JOINT DISTRIBUTION OF INSERTION ELEMENTS IS*4* AND IS*5* IN NATURAL ISOLATES OF *ESCHERICHIA COLI*

DANIEL E. DYKHUIZEN,* STANLEY A. SAWYER,*† LOUIS GREEN,*
RAYMOND D. MILLER* AND DANIEL L. HARTL*

**Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110 and*
†*Department of Mathematics, Washington University, St. Louis, Missouri 63130*

Manuscript received October 3, 1984
Revised copy accepted June 14, 1985

ABSTRACT

A reference collection of natural isolates of *Escherichia coli* has been studied in order to determine the distribution, abundance and joint occurence of DNA insertion elements IS*4* and IS*5*. Among these isolates, 36% were found to contain IS*4* and 30% were found to contain IS*5*. Among strains containing IS*4* the mean number of copies per strain was 4.4 ± 0.8; the comparable figure for IS*5* was 3.7 ± 1.0. Although the presence of the elements among the isolates was independent, among those isolates containing both IS*4* and IS*5*, there was a significant negative correlation in the number of copies of the elements.—The reference collection was also studied for the presence of the DNA sequences flanking the single copy of IS*4* in the chromosome of *E. coli* K12. Homologous sequences were found in only 26% of the isolates. The sequences flanking the IS*4* invariably occur together, and their presence is significantly correlated with the presence of IS*4*. In eight of the strains that carry these flanking sequences, an IS*4* is located between them, and the sequences are present at the homologous position as in the K12 strain. We suggest that IS*4* and its flanking sequences share a common mechanism of dissemination, such as plasmids, and we present evidence that they are included in a much larger transposable element.

THE discovery and study of transposable elements in prokaryotes and eukaryotes has resulted in a major shift in ideas about the genetic stability of the genome and mechanisms of mutation (CAMPBELL 1983; HARTL, DYKHUIZEN and BERG 1984). Once considered highly stable, with most mutations resulting from single nucleotide substitutions, the genome is now regarded as having the potential to undergo rapid change because of the activities of transposable elements, with many mutations resulting from the insertion of such elements within or near genes or from rearrangements induced by them. These changes in perspective have come about so suddenly and recently that the role of transposable elements in the whole sweep of evolution has yet to be evaluated. Indeed, it is unknown which position transposable elements should be assigned on the spectrum of possible evolutionary implications, although the extremes of the spectrum would seem to be, on the one hand, exclusively parasitic DNA (DOOLITTLE and SAPIENZA 1980; ORGEL and CRICK 1980) and,

on the other hand, potentially creative agents of genetic adaptation (CAMPBELL 1981; HARTL, DYKHUIZEN and BERG 1984; HARTL and DYKHUIZEN 1984).

Information on which to formulate or evaluate hypotheses about the evolutionary role and implications of transposable elements should include studies of their distribution within and among genomes in natural populations. Such studies have been initiated in eukaryotes, particularly Drosophila (STROBEL, DUNSMUIR and RUBIN 1979; MONTGOMERY and LANGLEY 1983), and in prokaryotes (HARSHMAN and RILEY 1980; NYMAN *et al.* 1983; GREEN *et al.* 1984). Most of the data from prokaryotes pertains to insertion sequences, which are DNA sequences, usually smaller than 2000 base pairs (bp), that are capable of transposition into several or many sites in the genome (CAMPBELL *et al.* 1979).

Seven distinct insertion sequences have been described in *E. coli*, which are designated IS*1*, IS*2*, IS*3*, IS*4*, IS*5*, IS*30*, and $\gamma\delta$ (CALOS and MILLER 1980; CASPERS *et al.* 1984). The distribution of insertion element IS*5* among natural isolates of *E. coli* has recently been described (GREEN *et al.* 1984). Despite its abundance in laboratory strain *E. coli* K12, in which there are approximately ten copies per genome (SCHONER and SCHONER 1981), IS*5* is absent from more than two-thirds of the natural isolates examined, and among the isolates containing IS*5* the average number of copies per genome is approximately three (GREEN *et al.* 1984).

The same collection of natural isolates examined with respect to IS*5* has now been examined with respect to IS*4*. The rationale for studying IS*4* is that it is unusual among insertion sequences in *E. coli* K12 because it is normally found in just one copy per genome (KLAER and STARLINGER 1980). Thus, substantial differences in the distributions of IS*4* and IS*5* among natural isolates would not be surprising; however, the distribution of IS*4* is very similar to that of IS*5*. The proportion of natural isolates lacking IS*4* is approximately two-thirds, and among strains in which IS*4* is present, the average number of copies per genome is approximately four. However, the joint distribution of IS*4* and IS*5* among the isolates suggests that the elements are not completely independent. Natural isolates were also examined for the DNA that flanks the IS*4* element present in the genome of *E. coli* K12. Surprisingly, this flanking DNA is present in only a minority of natural isolates. The distribution of this flanking DNA suggests that it is, itself, part of a transposable element.

MATERIALS AND METHODS

We studied 72 strains in a reference collection of *E. coli* (OCHMAN and SELANDER 1984), among which 67 independent isolates provided the data on the abundance of IS*4* and IS*5*. (Some strains were excluded because they were judged to be so closely related to others that retaining them might have biased the data.) Using the methods described in GREEN *et al.* (1984), DNA from strains containing either IS*4* or IS*5* were digested to completion with restriction enzyme *Eco*RI, subjected to agarose electrophoresis and hybridized with an appropriate nick-translated probe after transfer to nitrocellulose paper (SOUTHERN 1975). Strains lacking the sequences were previously identified by means of dot blots, using undigested DNA (GREEN *et al.* 1984). The probe for IS*5* was a sequence of 0.94 kb pairs obtained by double digestion of plasmid pMN4 with *Eco*RI and *Bgl*II followed by electrophoresis and isolation from agarose by means of the glass powder method (HARTL *et al.* 1983). This sequence is contained entirely within IS*5*. The probe for IS*4* was a 1.5 kb sequence obtained similarly from a *Hin*dII digest of plasmid pBRK10. This probe includes

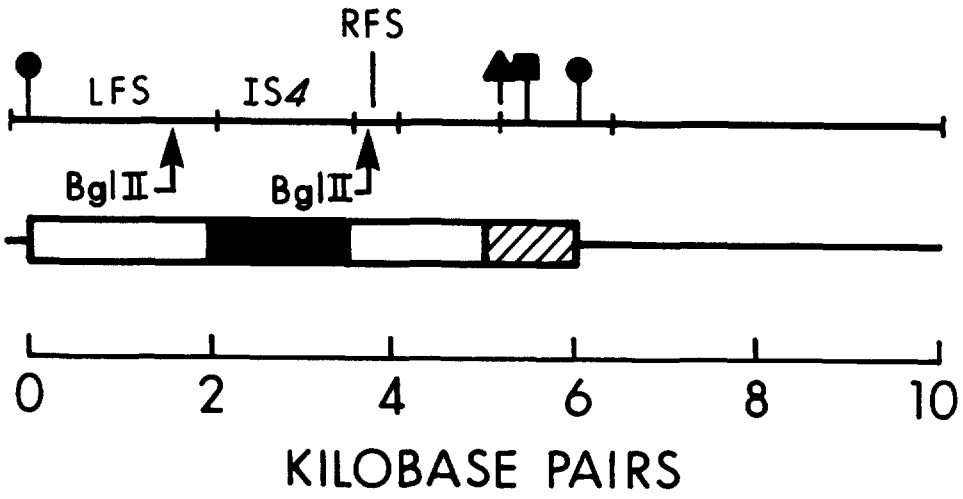

FIGURE 1.—Structure of pBRK10. The plasmid contains a 6.0 kb sequence of DNA from *E. coli* K12 (box) inserted into the *Hin*dII site of pBR322 (thin line). The insert contains all of IS*4* and part of IS2. The ticks on the line above represent *Hin*dII sites and the endpoints of sequences used as probes for IS*4* and the flanking sequences. Other symbols indicate restriction sites as follows: ● = *Hin*dIII, ■ = *Bgl*I, ▲ = *Hpa*I, and ↑ = *Bgl*II.

approximately 1400 bp of IS*4* plus a small amount of DNA (about 80 bp) that flanks the element on the right-hand side in the chromosome of *E. coli* K12. [Right and left are defined relative to the orientiation of the nucleotide sequence of IS*4* as published in figure 3 in KLAER *et al.* (1981), and in Figure 1 of this article.] This flanking DNA is insufficient to provide a significant hybridization signal under our conditions.

IS*4* contains no cleavage sites for *Eco*RI, thus each observed band should theoretically correspond to one copy of IS*4* in the genome. However, two potential complications arise. First, if two large restriction fragments of approximately the same size both carry IS*4*, they might not be sufficiently separated in the gel to be recognized as distict. Second, two copies of IS*4* might be contained in a single *Eco*RI restriction fragment. However, in both cases, the unresolved IS*4* would be detectable as a band with twice the expected intensity. Rarely have we seen such a situation and, consequently, conclude that our estimates of number of copies of IS*4* are not seriously biased by these potential complications. Four strains were observed to possess one or a few copies of a sequence that hybridized very weakly with IS*4*. These sequences are presumed to be related to, but distinct from, IS*4* and were not tabulated as *bona fide* IS*4* elements. Similar considerations pertaining to the estimates of the number of copies of IS*5*, and the finding of another element weakly homologous to IS*5*, have been discussed previously (GREEN *et al.* 1984).

Figure 1 illustrates plasmid pBRK10, which has a 6.0 kb sequence containing IS*4* (box) inserted into the *Hin*dIII site of pBR322 (thin line). The shaded region of the insert delimits IS*4*, and the hatched region indicates part of the sequence of an IS2 element present in the same insert. The ticks on the line across the top show the positions of *Hin*dII restriction sites. DNA fragments purified by means of the glass powder method (HARTL *et al.* 1983) that were used as probes for IS*4*, its left-hand flanking sequence (LFS) and its right-hand flanking sequence (RFS) are indicated. Strains of the reference collection containing the LFS were identified using dot blots, and DNA from a subset of strains, seven containing the LFS and 13 lacking the LFS, was digested with *Eco*RI, electrophoresed in agarose and hybridized with IS*4*, LFS or RFS probes after transfer to nitrocellulose (SOUTHERN 1975). Nitrocellulose filters that had initially been probed with IS*4* were stripped of probe and were rehybridized with the probe from the LFS. Stripping was accomplished by incubation of the filters at room temperature for 15 min in 200 ml of 20 mM NaOH, followed by neutralization with 250 ml of 0.5 M Tris-HCl pH 7.0. The LFS ws further studied by digestion

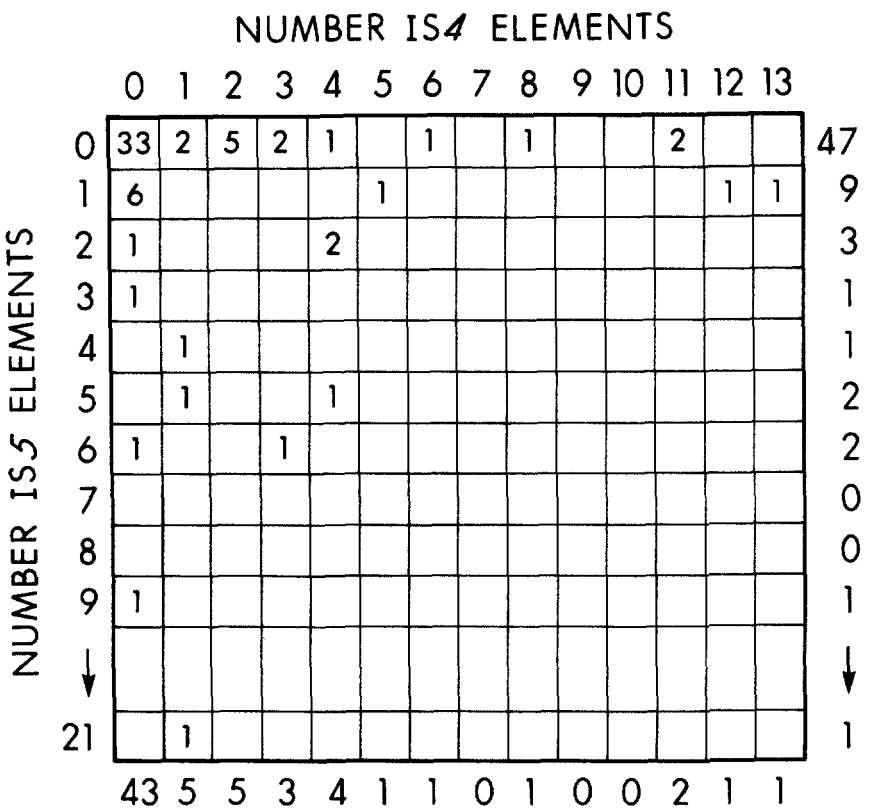

FIGURE 2.—Joint distribution of the number of copies of IS*4* and IS*5* in 67 reference strains (ECOR strains) of *E. coli*.

of the DNA of the strains containing LFS with *Eco*RI, *Xho*I or *Bam*HI; electrophoresis in agarose; transfer to nitrocellulose; and probing first with LFS and then with IS*4* after stripping. None of these enzymes cut within the cloned 6 kb fragment in pBRK10. The same procedure was also carried out with DNA restricted with *Bgl*II, which cuts twice within the cloned *E. coli* K12 fragment, once on either side of IS*4*, to yield a 2.6 kb fragment that will hybridize to both LFS and IS*4* probes (Figure 1).

To investigate the possibility that the LFS and IS*4* may occur in plasmids, DNA was prepared and subjected to agarose electrophoresis, as described in ANDERSON and MCKAY (1983), and was then transferred to nitrocellulose filters and probed by the usual methods.

Plasmid pBRK10 was obtained from P. STARLINGER through the laboratory of D. E. BERG, and it is virtually identical to pKS51, described by KLAER and STARLINGER (1980). Plasmid pMN4 was provided by R. E. Wolf, Jr., and is described in detail in NASOFF and WOLF (1980) and GREEN *et al.* (1984). The reference strains of *E. coli* (ECOR strains), described in OCHMAN and SELANDER (1984), were kindly provided by those authors. Many of these strains derive from a previous collection of MILKMAN (1973).

RESULTS

IS4 and IS5: The number of copies of IS*4* and IS*5* determined in each of 67 reference strains of *E. coli* is shown in Figure 2. Summary statistics of the marginal distributions are given in Table 1. Although the marginal distribu-

TABLE 1

Number of IS4 and IS5 elements

	Including 0 class	Excluding 0 class
IS4		
Mean	1.6	4.4
SEM	0.4	0.8
SD	3.1	3.8
IS5		
Mean	1.1	3.7
SEM	0.4	1.0
SD	3.0	4.7

tions are somewhat different, the differences are modest and not statistically significant. There were 43 strains (64%) containing no DNA sequences that hybridized with IS4, and 47 (70%) containing no DNA sequences that hybridized with IS5. Among those strains containing at least one element, the mean number of IS4 elements per strain was 4.4, and that of IS5 was 3.7. The standard deviation of the distribution of both elements is approximately equal to the mean. Both distributions deviate significantly from the Poisson distribution, principally because of an excess of strains that lack the element.

Turning to the joint distribution of IS4 and IS5, 33 strains had neither element, 14 had IS4 only, ten had IS5 only, and ten had both. As the chi-square test for independence yields a probability value of approximately 0.15, there is no reason to reject the hypothesis that presence or absence of IS4 is independent of the presence or absence of IS5.

Although the occurrence of the elements appears to be independent, the number of copies of each element apparently is not. Considering only those ten strains that contain at least one copy of each element, the Kendall nonparametric rank correlation coefficient, τ, equals -0.64, which is highly significant (SIEGEL 1956). The significant negative correlation implies that a large number of copies of IS4 is found in combination with a small number of copies of IS5, and vice versa, more often than would be expected from the marginal distributions.

The rank correlation in the number of elements calculated from all 67 strains is $\tau = 0.09$, which is nonsignificant. The lack of statistical significance in this case probably reflects the independence in the occurrence of the elements and the large number of strains that contain just one or neither of the elements.

IS4 flanking sequences: Probes for the chromosomal DNA sequence flanking IS4 on the left (LFS), and for the chromosomal sequence flanking IS4 on the right (RFS) in *E. coli* K12, were obtained from plasmid pBRK10 as diagramed in Figure 1. DNA from all 72 reference strains was probed in dot blots with the LFS. To our surprise, only 19 strains (26%) contained DNA sequences capable of hybridizing with the LFS (Figure 3). This suggests that the LFS might itself be part of an unrecognized transposable element present in the chromosome of *E. coli* K12.

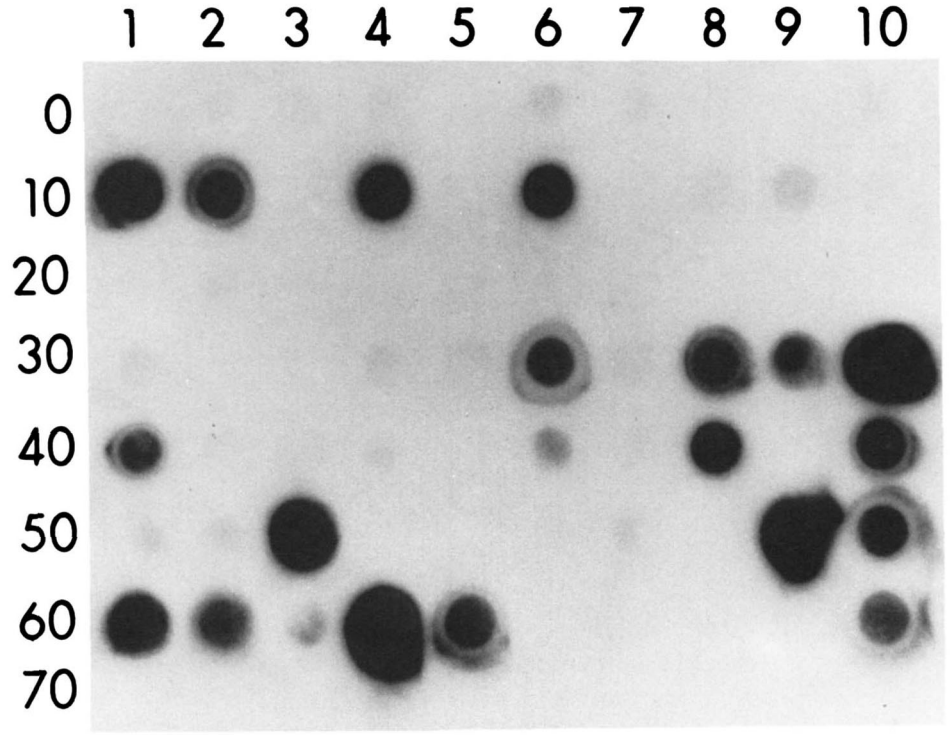

FIGURE 3.—Dot blot hybridizations showing presence of flanking sequences among ECOR strains. The ECOR strain number corresponds to the sum of the row and column numbers.

Coincidence in the occurrence of the LFS, the RFS and IS*4* was studied further in Southern transfers of *Eco*RI digests of DNA from seven strains containing the LFS and from 13 strains lacking the LFS. In these experiments some filters were hybridized serially with two probes, first with the IS*4* probe, and then, following autoradiography and removal of the previous probe, with the LFS. All seven strains containing the LFS had only a single copy of this sequence. Moreover, each of these strains also contained a sequence homologous with the RFS, and the molecular weight of the RFS-containing fragment was indistinguishable from that of the LFS-containing fragment. The 13 strains lacking the LFS also lack the RFS. Thus, the LFS and RFS occur simultaneously in the same strains, and they are close together in the genome. We conclude that LFS and RFS are equivalent probes for the fragment of DNA that had been interrupted by the insertion of IS*4* in *E. coli* K12, and all further work was done using only the LFS.

Southern transfers of *Eco*RI digests of the chromosomal DNA from all 19 strains containing the LFS were probed with the LFS. Fifteen strains gave single bands, indicating the presence of a single copy of this sequence. One strain [number 53 in the reference collection (OCHMAN and SELANDER 1984)] yielded two overlapping bands. DNA from this strain was also analyzed using the restriction enzymes *Xho*, *Bam*HI and *Bgl*II. Neither *Xho* nor *Bam*HI cut within the 6.0 kb sequence of DNA from *E. coli* K12 cloned into pBRK10, so

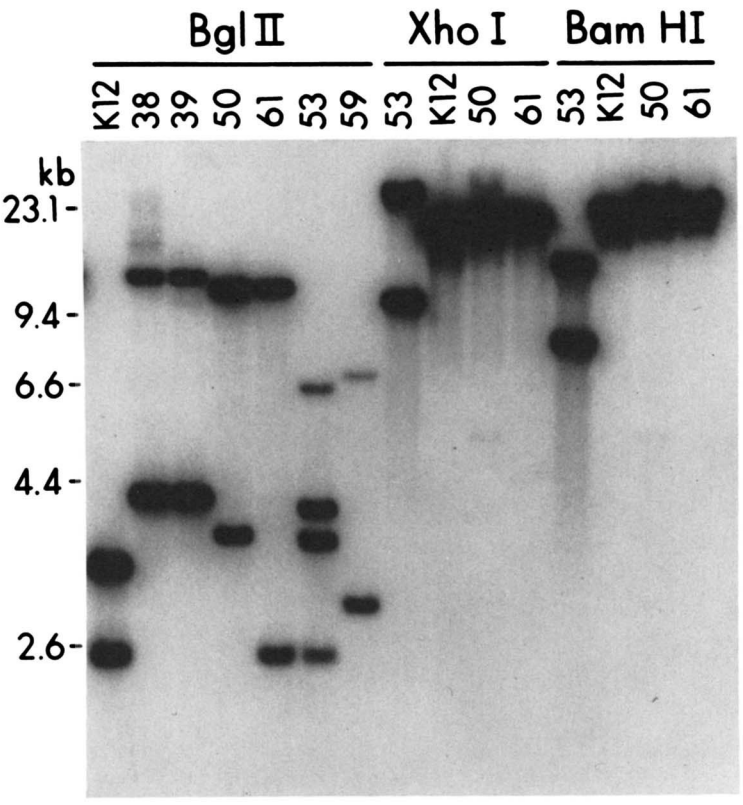

FIGURE 4.—Southern hybridization with the LFS probe, using DNA from a number of strains cleaved with several restriction enzymes. Note that strains 38 and 39 yield identical bands in digests with *Bgl*II.

a single copy of the LFS in a strain should yield a single band and two copies should yield two bands on a Southern transfer. On the other hand, *Bgl*II cleaves within the LFS, so a single copy of LFS should yield two bands and two copies should yield four bands on a Southern transfer of *Bgl*II restricted chromosomal DNA. The bands shown in Figure 4 are consistent with the presence of two different copies of the LFS in strain 53, but with only one copy in the other strains.

Altogether, 16 out of the 19 strains contained at least a single copy of the LFS in the chromosomal DNA; however, *Eco*RI digests of DNA from the other three strains produced little or no hybridization with the LFS. This suggested that, in these strains, the LFS was carried on a plasmid. To investigate this possibility, DNA from all 19 strains was prepared and subjected to agarose electrophoresis, as described in ANDERSON and MCKAY (1983), which leaves circular DNA molecules largely intact. With this procedure, plasmids smaller than ~25 kb (small plasmids) migrate faster than the band of linear DNA, which is mostly chromosomal DNA, and the large plasmids migrate slower than the chromosomal DNA. Southern transfers of this DNA were carried out and the DNA was hybridized with the LFS. The LFS hybridized to a large

TABLE 2

Strains containing the sequences that flank IS4 in E. coli K12

Strain[a]	Copies of IS4	Flanking sequences	
		Position[b]	IS4 association
K12	1	C	yes[c]
11	4	C	yes[c]
12	4	C	yes[c]
14	4	C	yes[c]
16	0	P	no
36	14	P	no
38	8	C	no
39	12	C	no
40	11	C	no
41	11	C	no
48	0	P	no
50	11	C	yes
53	4	C	yes[c], no
59	0	C	no
60	0	C	no
61	5	C	yes[c]
62	5	C	yes[c]
64	3	C	yes[c]
65	2	C	no
70	6	C	yes[c]

[a] Numbers are strain numbers in the ECOR collection.
[b] Presence in chromosome (C) or plasmid (P).
[c] Presence of IS4 within a 2.6 kb *Bgl*II fragment. (Strain 53 contains two copies of the flanking sequences.)

plasmid in all three strains that had shown poor hybridization with chromosomal DNA; no small plasmids were found to contain the LFS sequence.

The data presented above suggest that the LFS and the RFS might be part of a large transposable element, which in some isolates also contains IS4. Probing of *Eco*RI digests with both LFS and IS4 indicated that, in one-half of the strains, the restriction fragment containing the LFS does not contain an IS4 (Table 2). However, in most of the remaining strains the LFS and one copy of the IS4 are adjacent in the chromosome, as shown in probes of chromosomal DNA cleaved with *Bgl*II, which cuts twice within the cloned DNA from *E. coli* K12 (Figure 1), yielding a 2.6 kb fragment containing IS4. When blots of *Bgl*II-restricted chromosomal DNA were probed with LFS, all but one of the strains that seemed to have a copy of IS4 associated with the flanking sequences yielded a hybridization band of 2.6 kb, and an IS4 probe also hybridized to this same band. The concurrent appearance of IS4 and the LFS suggests that IS4 was brought into these strains along with the LFS. In the exceptional strain, ECOR number 50, both of the *Bgl*II fragments that hybridize to the LFS also hybridize to the IS4 probe, and both fragments are larger than 2.6 kb. Among the 16 strains that carry the LFS in the chromosome, eight distinct patterns of fragment size were found in *Bgl*II digests, of which four were

TABLE 3

Comparisons among electrophoretically identical strains

Strain	Copies of IS4	Presence of flanking sequence[a]
8	0	−
9	0	−
10	0	−
11	4	+
13	1	−
14	4	+
18	0	−
38	8	+
39	12	+
40	11	+
41	11	+
49	0	−
50	11	+
53	4	+
59	0	+
60	0	+
61	5	+
62	5	+

[a] Bracketed symbols indicate hybridizing fragments of identical size.

observed in only one strain. The fragment patterns in strains 38, 39, 40, and 41 are identical, and these strains are also identical in the electrophoretic profile for 11 enzyme loci (OCHMAN and SELANDER 1984). Strains 59 and 60 are identical to each other in fragment size and electrophoretic profile, as are strains 61 and 62; however, strains 11, 64 and 70 are identical to 61 and 62 in the size of the hybridizing *Bgl*II fragment, but they are different in electrophoretic profile. Similarly, strains 14 and 53 are alike in size of the hybridizing fragment, but they are different in electrophoretic profile. Conversely, several strains are identical in electrophoretic profile, but differ in the presence of the IS*4*-associated sequences (Table 3).

Occurrence of IS2 in plasmid pBRK10: In the initial hybridization experiments with the RFS, the probe contained a mixture of the nearly comigrating *Hin*dII fragments between coordinates 4.2 and 5.3 in Figure 1, and the *Hin*dII fragment between coordinates 5.3 and 6.5. Many strains were found to contain multiple DNA sequences that hybridized with this probe, suggesting that the probe contained part or all of another insertion sequence. Then, we noted from the description of plasmid pKS51 in KLAER and STARLINGER (1980) (pKS51 is virtually identical with pBRK10) that the *Hin*dII site at coordinate 5.3 in our Figure 1 is also a *Hpa*I site and that the DNA inserted into the plasmid terminates with a *Hin*dIII site at coordinate 6.0. These sites coincide with a *Hin*dII/*Hpa*I site at nucleotide 154, and a *Hin*dIII site at nucleotide

871, in the sequence of the 1327-bp element IS2 as reported in GHOSAL, SOMMER and SAEDLER (1979). We have also verified that the *Bgl*I site at nucleotide 363 in the published sequence of IS2 is present at the expected position in the element in pBRK10. In addition, as with IS2 (HU and DEONIER 1981), the element in pBRK10 does not occur in the genome of *E. coli* strain C. Furthermore, the probe from pBRK10 hybridizes with an *Eco*RI restriction fragment of 3.4 kb from our *E. coli* strain B, which is evidently identical in size with the fragment in strain B that hybridizes with IS2 reported by HU and DEONIER (1981). (Our strain B contains one additional copy of the sequence.) Finally, the probe from pBRK10 hybridizes with ten *Eco*RI restriction fragments in DNA from *E. coli* K12, which is within the range of number of copies of IS2 found in several K12 derivatives by HU and DEONIER (1981). Therefore, we conclude that the sequence between coordinates 5.1 and 6.0 in Figure 1 is identical with the first 871 bp of IS2.

DISCUSSION

Insertion sequences IS4 and IS5 appear to be present or absent independently in the reference strains; however, among strains in which both elements are present, a significant negative correlation in the number of copies of the elements is observed. The high degree of statistical significance of the Kendall nonparametric rank correlation coefficient suggests that the result is probably not spurious, in spite of the relatively small sample size ($N = 10$). The negative correlation in copy number is unexpected, and at present one can only speculate as to the possible mechanisms that cause it. One possibility is that IS4 and IS5, and perhaps other insertion sequences as well, respond to an overlapping set of molecules that regulate transposition, which might be produced by the elements themselves or by the host; however, the lack of significant sequence homology among the elements argues against this model. Another possible mechanism resulting in a negative correlation among elements is natural selection acting to eliminate strains with high numbers of copies of two or more elements. Major genome rearrangements resulting from activities of the elements themselves or from homologous recombination between elements have been reported (SAEDLER *et al.* 1980; DE MASSY *et al.* 1984; MILLER *et al.* 1984). Many rearrangements resulting directly from transposable elements or from recombination between them are probably lethal. To the extent that the risk of undergoing such rearrangements increases with the number of elements, there will be selection for strains that have not accumulated too many of the sequences. However, the selection pressure resulting from these processes is impossible to quantify at this time.

The relation between IS4 and the unique sequences that flank it in the genome of *E. coli* K12 was also unexpected. There is a significant association between the presence or absence of IS4 and the presence or absence of the LFS among strains of the reference collection (the contingency $\chi^2 = 17.4$ with 1 d.f., which is highly significant). There is also a significant association between the number of IS4 and the presence or absence of the LFS. Among strains containing IS4, strains also containing the LFS have an average of 7.13

± 0.95 copies of IS4, and those without the LFS have an average of 2.77 ± 1.02 copies of IS4.

The flanking sequences have the characteristics expected of sequences contained within a transposable element. They are found on plasmids, they are found in a minority of the strains and they occur in two copies in one strain. Although there is the formal possibility that the flanking sequences are present in all strains but have undergone sufficient divergence in some of them to prevent hybridization with the probes derived from *E. coli* K12, this seems very unlikely because the hybridization signal in Southern blots was either very strong or completely absent, with no cases of weak or ambiguous signals such as would be expected among sequences that had undergone varying degrees of genetic divergence.

The evidence given in Table 3 implies that the IS4-related sequences have transposed into some strains recently. In particular, strains 49 and 50 are closely related, as evidenced by the identity in the size of their IS5-containing restriction fragments (GREEN *et al.* 1984); yet strain 50 contains LFS and 11 copies of IS4, while strain 49 contains neither the flanking sequences nor IS4. Considering that the deletion of insertion sequences occurs at a rate several orders of magnitude smaller than the transposition rate (KLECKNER 1981), it is much more likely that strain 50 has acquired the LFS and IS4 than that strain 49 has lost both the flanking sequences and many copies of IS4. As the flanking sequences in strain 50 include IS4, it is probable that a plasmid carrying the flanking sequences and IS4 entered strain 50, the flanking sequences and IS4 were transposed into the chromosome and the plasmid was subsequently lost. There must also have been a proliferation of IS4 in the strain.

The finding that the 6.0 kb insert in pBRK10 contains DNA from two insertion sequences is significant. The genome of *E. coli* K12, approximately 3.9×10^6 bp (BRENNER *et al.* 1972), contains about 39 copies of the known insertion sequences (HU and DEONIER 1981; SCHONER and SCHONER 1981; CASPERS *et al.* 1984). These insertion sequences constitute perhaps 1% of the total size of the genome, so they may be considered as dimensionless geometrical points scattered in some spatial distribution around the chromosome. If insertion sequences were distributed at random, then the average number of elements in a random sequence of 3.9 kb would be 0.039. (The amount 3.9 kb is the approximate amount of non-IS DNA in the insert in pBRK10.) Since pBRK10 had been selected to contain IS4, we need to calculate the probability of finding two or more IS sequences in an insert of size 3.9 kb, given that the insert contains at least one IS sequence; this conditional probability is 0.038, assuming a random (Poisson) spatial distribution. The presence of both IS4 and IS2 in pBRK10 is therefore statistically significant at the 5% level. This finding could result if the DNA sequences in the IS4 region of the chromosome of *E. coli* K12 were a hot spot for the insertion of IS sequences. This hypothesis provides some biological significance to an observation of KLAER and STARLINGER (1980). In their colony hybridizations using an *E. coli* K12 library, which resulted in the recovery of the 6.0 kb sequence containing IS4 and part of IS2

(see Figure 1), another clone was recovered that was identical except that, at coordinate 1.5, it contained an insertion of IS*1*.

The total size of the postulated transposon containing IS*4* and the flanking sequences is unknown. Thus, it is not known whether the LFS includes DNA that is not part of the transposon. However, while there is considerable variation in fragment size among strains, there is also much similarity, which suggests that the restriction enzymes used cleave only within the transposon. For example, strains 11, 12 and 14 yield the same size of hybridizing fragment with *Eco*RI digestion, and strains 11, 61, 62, 64 and 70 all yield the same size hybridizing fragments with *Bgl*II digestion, even though these strains may be very different in their electrophoretic profiles. If the hybridized restriction fragments contain only transposon DNA, then the transposon in *E. coli* K12 must be larger than 10 kb, and in strains 38 and 39 it must be larger than 18 kb. Transposons of this size have been described (HEFFRON 1983). Because the transposable element seems to be a hot spot for insertion of other transposable elements, the diversity in size of hybridizing fragments might result from the incorporation of various other insertion sequences, which could then be disseminated as part of the larger transposon.

We thank H. OCHMAN, R. K. SELANDER, D. E. BERG, P. STARLINGER and R. E. WOLF, JR. for providing strains. The research was supported by National Institutes of Health grant GM 30201 to D. HARTL and D. DYKHUIZEN.

LITERATURE CITED

ANDERSON, D. G. and L. L. MCKAY, 1983 Simple and rapid method for isolating large plasmid DNA from lactic Streptococci. Appl. Environ. Microbiol. **46:** 549–552.

BRENNER, D. J., G. R. FANNING, F. J. SKERMAN and S. FALKOW, 1972 Polynucleotide sequence divergence among strains of *Escherichia coli* and closely related organisms. J. Bacteriol. **109:** 953–965.

CALOS, M. P. and J. H. MILLER, 1980 Transposable elements. Cell **20:** 579–595.

CAMPBELL, A., 1981 Evolutionary significance of accessory DNA elements in bacteria. Annu. Rev. Microbiol. **35:** 55–83.

CAMPBELL, A., 1983 Transposons and their evolutionary significance. pp. 258–279. In: *Evolution of Genes and Proteins*, Edited by M. NEI and R. K. KOEHN. Sinauer Associates, Sunderland, Massachusetts.

CAMPBELL, A., D. E. BERG, D. BOTSTEIN, E. M. LEDERBERG, R. P. NOVICK, P. STARLINGER and W. SZYBALSKI, 1979 Nomenclature of transposable elements in prokaryotes. Gene **5:** 197–206.

CASPERS, P., B. DALRYMPLE, S. IIDA and W. ARBER, 1984 IS*30*, a new insertion sequence of *Escherichia coli*. Mol. Gen. Genet. **196:** 68–73.

DE MASSY, B., J. PATTE, J.-M. LOUARN and J.-P. BOUCHE, 1984 *OriX*: a new origin of replication in *E. coli* (Correction). Cell **38:** 333.

DOOLITTLE, F. W. and C. SAPIENZA, 1980 Selfish genes, the phenotype paradigm and genome evolution. Nature **284:** 601–603.

GHOSAL, D., H. SOMMER, and H. SAEDLER, 1979 Nucleotide sequence of the transposable DNA-element IS2. Nucleic Acids Res. **6:** 1111–1122.

GREEN, L., R. D. MILLER, D. E. DYKHUIZEN and D. L. HARTL, 1984 Distribution of DNA

insertion element IS5 in natural isolates of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81:** 4500–4504.

HARSHMAN, L. and M. RILEY, 1980 Conservation and variation of nucleotide sequences in *Escherichia coli* strains isolated from nature. J. Bacteriol. **144:** 560–568.

HARTL, D. L. and D. E. DYKHUIZEN, 1984 The population genetics of *Escherichia coli*. Annu. Rev. Genet. **18:** 31–68.

HARTL, D. L., D. E. DYKHUIZEN and D. E. BERG, 1984 Accessory DNAs in the bacterial gene pool: playground for coevolution. Ciba Found. Symp. **102:** 233–245.

HARTL, D. L., D. E. DYKHUIZEN, R. D. MILLER, L. GREEN and J. DE FRAMOND, 1983 Transposable element IS*50* improves growth rate of *E. coli* cells without transposition. Cell **35:** 503–510.

HEFFRON, F., 1983 Tn*3* and its relatives. pp. 223–260. In: *Mobile Genetic Elements*, Edited by J. A. SHAPIRO. Academic Press, New York.

HU, M. and R. C. DEONIER, 1981 Comparison of IS*1*, IS2 and IS*3* copy number of *Escherichia coli* strains K-12, B and C. Gene **16:** 161–170.

KLAER, R., S. KUHN, E. TILLMANN, H.-J. FRITZ and P. STARLINGER, 1981 The sequence of IS*4*. Mol. Gen. Genet. **181:** 169–175.

KLAER, R. and P. STARLINGER, 1980 IS*4* at its chromosomal site in *E. coli* K12. Mol. Gen. Genet. **178:** 285–291.

KLECKER, N., 1981 Transposable elements in prokaryotes. Annu. Rev. Genet. **15:** 341–404.

MILKMAN, R., 1973 Electrophoretic variation in *E. coli* from natural sources. Science **182:** 1024–1026.

MILLER, R. D., D. E. DYKHUIZEN, L. GREEN and D. L. HARTL, 1984 Specific deletion occurring in the directed evolution of 6-phosphogluconate dehydrogenase in *Escherichia coli*. Genetics **108:** 765–772.

MONTGOMERY, E. A. and C. H. LANGLEY, 1983 Transposable elements in Mendelian populations. II. Distribution of copia-like elements in natural populations. Genetics **104:** 473–483.

NASOFF, M. S. and R. E. WOLF, JR., 1980 Molecular cloning, correlation of genetic and restriction maps and determination of the direction of transcription of *gnd* in *Escherichia coli*. J. Bacteriol. **143:** 731–741.

NYMAN, K., H. OHTSUBO, D. DAVIDSON and E. OHTSUBO, 1983 Distribution of insertion element IS*1* in natural isolates of *Escherichia coli*. Mol. Gen. Genet. **189:** 516–518.

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

ORGEL, L. E. and F. H. C. CRICK, 1980 Selfish DNA: the ultimate parasite. Nature **284:** 604–607.

SAEDLER, H., G. CORNELIS, J. CULLUM, B. SCHUMACHER and H. SOMMER, 1980 IS*1*-mediated DNA rearrangements. Cold Spring Harbor Symp. Quant. Biol. **45:** 93–98.

SCHONER, B. and R. G. SCHONER, 1981 Distribution of IS*5* in bacteria. Gene **16:** 347–352.

SIEGEL, S., 1956 *Nonparametric Statistics*. McGraw-Hill, New York.

SOUTHERN, E. M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 505–517.

STROBEL, E., P. DUNSMUIR and G. M. RUBIN, 1979 Polymorphism in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell **17:** 429–439.

Communicating editor: W. W. ANDERSON