* chromosome (5,6,46). SARA17 is another *Salmonella* IS*200* element (3), while ECOR8, ECOR51, ECOR63 and ECOR72 are IS*200* elements from *E.coli* (3). An IS*200* element from *Y.pestis* (4) is also included. Columns are shaded when more than half of the entries are identical.

Mu*d*-P22 prophages (10,45). Digestions were carried out with restriction enzymes that generate distinct restriction fragments in the elements IS*200*-HIS, IS*200*-VP and IS*200*-SAO: *Hpa*II, *Hae*III, *Hin*dIII, *Hph*I and *Eco*RI. Restriction fragments were then analyzed by Southern hybridization against IS*200* probes. The five elements proved to have identical restriction maps (data not shown). These results suggest that IS*200* elements from *S.typhimurium* have a highly conserved structure, as previously proposed (3). Copy IV was not included in these experiments because it has been already sequenced (46); it is extremely similar to both IS*200*-HIS and IS*200*-VP (see below).

## Sequence comparisons among IS*200* elements from various sources

Sequences of other IS*200* elements were retrieved from the EMBL databank and used for further comparisons. The sequences compiled were: (i) the complete sequence of an IS*200* element from *Yersinia pestis* (6); (ii) the complete sequence of IS*200* copy IV from the chromosome of *S.typhimurium* (46); the partial sequences obtained for an IS*200* element from *Salmonella* strain SARA17 (16) and for IS*200* elements from *E.coli* strains ECOR8, ECOR51, ECOR63 and ECOR72 (16). Relevant data are as follows.

(i) IS*200* copy IV is 710 bp (46), that is, 2 bp longer than IS*200*-SAO and IS*200*-VP and 3 bp longer than IS*200*-HIS. The IS*200* element from *Y.pestis* is 709 bp (4). The length of the IS*200* elements from strains SARA17, ECOR8, ECOR51, ECOR63 and ECOR72 is not known (16).

(ii) Only 1 bp change is found in the 50 bp of the left end of IS*200* copy IV, with respect to IS*200*-HIS and IS*200*-VP. Likewise, 1 bp change is found in the right end. The ends of the IS*200* element from *Y.pestis* are also extremely similar to those of IS*200*-HIS, IS*200*-SAO and IS*200*-VP: only 2 bp differences from each of the IS*200*-HIS ends; 3 and 2 bp, respectively, with the right and left ends of IS*200*-VP; 4 and 1 bp differences with the ends of IS*200*-SAO. The ends of the IS*200* elements found in strains SARA17, ECOR8, ECOR51, ECOR63 and ECOR72 have not been sequenced (3,16).

(iii) The IS*200* element from *Y.pestis* contains an ORF that potentially encodes a peptide of 151 amino acids, like IS*200*-SAO, IS*200*-VP and IS*200*-HIS. However the ORF from

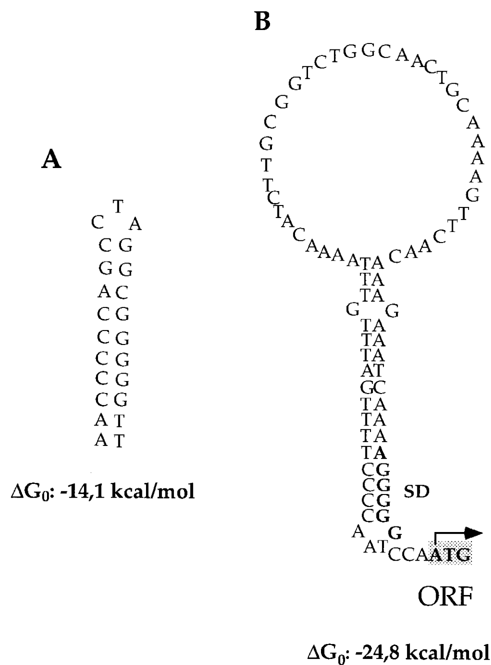


**Figure 2.** (**A**) Predicted structure of the highly conserved RNA hairpin located near the left end of IS*200*. Base pair numbers are for IS*200*-HIS. (**B**) Predicted structure of the stem–loop RNA structure that precedes the IS*200* ORF. 'SD' indicates the position of the putative ribosome-binding-site, whose bases are shown in boldface. The position of the start codon of the ORF is also shown.

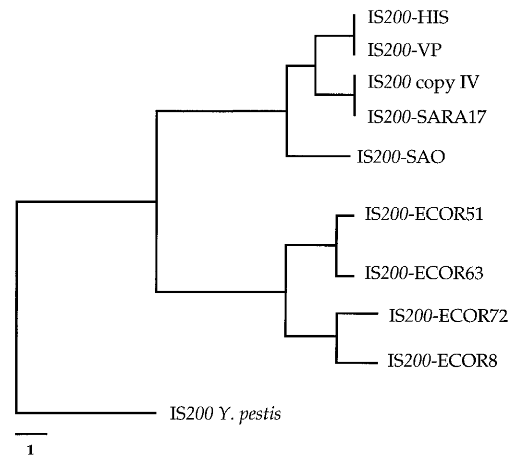


**Figure 3.** Evolutionary tree of IS*200* elements from *Salmonella, E.coli* and *Yersinia*, constructed using the predicted amino acid sequences of the peptides encoded by the IS*200* ORF. A key for the names and acronyms used to designate IS*200* elements is given in the legend of Figure 1. The scale bar corresponds to one substitution per 100 amino acids.

the *Y.pestis* IS*200* element shows 14 amino acid differences with the ORFs of IS*200*-HIS and IS*200*-VP (which are identical). The number of amino acid differences between IS*200* SARA17 and the pair IS*200*-HIS / IS*200*-VP is only five. In fact, four of these changes affect the last five amino acids of the predicted peptide and are caused by a single –1 frameshift in IS*200* SARA 17. The ORF of IS*200* copy IV is identical to that of IS*200* SARA 17.

(iv) The ORFs of the *E.coli* IS*200* elements (3) encode potential peptides of either 152 amino acids (ECOR8, ECOR51 and ECOR63) or 142 amino acids (ECOR72). The number of amino acid differences between these elements and the pair IS*200*-HIS/IS*200*-VP ranges from 9 to 15.

(v) Except when frameshifts are involved, the differences in amino acid composition of the ORF underestimate the nucleotide variation among IS*200* elements. For instance, the following nucleotide differences are found between the IS200-HIS ORF region and the ORFs of other elements: IS*200*-VP, 1 bp; IS*200* SARA17, 4 bp; IS*200*-SAO, 10 bp; IS*200* ECOR51, 48 bp; and IS*200* ECOR72, 49 bp; IS*200* ECOR 63, 52 bp; IS*200* ECOR8, 53 bp; IS*200* from *Y.pestis*, 75 bp.

Because the ORF is the only region sequenced in all IS*200* elements characterized from different hosts, we used the differences in the composition of their predicted peptides to obtain the phylogenetic tree shown in Figure 3. This tree reproduces the phylogeny of the bacterial species used as sources of IS*200* (*Salmonella*, *E.coli* and *Yersinia*: 14,15) and also reflects the relatedness among *Salmonella* species or serovars (47). These results are therefore supportive of the model that IS*200* is an ancestral element of the Enterobacteriaceae (16).

Outside the ORF, extremely high sequence conservation is also found in the hairpin and stem–loop regions of IS*200*: for instance no differences are found in these DNA tracts between IS*200* copy IV (46) and the three elements whose sequence is reported in this work (IS*200*-SAO, IS*200*-VP and IS*200*-HIS). The putative ribosome-binding-site is also 100% conserved in all IS*200* elements, irrespective of their origin.

### Characterization of IS*200* insertion sites

The DNA sequences flanking the IS*200* element inserted at *hisD* show two A–T pairs at each side of the element (Fig. 4), thus confirming that insertion of IS*200* at this site caused a 2 bp duplication, as previously reported (40). In turn, a databank search for sequences homologous to the regions which flank the IS*200* insertion in pSLT indicated that IS*200* had inserted into the fimbrial operon *pef* (48). An extra A–T pair is found at the IS*200* insertion site, suggesting that IS*200* insertion at this site has generated a 1 bp duplication.

The IS*200* element from the *S.abortusovis* chromosome is flanked by two A–T pairs at one side and eight A–T pairs at the other side (Fig. 4). Because an IS*200*-free version of this locus is not available, the occurrence of a duplication cannot be confirmed. However, the presence of A–T pairs at both sides suggests that duplication of either 1 or 2 bp may have occurred.

Support for the hypothesis that IS*200* may duplicate 1 or 2 bp at the insertion site is provided by additional data collected from the literature (4,8) and from databases (46). These data, included in Figure 4, are as follows.

(i) The IS*200* element inserted in the *gpt* gene is flanked by three A–T pairs at one side and four A–T pairs at the other side (8). Thus duplication of 1–3 bp may have occurred. The existence of a duplication cannot be confirmed because the wild-type *gpt* sequence is not available.

(ii) The IS*200* element inserted in the *invA* gene of *Y.pestis* is flanked by four A–T pairs at one side and two A–T pairs at the other side (4); thus duplication of 1–2 bp may have occurred. The wild-type sequence of the *invA* locus is not available. However,

| | | | | | |
|---|---|---|---|---|---|
| IS*200*-HIS | GCGCGTTTTT | GTC | IS*200* | TTA | TTACATTATC |
| IS*200*-VP | AGTATATAAT | GTC | IS*200* | TTA | TTCCTCTTCA |
| IS*200*-SAO | TTTTTTTTTT | GTC | IS*200* | TTA | TTCATTGCCG |
| IS200 copy IV | CCCGTTTATT | GTC | IS*200* | TTA | TTCTATCTGG |
| IS200 *gpt* | TGATGATTTT | GTC | IS*200* | TTA | TTTATGTTAT |
| IS200 *Y. pestis* | TTGCTTTATT | GTC | IS*200* | TTA | TTTTGGTAAA |

**Figure 4.** Junctions between IS*200* elements and host sequences. The three upper examples are from elements sequenced in this study; sequences for the other IS*200* elements have been retrieved from the literature (4,8) or from the EMBL databank (46).

*invA* loci have been sequenced in *Y.enterocolitica* (49,50) and *Y.pseudotuberculosa* (51). These loci are homologous to the *Y.pestis invA* gene (77 and 94%, respectively). When the comparison is carried out for the 30 bp that flank the insertion site, the homology is 100% and the only difference is the presence of two extra A–T pairs at the insertion site. Thus a 2 bp duplication must have occurred.

Alignment of IS*200* insertion sites does not disclose any significant homologies; however, all the IS*200* insertion sites appear to be A+T-rich: in the neighboring 13–25 bp, the percentages of A+T base pairs range between 76 and 95%. Thus we tentatively suggest that IS*200* may have regional specificity for A+T-rich DNA tracts, like other insertion elements (52–54).

## DISCUSSION

Sequence comparisons suggest that IS*200* was already present in the common ancestor of *E.coli*, *Salmonella* and *Yersinia*, a hypothesis that fulfills the predictions of Bisercic and Ochman (16). The recent discovery of an IS*200*-like element in *Vibrio* (41), a genus unrelated to enteric bacteria, might indicate either the existence of IS*200* in remote stages of bacterial evolution or the occurrence of horizontal transfer. The latter is regarded as a rare event among bacteria (55), but insertion sequences are potential exceptions (56). In the case discussed here, the high divergence between IS*200* elements from enterics and the *V.cholerae* IS*200*-like element seems to rule out recent horizontal transfer (e.g., the predicted amino acid sequence of the ORF of the *V.cholerae* IS*200* element shows only a 40% identity with IS*200*-HIS). Thus IS*200* may be an extremely old transposable element.

The IS*200* elements characterized in this study are 707–708 bp, while *S.typhimurium* copy IV is 710 bp (46) and the *Y.pestis* IS*200* element is 709 bp (4). A size of 707–710 bp makes IS*200* one of the smallest insertion elements found in bacteria (12). Another relevant feature, the lack of terminal repeats, is rare among prokaryotic insertion elements (12), albeit with well-known exceptions like bacteriophage Mu (57).

All IS*200* insertion sites show a high A+T content, but a consensus target site cannot be deduced from the insertion sites available. If the number of targets studied is significant, the preference of IS*200* for A+T-rich regions may correlate the behavior of IS*200* with that of elements that show regional specificity but not defined targets: for instance, IS*1* (52–54), IS*21* (58), IS*50* (59,60) and IS*186* (61,62). Regional specificity may be taken as evidence that the transposase of the element recognizes a topological DNA structure, rather than a defined target (63,64).

IS*200* generates duplications of 1–2 bp at the insertion site (Fig. 4), although the data available do not exclude the possibility of longer duplications (e.g., 3 bp). If confirmed, the formation of duplications of variable length would establish again a parallelism between IS*200* and IS*1* (54,65), IS*21* (58) and IS*186* (61,62).

The scarcity of IS*200*-induced mutations and the lack of reliable transposition assays raised the question of whether the *Salmonella* genome might contain inactive (i.e., degenerate) IS*200* copies (9). However, this study seems to rule out this possibility: IS*200* 'hops' and IS*200* elements residing in the *Salmonella* genome are similar or identical; thus most, perhaps all *Salmonella* IS*200* elements can be expected to be active. If such is the case, the extremely low frequency of IS*200* transposition (9,10) must reflect the normal behavior of the element. This hypothesis is supported by the observation that natural isolates of *Salmonella* show stable S*200* fingerprints (11).

Although the data available are largely descriptive, certain structural features of IS*200* suggest hypothetical mechanisms that might reduce the activity of the element. The putative transposase gene appears to be transcribed at extremely low rates (10). In addition, the hairpin located near the left end of IS*200* might prevent transcription from external promoters and the stem–loop that occludes the ribosome-binding-site of the putative transposase gene might keep translation low. Although the existence of these multiple controls awaits experimental confirmation, it provides a model to explain the paradoxical behavior of IS*200*, an insertion element that rarely 'hops'. A low frequency of transposition can be viewed as a self-restraint mechanism (66); thus the low activity of IS*200* might have actually favored its evolutionary persistence.

## ACKNOWLEDGEMENTS

This study was supported by grant PB93/649 from the Dirección General de Investigación Científica y Técnica (DGICYT) of

Spain. Additional funding was provided by the Regional Government of Andalusia (Junta de Andalucía). We are grateful to Angela Schiaffino for her contribution to the cloning of IS*200*-SAO, to Javier Ruiz Albert for advice about DNA sequencing, to Nicolás Prados for help with computer programs and to Gabriel Gutiérrez for advice on phylogenetic trees. We thank John Roth, Ken Haack, Fernando Govantes and Eduardo Santero for helpful discussions. We also acknowledge the efficient assistance received from José Córdoba and Luis Romanco.

## REFERENCES


1 Lam, S. T. and Roth J. R. (1983) *Cell*, **34**, 951–960.
2 Gibert, I., Barbé, J. and Casadesús, J. (1990) *J. Gen. Microbiol.*, **136,** 2555–2560.
3 Bisercic, M. and Ochman, H. (1993) *Genetics*, **133**, 449–454.
4 Simonet, M., Riot, B., Fortineau, N. and Berche, P. (1996) *Infect. Immun.*, **64,** 375–379.
5 Lam, S. T. and Roth, J. R. (1983) *Genetics*, **105,** 801–811.
6 Sanderson, K. E., Sciore, P., Liu, S. L. and Hessel, A. (1993) *J. Bacteriol.*, **175,** 7624–7628.
7 Casadesús, J. and Roth, J. R. (1989) *Mol. Gen. Genet.*, **216,** 210–216.
8 O'Reilly, C., Black, G. W., Laffey, R. and McConnell, D. J. (1990) *J. Bacteriol.*, **172,** 6599–6601.
9 Casadesús, J., Beuzón, C. R. and Gibert, I. (1992) *Genet. (Life Sci. Adv.)*, **11**, 179–186.
10 Beuzón, C. R. (1996) *Structural and functional analysis of insertion element IS200.* Ph. D. Thesis, University of Seville, Seville, Spain.
11 Schiaffino, A., Beuzón, C. R., Uzzau, S., Leori, G., Cappuccinelli. P., Casadesús, J. and Rubino, S. (1996) *Appl. Environ. Microbiol.*, **62,** 2375–2380.
12 Galas, D. J. and Chandler, M. (1989) In Berg, D. E. and Howe, M. M. (eds) *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 109–162.
13 Finnegan, D. J. (1989) *Trends Genet.*, **5**, 103–107.
14 Ochman, H. and Wilson, A. C. (1987) In Neidhardt. F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M. and Umbarger, H. E. (eds) *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology.* American Society for Microbiology, Washington, DC, pp. 1649–1654.
15 Lawrence, J. G., Hartl, D. L. and Ochman, H. (1991) *J. Gen. Microbiol.*, **137**, 1911–1921.
16 Bisercic, M. and Ochman, H. (1993) *J. Bacteriol.*, **175**, 7863–7868.
17 Colombo, M. M., Leori, G., Rubino, S., Barbato, A. and Cappuccinelli, P. (1992) *J. Gen. Microbiol.*, **138**, 725–731.
18 Inoue, H., Nojima, H. and Okayama, H. (1990) *Gene*, **93**, 26–28.
19 Yanisch-Perron, V., Vieira, J. and Messing, J. (1985) *Gene*, **33**, 103–119.
20 Garzón, A., Beuzón, C. R., Mahan, M. J. and Casadesús, J. (1996) *Mol. Gen. Genet.*, **250**, 570–580.
21 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning*; *A Laboratory Manual.* 2nd. edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22 Schmieger, H. (1972) *Mol. Gen. Genet.*, **119**, 75–88.
23 Tinge, S. A. and Curtiss, R. III (1990) *Infect. Immun.*, **58**, 3084–3092.
24 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1987) *Current Protocols in Molecular Biology.* Greene Publishing Associates & Wiley Interscience, New York, NY.
25 Stephen, D., Jones, C. and Schofield, P. J. (1990) *Nucleic Acids Res.*, **18**, 7463–7464.
26 Holmes, D. S. and Quigley, M. (1981) *Anal. Biochem.*, **114**, 193–197.
27 Southern, E. M. (1975) *J. Mol. Biol.*, **98**, 503–517.
28 Boyle, J. S. and Lew, A. M. (1995) *Trends Genet.*, **11**, 8.
29 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
30 Program Manual for the Wisconsin Package, Version 8 (1994).
31 Zuker, M. (1989) *Methods Enzymol.*, **180**, 262–288.
32 Swofford, D. L. (1993) PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL.
33 Sneath, P. H. A. and Sokal, R. R. (1973) *Numerical Taxonomy.* W. H. Freeman & Co, San Francisco.
34 Gibert, I., Carroll, K., Hillyard, D. R., Barbé, J. and Casadesús, J. (1991) *Nucleic Acids Res.*, **19**, 1343.
35 Winkler, M. E. (1996) In Neidhardt, F. C., Curtiss, R. III, Ingraham, J. L., Lin, A. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M. and Umbarger, H. E. (eds) *Escherichia coli and Salmonella. Cellular and Molecular Biology.* 2nd. edition. American Society for Microbiology, Washington, DC, pp. 485–505.
36 Arber, W., Iida, S., Jutte, H., Caspers, P., Meyer, J. and Hänni, C. (1978) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 1197–1208.
37 Green, L., Miller, R., Dykhuizen, D. E. and Hartl, D. L. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4500–4504.
38 Naas, T., Blot, M., Fitch, W. M. and Arber, W. (1994) *Genetics*, **136**, 721–730.
39 Anderson, R. P. and Roth, J. R. (1979) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 1083–1087.
40 Lam, S. T. and Roth, J. R. (1986) *J. Mol. Biol.*, **187**, 157–167.
41 Bik, E. M., Gouw, R. D. and Mooi, F. R. (1996) *J. Clin. Microbiol.*, **34**, 1453–1461.
42 Haack, K. R. and Roth, J. R. (1995) *Genetics*, **141**, 1245–1252.
43 Shine, J. and Dalgarno, L. (1975) *Nature*, **254**, 34–38.
44 Draper, D. E. (1996) In Neidhardt, F. C., Curtiss, R. III, Ingraham, J. L., Lin, A. C. C., Low, K. B., Magasanik, B., Reznikoff, W, S., Riley, M., Schaechter, M. and Umbarger, H. E. (eds) *Escherichia coli and Salmonella. Cellular and Molecular Biology.* 2nd. edition. American Society for Microbiology, Washington, D. C., pp. 902–908.
45 Youderian, P., Sugiono, P., Brewer, K., Higgins, N. P. and Elliott, T. (1988) *Genetics*, **118**, 581–592.
46 Burnens, A. P., Stanley, J., Hunkizer, P., Brodard, I. and Nicolet, J. (unpublished) Z54217.
47 Brenner, D. J. (1984) In Krieg. N. R. and Holt, J. G. (eds) *Bergey's Manual of Systematic Bacteriology.* Williams & Wilkins, Baltimore and London, pp. 408–430.
48 Friedrich. M. J., Kinsey, N. E., Vila, J. and Kadner, R. J. (1993) *Mol. Microbiol.*, **8**, 543–558.
49 Young, V. B., Miller, V. L., Falkow, S. and Schoolnik, G. K. (1990) *Mol. Microbiol.*, **4**, 1119–1128.
50 Pepe, J. C., Badger, J. L. and Miller, V. L. (1994) *Mol. Microbiol.*, **11**, 123–135.
51 Isberg, R. R., Voorhis, D. L. and Falkow, S. (1987) *Cell*, **50**, 769–778.
52 Meyer, J., Iida, S. and Arber, W. (1980) *Mol. Gen. Genet.*, **178,** 471–473.
53 Galas, D. J., Calos, P. and Miller, J. H. (1980) *J. Mol. Biol.*, **144,** 19–41.
54 Zerbib, D., Gamas, P., Chandler, M., Prentki, P., Bass, S. and Galas, D. (1985) *J. Mol. Biol.*, **185**, 517–524.
55 Maynard Smith, J., Dowson, C. G. and Spratt, B. G. (1991) *Nature*, **349**, 29–31.
56 Lawrence, J. G., Ochman, H. and Hartl, D. L. (1992) *Genetics*, **131**, 9–20.
57 Chaconas, G. (1987) In Symonds, N., Toussaint, A., van de Putte, P. and Howe, M. M. *Bacteriophage Mu.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 137–157.
58 Reimmann, C., Moore, R., Little, S., Savioz, A., Willetts, N. S. and Haas, D. (1989) *Mol. Gen. Genet.*, **215**, 416–424.
59 Berg, D. E., Schmandt, M. A. and Lowe, J. B. (1983) *Genetics*, **105**, 813–828.
60 Lupski, J. R., Gershon, P., Osaki, L. S. and Godson, G. N. (1984) *Gene*, **30**, 99–106.
61 Kothary, R. K., Jones, D. and Candido, P. E. (1985) *J. Bacteriol.*, **164**, 957–959.
62 Sengstag, C., Iida, S., Hiestand-Nauer, R. and Arber, W. (1986) *Gene*, **49**, 153–156.
63 Stellwagen, N. C. (1983) *Biochemistry*, **22**, 6186–6193.
64 Gamas, P., Chandler, M. G., Prenkti, P. and Galas, D. J. (1987) *J. Mol. Biol.*, **195**, 261–272.
65 Iida, S., Hiestand-Nauer, R. and Arber, W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 839–843.
66 Doolittle, W. F., Kirkwood, T. B. L. and Dempster, M. A. H. (1984) *Nature*, **307**, 501–502.