Bacterial Interspersed Mosaic Elements (BIMEs) Are a Major Source of Sequence Polymorphism in *Escherichia coli* Intergenic Regions Including Specific Associations With a New Insertion Sequence

Sophie Bachellier,[1] Jean-Marie Clément, Maurice Hofnung and Eric Gilson*

*Unité de Programmation Moléculaire et Toxicologie Génétique, CNRS URA 1444, Institut Pasteur, Paris, France and *Laboratoire de Biologie Moléculaire et Cellulaire de l'Ecole Normale Supérieure de Lyon, UMR49 CNRS/ENS, France*

Manuscript received June 27, 1996
Accepted for publication November 18, 1996

ABSTRACT

A significant fraction of *Escherichia coli* intergenic DNA sequences is composed of two families of repeated bacterial interspersed mosaic elements (BIME-1 and BIME-2). In this study, we determined the sequence organization of six intergenic regions in 51 *E. coli* and Shigella natural isolates. Each region contains a BIME in *E. coli* K-12. We found that multiple sequence variations are located within or near these BIMEs in the different bacteria. Events included excisions of a whole BIME-1, expansion/deletion within a BIME-2 and insertions of non-BIME sequences like the boxC repeat or a new IS element, named IS*1397*. Remarkably, 14 out of 14 IS*1397* integration sites correspond to a BIME sequence, strongly suggesting that this IS element is specifically associated with BIMEs, and thus inserts only in extragenic regions. Unlike BIMEs, IS*1397* is not detected in all *E. coli* isolates. Possible relationships between the presence of this IS element and the evolution of BIMEs are discussed.

NONCODING DNA regions in prokaryotic genomes are usually short and are mainly composed of regulatory elements essential for proper gene expression. However, in the last decade, it has become clear that bacterial intergenic regions also contain various families of repetitive elements. These repeated sequences are shorter than insertion sequences (IS) and do not usually encode proteins; their function is still elusive. They include bacterial interspersed mosaic elements (BIMEs), boxC, IRU (also called ERIC), RSA, iap and Ter sequences (reviewed in BACHELLIER *et al.* 1996).

About 500 BIMEs are scattered over the genome of *Escherichia coli* (GILSON *et al.* 1991a). They are found exclusively in extragenic positions, at the 3′ end of operons or between cotranscribed genes (GILSON *et al.* 1991b; DIMRI *et al.* 1992). They are composed of a mosaic combination of different conserved motifs: the palindromic unit (PU) (HIGGINS *et al.* 1982; GILSON *et al.* 1984), also called repetitive extragenic palindromic (REP) sequence (STERN *et al.* 1984), and seven PU-flanking sequences, called S, L, s, l, r, A and B (GILSON *et al.* 1991a,b). PUs were further subdivided into three variants differing slightly in sequence: Y, Z^1 and Z^2 (GILSON *et al.* 1991b; BACHELLIER *et al.* 1994). Specific combinations of these motifs define two major families of *E. coli* BIMEs, called BIME-1 and BIME-2 (BACHELLIER *et al.* 1994). BIME-1 contains two PUs (Y and Z^1), while BIME-2 contains from two to 12 PUs (Y and Z^2) and are direct repetitions of an assembly of motifs (BACHELLIER *et al.* 1994). The functions of BIMEs appear to be diverse, including mRNA stabilization (NEWBURY *et al.* 1987a,b), transcription termination (GILSON *et al.* 1986), translational control (STERN *et al.* 1988) and genomic rearrangements (SHYAMALA *et al.* 1990). Their specific interaction with integration host factor (IHF) (BOCCARD and PRENTKI 1993; OPPENHEIM *et al.* 1993), DNA gyrase (YANG and AMES 1988) and DNA polymerase I (GILSON *et al.* 1990a) suggests a role in the functional organization of the bacterial chromosome.

BIMEs have been identified in several enterobacteria [*E. coli, S. typhimurium* and several species from the Klebsiella tribe (BACHELLIER *et al.* 1993)] in which they are not necessarily present at the same chromosomal locations (GILSON *et al.* 1987). When BIMEs share the same intergenic region (orthologous BIMEs), their structures can vary; for example, the BIME between *lamB* and *malM* (DAHL *et al.* 1989; BACHELLIER *et al.* 1993) or between *araA* and *araD* (LIN *et al.* 1985; LEE *et al.* 1986) are different between *E. coli* and *S. typhimurium*. This polymorphism makes unclear whether orthologous BIMEs originate from a common ancestor, from transposition or from a horizontal transfer event. The aim of this study was to determine the sequence organization of orthologous BIMEs in 51 closely related enterobacteria, including 43 *E. coli* isolates of known phylogeny, to draw inferences about the dynamics of BIME evolution. We chose six intergenic regions containing

Corresponding author: Maurice Hofnung, Unité de Programmation Moléculaire et Toxicologie Génétique, Institut Pasteur, 25 rue du Docteur Roux, F75645 Paris Cedex 15, France.
E-mail: mhofnung@pasteur.fr

[1] *Present address:* CRC Nucleic Acid Structure Research Group, Department of Biochemistry, Medical Sciences Institute, The University, Dundee, United Kingdom.

a BIME in *E. coli* K-12, and analyzed their organizations in other isolates by PCR and sequencing. BIMEs showed a high level of size and sequence polymorphism in intergenic regions, indicating that various types of sequence rearrangements had occurred. In addition, this work revealed the existence of a new IS that appears to be specifically associated with BIMEs.

MATERIALS AND METHODS

Bacterial strains: Fifty-one *E. coli* strains were used: 43 natural isolates originated from the ECOR collection obtained from T. WHITTAM (NUMBERS 3, 8, 11, 13, 14, 21, 22, 26–29, 31, 37–39, 41–44, 46–66, 68, 69 and 71, OCHMAN and SELANDER 1984). Strains CI-3 (#191) and CI-4 (#207) were obtained from the Unité des Entérobactéries (Institut Pasteur, France); strains CI-1 and CI-2 were described in GILSON *et al.* (1990b); strains EPEC 24 and EPEC 25 were kindly provided by W. MAAS (LIM *et al.* 1990). We also used *Shigella flexneri* SC5700 (SANSONETTI and ARONDEL 1989) and *S. sonnei* S60-80. As a reference strain, we chose *E. coli* K-12 NCTC 10538 (wild type).

DNA manipulations and preparations: Total bacterial DNA samples were prepared as described in GILSON *et al.* (1990b). The PCR experiments were carried out in a final volume of 100 μl as follows: 10 pmol of each primer were added to 100 ng of genomic DNA, with 1.25 mM of each dNTP and 2.5 U of AmpliTaq DNA polymerase (Perkin-Elmer Cetus or Amersham). PCR reactions were performed in an automated thermocycler (Hybaid) as follows. DNA polymerase and 100 μl of mineral oil were added after an initial step of denaturation (8 min at 95°); annealing was performed during 2 min at 55°, and the extension step was 3 min at 72°. After 30 cycles, the last step of polymerization was extended to 10 min.

The sizes of amplified DNAs were determined after agarose gel electrophoresis; when they differed from those of corresponding fragments of *E. coli* K-12, the PCR was performed again using primers phosphorylated with T4 polynucleotide kinase for 1 hr at 37°. PCR fragments were purified on centricon units (purchased from Amicon), ligated in M13mp10 treated with *Sma*I and phosphatase (Amersham), and the ligation mix used to transform *E. coli* TG1. The presence of an insert was determined upon electrophoretic migration of single-stranded DNAs from white plaques as compared to M13mp18 single-stranded DNA.

Ligation-mediated PCR: Genomic DNA of EPEC 25 (1 μg) was digested by *Hin*dIII and *Xba*I; denaturation and elongation starting from an IS*1397*-specific primer (primerA: 5′GCCGTAGAAATGATGCCTGC3′) were performed as described in MUELLER and WOLD (1989), except that 0.2 mM of each dNTP were used. A double-double-stranded linker (linker1: 5′GCGCTACCGCGCGAGATCTGC3′; linker2: 5′GCAGATCTCGCGCGG3′) was ligated to the blunt ends generated in the previous step. After ethanol precipitation, the ligation products were resuspended in water and submitted to PCR amplification as described (MUELLER and WOLD 1989; see above for PCR reaction conditions) using, as primers, linker1 and a second IS*1397*-specific oligonucleotide (primerB: 5′GCCAGGTAATGATTTACAGC3′). PCR products were purified on centricon 100 units, ligated in the pCR-Script SK(+) vector (Stratagene) and *E. coli* XL1 cells were transformed as recommended. Ampicillin-resistant clones were screened by colony hybridization (as described in SAMBROOK *et al.* 1989) with ^{32}P-labeled primer B as a probe.

EPEC 25 library: EPEC 25 genomic DNA was prepared as described in SILHAVY *et al.* (1984), without lysozyme incubation. *Hin*dIII-digested EPEC 25 DNA (0, 50, 150, 350, 550 and 750 ng) were ligated with pUC19 treated with *Hin*dIII and phosphatase (Pharmacia) in a final volume of 10 μl. Competent JM109 *E. coli* cells (Promega) were transformed according to the manufacturer's indications with 1 μl of each ligation mixture. Five thousand eighty-eight individual ampicillin-resistant white colonies were seeded into 100 μl LB in 96-wells culture plates and incubated overnight at 37°. One hundred microliters of LB containing 20% DMSO were added to each well before storage of the plates at −80°. Culture plates were replicated together with a positive control (TG1 containing the ECOR 49 *araA-D* intergenic region) onto Hybond N+ membranes (Amersham) and grown overnight on LB-ampicillin plates at 37°. Cell lysis and DNA binding on membranes were performed as described (SAMBROOK *et al.* 1989). A 1262-bp fragment was obtained by PCR amplification of ECOR 49 *araA-D* DNA with primers 5orfA (ATGAAACATTCATTTGAAGTAAAAC) and primer2′ (see sequence below). Membranes were hybridized in the presence of the PCR ^{32}P-labeled IS*1397*-specific probe as described (GILSON *et al.* 1990b). Autoradiographic exposure of the filters revealed 43 positive colonies that were reisolated. Minilysates were analyzed by restriction analysis. Eight different patterns were identified and one clone of each class was further analyzed.

DNA sequencing: M13 single-strand sequencing of BIME-containing regions was performed using the Taq Track sequencing kit (Promega) with the universal primer, except for EPEC 25 *mtlA-mtlD* IS sequence determination, for which we used a series of primers: primer1, GCAGGCATCATTTCTACGG; primer2′, TATGTATTCGTTAACGGC; primer3, CAGAGCCTGAGGTGCGGAC; primer4, CTTCAGCAAGGCTTTCCAG; primer5, GCCATGCAGGAGCATCTTC; primer6, TGTTGCATCAGTTTTCGCAC; primer7, AAACGGTGCGAAAACTGATG; primer8, CCTGATAAAGGCCTCATTAC; primer9, GGGCAGCTGTCTTGATAATG. Double-strand plasmid DNAs were sequenced using Amersham's Sequenase kit as described (HSIAO 1991); primers used for sequencing of ligation-mediated PCR clones were both linker1 and primerB, while primers used for sequencing, respectively, left and right junctions of IS in the eight clones originating from the EPEC 25 library were primerA and primer11 (GAAAATGCCGTTAACGAATACAT). In six out of these eight clones, we totally resequenced IS*1397*. The reactions were submitted to electrophoresis on hydrolink 5%, and for a best readability, a salt gradient was created after 1 hr of run by adding sodium acetate to a concentration of 1 M to the lower buffer.

Southern hybridization: Total bacterial DNAs were digested with *Hin*dIII and *Hpa*I, separated on a 0.8% agarose gel and transferred overnight onto Hybond N$^+$ membrane (Amersham) in 0.4 N NaOH. The membrane was rinsed with 2× SSC and hybridized as described in GILSON *et al.* (1990b), except that hybridization and wash temperatures were 65°. The probe was a 1.1-kb fragment originating from an *Ava*II-*Hpa*I digest of IS*1397* ^{32}P-labeled with the Amersham nick-translation kit.

Oligonucleotides: The oligonucleotides used as primers in PCR reactions were obtained from the Laboratoire de Chimie Organique (J. IGOLEN), Institut Pasteur, Paris. Their sequences are the following:

ara1 (*araA* 3′ end): 5′GGCGCGCACCATACCGTCTTCAGCCATGCGCTG3′
ara2 (*araD* 5′ end): 5′TACCTGGCGTTTGAGATCTTCTAACAT3′
intB1 (*lamB* 3′ end): 5′GGTGCCCAGATGGAAATCTGGTCGTAAT3′
intB2 (*malM* 5′ end): 5′CAGAGGACGACGAGACTTTTCTTCATTTTCAT3′
intE1 (*malF* 5′ end): 5′CGTCGCTTTGCCACCAATGTTTCTTTTTAATGACATCCAT3′
intE2 (*malE* 3′ end): 5′CTGAAAGACGCGCAGACTCGTATCACCAAGTAA3′

fda1 (*fda* 3′ end): 5′CCAGGAACTGAACGCGATCGACGTTCTG3′
fda2 (upstream region from urf4): 5′TATGACGGGCAGTCTGACAAGGG3′
fepA (3′ end of *fepA*): 5′TGGTATATGAGCGTAAACACCCACTTCTGA3′
entD (5′ end of *entD*): 5′CAAAGGGGAGGGAGGTATGCGTAGTTTTCAT3′
mtlA (3′ end of *mtlA*): 5′GGAACTGCTGGCAGGTCGTAAGTAA3′
mtlD (5′ end of *mtlD*): 5′ CTGCGCCAAAATGTAATGCTTTCAT 3′.

RESULTS

To assay polymorphism occurring in orthologous BIMEs, six intergenic regions known to contain a BIME sequence in *E. coli* K-12 were amplified by PCR from 51 *E. coli* and Shigella isolates, including 43 ECOR strains, and analyzed by agarose gel electrophoresis, cloning and DNA sequencing. Since clonal relationships between ECOR strains have been established (MILLER and HARTL 1986; SELANDER *et al.* 1987; HERZER *et al.* 1990), the fate of a given BIME can be inferred from comparisons between its polymorphism pattern and the ECOR phenogram.

Collection of BIME polymorphism patterns: Six intergenic regions, known to contain either a BIME-2 sequence (*araA-araD, lamB-malM* and *malE-malF*) or a BIME-1 sequence (*fepA-entD, fda*-urf4, and *mtlA-mtlD*) in *E. coli* K-12, were each amplified by PCR from the genomic DNA of 43 ECOR strains, six other *E. coli* isolates and two Shigella species (see MATERIALS AND METHODS) (Figure 1). In five out of the six regions, we obtained fragments of a size different from that of K-12. Thirty-one PCR fragments were cloned and sequenced (marked by a star in Figure 2): in the seven sequenced regions having a size identical to that of the *E. coli* K-12 region, the sequence was highly conserved as compared to the corresponding *E. coli* K-12 BIME (data not shown). Similarly, the sequences of PCR fragments having a size different from *E. coli* K-12 but similar between different strains were highly conserved (see for example ECOR 63 and CI-1 for *lamB-malM,* ECOR 69, CI-1 and CI-4 for *fda-urf4* or ECOR 53, CI-1 and CI-4 for *mtlA-mtlD* in Figures 1 and 2). In these cases, the differences from *E. coli* K-12 are located within or near the BIME (see for example the *araA-araD* intergenic region, Figure 1A). From these results, we assumed that, for a given intergenic region, PCR fragments of identical length correspond to a similar BIME organization (Figures 1 and 2).

BIME polymorphism patterns of the ECOR strains are roughly congruent with their phenogram (HERZER *et al.* 1990; Figure 2). For example, the high level of BIME polymorphism observed in the *mtlA-D* BIME (Figure 1F) is restricted to all the strains of the B2 group, and the distribution of BIME patterns correlates with clusters of strains within this group (Figure 2). These results are in agreement with previous PU fingerprinting experiments suggesting that BIME polymorphism has clonal relationships (DIMRI *et al.* 1992). However, some of our data are inconsistent with the tree. (1) In the *araA-D* intergenic region, bacteria with the same BIME structure are found in different groups, and all groups contain bacteria with a different structure. This can be explained either by a high rate of variations occurring within that BIME or by a process of horizontal transfer. (2) In strains belonging to groups A, B1 and B2, and in ECOR 42 and 43, a second repetitive element, called boxC, is found associated with the BIME in this region. Like BIMEs, boxCs are extragenic, transcribed, and both elements are sometimes found in the same intergenic regions (for reviews on boxCs, see BACHELLIER 1995; BACHELLIER *et al.* 1996). This pattern is most likely explained by horizontal transfers since independent boxC insertions leading to identical BIME rearrangement are unlikely to occur at such a high frequency (unless a specific mechanism exists). (3) For ECOR 69, three intergenic regions (*lamB-malM, fda-urf4* and *mtlA-mtlD*) exhibited a more pronounced BIME polymorphism as compared to the other strains of the B1 group, to which it is linked (Figure 2). Thus, either this strain does not belong to the B1 group or BIME rearrangements rapidly occurred, for example by a high frequency of horizontal transfer events.

Absence/presence of a whole BIME-1: The *fepA-entD* intergenic region, containing a BIME-1 in *E. coli* K-12, is highly conserved in all the tested strains: the PCR fragments obtained from the various *E. coli* and Shigella isolates have an identical size (Figures 1 and 2) and the sequence of this region obtained from ECOR 38 was highly conserved as compared to *E. coli* K-12, in agreement with our hypothesis (not shown).

In contrast, we found that the intergenic regions between *fda* and urf4 and between *mtlA* and *mtlD* were shorter than in *E. coli* K-12 in, respectively, 29 and nine tested strains. Sequence analysis in five strains showed that these small intergenic regions do not contain any BIME motif. Remarkably, in these two regions but not in *fepA-entD,* BIME-1 is bracketed by directly repeated sequences (schematically drawn on Figure 1). In *mtlA-mtlD,* the 28-bp repeats are two truncated boxC sequences. In *fda*-urf4, the repeats are 36 and 37 bp long, and do not exhibit any sequence similarities with known sequences.

Complex expansion/deletion events within BIME-2: For two BIME-2 containing regions (*lamB-malM* and *araA-araD*), a high level of length polymorphism is observed among the tested bacteria: the number of motifs is variable from one strain to another, ranging from one to seven PUs in the *lamB-malM* region, and from two to six PUs in *araA-araD* region (Figures 1 and 2). In the *mtlA-mtlD* regions, the presence/absence of the whole BIME-1 sequence (see above) is not the only observed variation: several regions contain an unusual BIME, including motifs typical of BIME-1 (Z^1) and BIME-2 (Z^2, I, S), separated by a 20-bp atypical motif

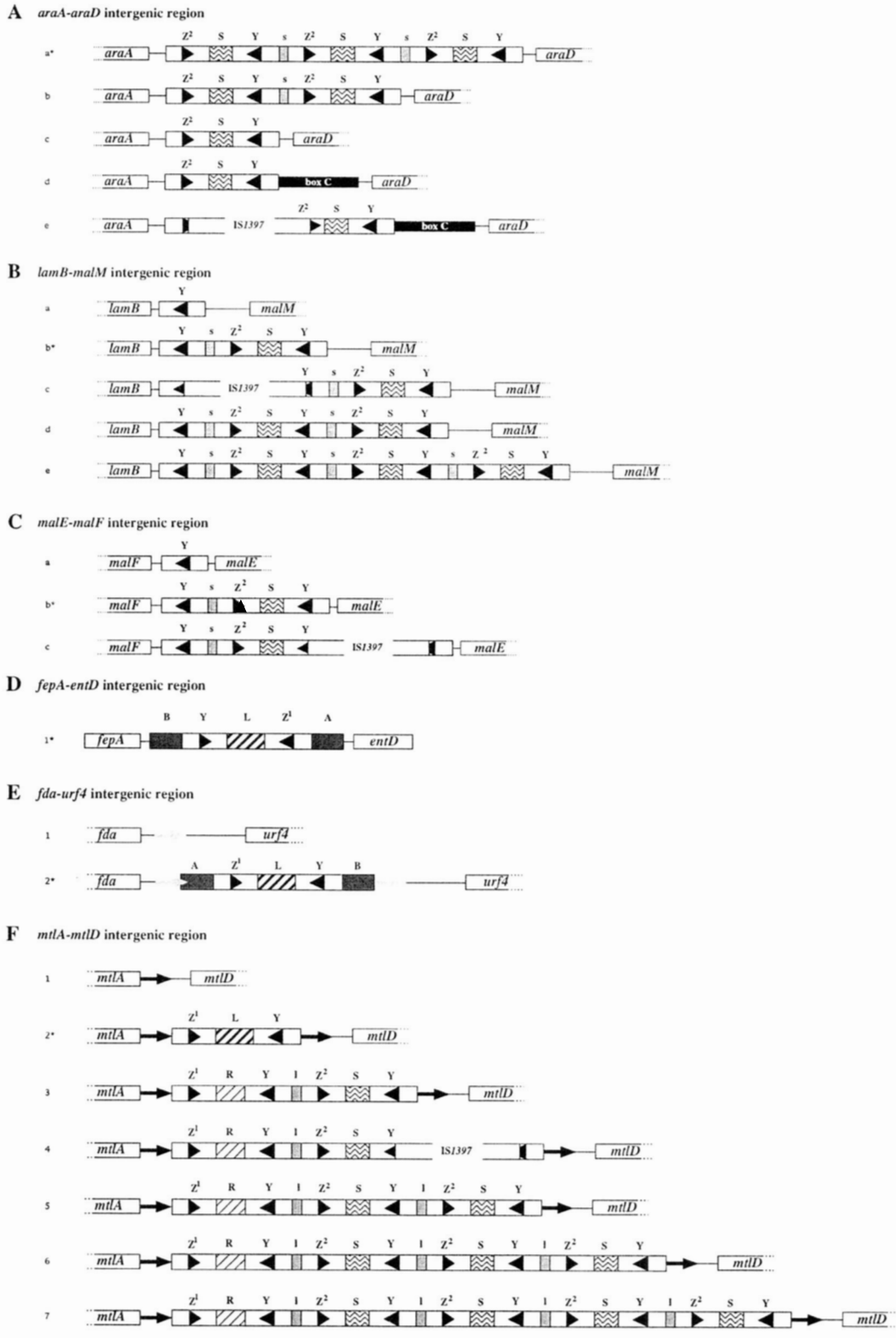
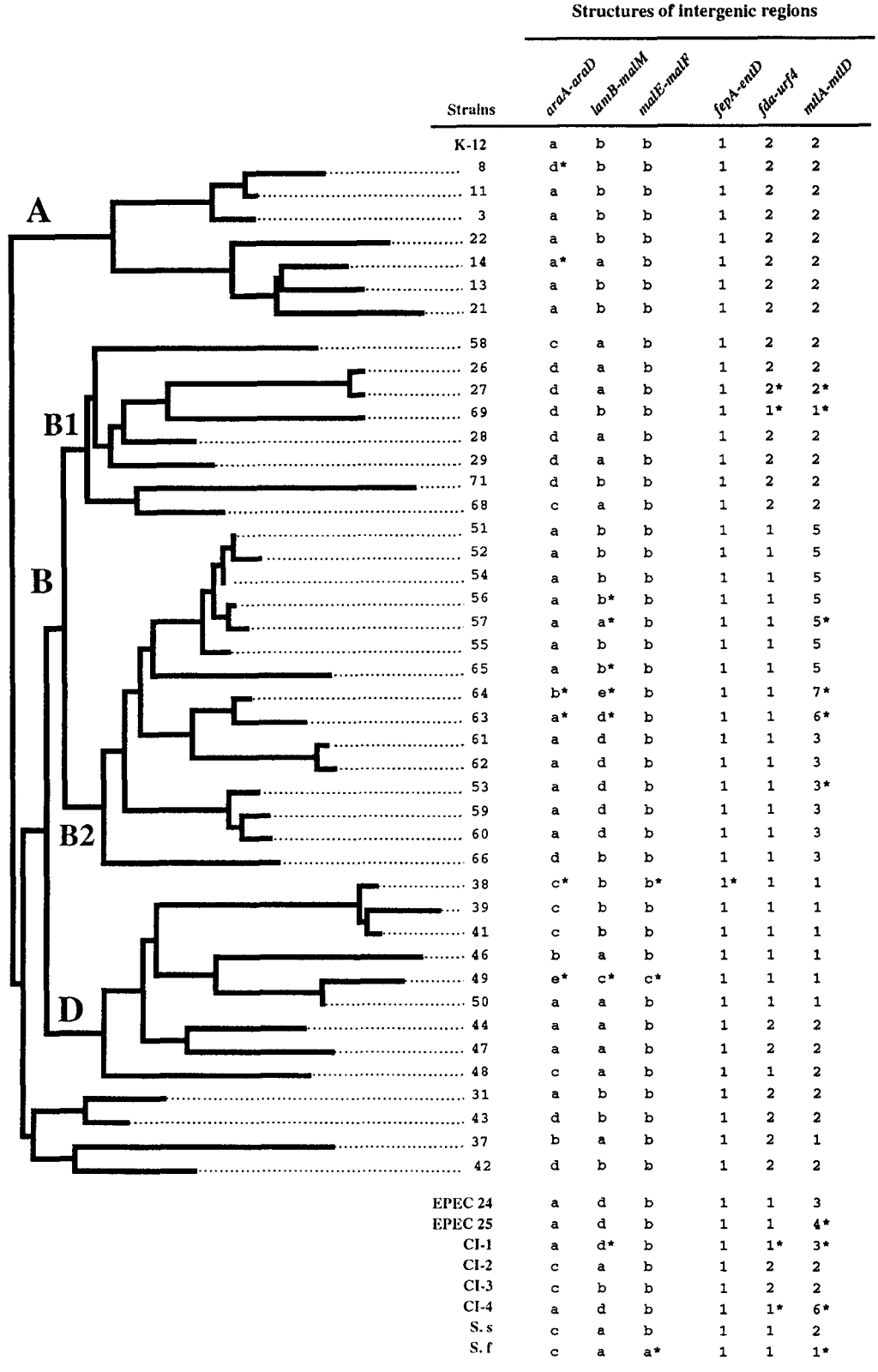


FIGURE 1.—Schematic drawing of the variations observed for the six BIME-containing regions. (A–C) Variations of BIME-2, present in *araA-araD* (A), in *lamB-malM* (B) and in *malF-malE* (C) intergenic regions. (D–F) Variations of three BIME-1, located between *fepA* and *entD* (D), *fda* and *urf4* (E) and *mtlA* and *mtlD* (F). The structures of the six intergenic regions are identified by letters (BIME-2) or numbers (BIME-1); *E. coli* K-12 BIME organizations are indicated by stars. In *mtlA-D* and *fda-urf4* intergenic regions, BIMEs are bracketed by direct repeats, symbolized by arrows. The repeats of the two regions are different both in sizes and in nucleotide sequences. The distances between the BIMEs and the ends of the flanking genes are drawn to scale. The orientations of PUs are indicated by the triangles included in the boxes. The names of the BIME motifs are given above the drawings, the new one appearing in rearranged *mtlA-D* BIME being called "R."



Strains	Structures of intergenic regions					
	araA-araD	*lamB-malM*	*malE-malF*	*fepA-entD*	*fda-urf4*	*mtlA-mtlD*
K-12	a	b	b	1	2	2
8	d*	b	b	1	2	2
11	a	b	b	1	2	2
3	a	b	b	1	2	2
22	a	b	b	1	2	2
14	a*	a	b	1	2	2
13	a	b	b	1	2	2
21	a	b	b	1	2	2
58	c	a	b	1	2	2
26	d	a	b	1	2	2
27	d	a	b	1	2*	2*
69	d	b	b	1	1*	1*
28	d	a	b	1	2	2
29	d	a	b	1	2	2
71	d	b	b	1	2	2
68	c	a	b	1	2	2
51	a	b	b	1	1	5
52	a	b	b	1	1	5
54	a	b	b	1	1	5
56	a	b*	b	1	1	5
57	a	a*	b	1	1	5*
55	a	b	b	1	1	5
65	a	b*	b	1	1	5
64	b*	e*	b	1	1	7*
63	a*	d*	b	1	1	6*
61	a	d	b	1	1	3
62	a	d	b	1	1	3
53	a	d	b	1	1	3*
59	a	d	b	1	1	3
60	a	d	b	1	1	3
66	d	b	b	1	1	3
38	c*	b	b*	1*	1	1
39	c	b	b	1	1	1
41	c	b	b	1	1	1
46	b	a	b	1	1	1
49	e*	c*	c*	1	1	1
50	a	a	b	1	1	1
44	a	a	b	1	2	2
47	a	a	b	1	2	2
48	c	a	b	1	1	2
31	a	b	b	1	2	2
43	d	b	b	1	2	2
37	b	a	b	1	2	1
42	d	b	b	1	2	2
EPEC 24	a	d	b	1	1	3
EPEC 25	a	d	b	1	1	4*
CI-1	a	d*	b	1	1*	3*
CI-2	c	a	b	1	2	2
CI-3	c	b	b	1	2	2
CI-4	a	d	b	1	1*	6*
S. s	c	a	b	1	1	2
S. f	c	a	a*	1	1	1*

FIGURE 2.—*E. coli* natural isolates BIMEs variations. The phenogram of the 43 analyzed ECOR strains is redrawn from HERZER *et al.* (1990); major phylogenetic groups are labeled A, B, and D, the B group being subdivided in B1 and B2 subgroups. The structures of the intergenic regions are listed by letters (BIME-2) or numbers (BIME-1) according to Figure 1. The fragments that have been sequenced are marked by a star. Nucleotide sequences have been deposited at EMBL under the following accession numbers: *araA-araD:* ECOR 8, Z75411; ECOR 14, Z75409; ECOR 38, X74279; ECOR 49, Z75436 and Z75437; ECOR 63, Z75408; ECOR 64, Z75410. *lamB-malM:* ECOR 49, Z75440 and Z75441; ECOR 56, Z75414; ECOR 57, Z75413; ECOR 63, Z75415; ECOR 64, Z75417; ECOR 65, Z75416; CI-1, Z75412. *malE-malF:* ECOR 38, Z75418; ECOR 49, Z75438 and Z75439; *S.flexneri:* Z75419. *fepA-entD:* ECOR 38, Z75420. *fda*-urf4: ECOR 27, Z75423; ECOR69, Z75422; CI-1, Z75425; CI-4, Z75421. *mtlA-mtlD:* ECOR 27, Z75424; ECOR53, Z75435; ECOR 57, Z75434; ECOR 63, Z75432; ECOR 64, Z75431; ECOR 69, Z75430; EPEC 24, Z75427; EPEC 25, Z75426; CI-1, Z75428; CI-4, Z75429; *S. flexneri:* Z75433.

(called R), which exhibits 70% identity with the 5′ part of the original L motif (see Figure 1F). Recombination events between distant BIMEs leading to operon fusions have already been reported (SHYAMALA *et al.* 1990), but in the present case, BIMEs-flanking sequences are the same in *E. coli* K-12 and in strains containing rearranged BIMEs. Since four regions containing both Z^1 and Z^2 motifs exist in the *E. coli* K-12 genome (listed in BACHELLIER *et al.* 1996), interfamily genetic exchanges relying, for example, on conversion or reciprocal recombination mechanisms may have occurred at low frequency. The number of motifs in *mtlA-mtlD* composite BIMEs is variable from one strain to another, ranging from four to 10 PUs. When composed of more than four PUs, the composite BIME results from an expansion of its BIME-2 part (compare variants 3, 5, 6 and 7, Figure 1F). Thus, we conclude that expansion/deletion of motifs within a BIME is specific for BIME-2 sequences. It is worth noting that the motif sequences composing a given BIME-2 are highly conserved rendering unlikely ectopic exchanges or gene conversion with another BIME. Thus, we favor local mechanisms for expansion/deletion events including unequal sister chromatid exchange or intrachromatid recombination between homologous motifs or replication slippage (GILSON *et al.* 1991a).

Associations of BIMEs with boxC and a novel type of IS: In some bacteria, we observed the insertion of repeated sequences unrelated to BIMEs. This is the case for the *mtlA-mtlD* BIMEs bracketed by truncated boxC sequences (see above and Figure 1F). In *araA-araD,* a complete boxC sequence is flanking the tail end of the Y PU in 11 isolates (Figure 1A).

Surprisingly, in ECOR 49, the *araA-araD, lamB-malM* and *malE-malF* intergenic regions are unusually large, reaching more than 2 kb (Figure 1, A–C). Such a large intergenic region is also found in the *mtlA-mtlD* region of strain EPEC 25 (Figure 1F). We first sequenced the *mtlA-mtlD* region from strain EPEC 25, and showed that an insertion-like sequence (IS) was inserted in the center of the last PU leading to a 4-bp duplication (Figures 3 and 4). This element had not been described before and was named IS*1397.* Southern hybridization experiments showed that the same sequence was present in the *araA-araD, lamB-malM* and *malE-malF* regions from strain ECOR 49 (not shown); partial nucleotide sequencing of the three regions revealed that the insertion sites of the IS-like element IS*1397* were in the PU central regions, generating a 4-bp duplication in *araA-araD* and *malE-malF,* and a 3-bp duplication in *lamB-malM* (Figure 3). Interestingly, in the *araA-araD* intergenic region of strain ECOR 49, both a boxC and the IS-like element are associated to the BIME-2 (Figure 1A). This element had not been described before and was named IS*1397* for the following reasons: (1) it is present a number of times in lineages that lacked it at particular sites, (2) its insertion is associated with target site repeats and (3) its sequence shows sequence similarity to known IS elements from the IS*3* family (see below).

IS*1397* is a new member of the IS*3* family: The complete 1432-bp nucleotide sequence of IS*1397* is shown on Figure 4. This IS ends with 25-bp-long imperfect terminal inverted repeats containing four mismatches (IRL and IRR, indicated with arrows on Figure 4). Two long open reading frames (ORF) are found on the same strand. The first one, called *orfA,* starting at nucleotide 52 (shaded ATG codon) and ending at nucleotide 573, could encode a 173-amino acid (aa) protein, called ORFA. A potential ribosome binding site (RBS) is found at a distance of 8 bp from the ATG codon (underlined on Figure 4), and possible promoter sequences are found upstream from the RBS, in the left inverted repeat (−35 and −10 possible sequences are underlined with wavy lines on Figure 4). ORFA contains a putative α-helix-turn-α-helix motif (HTH), from aa 22 to aa 41 (boxed on Figure 4), with a good probability according to DODD and EGAN (1990): the SD (standard deviation) score obtained by comparison with a weight matrix was 3.45, corresponding to a probability near 70%. A putative leucine zipper (Met 135, Leu 142, Leu 149, Leu 156, boxed on Figure 4) is located near its carboxy terminal end (LANDSCHULTZ *et al.* 1988). The second ORF, called *orfB,* is located between coordinates 678 (GTG codon included in a shaded box on Figure 4) and 1400 (TAA), and could encode a protein of 240 aa (ORFB). *orfB* is in −1 frame with respect to *orfA.* It contains a motif for a nucleotide binding site [GXXXXGK(T,S), circled aa on Figure 4; WALKER *et al.* 1982)]. The carboxy terminal part of ORFB contains a conserved region found in other ORFB proteins from members of IS*3* family (see Figure 5 and below).

The GenBank nucleotide data bank was screened for similarity to either the IS*1397* nucleotide sequence or the IS*1397* putative encoded proteins. Nucleotide sequence comparisons gave best scores with several members of the IS*3* family; for example, 56.6% identities were found with IS*150* and 51.1% with IS*1223* from Lactobacillus spp. A significant score was also obtained with a small region of 65 bp at the end of *RhsB* element of the *E. coli* K-12 chromosomal DNA (ZHAO *et al.* 1993). *Rhs* are *E. coli* accessory elements found in variable copy numbers, and their limits are defined by comparison between *Rhs*-containing and *Rhs*-free regions (SADOSKY *et al.* 1989; HILL *et al.* 1994). This 65-bp sequence exhibited 96.9% identity with the right end of IS*1397,* from nucleotide 1368 to nucleotide 1432 (Figure 6). It is adjacent on one side to a BIME-1 (with only a half of the first PU) that is part of the *RhsB* region (as defined by comparison to the *RhsB*° region of *E. coli* ECOR 32; ZHAO *et al.* 1993), and, on the other side, to an H-repeat, which is an IS-like element found associated with some *Rhs* elements in *E. coli* (GUSTAFSON *et al.* 1994; HILL *et al.* 1994; XIANG *et al.* 1994).

The two putative proteins encoded by IS*1397* were screened against protein sequences obtained by transla-

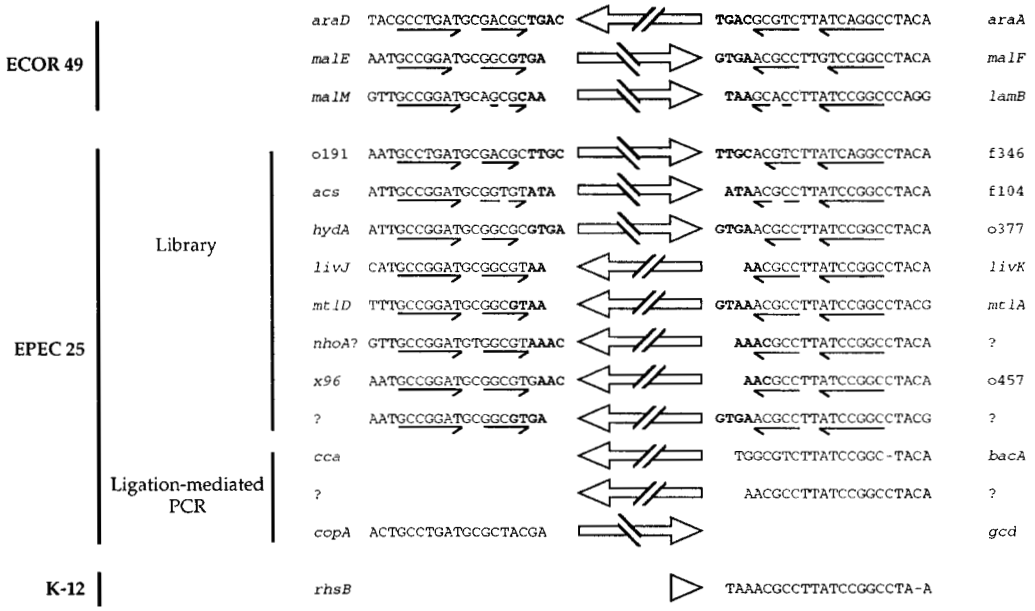


FIGURE 3.—IS insertion sites. The nucleotide sequences of palindromic units containing IS*1397* are shown. The first three sequences (found in *araA-araD, malE-malF* and *lamB-malM* intergenic regions) originate from strain ECOR 49. The other ones are from strain EPEC 25, except the last one, which originates from K-12. The names of flanking genes are indicated on both sides; the sequence labeled *nhoA*? is unknown in *E. coli*, but similar to *S. typhimurium nhoA* gene. The PUs are oriented from tail to head, the orientation of IS*1397* is indicated by a central arrow (from IRL to IRR). PUs dyad symmetries are underlined. Duplications of the target sites are in bold. The three last sequences of strain EPEC 25 were obtained by ligation-mediated PCR, hence only the sequence on the left side of the insertion is known. In the *rhsB* element, only the very right end of IS*1397* is present, hence only the sequence at the right side of the insertion is available.

tion of GenBank DNA sequences. ORFB presented sequence similarities with proteins from several IS*3* family members (namely at the conserved DDE motif, see below), while similarities were found between ORFA and proteins from several IS (IS*1223*, 30.4% identity; *Lactococcus lactis upp*-flanking IS, 26.2% identity; IS*150*, 31.9% identity; and an uncharacterized *N. gonorrhoeae* IS, 28.8% identity). Sequence similarities between ORFA sequences are more limited than between ORFB, but were already observed among members of the IS*3* family (MACADAM *et al.* 1990; SKAUGEN and NES 1994).

IS*1397* exhibited other structural similarities with members of the IS*3* family. The two 25-bp-long inverted repeats begin with the trinucleotide $^{5'}$TGA$^{3'}$, followed by a stretch of Cs and then by an A + T-rich region, as it has been described for most of the members of the IS*3* family (PRÈRE *et al.* 1990). Two open reading frames, *orfA* and *orfB*, were found on the same strand, and the reading frames were 0 and −1, respectively. The two ORFs were separated by 104 bp, which is unusual for the IS*3* family. *OrfB* exhibited a D,D(35)E motif (Figure 5) found in integrase domains of retroviruses and in transposases of IS*3* family members (KULKOSKY *et al.* 1992; reviewed in POLARD and CHANDLER 1995). An A_6G sequence (overlined in Figure 4) was found before the end of *orfA* and followed by a dyad symmetry (overlining arrows on Figure 4). Such a structure (called the frameshift window) has been shown to favor frameshifting, leading, in the case of IS*3* family members, to the formation of a fusion protein between ORFA and ORFB, called ORFAB (POLARD *et al.* 1991; VÖGELE *et al.* 1991; CHANDLER and FAYET 1993; SEKINE *et al.* 1994). In IS*1397*, ORFAB would contain a short amino acid sequence originating from the 104 bp located between ORFA and ORFB, hence specific to the fusion protein. This short sequence contains a putative HTH [SD score of 4.43, corresponding to a probability of HTH occurrence >90%, according to DODD and EGAN (1990)].

IS*1397* is present only in a subset of *E. coli* isolates: We investigated the distribution of IS*1397* in several *E. coli* natural isolates, in two Shigella strains and in *E. coli* K-12, by Southern hybridization. As a probe, we used an IS*1397* 1.1-kb *Ava*II-*Hpa*I fragment (see MATERIALS AND METHODS). Genomic DNAs were digested with *Hin*dIII (which has no site within IS*1397*) and *Hpa*I, which cleaves twice at the right end of IS*1397*. Five out of the 10 strains we tested did not yield hybridization signals (K-12, strain CI-3, the two Shigella strains and strain ECOR 8, Figure 7). Our probe was specific for IS*1397* since it did not cross-hybridize with other IS elements present in the *E. coli* K-12 or ECOR 8 genomes (SAWYER *et al.* 1987; BIRKENBIHL and VIELMETTER 1989). The five other strains (CI-4, EPEC 24, EPEC 25, ECOR 49 and ECOR 50) exhibited variable numbers of cross-hybridizing bands: four in strain CI-4, eight in strain ECOR 50, 10 in strain EPEC 24, ~15 in strain ECOR 49 and >25 in strain EPEC 25 (Figure 7). Strain ECOR 49 DNA and in a lesser extent strain EPEC 25 DNA exhibited very wide and intense bands, which may be explained by overrepresentation of IS-containing frag-

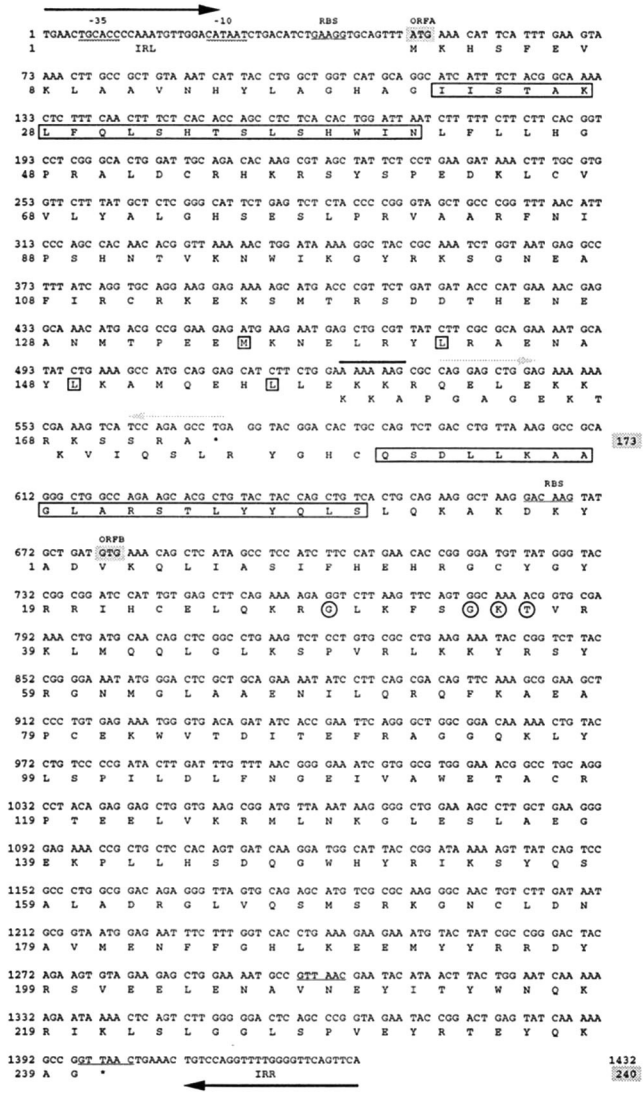


FIGURE 4.—Nucleotide sequence of IS*1397* and predicted amino acids sequences of putative ORFs. Left and right imperfect terminal inverted repeats (IRL and IRR) are indicated by black arrows. Possible start codons are in shaded boxes, preceded by putative ribosome binding sites (RBS), which are underlined. Termination codons are indicated by stars. Potential promoter sequences are indicated by wavy lines. A possible frameshift window (A_6G) is overlined; it is immediately followed by a palindrome, indicated by overlining gray arrows. Amino acid sequences deduced from two open reading frames, *orfA* and *orfB*, are given below the nucleotide sequence, as well as the amino acid sequence unique to the putative fusion protein, ORFAB (see text). ORFA contains a putative α-helix-turn-α-helix motif (aa 22–41, boxed) and a potential leucine-zipper (Met_{135}, Leu_{142}, Leu_{149} and Leu_{156}, boxed). A putative nucleotide binding-site motif (circled amino acids) is found in ORFB. The fusion protein contains an α-helix-turn-α-helix motif (aa 180–199, boxed) in its specific part (see text). In the left margin are given the coordinates of nucleotides and amino acids. The length of ORFA and ORFB are indicated in shaded boxes in the right margin. The two *Hpa*I restriction sites located at the 3′ end of the IS are underlined. IS*1397* nucleotide sequence has been deposited at EMBL under the accession number X92970.

ments with the same or similar sizes. Both strains contained natural plasmids, none of which harbor an IS*1397* insertion (not shown).

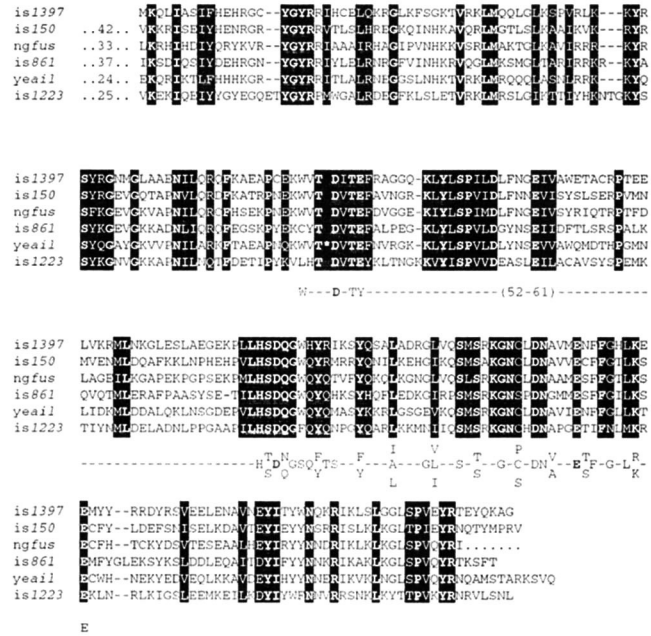


FIGURE 5.—Alignment of IS*3* family members ORFB proteins. ORFB sequences from four IS (labeled is1397, is150, is861 and is1223) and two IS-like (ngfus, from *Neisseria gonorrhoeae*, and yeail, from *Yersinia enterocolitica*) have been aligned with CLUSTALW (HIGGINS *et al.* 1991). Boxes indicate when the six ORFB sequences contained identical (bold letters) or similar amino acids, according to Blosum 50 matrix. The five insertion sequences compared with IS*1397* are longer, and the distance at which the first position is found is indicated in number of amino acids. In the IS-like found near *Y. enterocolitica ail* locus, a −1 frameshift, indicated by a star, has been introduced to maximize the alignment. Dots at the end of ngfus indicate that the end of the sequence is not known. Below the sequences is given the consensus for the transposase domain from the IS*3* family (POLARD and CHANDLER 1995).

IS*1397* is specifically associated with BIMEs: We investigated the sequence specificity of IS*1397* localization in strain EPEC 25 by the two following strategies.

Strain EPEC 25 genomic DNA was amplified by ligation-mediated PCR (MUELLER and WOLD 1989) to generate a set of fragments containing one end of IS*1397* and chromosomal flanking regions. Three IRL-containing clones were obtained, two of them carrying regions known in *E. coli* K-12: the *bacA-cca* intergenic region, containing a BIME-2 with the structure Y s Z^2 S Y, and the region after *gcd*, containing a BIME-1. In all three cases, the left end of the IS was flanked by a PU, either a Y PU (in *bacA-cca* region and in the unidentified clone) or a Z^1 PU (in the BIME-1 after *gcd*).

Since the previous method gives the sequences flanking one end of the IS, we also constructed a library with strain EPEC 25 genomic DNA. Briefly, chromosomal DNA was digested by *Hin*dIII and inserted into *Hin*dIII-digested pUC19. The ligation was used to transform *E. coli* K-12 JM109: the resulting IS*1397*-containing clones were screened by colony hybridization, and the IS insertion sites were sequenced. We obtained eight different clones, one of them contained the previously identified

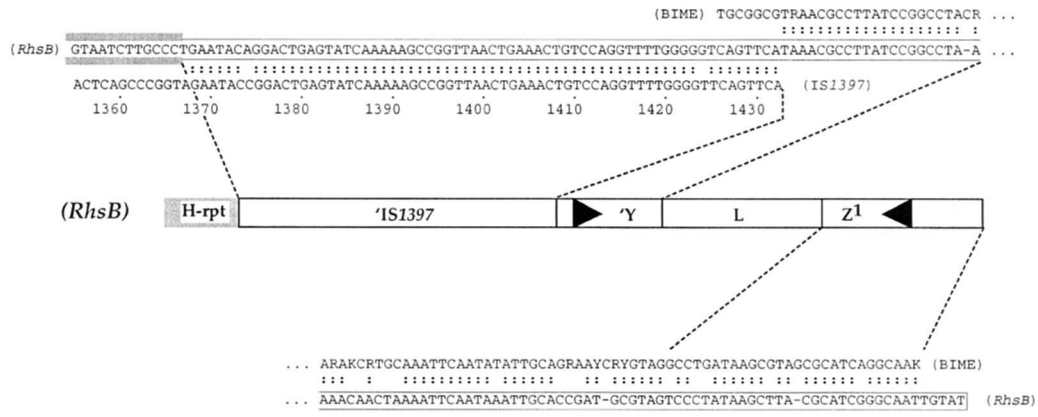


FIGURE 6.—*RhsB* matching with IS*1397* right end. (Top) Sequence alignment between *rhsB* (middle sequence, boxed) and IS*1397* left end (lower sequence) or BIME Y consensus (upper sequence). IS*1397* nucleotides are numbered as in Figure 4. (Bottom) Sequence alignment between *rhsB* (lower sequence, boxed) and BIME-1 L and Z^1 consensus (upper sequence). (Middle) Schematic structure of *rhsB*.

mtlA-mtlD region. Remarkably, in each clone, IS*1397* was inserted in the central part of a PU, either Y, Z^1 or Z^2 (Figure 3). In all but one region, the insertion occurred in the last PU of the BIME.

DISCUSSION

BIMEs are intergenic repeated DNA sequences composed of mosaic combinations of small motifs, present in various enterobacteria. Their function and their contribution to bacterial genome plasticity are still poorly understood. To find clues about their sequence dynamic, we have examined intraspecific polymorphism within BIME-containing regions by PCR amplification and DNA sequencing. Among 51 *E. coli* and Shigella isolates, various patterns of BIME rearrangements were observed in six orthologous intergenic regions. Most of these events exhibit a clonal inheritance and some of them appear to have occurred relatively recently in *E. coli* populations. Only a minority of them seems to have arisen by a horizontal transfer process. Therefore, we conclude that orthologous BIMEs have undergone sequence rearrangements after the divergence of natural *E. coli* strains. These rearrangements are likely to represent a rapid source of sequence polymorphism in intergenic regions since the extra-BIME intergenic sequences remain largely unchanged (unpublished results). The nature of the variations differs between the two previously described BIME families (GILSON *et al.* 1991b; BACHELLIER *et al.* 1994), including presence/absence of a whole BIME-1, various expansion/deletion events within BIME-2 and insertions of other types of repeated DNA at or near the BIME. These differences in evolutionary dynamic observed between the two BIME families further highlight their structural and functional diversity (BACHELLIER *et al.* 1994).

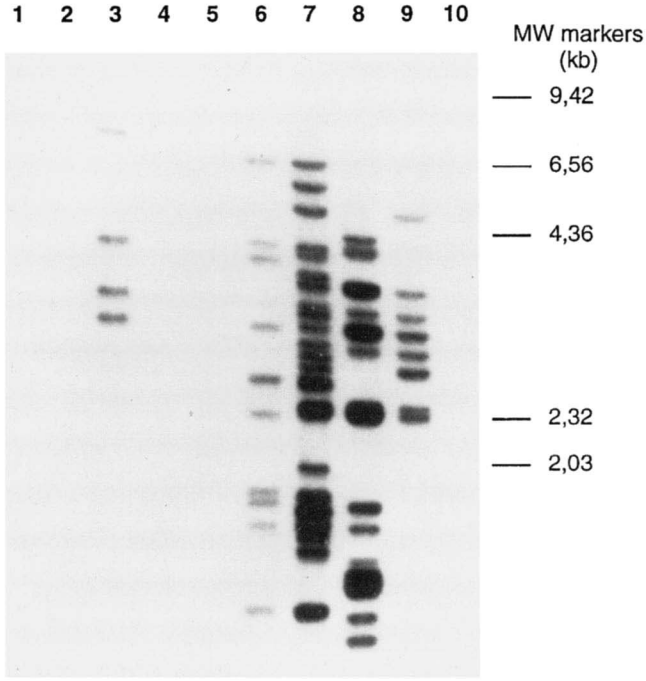


FIGURE 7.—Distribution of IS*1397* in several *E. coli* strains. Total *Hpa*I-*Hin*dIII-digested DNA from the following strains: *E. coli* K-12 C600 (lane 1), CI-3 (lane 2), CI-4 (lane 3), *S. sonnei* (lane 4), *S. flexneri* (lane 5), EPEC 24 (lane 6), EPEC 25 (lane 7), ECOR 49 (lane 8), ECOR 50 (lane 9) and ECOR 8 (lane 10) was electrophoresed on agarose 0.8%, transferred to a nylon membrane and probed with an internal IS fragment (see MATERIALS AND METHODS).

Interestingly, BIME-1-containing regions differ in their rate of evolution. One region (*fepA-entD*) appears highly stable over the whole panel of bacteria used in this study. In the two other regions (*fda*-urf4 and *mtlA-mtlD*) BIME-1 appears to be frequently subject to insertions or deletions. In these two cases the BIME-1 are bracketed by direct repeats of short sequences unrelated to any known BIME motif. A search in sequence databases revealed that six out of the 55 known BIME-1 were flanked by direct repeats: in the *gusC*-ORF, *mgtA*-f141, *narP-yejP*, o310-f541 and o499-f111 intergenic regions and after the *artJ* gene (unpublished results). These repeats are all different, have a size ranging from 25 to 31 bp, and exhibit dyad symmetry. We failed to detect any repetition around BIME-2 sequences. Interestingly, the BIME-1 sequences appear more homogeneous at the nucleotide level between orthologous regions than between intergenic regions present in a given strain, suggesting that most of the BIME-1 insertion events occurred before the divergence of the studied *E. coli* strains. Therefore, it seems likely that the presence/

absence of BIME-1 sequences reflects rather their precise excision than their insertion. This excision event can be explained by an intrachromatid recombination between the direct flanking repeats characterizing the unstable BIME-1 (see above). Since these direct repeats are also reminiscent of duplications flanking transposons, we cannot exclude that transposition events may also occur but at a low frequency. Overall, we conclude that the present dynamic of BIME-1 in *E. coli* seems to lead toward their elimination. Interestingly, we failed to detect any BIME-1-like sequence in *S. typhimurium.*

Two BIME-2 with the same structures in *E. coli* K-12 (*lamB-malM* and *malE-malF,* Figure 1), and that are separated by only 4.1 kb on the chromosome, do not behave similarly. In 49 out of 51 tested bacteria, the *malE-malF* region remains unchanged while various events of expansion/deletion occurred within the *lamB-malM* BIME. The reason for such a discrepancy in the rate of variation is not known and may be imposed by functional constraints. For example, the expression level of flanking genes may be affected by the structure of the intergenic region. Indeed, it was shown that the presence of a BIME is important for the relative expression between *malE* and the downstream genes, *malF* and *malG* (NEWBURY *et al.* 1987a,b). This may not be the case for *lamB* and *malM.*

As a result of this study, a novel type of IS-like element was identified and named IS*1397.* A single copy of IS*1397* is present into the *mtlA-mtlD* BIME of EPEC 25 and the *araA-araD, lamB-malM* and *malE-malF* BIMEs of ECOR 49; in the four regions, IS*1397* is inserted within a PU motif. According to sequence comparison and general organization, this element is related to the IS*3* family. Its presence is restricted to a subset of natural *E. coli* isolates and it is absent from *E. coli* K-12 as reflected both by Southern analysis (Figure 7) and by the presence of a binding activity specific for the IS*1397* terminal inverted DNA detected in EPEC 25 and ECOR 49 but not in *E. coli* K-12 (unpublished results).

Interestingly, in all the known insertion sites, *i.e.,* four determined from the systematic study of BIME-containing intergenic regions and 10 randomly chosen IS*1397* insertion sites from EPEC 25 genomic DNA, the IS is always located within BIMEs, flanked by either a 4- or 3-bp duplication. This clearly shows a highly specific association between IS*1397* and BIME sequences. The fact that IS*1397* is flanked by direct repeats of BIME sequences argues that BIMEs are specific target sites for IS*1397* insertions. The PU part of the BIME appears to play a critical role in this integration specificity since IS*1397* insertion sites are always at the center of PU sequences and since BIMEs composed of different PU-flanking sequences can be used as a target sequences. Such a sequence-specificity in targeting is rather unusual among IS*3* family members (reviewed in GALAS and CHANDLER 1989) and suggests a mechanism of recognition between the IS and PU sequences. For example, one of the proteins encoded by the IS may be able to preferentially bind PU sequences or host proteins bound to PU DNA. To our knowledge, IS*1397* is the first example of a prokaryotic mobile genetic element that appears to move exclusively in extragenic locations.

The association of IS*1397* with BIMEs reported here further strengthens the link between mobile genetic elements, repeated sequences and genome plasticity. Specific associations between transposable elements and repeated DNA elements appear widespread among both eukaryotes and prokaryotes. For example, in *E. coli,* IS-like elements are associated with repeated *Rhs* sequences (ZHAO *et al.* 1993; HILL *et al.* 1994, 1995; ZHAO and HILL 1995) and, in yeast, the Ty5 retrotransposon is preferentially inserted within subtelomeric repeats (ZOU *et al.* 1996). At least in some cases, the activity of these transposable elements can influence the fate of their flanking repeated sequences. For example, in Drosophila, mobilization of a *P*-transposable element drives gene conversion within a flanking tandem repeat (THOMPSON-STEWART *et al.* 1994). Thus, it is tempting to speculate that IS*1397* may play an important role in the evolution of BIMEs.

We thank T. WHITTAM, W. MAAS, P. GRIMONT and P. SANSONETTI for supplying strains, and D. PERRIN for helpful discussions and his constant interest in this work.

LITERATURE CITED

BACHELLIER, S., 1995 Les BIMEs: un exemple de séquences répétées chez les entérobactéries. Bull. Inst. Pasteur **93:** 97–152.

BACHELLIER, S., D. PERRIN, M. HOFNUNG and E. GILSON, 1993 Bacterial Interspersed Mosaic Elements (BIMEs) are present in the genome of *Klebsiella.* Mol. Microbiol. **7:** 537–544.

BACHELLIER, S., W. SAURIN, D. PERRIN, M. HOFNUNG and E. GILSON, 1994 Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs). Mol. Microbiol. **12:** 61–70.

BACHELLIER, S., E. GILSON, M. HOFNUNG and C. W. HILL, 1996 Repeated sequences, pp. 2012–2040 in *Escherichia coli and Salmonella: Cellular and Molecular Biology,* edited by F. C. NEIDHARDT, R. CURTISS, III, J. L. INGRAHAM, E. C. C. LIN, K. B. LOW *et al.* American Society for Microbiology, Washington, DC.

BIRKENBIHL, R. P., and W. VIELMETTER, 1989 Complete maps of IS*1*, IS*2*, IS*3*, IS*4*, IS*5*, IS*30* and IS*150* locations in *Escherichia coli* K12. Mol. Gen. Genet. **220:** 147–153.

BOCCARD, F., and P. PRENTKI, 1993 Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3′ end of transcription units. EMBO J. **12:** 5019–5027.

CHANDLER, M., and O. FAYET, 1993 Translational frameshifting in the control of transposition in bacteria. Mol. Microbiol. **7:** 497–503.

DAHL, M. K., E. FRANCOZ, W. SAURIN, W. BOOS, M. D. MANSON *et al.,* 1989 Comparison of sequences from the *malB* regions of *Salmonella typhimurium* and *Enterobacter aerogenes* with *Escherichia coli* K12: a potential new regulatory site in the interoperonic region. Mol. Gen. Genet. **218:** 199–207.

DIMRI, G. P., K. E. RUDD, M. K. MORGAN, H. BAYAT and G. FERRO-LUZZI AMES, 1992 Physical mapping of repetitive extragenic palindromic sequences in *Escherichia coli* and phylogenetic distribution among *Escherichia coli* strains and other enteric bacteria. J. Bacteriol. **174:** 4583–4593.

DODD, I. B., and J. B. EGAN, 1990 Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18:** 5019–5026.

GALAS, D. J., and M. CHANDLER, 1989 Bacterial insertion sequences, pp. 109–162 in *Mobile DNA,* edited by D. E. BERG and M. M. HOWE. American Society for Microbiology, Washington, DC.

GILSON, E., J.-M. CLÉMENT, D. BRUTLAG and M. HOFNUNG, 1984 A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. EMBO J. **3:** 1417–1421.

GILSON, E., J.-P. ROUSSET, J.-M. CLÉMENT and M. HOFNUNG, 1986 A subfamily of *E. coli* palindromic units implicated in transcription termination? Ann. Inst. Pasteur/Microbiol. **137B:** 259–270.

GILSON, E., J.-M. CLÉMENT, D. PERRIN and M. HOFNUNG, 1987 Palindromic units: a case of highly repetitive DNA sequences in bacteria. Trends Genet. **3:** 226–230.

GILSON, E., D. PERRIN and M. HOFNUNG, 1990a DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic units highly repetitive DNA: implications for bacterial chromosome organization. Nucleic Acids Res. **18:** 3941–3952.

GILSON, E., S. BACHELLIER, S. PERRIN, D. PERRIN, P. A. D. GRIMONT *et al.*, 1990b Palindromic units highly repetitive DNA sequences exhibit species specificity within *enterobacteriaceae*. Res. Microbiol. **141:** 1103–1116.

GILSON, E., W. SAURIN, D. PERRIN, S. BACHELLIER and M. HOFNUNG, 1991a Palindromic units are part of a new bacterial interspersed mosaic element (BIME). Nucleic Acids Res. **19:** 1375–1383.

GILSON, E., W. SAURIN, D. PERRIN, S. BACHELLIER and M. HOFNUNG, 1991b The BIME family of bacterial highly repetitive sequences. Res. Microbiol. **142:** 217–222.

GUSTAFSON, C. E., S. CHU and T. J. TRUST, 1994 Mutagenesis of the paracrystalline surface protein array of *Aeromonas salmonicida* by endogenous insertion elements. J. Mol. Biol. **237:** 452–463.

HERZER, P. J., S. INOUYE, M. INOUYE and T. S. WHITTAM, 1990 Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172:** 6175–6181.

HIGGINS C. F., G. FERRO-LUZZI AMES, W. M. BARNES, J.-M. CLÉMENT and M. HOFNUNG, 1982 A novel intercistronic regulatory element of prokaryotic operons. Nature **298:** 760–762.

HIGGINS, D. G., A. J. BLEASBY and R. FUCHS, 1991 CLUSTAL V: improved software for multiple sequence alignment. CABIOS **8:** 189–191.

HILL, C. W., C. H. SANDT and D. A. VLAZNY, 1994 *Rhs* elements of *Escherichia coli:* a family of genetic composites each encoding a large mosaic protein. Mol. Microbiol. **12:** 865–871.

HILL, C. W., G. FEULNER, M. S. BRODY, S. ZHAO, A. B. SADOSKY *et al.*, 1995 Correlation of *Rhs* elements with *Escherichia coli* population structure. Genetics **141:** 15–24.

HSIAO, K., 1991 A fast and simple procedure for sequencing double stranded DNA with Sequenase. Nucleic Acids Res. **19:** 2787.

KULKOSKY, J., K. S. JONES, R. A. KATZ, J. P. G. MACK and A. M. SKALKA, 1992 Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. **12:** 2331–2338.

LANDSCHULTZ, W. H., P. F. JOHNSON and S. L. MCKNIGHT, 1988 The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240:** 1759–1764.

LEE, N., W. GIELOW, R. MARTIN, E. HAMILTON and A. FOWLER, 1986 The organization of the *araBAD* operon of *Escherichia coli*. Gene **47:** 231–244.

LIM, D., T. A. T. GOMES and W. K. MAAS, 1990 Distribution of msDNA among serotypes of enteropathogenic *Escherichia coli* strains. Mol. Microbiol. **4:** 1711–1714.

LIN, H.-C., S.-P. LEI, G. STUDNICKA and G. WILCOX, 1985 The *araBAD* operon of *Salmonella typhimurium* LT2. III. Nucleotide sequence of *araD* and its flanking regions, and primary structure of its product, L-ribulose-5-phophate 4-epimerase. Gene **34:** 129–134.

MACADAM, R. A., P. W. M. HERMANS, D. VAN SOOLINGEN, Z. F. ZAINUDDIN, D. CATTY *et al.*, 1990 Characterization of a *Mycobacterium tuberculosis* insertion sequence belonging to the IS*3* family. Mol. Microbiol. **4:** 1607–1613.

MILLER, R. D., and D. L. HARTL, 1986 Biotyping confirms a nearly clonal population structure in *Escherichia coli*. Evolution **40:** 1–12.

MUELLER, P. R., and B. WOLD, 1989 *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. Science **246:** 780–786.

NEWBURY, S. F., N. H. SMITH and C. F. HIGGINS, 1987a Differential mRNA stability controls relative gene expression within a polycistronic operon. Cell **51:** 1131–1143.

NEWBURY, S. F., N. H. SMITH, E. C. ROBINSON, I. D. HILES and C. F. HIGGINS, 1987b Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell **48:** 297–310.

OCHMAN, H., and R. K. SELANDER, 1984 Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:** 690–693.

OPPENHEIM, A. B., K. E. RUDD, I. MENDELSON and D. TEFF, 1993 Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli*. Mol. Microbiol. **10:** 113–122.

POLARD, P., and M. CHANDLER, 1995 Bacterial transposases and retroviral integrases. Mol. Microbiol. **15:** 13–23.

POLARD, P., M. F. PRÈRE, M. CHANDLER and O. FAYET, 1991 Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS*911*. J. Mol. Biol. **222:** 465–477.

PRÈRE, M.-F., M. CHANDLER and O. FAYET, 1990 Transposition in *Shigella dysenteriae:* isolation and analysis of IS*911*, a new member of the IS*3* group of insertion sequences. J. Bacteriol. **172:** 4094–4099.

SADOSKY, A. B., A. DAVIDSON, R.-J. LIN and C. W. HILL, 1989 *rhs* gene family of *Escherichia coli* K-12. J. Bacteriol. **171:** 636–642.

SAMBROOK, J., E. F. FRISCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, New York.

SANSONETTI, P. J., and J. ARONDEL, 1989 Construction and evaluation of a double mutant of *Shigella flexneri* as a candidate for oral vaccination against shigellosis. Vaccine **7:** 443–450.

SAWYER, S. A., D. E. DYKHUISEN, R. F. DUBOSE, L. GREEN, T. MUTANGADURA-MHLANGA *et al.*, 1987 Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. Genetics **115:** 51–63.

SEKINE, Y., N. EISAKI and E. OHTSUBO, 1994 Translational control in production of transposase and in transposition of insertion sequence IS*3*. J. Mol. Biol. **235:** 1406–1420.

SELANDER, R. K., D. A. CAUGANT and T. S. WHITTAM, 1987 Genetic structure and variation in natural populations of *Escherichia coli*, pp. 1625–1648 in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, edited by J. L. INGRAHAM, K. B. LOW, B. MAGASANIK, M. SCHAECHTER and H. E. UMBARGER. American Society for Microbiology, Washington, DC.

SHYAMALA, V., E. SCHNEIDER and G. FERRO-LUZZI AMES, 1990 Tandem chromosomal duplications: role of REP sequences in the recombination event at the join-point. EMBO J. **9:** 939–946.

SILHAVY, T. J., M. L. BERMAN and L. W. ENQUIST, 1984 *Experiments With Gene Fusions*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

SKAUGEN, M., and I. F. NES, 1994 Transposition in *Lactobacillus sake* and its abolition of lactocin S production by insertion of IS*1163*, a new member of the IS*3* family. Appl. Environ. Microbiol. **60:** 2818–2825.

STERN, M. J., G. FERRO-LUZZI AMES, N. H. SMITH, E. C. ROBINSON and C. F. HIGGINS, 1984 Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37:** 1015–1026.

STERN, M. J., E. PROSSNITZ and G. FERRO-LUZZI AMES, 1988 Role of the intercistronic region in post-transcriptional control of gene expression in the histidine transport operon of *Salmonella typhimurium:* involvement of REP sequences. Mol. Microbiol. **2:** 141–152.

THOMPSON-STEWART, D., G. H. KARPEN and A. C. SPARDLING, 1994 A transposable element can drive the concerted evolution of tandemly repetitious DNA. Proc. Natl. Acad. Sci. USA **91:** 9042–9046.

VÖGELE, K., E. SCHWARTZ, C. WELZ, E. SCHILTZ and B. RAK, 1991 High-level ribosomal frameshifting directs the synthesis if IS*150* gene products. Nucleic Acids Res. **19:** 4377–4385.

WALKER, J. E., M. SARASTE, M. J. RUNSWICK and N. J. GAY, 1982 Distantly related sequences in the α- and β- subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:** 945–951.

XIANG, S.-H., M. HOBBS and P. R. REEVES, 1994 Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. J. Bacteriol. **176:** 4357–4365.

YANG, Y., and G. FERRO-LUZZI AMES, 1988 DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. Proc. Natl. Acad. Sci. USA **85:** 8850–8854.

ZHAO, S., and C. W. HILL, 1995 Reshuffling of *Rhs* components to create a new element. J. Bacteriol. **177:** 1393–1398.

ZHAO, S., C. H. SANDT, G. FEULNER, D. A. VLAZNY, J. A. GRAY *et al.* 1993 *Rhs* elements of *Escherichia coli* K-12: complex composites of shared and unique components that have different evolutionary histories. J. Bacteriol. **175:** 2799–2808.

ZOU, S., N. KE, J. M. KIM and D. F. VOYTAS, 1996 The *Saccharomyces* retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev. **10:** 634–644.

Communicating editor: N. L. CRAIG