JOURNAL OF BACTERIOLOGY, July 1990, p. 4090–4099
0021-9193/90/074090-10$02.00/0

Vol. 172, No. 7

Transposition in *Shigella dysenteriae*: Isolation and Analysis of IS*911*, a New Member of the IS*3* Group of Insertion Sequences

M.-F. PRÈRE, M. CHANDLER, AND O. FAYET*

Centre de Biochimie et Génétique Cellulaires du Centre National de la Recherche Scientifique, 118 Route de Narbonne, 31062 Toulouse Cedex, France

Received 5 February 1990/Accepted 25 April 1990

Twenty-nine clear-plaque mutants of bacteriophage λ were isolated from a *Shigella dysenteriae* lysogen. Three were associated with insertions in the *c*I gene: two were due to insertion of IS*600*, and the third resulted from insertion of a new element, IS*911*. IS*911* is 1,250 base pairs (bp) long, carries 27-bp imperfect terminal inverted repeats, and generates 3-bp duplications of the target DNA on insertion. It was found in various copy numbers in all four species of *Shigella* tested and in *Escherichia coli* K-12 but not in *E. coli* W. Analysis of IS*911*-mediated cointegrate molecules indicated that the majority were generated without duplication of IS*911*. They appeared to result from direct insertion via one end of the element and the neighboring region of DNA, which resembles a terminal inverted repeat of IS*911*. Nucleotide sequence analysis revealed that IS*911* carries two consecutive open reading frames which code for potential proteins showing similarities to those of the IS*3* group of elements.

In contrast to many strains of *Escherichia coli*, the genome of members of the closely related genus *Shigella* carries relatively large numbers of a variety of insertion sequences. It has been estimated, for example, that the four *Shigella* species carry 40 to 200 copies of insertion sequences closely related to IS*1* (27). This elevated number of IS-like sequences is not restricted to IS*1*; it also includes IS*2* and IS*4* (15; Fayet and Prère, unpublished observations; see reference 11 for review).

In order to assess transposition activity from the *Shigella dysenteriae* genome, we used a simple technique described by Lieb (16) which employs bacteriophage λ *c*I *ind* as a trap for insertion sequences. By screening for clear-plaque mutants among phages released spontaneously from an *E. coli* lysogen, it was determined that of 25 mutants, 15 were the result of insertion of IS elements into the *c*I gene; of these, 7 insertions were of IS*1*, 5 were of IS*5*, and 3 were of IS*3*.

Since some strains of *S. dysenteriae* are sensitive to lambdoid phages, we used this technique to obtain some indication of the relative activity of the various IS elements carried by the *S. dysenteriae* chromosome. We found that the spectrum of *c*I mutants obtained in *S. dysenteriae* is very different from that observed in *E. coli*. No example of an IS*1* insertion was found, and only 3 of the 29 independent mutants were found to carry insertions in the *c*I gene. This result suggests that the chromosomal copies of IS*1* in *S. dysenteriae* may be less active than those of *E. coli*. Of the three insertions obtained, two were examples of IS*600*, an element previously isolated by physical means from the *Shigella sonnei* chromosome (21), while the third (IS*911*) proved to be an element not described previously.

We present here the nucleotide sequence of IS*911* (included in the European Molecular Biology Laboratory sequence data base under accession number X17613), assess its distribution within *E. coli* K-12, *E. coli* W, and the four species of *Shigella*, and analyze some of its transposition properties. Sequence analysis of IS*911* indicates that it is a member of the IS*3* group of elements (32, 33). This analysis also revealed that IS*3411* (12), IS*476* (14), and IS*R1* (29) also belong to this group. The disposition of open reading frames and possible control of expression of transposition functions of IS*911* and other members of the group are also discussed.

MATERIALS AND METHODS

Bacterial strains and culture conditions. Six *Shigella* strains were used. Four were clinical isolates of *S. sonnei* (OFB759), *S. boydii* (OFB760), *S. flexneri* (OFB761), and *S. dysenteriae* (OFB762). The two other strains were *S. dysenteriae* ATCC 11456 and its φ80 *imm*$^{\lambda}$ *c*I *ind* lysogenic derivative.

The *E. coli* strains are K-12 derivatives except for OFB401, which is an isolate of the original W strain (NCIB 8666). MGC182 (C600 *hsdR recA*) was used as the recipient in all cloning experiments, and 1106 (*lac met supE supF hsdS*) served for phage propagation. Mating assays were carried out between MGC182(pOX38Km)(pOFT139) as the donor and OF3 (C600 *rpsL*) as the recipient.

Bacteria were grown in LB medium (22) supplemented, when necessary, with ampicillin (50 μg/ml) plus methicillin (500 μg/ml), kanamycin (25 μg/ml), and streptomycin (200 μg/ml). Tryptone agar medium was used for phage production (1).

Lysogenization of *S. dysenteriae* and selection of IS-containing phages. The *S. dysenteriae* ATCC 11456 strain is resistant to λ but sensitive to φ80. We therefore used a hybrid phage, φ80 imm$^{\lambda}$ *c*I *ind*, to construct a lysogenic derivative by the phage-spotting procedure of Arber et al. (1). This lysogen was then used to select a series of 29 independent clear-plaque phage mutants (16). DNA from each phage was prepared by the plate lysate method of Maniatis et al. (20). Three insertion-containing phages were obtained: φ80 *imm*$^{\lambda}$ *c*I *ind*::IS*911*, φ80 *imm*$^{\lambda}$ *c*I *ind*::IS*600*-1, and φ80 *imm*$^{\lambda}$ *c*I *ind*::IS*600*-2.

DNA manipulation and Southern hybridization. Large-scale or minipreparation of plasmid DNA and all cloning procedures were done as described before (20). Restriction enzymes, *E. coli* DNA polymerase Klenow fragment, and T4 DNA ligase were purchased from Amersham, New England BioLabs, or Bethesda Research Laboratories and used as recommended.

* Corresponding author.

TABLE 1. Insertion sequences of the IS3 family[a]

IS	Reference	Length (bp)	ORFA			ORFB		
			Coordinates (bp)	Size (aa)	Other name	Coordinates (bp)	Size (aa)	Other name
IS2	31	1,331	91–454	121	ORF1	(397)–415–1317	301	ORF2
IS3	36	1,258	66–362	99	ORFII	(311)–362–1225	288	ORFI
IS51	39	1,311	51–371	108	ORF3	(323)–371–1270	300	ORF4
IS150	33	1,443	48–566	173		(511)–566–1414	283	
IS476[b]	14	1,226	65–328	88		(49)–364–1194	277	
IS600	21	1,264	65–364	100		(304)–403–1218	272	
IS861	32	1,442	176–588	141	ORF1	(481)–577–1407	277	ORF2
IS911	This study	1,250	73–372	100		(318)–402–1217	272	
IS3411[b]	12	1,310	55–378	108	ORFI	(315)–378–1222	282	ORFV
ISR1[b]	29	1,259	64–327	88	ORFA3	(249)–324–1238	305	

[a] Coordinates are expressed as base pairs from the left end of the IS. For ORFA, the first coordinate indicates the position of the first potential initiation codon (ATG or GTG). The size of the ORF is given as coding capacity in amino acids (aa). Other names refer to designations used in the indicated references. For ORFB, the first coordinate (in parentheses) corresponds to the position of the first amino acid codon of the ORF. The second gives the position of the first potential initiation codon except for IS476, where the start following ORFA and preceded by a potential ribosome-binding site was used. In that case, there are two other possible starts (at 145 and 184).

[b] For reasons presented in the Results section, the published sequences of these three insertion sequences had to be modified: one base was added to IS476 (a C after nucleotide 221) and IS3411 (a T after nucleotide 938), whereas three changes were introduced in the case of ISR1 (deletion of nucleotides 461 and 664 and addition of a T after nucleotide 916).

Southern hybridizations were performed with Hybond CE membrane (Amersham), using the protocols provided by the manufacturer. Total DNA from *E. coli* or the *Shigella* strains was prepared by the method of Fayet et al. (9). DNA fragments used as probes were purified from agarose gels with a Geneclean kit (Bio 101). Labeling was achieved by nick translation in the presence of [α-^{32}P]dATP; the labeling kit and the radioactive product were both from Amersham.

Plasmids. pOX38Km is a conjugative plasmid derived from the F factor (6), and pAT153 is a deleted derivative of the pBR322 vector (37).

Plasmid pOFT139 was constructed in several steps, briefly described below. First, a 3.64-kilobase-pair (kb) *Bgl*II fragment from ϕ80 *imm*$^{\lambda}$ *c*I *ind*::IS911 was cloned into the *Bam*HI site of pAT153. This fragment contains the *c*I region of λ (between coordinates 38013 and 35711 [7]) and IS911 (inserted at position 37437). λ DNA was then partly eliminated by using a *Hin*dIII site on one side (coordinate 37459) and by a DNase I-promoted deletion (17) on the other. The insert of pOFT139, present between the *Hin*dIII and *Bam*HI sites of the vector, is composed of a short λ DNA stretch (from the *Hin*dIII site at 37459 to position 37435) followed by IS911 (1,250 base pairs [bp]) and 197 bp of λ DNA (from coordinates 37437 to 37241). The left end of IS911, defined by analogy with other IS elements (see below), follows the shorter stretch of λ DNA.

Mating-out assay. The frequency of cointegrate formation was determined by a standard mating-out assay (10) with the conjugal plasmid pOX38Km as a target replicon. Strain MGC182(pOX38Km)(pOFT139) was crossed (in five independent experiments) with the recipient strain OF3 (Sm^r) at 37°C for 90 min. Suitable dilutions were plated on LB agar plates supplemented with kanamycin and streptomycin (conjugation frequency) and with ampicillin plus methicillin and kanamycin (cointegrate frequency).

DNA sequence determination and analysis. The sequence of both strands of the whole insert of pOFT139 was determined by the dideoxynucleotide method adapted for the modified form of the T7 DNA polymerase (35). Sequencing reactions were performed on double-stranded templates with the enzyme, reagents, and protocols from a Sequenase kit (U.S. Biochemicals). The sequencing strategy involved the construction of two sets of unidirectional DNase I-mediated deletions into the IS911 part of pOFT139 (17). These deleted plasmids were then sequenced with two pBR322 primers: in a clockwise direction from the *Hin*dIII site (5′-CATGTT TGACAGCTTATC-3′) and in a counterclockwise direction from the *Bam*HI site of the vector (5′-CGGCCACGATG CGTCCGG-3′). The latter was also used together with primers from both ends of IS911 (5′-GTGAGCATATCA CCTCTG-3′ and 5′-TGGAAAAAACTCTAACTC-3′) to determine the sequence of the cointegrate junctions.

Sequences (DNA, RNA, and proteins) were analyzed with various programs contained in the PC-gene package (Genofit, S.A.). The search for Cro-like binding domains was carried out with a program written by D. Zerbib and based on the Dodd and Egan matrix (8).

Detailed comparisons were performed within a group (the IS3 family) comprising IS911 and nine other insertion sequences which, on the basis of similarities in their primary structure (DNA and translated open reading frames [ORFs]) and overall organization, appear to be related (see Results) (32, 33). They all contain two consecutive and partly overlapping ORFs which cover most of the length of one strand. Following the nomenclature of Schwartz et al. (33), the first and shortest ORF is called ORFA, and the second, in the −1 frame relative to ORFA, is called ORFB. The size and coordinates (in base pairs from the left end of each IS) of these ORFs are given in Table 1. We defined the left end of each IS as the extremity close to the amino terminus of ORFA. For reasons discussed in the Results section, minor modifications were introduced in the published sequences of IS3411, IS476, and ISR1.

RESULTS

Isolation of insertion sequences from the *S. dysenteriae* genome. DNA from stocks of 29 independent clear-plaque mutants of phage ϕ80 *imm*$^{\lambda}$ *c*I *ind* isolated from an *S. dysenteriae* lysogen (Materials and Methods) was analyzed by digestion with *Bam*HI, since this enzyme generates a distinct DNA fragment carrying the *c*I region (data not shown). In contrast to the results of Lieb (16) with *E. coli*, only three of these mutants were found to carry DNA insertions. It is not clear whether this difference is due to a higher frequency of point mutations or to a lower transposi-

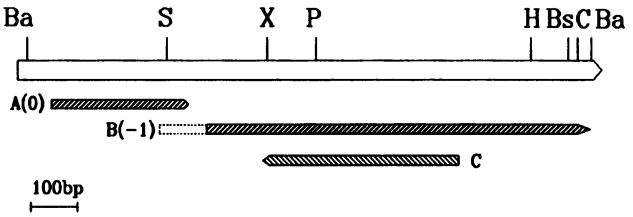

FIG. 1. Restriction map of IS*911*. The cleavage sites for *Bal*I (Ba), *Bst*EII (Bs), *Cla*I (C), *Hin*dIII (H), *Pvu*II (P), *Xma*III (X), and *Ssp*I (S) are indicated together with three ORFs, A, B, and C. Numbers in parentheses indicate the relative reading phase of the ORFA and ORFB frames. The open region attached to ORFB indicates a possible position at which translational frameshifting could occur to generate a fusion protein between ORFA and ORFB.

tion activity of IS elements in *S. dysenteriae*. Moreover, all three insertions appeared to be approximately 1.3 kb in length. Thus, in spite of the frequent occurrence of IS*1*-like sequences in the *Shigella* chromosome and the frequency at which they generate *c*I mutations in *E. coli*, none of these mutations could be attributed to insertion of IS*1* (768 bp).

Analysis of the insertions. To characterize these sequences further, a *Bgl*II fragment carrying each insertion was cloned in the *Bam*HI site of pAT153. Restriction analysis of the resulting clones (data not shown) indicated that two carried similar sequences. The nucleotide sequence of these two insertions (data not shown) identified them as copies of IS*600*, a sequence previously isolated from *S. sonnei* by physical means but not previously demonstrated to be capable of transposition in a cointegration assay (21). The two insertions occurred at λ coordinates 37376 and 37601 (7) and were accompanied by a duplication of 3 bp in the target DNA (bp 37374 to 37376 and 37599 to 37601). Our IS*600* isolates were also unable to mediate the formation of cointegrates in our standard transposition assay (Materials and Methods) with the pOX38Km conjugative plasmid (data not shown).

The restriction map of the third insertion (Fig. 1) did not correspond to that of any known IS. This was confirmed by nucleotide sequence analysis, and the IS was registered as IS*911*. Its entire nucleotide sequence, determined from both strands, is shown in Fig. 2. Insertion in this case occurred at λ coordinate 37437, generating a 3-bp duplication in the target DNA (from bp 37435 to 37437 [7]). No overall homology was detected between IS*911* and any other known IS element, although some regional homologies were observed with IS*2*, IS*3*, IS*51*, IS*150*, IS*600* (33), and IS*861* (32).

Distribution of IS*911* in *Shigella* and *E. coli* strains. We investigated the distribution of IS*911* in the original *S. dysenteriae* (ATCC 11546) strain, in other *Shigella* species, and in *E. coli* by Southern analysis with a *Hin*dIII fragment of pOFT139 (Materials and Methods), which carries 1,091 bp of IS*911* and 25 bp of λ, as the probe. Total DNA was digested with *Pst*I and *Bgl*II, two enzymes which do not cleave the original IS*911* sequence. Bands with homology to IS*911* were observed in all four *Shigella* species (Fig. 3, lanes 1 to 5) and in *E. coli* K-12 (lanes 6 and 7) but not in *E. coli* W (lane 8). While *S. sonnei* (lane 1) and *E. coli* K-12 showed relatively few bands, the other *Shigella* strains showed many bands. Keeping in mind that these experiments do not distinguish between copies carried by resident plasmids and chromosomal copies and do not take into account the possibility that certain ("iso") copies may carry cleavage sites for the enzymes *Pst*I and *Bgl*II, we estimate that *S. boydii*, *S. flexneri*, and *S. dysenteriae* could carry between 10 and 20 copies each. It is also interesting that not only did the four *Shigella* species show significant differences in hybridization pattern, but, more surprisingly, clear differences were also found between the two *S. dysenteriae* strains (lanes 4 and 5).

Transposition activity of IS*911*. We determined the ability of IS*911* to generate replicon fusions. Plasmid pOFT139, a pAT153 derivative carrying a cloned copy of IS*911* together with flanking λ DNA, was used as the donor plasmid in a standard mating-out assay with the conjugal plasmid pOX38Km as the recipient replicon (Materials and Methods). Transfer of the pOFT139 Ap^r marker, a measure of the frequency of cointegrate formation, occurred at 1.0×10^{-5} per pOX38Km exconjugant.

To determine the structure of the cointegrate molecules, plasmid DNA from 16 independent Ap^r Km^r exconjugants was analyzed by digestion with a suitable restriction enzyme and by Southern hybridization. Plasmid DNA was digested with *Bam*HI prior to electrophoresis and transferred to a nitrocellulose membrane. This enzyme cleaves once within pOFT139 and produces three visible fragments from pOX38Km and therefore can be used to separate the junctions between donor and recipient replicons in the cointegrates. Since donor and recipient replicons should be separated by a single copy of the element at each junction (10, 11), the sum of the junction fragments should be equivalent to the sum of the initial pOX38Km fragment, pOFT139 (4.7 kb), and an additional copy of IS*911* (1.2 kb). The pattern of digestion is shown in Fig. 4a. In all cointegrates except nos. 6 and 8 (in which the insertion appeared to have taken place in the large pOX38Km fragment), insertion occurred within the 11.9-kb pOX38Km fragment. In no. 14, the sum of the sizes of the two junction fragments (17.8 kb) was equivalent to that of the original 11.9-kb pOX38Km fragment together with a copy of pOFT139 and a second copy of IS*911*, as expected for a cointegrate molecule. Moreover, a fragment of the size of pOFT139 was also present at relatively high levels, suggesting that recombination could be occurring between directly repeated IS*911* elements in the cointegrate to regenerate the original donor plasmid. The sum of the sizes of the junction fragments of the remaining plasmids (16.6 kb), however, suggests that insertion of pOFT139 into pOX38Km occurred without duplication of the IS*911* element.

Due to the position of cleavage within pOFT139 (between the pAT153 sequences and one end of the IS*911*-carrying λ fragment), only one of the junction fragments should carry DNA homologous to pAT153. This was confirmed by the results of hybridization with a pAT153 probe (Fig. 4b). For cointegrates in which two copies of the transposable element flanked the insertion, both junction fragments should hybridize with a labeled IS911 probe. This was the case for cointegrate 14 (Fig. 4c). However, for the remaining 15 cointegrates, only one fragment showing homology with IS*911* was detected, demonstrating that only a single copy was present. This was the same fragment that exhibited homology with the pAT153 probe.

One possible mechanism for the formation of these IS*911*-promoted cointegrates involves nonreplicative integration via a single end of the element (Fig. 5a). To assess this model, we determined the nucleotide sequence at the junctions of six of the cointegrates with the oligonucleotides shown in Fig. 5a. The results are shown in Fig. 5b and can be summarized as follows. In each case, the right end of IS*911* (IRR) formed a new junction with sequences of the pOX38Km target, while the left end remained associated with the original λ sequences. The second junction between

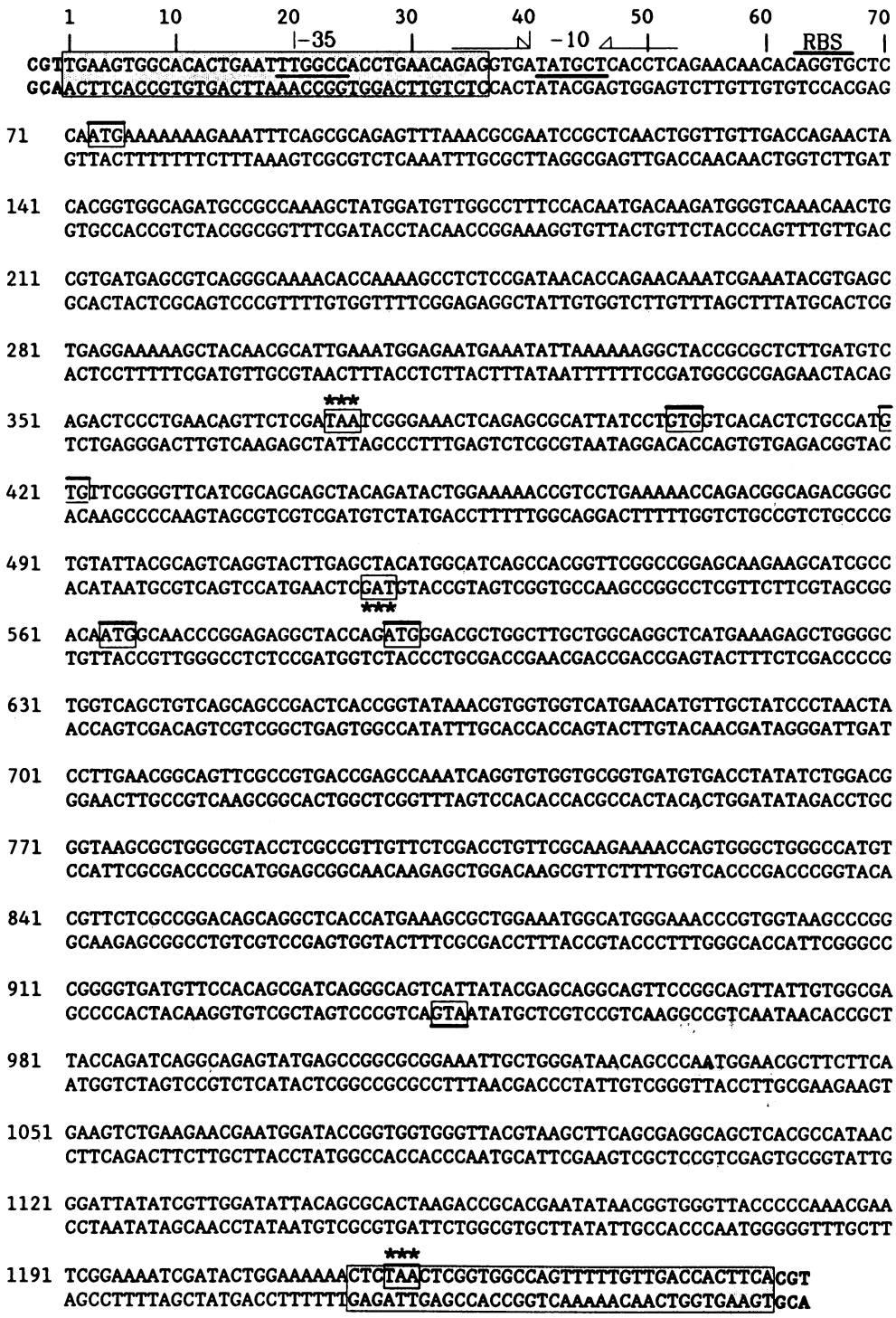

FIG. 2. Nucleotide sequence of IS*911*. Boldface letters indicate 3-bp duplication generated in the phage λ target DNA. Shaded boxes, Imperfect terminal inverted repeats; boxed and overlined letters, potential initiation codons; boxed letters indicated by ***, termination codons; overlined sequence labeled RBS, potential ribosome-binding site. The −35 and −10 regions of a potential promoter are also shown. This sequence will appear in the European Molecular Biology Laboratory, GenBank, and DNA Data Bank of Japan nucleotide sequence data bases under accession number X17613.

pOFT139 and pOX38Km was formed by using the λ sequence originally attached to IRR. The 3 bp (XXX in the figure) which formed the direct target repeat in the original insertion were no longer present at this junction but were replaced by 3 bp of pOX38Km, also repeated at the other junction (*** in the figure). Thus, like the original transposition event into λ, this unusual type of insertion also generated a 3-bp direct repeat in the target DNA. Cointegrate formation in these cases therefore appeared to result from direct insertion of pOFT139 via the IRR and proximal

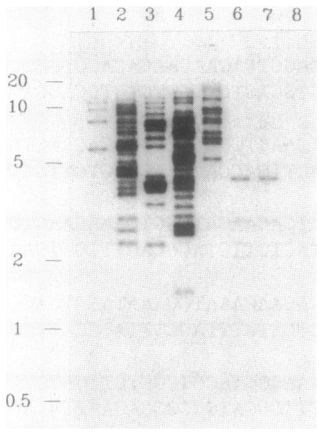

FIG. 3. Distribution of IS*911* among various bacterial strains. Total DNA from *S. sonnei* (lane 1), *S. boydii* (lane 2), *S. flexneri* (lane 3), *S. dysenteriae* (lane 4), *S. dysenteriae* ATCC 11546 (lane 5), *E. coli* 1106 (lane 6), *E. coli* OF3 (lane 7), and *E. coli* W (lane 8) was digested with *Pst*I and *Bgl*II, separated on an 0.8% agarose gel, transferred to a nitrocellulose membrane, and hybridized with a labeled IS*911* fragment (Materials and Methods). Sizes are shown in kilobase pairs.

λ DNA. It is interesting that the λ sequence (TGACCCT GAGCAGGCTGTTGAGCCAGGTGA...) involved in the formation of this junction showed some homology with the right and left ends of IS*911* (TGAagtgGtcaAcaaaaactgGC CAccgag... and TGAagtgGcaCActgaaTTtgGCCAccTGA..., respectively, where lowercase letters indicate bases which differ from the above λ sequence), raising the possibility that it is recognized by the transposition enzymes.

Features of the sequence of IS*911*. IS*911* is 1,250 bp long and carries imperfect terminal inverted repeats of 27 bp (Fig. 2). Computer analysis of the sequence revealed a potential promoter partially located in the left terminal repeat in which the −10 region is flanked by a 7-bp inverted repeat. Several relatively large reading frames of greater than 90 amino acids were detected. These are depicted schematically in Fig. 1, and their exact positions are shown in Fig. 2. ORFA is located between coordinates 73 (ATG) and 372, terminates with a TAA, and would specify a basic protein of 11,524 daltons (pI, 10.57). It is preceded by a potential ribosome-binding site. The second, ORFB, occurs in frame −1 with respect to ORFA. It ends at coordinate 1217, just within the terminal inverted repeat, and terminates with TAA. This reading frame, starting at position 318, exhibits several possible initiation codons: GTG at 402, GTG at 420, ATG at 564, and ATG at 588. In view of the results of the sequence comparisons presented below, we chose the GTG at 402 as the initiation codon. This ORF would specify a protein of 31,620 daltons (pI, 10.4).

A single ORF with a coding capacity of greater than 90 amino acids was detected on the complementary strand between positions 944 and 519. ORFC could encode a protein of 15,562 daltons (pI, 6.67) and terminates with TAG.

IS*911* is a member of the IS*3* group of elements. Certain features suggest that IS*911* is related to the IS*3* group of elements, which includes IS*2*, IS*3*, IS*51*, IS*150*, IS*600*, IS*629* (33), and IS*861* (32). Three additional elements, IS*3411* (12), IS*R1* (29), and IS*476* (14), are also members of this group (see below). They exhibit a limited degree of similarity at the nucleotide sequence level (33) (data not shown). The terminal inverted repeats of these elements (Fig. 6a) shows some similarity. In addition to the TG(A) located at the end, most of them (but not IS*911*) presents a cluster of C residues followed by an A+T-rich region. Moreover, it has been pointed out that members of this family exhibit two ORFs (corresponding to ORFA and ORFB) disposed similarly on the element (11, 33) (Table 1) and that the downstream reading frame shows significant blocks of similarity among members of the group (33). This is also true in the case of IS*911*. A comparison of its ORFB with the corresponding frame of IS*3* (30.5% identity, 14.3% similarity), IS*600* (34.2% identity, 14.0% similarity), and IS*150* (28.3% identity, 14.7% similarity) by the procedure of Myers and Miller (25) is shown in Fig. 6b. All pairwise comparisons between ORFA and between ORFB of the members of this group were obtained by the method of Needleman and Wunsch (26) (data not shown). In most cases the level of similarity is significant for both the A and B frames. However, ORFB is the better conserved of the two, especially in its carboxy-terminal region.

Interestingly, as in the case of the unrelated element IS*1*, the smaller upstream ORF of all the members of this group carries a sequence which could potentially form an α-helix–turn–α-helix motif typical of several DNA-binding proteins which function as repressors (28), although in two cases, the presence of a proline would destroy the integrity of the first α-helix (3). The results of an analysis by the technique of Dodd and Egan (8) are shown in Fig. 6c.

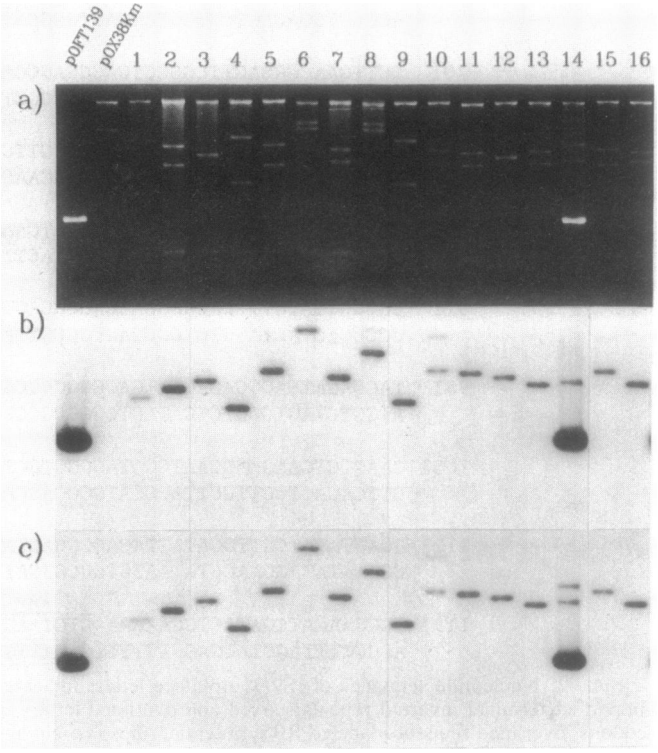

FIG. 4. Analysis of 16 IS*911*-mediated cointegrates. (a) Agarose gel electrophoresis of *Bam*HI-digested cointegrates. The first two lanes contain *Bam*HI-digested pOFT139 (4.7 kb) and pOX38Km (3.5, 11.9, and 40.4 kb), respectively. Lanes 1 to 16 each contained a different cointegrate. (b) Southern hybridization with pAT153 as a probe. (c) Southern hybridization with IS*911* as a probe (1,206-bp *Ball* fragment; see Fig. 1). for cointegrates 2, 8, 10, and 15, the faint upper band seen in panels b and c corresponds to a partial digestion product.

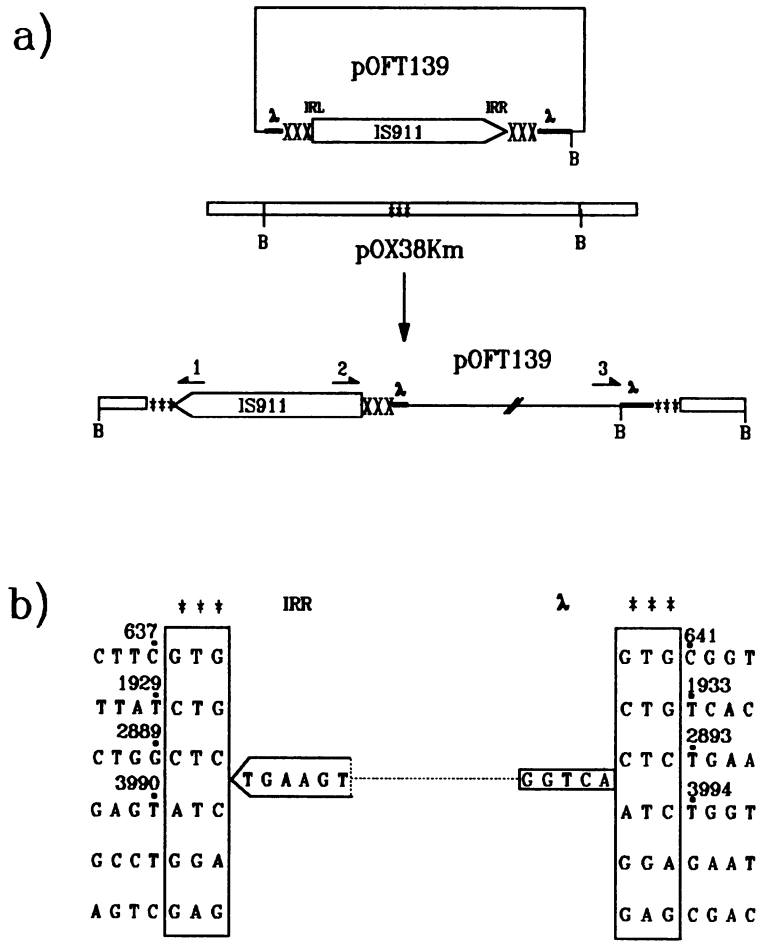

FIG. 5. IS*911* cointegrate formation. (a) A model for IS*911*-mediated cointegrate formation between pOFT139 and pOX38Km. IRL and IRR, Left and right IS*911* terminal inverted repeats, respectively; ═, phage λ target DNA; XXX, 3-bp duplication of the original target; B, *Bam*HI restriction sites; ***, duplicated 3 bp in the pOX38Km target DNA; →, oligonucleotides used as sequencing primers in the analysis of cointegrate molecules (Materials and Methods). (b) Nucleotide sequence of the cointegrate junctions. The junction between pOX38Km and the right end of IS*911* (IRR) and between pOX38Km and λ DNA for six pOFT139::pOX38Km cointegrates is shown; 3-bp pOX38Km target duplications (***) are boxed. The IS*911*-mediated insertion in the first four cointegrates occurred within a region of pOX38Km of known sequence. The coordinates indicated were calculated from the overlapping mini-F sequences determined by Morutsu et al. (24) for positions 1 to 2248 and by Mori et al. (23) for positions 1776 to 5558. Insertion in the last two cases occurred in a region of F for which the nucleotide sequence is not available.

Emendation of the nucleotide sequences of IS*3411*, IS*476*, and IS*R1*. Comparison of the published nucleotide sequences of IS*3411* (12), IS*476* (14), and IS*R1* (29) with those of various members of the IS*3* family indicated that these elements are related. IS*476* and IS*R1* are similar at the nucleotide sequence level, as are IS*3411* and IS*51* (data not shown). The ORFs of these elements, however, showed significant divergence from other members of the group. This divergence occurred abruptly in the carboxy terminus of ORFA (IS*476*) and ORFB (IS*3411* and IS*R1*). The similarity could, however, be extended over the entire protein sequence for IS*476* and IS*3411* by the addition of a single base pair. In the case of IS*R1*, three changes were required to generate a suitable ORFB. Resequencing of these elements revealed the following modifications to the published sequences: insertion of T following nucleotide 938 in IS*3411* (Ishiguro, personal communication); insertion of C following nucleotide 221 in IS*476* (Kearney, personal communication); and deletion of nucleotides 461 (G) and 664 (C) and addition of T following nucleotide 916 in IS*R1* (Kalinowski, personal communication).

DISCUSSION

In the experiments reported here, we have isolated and characterized a previously unknown insertion sequence. This sequence, IS*911*, is 1,250 bp in length and generates a 3-bp duplication in the target DNA on insertion. Southern analysis indicated that it is present in *E. coli* K-12 and all four species of *Shigella* but not in *E. coli* W. In the case of *S. boydii*, *S. flexneri*, and *S. dysenteriae*, it is present at 10 to 20 copies per cell.

The results of an analysis of the transposition properties of IS*911* indicated that it is capable of generating cointegrate molecules between the pAT153 derivative pOFT139 (into which it had been cloned with flanking λ DNA) and the conjugal pOX38Km plasmid at a frequency of 1.0×10^{-5}. Southern hybridization revealed, however, that only 1 of the 16 cointegrates analyzed carried a copy of IS*911* at each junction between the donor and recipient replicons, characteristic of most IS-mediated cointegrates. The remaining 15 examples carried only a single copy (at one of the two junctions). This suggests that transposition may not occur by

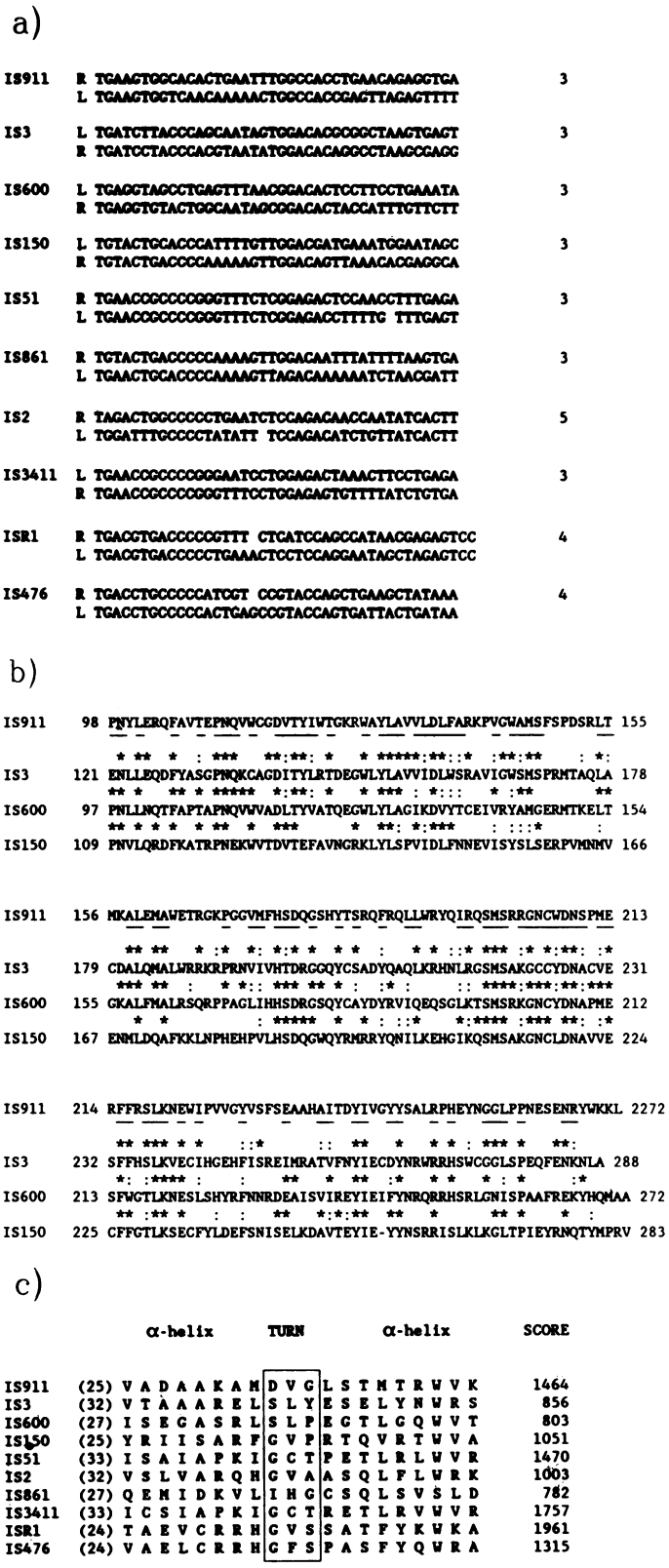

FIG. 6. Sequence comparisons among members of the IS*3* family. (a) Terminal inverted repeats. L and R indicate the left and right extremities, respectively, as defined in Materials and Methods. Specific references for the sequence of each member of the IS*3* group are given in Table 1. Numbers to the right of the figure indicate the size (in base pairs) of the target duplication induced by each insertion element. (b) Comparison of ORFB of IS*911* with the equivalent reading frames of IS*3*, IS*600*, and IS*150* by the method of Myers and Miller (25). Identical matches (*) with IS*911* and similar amino acids (:) are indicated. Positions underlined on the IS*911* sequence indicate amino acids which occur in at least two of the three other sequences. (c) Potential α-helix–turn–α-helix motif in ORFA. The position in the protein of the first amino acid is given in parentheses. The score of each motif was calculated with the weight matrix of Dodd and Egan (8).

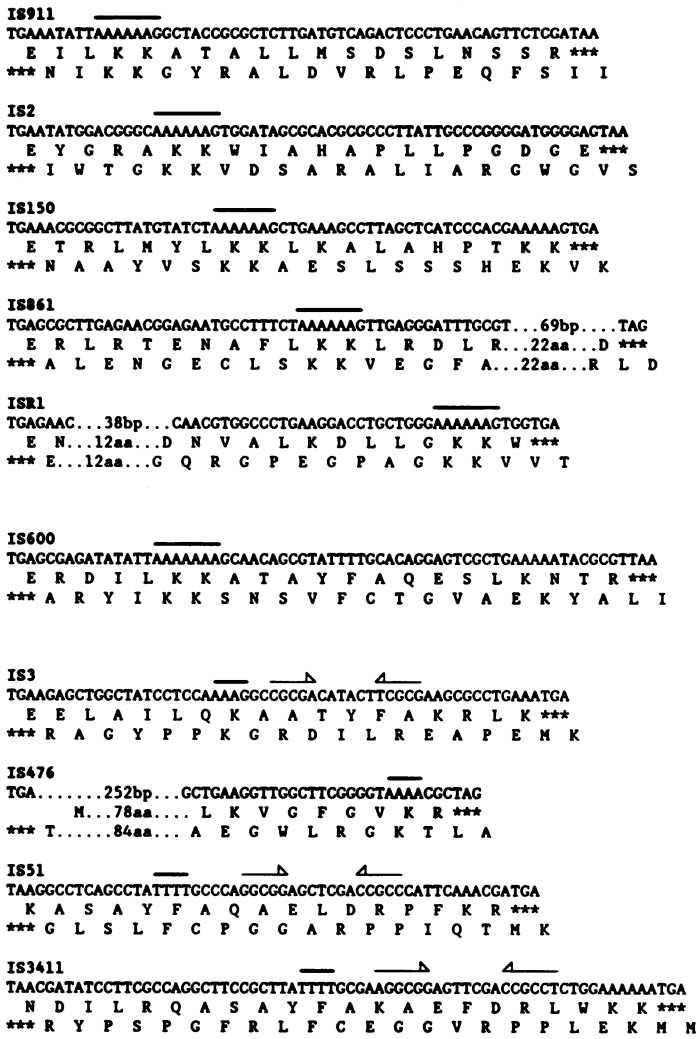

FIG. 7. Overlap region where ribosomal rephasing could occur between ORFA and ORFB in the different members of the IS*3* family. For each IS, the first line shows the nucleotide sequence from the stop codon (***) preceding ORFB to the stop codon at the end of ORFA. The second and third lines give the amino acid (aa) sequences of ORFA and ORFB, respectively. Potential ribosomal frameshift motifs are overlined. In three cases, inverted repeats are marked by arrows. For the other ISs, potential stem-loop structures (involving larger regions and therefore not indicated) can also be found downstream of the putative frameshift sites.

a replicative process and, as has been suggested for transposon Tn*5*-mediated cointegrates (2), that the single "true" cointegrate may have occurred by direct transposition from a plasmid dimer. The nucleotide sequence of the junctions of six cointegrates showed that one new junction had been formed between IRR and pOX38Km sequences. The second junction was formed between pOX38Km and λ DNA located 3 bp (the original target duplication) from IRR in pOFT139. Insertion resulted in loss of the intervening 3 bp and generation of a 3-bp duplication of the pOX38Km target DNA. The λ sequence shows some resemblance, in inverted orientation, to the terminal repeats of IS*911* and is presumably recognized by the cognate transposase. Similar types of cointegrate have been obtained with an unrelated element, IS*21*. In this case, however, fusion involves the neighboring ends of a tandem pair which are separated by 3 bp. Here too, cointegrate formation is accompanied by loss of the intervening 3 bp (30). The use of external DNA of sequence similar to the terminal inverted repeats of other elements has also been reported (19).

Sequence comparisons with other IS elements strongly suggest that IS*911* is a member of the IS*3* family of elements (33), which also includes IS*861*, isolated from a type B streptococcus (32), IS*3411* (12), IS*476* (14), and IS*R1* (29). Its terminal inverted repeats show significant similarity with other members of this group. Like the majority of these elements, it generates 3-bp duplications in the target DNA on insertion. More strikingly, like all the other IS elements of this family, it exhibits a small ORF (ORFA) followed by a longer ORF (ORFB) in the −1 frame. ORFB is the more conserved of the two within the group, and it has been proposed that it could carry the transposition-associated nicking-joining functions (33). Interestingly, a potential α-helix–turn–α-helix motif can be found within most of the ORFA but not within the ORFB putative proteins. This suggests that the former may bind to the cognate element in a sequence-specific manner, presumably at the terminal inverted repeats. Consistent with this possibility and with the differences in nucleotide sequence of the ends, the amino acid sequence within this motif is quite different from element to element.

On the basis of a comparative analysis and of preliminary

data on the proteins expressed by IS*911*, we propose that the actual transposase of at least six members of the IS*3* group is an ORFA-ORFB fusion protein generated by a translational frameshift. Such an event, if the above assumptions about ORFA and ORFB are correct, would bring together in a single polypeptide a function for recognition of the ends of the IS and the essential catalytic transposition functions. The evidence is as follows.

In all the elements considered above, ORFB lies in frame −1 with respect to ORFA. A similar arrangement is seen with the unrelated element IS*1*. In this case the product of the small upstream ORF binds to the terminal inverted repeats (40; D. Zerbib, P. Prentki, P. Gamas, E. Freund, D. J. Galas, and M. Chandler, Mol. Microbiol., in press) and both represses its own synthesis and inhibits transposition (18, 41). It is thought that the transposase is generated by a −1 translational frameshift at the 3′ end of the first gene to yield a fusion product with the protein specified by the larger downstream ORF (34). This region carries a motif (A—AAA—AAC) identical to that found in certain retroviruses, where it provokes, possibly in conjunction with downstream secondary structures (stem-loop and pseudoknot), an efficient −1 ribosomal frameshift (4, 5, 13). Weiss et al. (38) demonstrated that this motif also promotes ribosomal frameshifting in *E. coli*, as do the sequences A—AAA—AAA and A—AAA—AAG. In that situation, the downstream structure serves as a dispensable enhancer. The A—AAA—AAG motif, very rarely found in *E. coli* genes (R. Weiss and J. Atkins, personal communication), is an even more efficient rephasing signal than the other two (38). The downstream reading frame of all members of the IS*3* group can be extended upstream to overlap the 3′ terminus of ORFA (Table 1). It is therefore noteworthy that six members of the family carry a tandem "shifty" codon motif (five A—AAA—AAG and one A—AAA—AAA) in the overlap region (Fig. 7). The four other members of the group do not exhibit such a strong motif. Direct evidence for the production of a fusion protein has been obtained for IS*150* (B. Rak, personal communication). Our preliminary experiments also strongly suggest that expression of IS*911* genes involves a frameshift event. This element can express three proteins from the ORFA-ORFB coding strand. A 12-kilodalton protein is attributible to ORFA, a 30-kilodalton product corresponds to ORFB, and a 40-kilodalton product represents the ORFA-ORFB fusion protein (P. Polard, M.-F. Prère, O. Fayet, and M. Chandler, unpublished data).

It now remains to be determined whether the observed fusion protein is indeed the functional transposase and whether all members of the IS*3* family, in particular those which do not have a "strong" frameshift motif, control their level of transposition in such a way.

ACKNOWLEDGMENTS

We thank N. Ishiguro, B. Kearney, and J. Kalinowski for communicating modifications to their published nucleic acid sequences, J. Atkins, R. Gesteland, and R. Weiss for communicating their unpublished data, E. Barbey for the photography, and P. Polard and the other members of the Mobile Genetic Element Group in Toulouse as well as D. Lane for stimulating discussions.

This work was supported in part by a North Atlantic Treaty Organization travel grant (to M.C. and D. Galas).

LITERATURE CITED

1. **Arber, W., L. Enquist, B. Hohn, N. Murray, and K. Murray.** 1983. Experimental methods for use with lambda, p. 433–466. *In* R. Hendrix, J. Roberts, F. Stalh, and R. Weisberg (ed.), Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
2. **Berg, D. E.** 1988. Structural requirement for IS*50*-mediated gene transposition. Proc. Natl. Acad. Sci. USA **80**:792–796.
3. **Brennan, R. G., and B. W. Matthews.** 1989. The helix-turn-helix DNA binding motif. J. Biol. Chem. **264**:1903–1906.
4. **Brierly, I., M. E. G. Boursnell, M. M. Binns, B. Bilimoria, V. C. Blok, T. D. K. Brown, and S. C. Inglis.** 1987. An efficient ribosomal frame-shifting signal in the polymerase encoding region of the coronavirus IBV. EMBO J. **12**:3779–3785.
5. **Brierly, I., P. Digard, and S. C. Inglis.** 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement of an RNA pseudoknot. Cell **57**:537–547.
6. **Chandler, M., and D. J. Galas.** 1983. IS*1*-mediated tandem duplication of plasmid pBR322. Dependence on RecA and on DNA polymerase I. J. Mol. Biol. **165**:183–190.
7. **Daniels, D., J. Schroeder, W. Szybalski, F. Sanger, A. Coulson, G. Hong, D. Hill, G. Petersen, and F. Blattner.** 1983. Complete annotated lambda sequence, p. 519–676. *In* R. Hendrix, J. Roberts, F. Stalh, and R. Weisberg (ed.), Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. **Dodd, I. B., and J. B. Egan.** 1987. Systematic method for the detection of potential λ Cro-like DNA-binding regions in proteins. J. Mol. Biol. **194**:557–564.
9. **Fayet, O., J. M. Louarn, and C. Georgopoulos.** 1986. Suppression of the *Escherichia coli dnaA46* mutation by amplification of the *groES* and *groEL* genes. Mol. Gen. Genet. **202**:435–445.
10. **Galas, D. J., and M. Chandler.** 1982. Structure and stability of Tn*9*-mediated cointegrates. Evidence for two pathways of transposition. J. Mol. Biol. **154**:245–272.
11. **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.
12. **Ishiguro, N., and G. Sato.** 1988. Nucleotide sequence of insertion IS*3411*, which flanks the citrate utilization determinant of transposon Tn*3411*. J. Bacteriol. **170**:1902–1906.
13. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell **55**:447–458.
14. **Kearney, B., and B. J. Staskawicz.** 1990. Characterization of IS*476* and its role in bacterial spot disease of tomato and pepper. J. Bacteriol. **172**:143–148.
15. **Lam, S., and J. R. Roth.** 1983. IS*200*: a *Salmonella*-specific insertion sequence. Cell **34**:951–960.
16. **Lieb, M.** 1981. A fine-structure map of spontaneous and induced mutations in the lambda repressor gene, including insertions of IS elements. Mol. Gen. Genet. **184**:364–371.
17. **Lin, H. C., S. P. Lei, and G. Wilcox.** 1985. An improved DNA sequencing strategy. Anal. Biochem. **147**:114–119.
18. **Machida, C., and Y. Machida.** 1989. Regulation of IS*1* transposition by the *insA* gene product. J. Mol. Biol. **208**:567–574.
19. **Machida, Y., C. Machida, and E. Ohtsubo.** 1982. A novel type of transposon generated by insertion element IS*102* present in a pSC101 derivative. Cell **30**:29–36.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Matsutani, S., H. Ohtsubo, Y. Maeda, and E. Ohtsubo.** 1987. Isolation and characterization of IS elements repeated in the bacterial chromosome. J. Mol. Biol. **196**:445–455.
22. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Mori, H., A. Kondo, A. Ohshima, T. Ogura, and S. Hiraga.** 1986. Structure and function of the F plasmid genes essential for partitioning. J. Mol. Biol. **192**:1–15.
24. **Murotsu, T., K. Matsubara H. Sugisaki, and M. Takanami.** 1981. Nine unique repeating sequences in a region essential for

replication and incompatibility of the mini-F plasmid. Gene **15**:257–271.
25. **Myers, E. W., and W. Miller.** 1988. Optimal alignments in linear space. Comput. Appl. Biosci. **4**:11–18.
26. **Needlemann, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–453.
27. **Nyman, K., K. Nakamura, H. Ohtsubo, and E. Ohtsubo.** 1981. Distribution of the insertion element IS*1* in gram-negative bacteria. Nature (London) **289**:516–518.
28. **Pabo, C. O., and R. T. Sauer.** 1984. Protein DNA recognition. Annu. Rev. Biochem. **53**:293–321.
29. **Priefer, U. B., J. Kalinowski, B. Ruger, W. Heumann, and A. Puhler.** 1989. ISR1, a transposable DNA sequence resident in *Rhizobium* class IV strains, shows structural characteristics of classical insertion elements. Plasmid **21**:120–128.
30. **Reimmann, C., R. Moore, S. Little, A. Savioz, N. S. Willets, and D. Haas.** 1989. Genetic structure, function and regulation of the transposable element IS*21*. Mol. Gen. Genet. **215**:416–424.
31. **Ronecker, H. J., and B. Rak.** 1987. Genetic organisation of insertion element IS*2* based on a revised nucleotide sequence. Gene **59**:291–296.
32. **Rubens, C. E., L. M. Heggen, and J. M. Kuypers.** 1989. IS*861*, a group B streptococcal insertion sequence related to IS*150* and IS*3* of *Escherichia coli*. J. Bacteriol. **171**:5531–5535.
33. **Schwartz, E., M. Kroger, and B. Rak.** 1988. IS*150*: distribution, nucleotide sequence and phylogenetic relationships of a new *E. coli* insertion element. Nucleic Acids Res. **16**:6789–6802.
34. **Sekine, Y., and E. Ohtsubo.** 1989. Frameshifting is required for production of the transposase encoded by insertion sequence 1. Proc. Natl. Acad. Sci. USA **86**:4609–4613.
35. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
36. **Timmerman, D. P., and C. D. Tu.** 1985. Complete sequence of IS*3*. Nucleic Acids Res. **6**:2127–2139.
37. **Twigg, A. J., and D. Sherratt.** 1980. *trans*-Complementable copy number mutants of plasmid ColE1. Nature (London) **283**:216–218.
38. **Weiss, R. B., D. M. Dunn, M. Shuh, J. F. Atkins, and R. F. Gesteland.** 1990. *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. New Biol. **1**:159–169.
39. **Yamada, T., P. D. Lee, and T. Kosuge.** 1986. Insertion sequence elements of *Pseudomonas savastanoi*: nucleotide sequence and homology with *Agrobacterium tumefaciens* transfer DNA. Proc. Natl. Acad. Sci. USA **83**:8263–8267.
40. **Zerbib, D., M. Jakowec, P. Prentki, D. J. Galas, and M. Chandler.** 1987. Expression of proteins essential for IS*1* transposition: specific binding of InsA to the ends of IS*1*. EMBO J. **6**:3163–3169.
41. **Zerbib, D., P. Polard, J. M. Escoubas, D. J. Galas, and M. Chandler.** 1990. The regulatory role of the IS*1*-encoded InsA protein in transposition. Mol. Microbiol. **4**:471–477.