# IS*103*, a New Insertion Element in *Escherichia coli*: Characterization and Distribution in Natural Populations

**Barry G. Hall,* Laura L. Parker,* Paul W. Betts,* Robert F. DuBose,† Stanley A. Sawyer† and Daniel L. Hartl†**

**Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269-3044, and †Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110*

Manuscript received July 15, 1988
Accepted for publication November 25, 1988

ABSTRACT

IS*103* is a previously unknown insertion sequence found in *Escherichia coli* K12. We have sequenced IS*103* and find that it is a 1441-bp element that consists of a 1395-bp core flanked by imperfect 23-bp inverted repeats. IS*103* causes a 6-bp duplication of the target sequence into which it inserts. There is a single copy of IS*103* present in wild-type *E. coli* K12 strain HfrC. In strain χ342 and its descendents there are two additional copies, one of which is located within the *bglF* gene. IS*103* is capable of excising from within *bglF* and restoring function of that gene. IS*103* exhibits 44% sequence identity with IS*3*, suggesting that the two insertion sequences are probably derived from a common ancestor. We have examined the distribution of IS*103* in the chromosomes and plasmids of the ECOR collection of natural isolates of *E. coli*. IS*103* is found in 36 of the 71 strains examined, and it strongly tends to inhabit plasmids rather than chromosomes. Comparison of the observed distribution of IS*103* with distributions predicted by nine different models for the regulation of transposition according to copy number and of the effects of copy number on fitness suggest that transposition of IS*103* is strongly regulated and that it has only minor effects on fitness. The strong clustering of IS*103* within one phylogenetic subgroup of the *E. coli* population despite its presence on plasmids suggests that plasmids tend to remain within closely related strains and that transfer to distantly related strains is inhibited.

INSERTION sequences are mobile genetic elements that are found in prokaryotes. They range in size from about 1 to 2 kilobase pairs (kb) and consist of short inverted repeat sequences that flank at least one open reading frame that, in some cases, is known to encode a transposase, the enzyme that mediates transposition. Transposons are more complex mobile structures, many of which consist of a pair of insertion sequences (IS elements) that flank a central region of DNA that often contains an antibiotic resistance gene; transposons are capable of transposing as a unit (CRAIG and KLECKNER 1987). Both IS elements and transposons are capable of inserting into and thereby inactivating genes. IS elements are also capable of activating cryptic genes (REYNOLDS, FELTON and WRIGHT 1981).

Insertion sequences have been studied most thoroughly in *E. coli* K12, within which seven IS elements, IS*1*, IS*2*, IS*3*, IS*4*, IS*5*, IS*30* and IS*186* have been identified. A larger (5.7 kb) mobile element that carries no known resistance genes, γδ or Tn*1000*, is present on the F- plasmid of male strains. See DEONIER (1987) for a recent review of the number and locations of these elements in *E. coli* K12.

IS elements are of particular interest to evolutionary biologists because they can, due to transposition, increase in frequency in the genome; *i.e.*, they may behave as selfish or parasitic DNA. It is presumed that as the copy number of an IS element increases there is eventually a deleterious effect on fitness, either because IS elements disrupt genes into which they insert, or for other reasons. The relative roles of parasitism, selection, and regulation of transposition in determining copy number are poorly understood, but recent studies of the distribution of IS elements in natural populations have begun to shed some light on these topics (GREEN *et al.* 1984; DYKHUIZEN *et al.* 1985; SAWYER and HARTL 1986; SAWYER *et al.* 1987).

We recently reported an unusual case in which the cryptic *bgl* operon of *E. coli* K12 was activated by the excision of a 1.4-kb insert from within the *bglF* gene (PARKER, BETTS and HALL 1988). Because a spontaneous precise excision of the insert activated the *bgl* operon, we speculated that the insert might be an IS. Because the pattern of restriction sites within the insert was distinct from that of all known *E. coli* IS elements we suggested that it might be a new IS element. Here we report that the 1.4-kb insert is a previously unidentified IS element, and we discuss the properties and distribution of that element in natural *E. coli* populations.

## MATERIALS AND METHODS

**Culture media and conditions:** Cultures were grown at 30° or 37° with aeration either in L-broth (MILLER 1972) or in minimal media. Minimal media contained phosphate buffered mineral salts solution (HALL and BETTS 1987) with 1.5% agar as a solidifying agent. Sugars were added at a concentration of 0.1% (w/v) except for glucose which we used at 0.2% (w/v). When required, amino acids were added at a concentration of 100 μg/ml, kanamycin was used at a concentration of 75 μg/ml, and ampicillin was used at a concentration of 100 μg/ml. MacConkey salicin indicator plates contained 1% (w/v) salicin and were prepared according to the manufacturer (Difco).

**Molecular techniques:** Methods for the preparation of genomic DNA, plasmid DNA, isolation and labeling of probe DNA, and DNA/DNA hybridizations were previously described (HALL and BETTS 1987). Preparation of competent cells and transformations were as previously described (HALL, BETTS and KRICKER 1986). DNA fragments to be sequenced were subcloned into plasmids pBluM and pBluP. Fragments were sequenced directly from the double-stranded DNA using primers provided by Stratagene, Inc., by a modification of the dideoxy method of SANGER, NICKLEN and COULSON (1977) using $^{35}$S-dATP.

***In vivo* cloning:** The mini-Mu derivative Mud5005 was used as described in GROISMAN and CASADABAN (1986). The helper phage employed was Mu cts62 pAp5 which carries an ampicillin resistance determinant (AKROYD *et al.* 1984).

**Other plasmids:** The plasmids pBluP and pBluM are 3.0-kb high copy number vectors that carry multiple cloning site cartridges and are products of Stratagene, Inc., of San Diego, California. Plasmid pAR6 is a pBR322 based plasmid that carries the wild type *bgl* operon (REYNOLDS, FELTON and WRIGHT 1981).

**Genetic nomenclature for the *bgl* operon:** Until recently *bglC* was used to designate the gene specifying the β-glucoside transport system, and *bglS* the gene specifying a positive regulatory locus (PRASAD and SCHAEFLER 1974). Two recent papers (MAHADEVAN, REYNOLDS and WRIGHT 1987; SCHNETZ, TOLOCZYKI and RAK 1987) reversed those definitions. In order to avoid further confusion, BARBARA J. BACHMANN, curator of the *E. coli* Genetic Stock Center, after consultation with the CGSC Advisory Committee, has renamed these loci (B. BACHMANN, personal communication). In this paper we use the new designations as they will appear in the next edition of the linkage map of *E. coli*. *bglF* now designates the β-glucoside-specific enzyme II of the phosphotransferase system, *i.e.*, the transport protein. *bglG* now designates the gene encoding the positive regulator of expression of the *bgl* operon.

***E. coli* strains:** The *E. coli* K12 strains used in this study were: strain HfrC (HfrC *metB1 relA spoT*) and strain χ342 (HfrC *proC metB1 relA spoT* λ$^-$ *bglF*::IS*103*), both obtained from the *E. coli* Genetic Stock Center, strain JF201 (F$^-$ *lacZ*Δx74, Δ*bgl-pho*, *ara*, *thi*, *gyrA*) obtained from A. WRIGHT, and strain 1011A (a *bglR*$^+$ mutant derived from χ342 (PARKER, BETTS and HALL 1988). The natural isolates of *E. coli* that make up the ECOR collection are described in detail in OCHMAN and SELANDER (1984).

**Screening the ECOR collection for IS*103*:** Chromosomal and plasmid DNA were prepared from strains of the ECOR collection, digested with restriction endonuclease *EcoRI*, and the fragments were separated by agarose gel electrophoresis. The fragments were transferred to nitrocellulose or nylon membranes by blotting. The preparation of the filters was described in detail in SAWYER *et al.* (1987). The identical filters used in that study were reprobed with a fragment of IS*103* for the present study.

## RESULTS

**Cloning IS*103*:** The *bglF* gene of strain χ342 and its descendents contains a 1.4-kb insertion of foreign DNA (PARKER, BETTS and HALL 1988). For the sake of convenience we refer to that insert as IS*103*, and evidence presented later in this paper shows that it is indeed an authentic insertion sequence.

To clone IS*103* strain χ342 was lysogenized with the mini-mu vector Mud5005 and with the helper phage MupAp5 by selecting for kanamycin and ampicillin resistance respectively. Lysates produced by thermal induction were used to infect the Mu sensitive Salmonella strain SL4213 (GROISMAN and CASADABAN 1987), and kanamycin resistant transfectants were selected. Kanamycin-resistant colonies all carry a Mud5005 based plasmid that contains, on average, about 15 kb of insert DNA. Kanamycin-resistant colonies were transferred to Gene Screen Plus filters and hybridized to a 3.5-kb *Hin*dIII fragment of the *bgl* operon (derived from plasmid pAR6) that had been labeled with $^{35}$S-dATP. Colonies that hybridized to the 3.5 kb *Hin*dIII probe were subsequently hybridized to a 1.0 kb *Hin*dIII fragment and a 3.0 kb *Hpa*I fragment that flank the *bglF* gene. One colony that hybridized to all three probes and therefore contained *bglF*, *bglG* and *bglB* was retained and the plasmid was designated pUF713.

Plasmids pAR6 and pUF713 were digested with the restriction endonucleases *Nci*I, *Hin*cII, and *Dde*I and with pairs of the three enzymes. Fragments were separated by agarose gel electrophoresis, transferred to Gene Screen Plus blotting membrane, and probed with a *Hin*dIII-*Hpa*I fragment of pAR6 that covers bp 1771–4783 of the *bgl* operon [base pairs of the *bgl* operon are numbered as in SCHENTZ, TOLOCZYKI and RAK 1987]. It was determined that IS*103* was inserted between the *Hin*fI site at bp 2637 and the *Dde*I site at bp 2971 of the *bgl* operon. IS*103* contained no *Nci*I sites, and was entirely contained with a 3.1-kb *NciI* fragment bounded by bp 1904 and 3642 of the *bgl* operon. That *Nci*I fragment was excised from pUF713, the ends were filled in with Klenow enzyme, and the fragment was ligated into the *Eco*RV site of plasmid pBluM. The resulting plasmid was designated pUF720.

Mapping of pUF720 showed that IS*103* was entirely contained within a *Hin*cII fragment that extends from bp 2675–2814 of the *bgl* operon. That fragment was excised from pUF720 and ligated into the *Hin*cII site of pBluM to produce plasmid pUF722. A restriction map of pUF722 showing the sites that were used for sequencing is shown in Figure 1.

**Sequencing IS*103*:** The sequence of the entire *bgl*

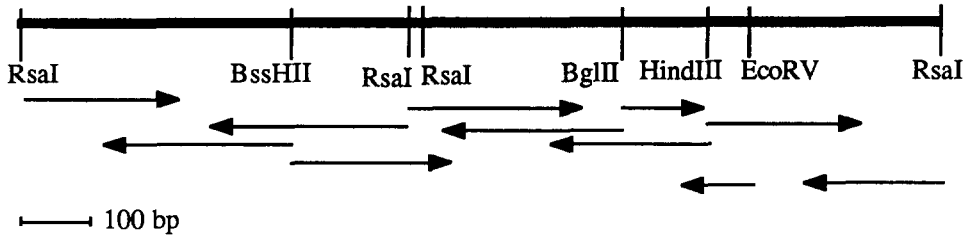


FIGURE 1.—Restriction map of IS*103*. The entire sequence is 1441 bp long. Only those sites used for sequencing are shown. Arrows below the map show the direction and extent of sequencing from each site.

operon is known (SCHENTZ, TOLOCZYKI and RAK 1987). Sequencing of fragments derived from pUF722 showed that IS*103* had inserted into the *bglF* gene between bp 2713 and 2718 of the operon. The complete 1441-bp sequence of IS*103* is shown in Figure 2. Inspection of this sequence and comparison with that of the wild-type *bglF* gene reveals that IS*103* exhibits two properties that are characteristic of insertion elements: the termini of the element are 23-bp imperfect inverted repeats, and the element has generated a 6-bp duplication of target DNA sequences during insertion.

**Comparison with known insertion sequences:** The sequence of IS*103* is distinct from the sequence of any of the previously identified IS elements of *E. coli*. We have used the GAP program of the University of Wisconsin Genetics Computer Group (DEVEREUX, HAEBERLI and SMITHIES 1984) to align IS*103* with IS*1*, IS*2*, IS*3*, IS*4*, IS*5*, IS*30* and IS*186*. Significant similarity was detected only with IS*3*. IS*103* exhibited 44% DNA sequence similarity with IS*3* over 1221 bp (gaps were excluded from the comparison), suggesting that IS*103* and IS*3* are derived from a common ancestor.

MATSUTANI *et al.* (1987) have recently reported the discovery of several novel IS elements in *Shigella sonnei*. One of those elements, IS*600*, was 1264 bp long and showed 44% sequence similarity with IS*3*. The sequence of IS*103* is very different from that of IS*600* and the ends of IS*103* and IS*600* bear no resemblance to each other. These results suggest that IS*3*, IS*103* and IS*600* are members of a single family of IS elements that exist within members of the family Enterobacteriaceae.

**Copy number of IS*103* in *E. coli* K12:** The copy number of IS*103* was determined from blots of genomic DNA from two *E. coli* K12 strains that had been digested with two restriction enzymes that do not cut within IS*103* (Figure 3). The results show that there is one copy of IS*103* in *E. coli* K12 strain HfrC, and three copies in strain χ342. *Nci*I digests of other *E. coli* K12 strains have shown from 2 to 6 copies of IS*103* present with the 3.3 kb *Nci*I fragment being the only band that is always present (data not shown). The observation that we detect a single copy of IS*103* in strain HfrC shows that, despite the limited homology between IS*103* and IS*3*, our probe does not cross-hybridize to IS*3* which is present in 5 to 6 copies in *E. coli* K12 (DEONIER 1987).

Strain χ342 is two mutational steps removed from strain HfrC (PARKER, BETTS and HALL 1988), both of those steps involved UV mutagenesis. The observations that: (1) χ342 contains two additional copies of IS*103*, and (2) other *E. coli* K12 strains with unknown histories, but which may well have been subjected to UV mutagenesis, have additional copies of IS*103*, raise the intriguing possibility that IS*103* may transpose in response to UV or to induction of the SOS system. A plausible alternative is that the transpositions occurred during long-term storage in agar stabs as has been reported for other IS elements (RALEIGH and KLECKNER 1984).

The observation that strain χ342 contains two additional copies of IS*103* when compared with its grandparent strain HfrC indicates that it undergoes transposition and provides additional evidence that IS*103* is indeed an insertion sequence. The designation IS*103* was assigned by the Plasmid Reference Center, Stanford University.

**Distribution of IS*103* in natural isolates of *E. coli*:** The ECOR collection consists of 72 natural isolates of *E. coli* obtained from a variety of human and animal hosts and from a diverse set of geographic locations (OCHMAN and SELANDER 1984). The strains have been classified and phylogenetic relationships determined based upon the electrophoretic mobilities of several enzymes (SELANDER, CAUGENT and WHITTAM 1987). Recently SAWYER *et al.*( 1987) studied the distribution and abundance of IS*1*, IS*2*, IS*3*, IS*4*, IS*5*, and IS*30* in 71 strains of the ECOR collection by probing *Eco*RI digests of chromosomal and of plasmid DNA from each strain. The same filters that were used in that study have now been probed with the 800-bp fragment of IS*103* to determine its distribution and abundance in natural isolates. Table 1 shows the distribution of IS*103* copies in the chromosome and in plasmids from these strains.

Figure 4 shows the distribution of IS*103* mapped onto that phylogeny of the ECOR collection (SELANDER, CAUGENT and WHITTAM 1987). IS*103* is strongly based toward group A of that phylogeny, with 22 of the 25 members of that group carrying IS*103*, while only 14 of the remaining 46 strains outside of group A carry IS*103*.

```
   1  GTACTGCACC CATTTTGTTG GACGATGAAA TGGAATAGCC CCTAATATGT
  51  CAAAGCCAAA ATACCCTTTT GAAAAGCGCC TTGAAGTCGT GAATCACTAC
 101  TTCACAACTG ATGATGGTTA CAGGATCATC TCGGCACGTT TTGGTGTCCC
 151  CCGAACCCAG GTCAGGACAT GGGTTGCCCT CTATGAAAAA CATGGAGAAA
 201  AAGGTTTAAT TCCCAAACCT AAAGGCGTTA GTGCTGATCC AGAGTTGCGT
 251  ATTAAGGTCG TGAAAGCTGT GATCGAGCAG CACATGTCCC TTAATCAGGC
 301  TGCTGCTCAC TTTATCCTTG CTGGTAGTGG TTCTGTAGCC AGGTGGCTGA
 351  AGGTCTATGA AGAGCGCGGA GAAGCTGGTT TACGCGCGCT CAAGATTGGC
 401  ACCAAAAGAA ACATTGCAAT ATCAGTTGAT CCAGAAAAAG CGGCATCAGC
 451  ATTGGAGCTG TCAAAAGACC GACGCATTGA GGATCTTGAA AGGCAAGTTC
 501  GATTTCTTGA AACGCGGCTT ATGTATCTAA AAAAGCTGAA AGCCTTAGCT
 551  CATCCCACGA AAAAGTGAAA GTACTCAACG AGCTAAGGCA GTTTTATCCT
 601  CTTGATGAGC TTCTCAGGGC TGCGGAGATA CCGCGCAGTA CGTTTTATTA
 651  TCATCTAAAG GCTCTCAGCA AGCCTGACAA GTATGCGGAC GTTAAAAAGC
 701  GTATTAGTGA GATTTATCAC GAGAATAGAG GCCGATACGG ATACCGTAGG
 751  GTAACGCTGT CTCTTCATCG AGAAGGGAAA CAGATTAACC ATAAAGCTGT
 801  TCAGCGCCTG ATGGGAACCC TCTCACTTAA AGCAGCGATT AAGGTCAAGC
 851  GATACCGCTC TTACAGAGGA GAGGTAGGGC AAACCGCCCC TAATGTTCTC
 901  CAAAGAGATT TCAAGCTACG CGGCCAAACG AGAAGTGGGT TACCGATGTT
 951  ACTGAATTTG CAGTCAATGG GCGCAAGCTG TATTTGTCTC CAGTAATAGA
1001  TCTCTTCAAC AACGAAGTTA TTTCTTACAG CCTTTCGGAA AGACCAGTGA
1051  TGAACATGGT TGAGAATATG CTCGATCAGG CATTCAAAAA GCTTAACTCC
1101  TCACGAGCAT CCTGTTCTGC ACTCTGACCA GGGATGGCAG TATCGTATGA
1151  GAAGATATCA AAATATCCTT AAAGAACATG GTATTAAACA AAGCATGTCC
1201  AGAAAAGGCA ATTGTCTGGA TAATGCTGTG GTGGAGTGTT TCTTTGGAAC
1251  CTTAAAGTCG GAGTGTTTTT ATCTTGATGA GTTCAGTAAT ATAAGCGAAC
1301  TGAAGGATGC TGTTACGGAA TATATTGAAT ACTACAACAG CAGAAGAATT
1351  AGCCTGAAAT TAAAAGGTCT GACTCCAATT GAATATCGGA ATCAGACCTA
1401  TATGCCTCGT GTTTAACTGT CCAACTTTTT GGGGTCAGTA C
```

FIGURE 2.—Sequence of IS*103*. The arrows at the termini indicate the inverted repeat regions, and the 6-bp perfect repeat termini are indicated by a thicker line.

IS*103* occurs much more frequently in plasmids than do other IS sequences studied in the ECOR collection. For IS*1–5* + IS*30*, 1056 elements were found in chromosomes *vs.* 117 elements on plasmids. For IS*103*, 53 elements are found in chromosomes and 79 on plasmids. For IS*103* the $\chi^2 = 231$ with 1 d.f. and $P$, the probability that IS*103* is randomly distributed between chromosomes and plasmids, is $\ll 10^{-6}$. The only other insertion sequence that is biased toward plasmids is $\gamma\delta$, an element that was found only in plasmids but that was present in only four strains.

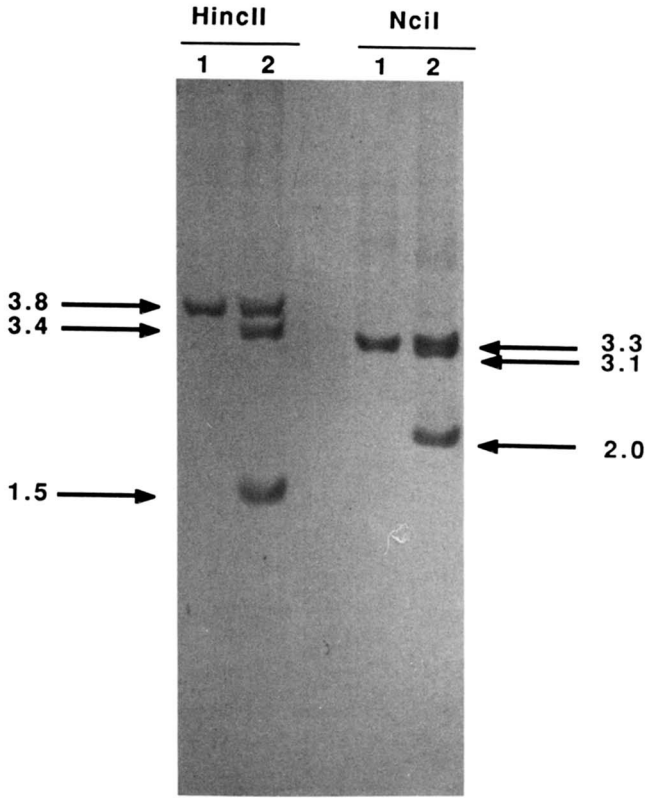


FIGURE 3.—Copy number of IS*103* in *E. coli* K12. Genomic DNA digested with indicated enzymes was electrophoresed, transferred to a nylon membrane, and hybridized to the $^{35}$S-labeled 801-bp *Rsa*I fragment of IS*103*. In each set lane 1 is strain HfrC, and lane 2 is strain 1011A (*bglF*::IS*103*). Sizes of the bands were estimated by comparison with *Hin*dIII fragments of λ DNA.

SAWYER *et al.* (1987) used a continuous-time multitype branching process to model copy number of insertion sequences with hosts. They considered models in which the transposition rate $T(n)$ [$T$ = transpositions per cell per generation over all copies of the element] varied according to the copy number, $n$, of the IS element. The choices considered were $T(n) = T$ (constant), $T(n) = Tn^{1/2}$, $T(n) = Tn$, $T(n) = Tn^2$, $T/n$, and $T/n^{1/2}$. Individuals carrying insertion sequences were assumed to have either a slower growth rate or higher death rate (or both) than uninfected individuals, and death rate $D(n)$ was taken to vary according to copy number using choices analogous to transposition rate choices. In all cases the potential number of target sites for transposition is assumed to be much higher than the number of elements, i.e. the targets are not near saturation; and any regulation of the element is assumed to be the same in all of the strains examined. Nine models with different choices for $T(n)$ and $D(n)$ were considered, and the observed distribution of IS copies in the ECOR collection was compared with that predicted by each of the models. These comparisons were based upon the chromosomal copies of the IS elements. For IS*1–5* + IS*30* the inclusion of plasmid copies made essentially no difference, in part because the plasmid copies were so much less frequent than chromosomal copies.

TABLE 1

**Distribution of IS*103* in the ECOR collection**

| ECOR strain No. | Chromosomal copies | Plasmid copies | ECOR strain No. | Chromosomal copies | Plasmid copies |
|---|---|---|---|---|---|
| 1 | 1 | 4 | 23 | 1 | 1 |
| 2 | 1 | 0 | 24 | 0 | 4 |
| 3 | 6 | 0 | 25 | 1 | 3 |
| 4 | 4 | 3 | 31 | 0 | 5 |
| 5 | 2 | 1 | 35 | 1 | 6 |
| 6 | 4 | 1 | 36 | 1 | 4 |
| 10 | 1 | 7 | 38 | 1 | 1 |
| 11 | 1 | 0 | 39 | 1 | 1 |
| 12 | 1 | 2 | 40 | 1 | 4 |
| 13 | 2 | 0 | 41 | 1 | 4 |
| 14 | 2 | 1 | 46 | 3 | 0 |
| 15 | 5 | 1 | 50 | 1 | 7 |
| 16 | 1 | 0 | 59 | 1 | 0 |
| 17 | 2 | 1 | 63 | 0 | 2 |
| 18 | 1 | 2 | 64 | 0 | 2 |
| 19 | 2 | 1 | 71 | 0 | 2 |
| 20 | 1 | 3 | 72 | 0 | 2 |
| 21 | 1 | 2 | All others | 0 | 0 |
| 22 | 2 | 2 | **Totals** | **53** | **79** |

Considering the high frequency of IS*103* in plasmids, on the other hand, it seems reasonable to analyze the data based upon chromosomal + plasmid copies. One problem with this approach is that it does not correct for differences in plasmid abundance (*e.g.*, high copy *vs.* low copy number plasmids). It is reasonable to treat each plasmid band as a single copy in the genome because most IS elements in IS*1–5* + IS*30* occur in large, presumably single copy, plasmids (SAWYER *et al.*, 1987). Under that assumption, the best fitting models are the one in which $T(n) = T/n$, and $D(n)$ = constant (model HC), and the one in which $T(n) = T/n^{1/2}$ and $D(n) = Dn^{1/2}$ (model DR). For IS*103* all other models except CR [a model in which $T(n)$ is a constant, and $D(n) = Dn^{1/2}$] can be rejected, and CR does not fit nearly as well as do HC and DR. The differences in log likelihood among the three models that fit are: 0.00 (HC), 0.17 (DR) and 2.05 (CR). These correspond to the numbers in SAWYER *et al.* (1987) Table 5B. Table 2 gives the values of parameters corresponding to Table 6 of SAWYER *et al.* (1987) for IS*103*, and Table 3 gives the fitted distributions of copy number for IS*103* using the maximum likelihood estimates of the parameters.

## DISCUSSION

The DNA inserted into the middle of the *bglF* gene in strain χ342 and its descendents (PARKER, BETTS and HALL 1988) is an authentic insertion sequence by several criteria: (1) it inserted into *bglF* by transposi-

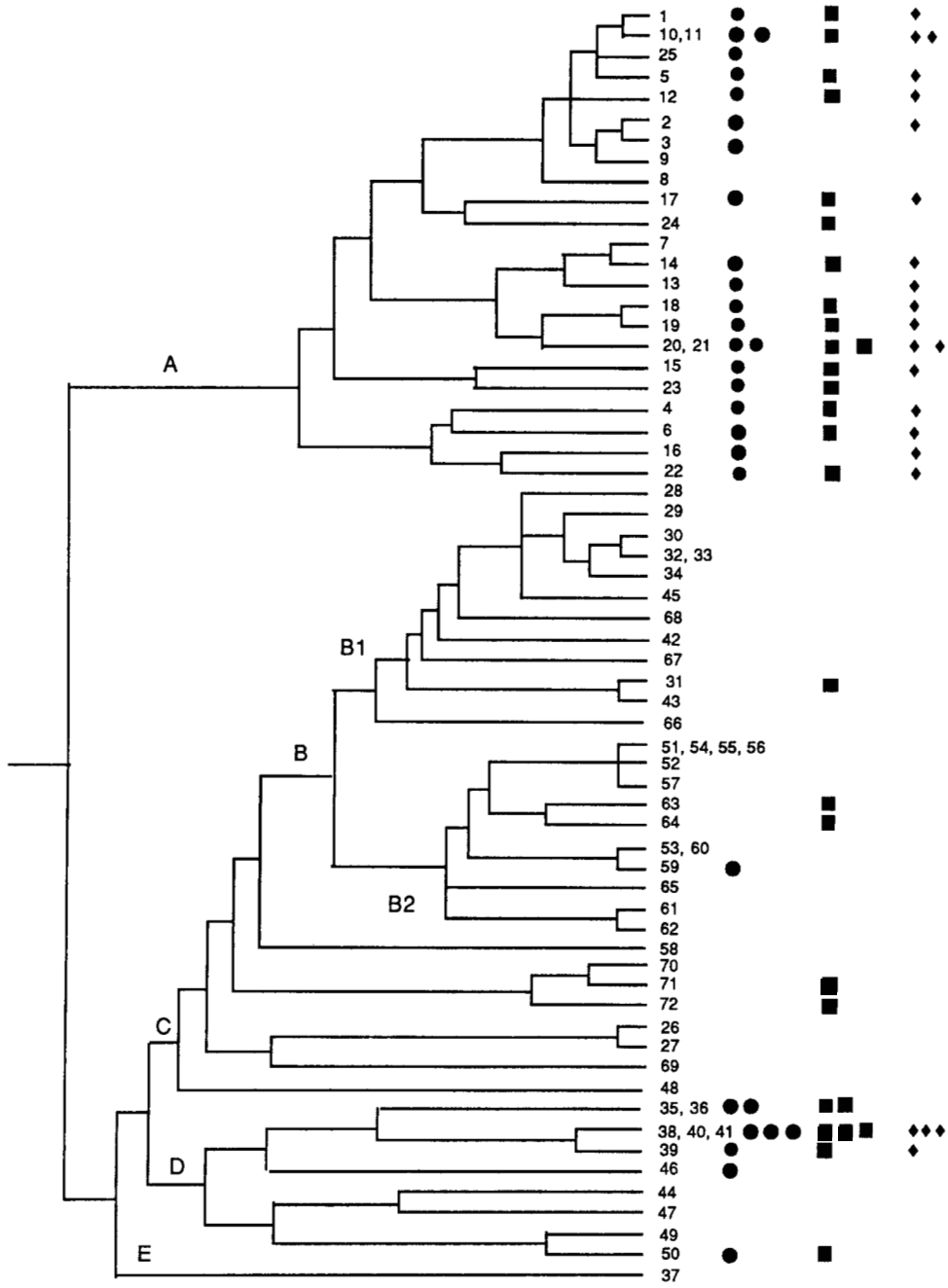


FIGURE 4.—Distribution of IS*103* within the ECOR collection of natural isolates of *E. coli*. The dendrogram is taken from SELANDER, CAUGENT and WHITTAM 1987. Circles indicate the presence of at least one copy of IS*103* in the chromosome, squares indicate at least one copy of IS*103* on a plasmid. Diamonds indicate the presence of IS*103* within a 13.5-kb *Eco*RI fragment of the chromosome (see text).

**TABLE 2**

**Maximum likelihood estimates for grouped data**

| Model | $\mu$/D | $\mu$/AvD | T/D | T/$\mu$ |
|---|---|---|---|---|
| HC | 0.50 | 0.50 | 3.70 | 7.70 |
| DR | 0.94 | 0.51 | 6.53 | 6.98 |

$\mu$ = probability of infection by IS*103*, *i.e.*, transformation rate from 0 to 1 copy of IS*103*. $T$ = transformation rate from $n$ to $n+1$ copies of the element. $D$ = death rate for hosts.

**TABLE 3**

**Fitted distribution for grouped data**

| | Copy No. | | | | | | |
|---|---|---|---|---|---|---|---|
| Model | 0 | 1 | 2 | 3 | 4 | >5 | *P* value |
| Observed | 41 | 19 | 6 | 0 | 2 | 2 | |
| HC | 41.0 | 14.3 | 9.7 | 3.7 | 1.0 | 0.3 | 0.78 |
| DR | 41.5 | 14.2 | 9.1 | 3.7 | 1.1 | 0.3 | 0.77 |

tion as demonstrated by the additional copy present in $\chi$342 compared with wild-type *E. coli* K12 and with strain HfrC, (2) it consists of a central core surrounded by a 23-bp inverted repeat of which the outer 6 bp

form a perfect inverted repeat, and (3) it causes a 6-bp flanking duplication of target DNA at the point of insertion.

IS*103* was found to be similar to IS*3* by two independent criteria: (1) it exhibits sufficient (44%) sequence similarity to IS*3* that it is reasonable to believe that IS*3* and IS*103* are descended from a common ancestor, and (2) a statistical analysis of the distribution of IS*103* within the ECOR collection, done independently of any knowledge of sequence similarity, indicates that both are subject to strong regulation of transposition.

IS*103* is present in 36 of the 71 ECOR collection *E. coli* strains examined. Among those strains that possess IS*103* there are an average of 1.47 copies per chromosome and 2.19 copies on plasmids. IS*103* shows an unusually strong preference for plasmids in that 60% of the copies of IS*103* were found on plasmids. In contrast, IS*1–5* and IS*30* averaged only 10% of the copies on plasmids, with the range being 3–16% (SAWYER *et al.* 1987).

IS*103* is highly, but not exclusively, concentrated in phylogenetic group A strains of SELANDER, CAUGENT and WHITTAM (1987). A number of the strains carry a copy of IS*103* within a 13.5-kb fragment of the chromosome (Figure 4). This particular copy of IS*103* is present in all but three of the members of group A that have IS*103* in the chromosome. Outside of group A, it is present only in strains 38, 39, 40 and 41 which also constitute a cluster. A reasonable interpretation of this distribution is that IS*103* invaded *E. coli*, presumably from a plasmid, shortly after the divergence of group A from the remainder of *E. coli*. Prior to divergence within that group a copy of IS*103* transposed to the chromosome. The presence of the 13.5-kb fragment copy in strains 38–41 suggests that the immediate progenitor of those strains received that portion of the chromosome as the result of recombination with a member of group A.

Plasmids are the most likely vector for the spread of IS elements, thus it is not surprising that only 2 of the IS*103* bearing strains outside of group A have IS*103* exclusively on the chromosome. These can be presumed to have acquired IS*103* via plasmid transfer and subsequently lost the plasmid.

Given the strong tendency of IS*103* to be found on plasmids, it is very surprising that it has spread so little into strains outside of group A. Because of the rapid spread of plasmids carrying antibiotic resistance determinants we tend to think of plasmids as spreading rapidly and freely throughout the *E. coli* population. The observed distribution of IS*103* strongly implies that plasmids tend to remain associated with closely related strains, and that their transfer to more distantly related strains is inhibited. This is consistent with the finding (HARTL *et al.* 1986) that plasmids are not distributed randomly with respect to bacterial genotype, and that it is very unclear how frequently conjugative plasmids transfer among nonpathogenic strains in natural environments.

The cluster consisting of strains 35, 36, 38–41 and 46 is a curiosity. It might have been the result of an invasion at the root of the upper portion of group D, but this is not consistent with the finding that strains 38–41 all carry the 13.5-kb band containing IS*103*. A more reasonable alternative interpretation is that IS*103* was introduced into the common ancestor of strains 38–41 by chromosomal recombination with a member of group A, and that subsequently IS*103* transposed to a resident plasmid which then spread to closely related strains 35, 36, 39 and 46.

Considering the distribution and abundance of IS*103* in plasmids and chromosomes together, the models that fit the data best are HC in which $T(n) = T/n$, and $D(n) =$ constant, and DR in which $T(n) = T/n^{1/2}$ and $D(n) = Dn^{1/2}$. These are the same models that fit best for IS*3*. Both models imply strong copy number regulation of transposition and little, if any, effect of copy number on fitness.

The mathematical models that imply strong regulation of transposition are consistent with the observation that another form of mobility, excision, appears to be strongly regulated by (unknown) physiological or environmental factors that make excision undetectable in growing cultures, but a frequent event in colonies that have reached a maximum size on MacConkey salicin plates.

Recently it was shown that IS*103* excises precisely from within *bglF* at a frequency of $<2 \times 10^{-12}$ per cell division in growing cells, and that excision is not detectable in old colonies on MacConkey plates without an added sugar. However, in old colonies on MacConkey plates containing salicin the excision frequency is in the range of $10^{-2}$ to $10^{-1}$ (HALL 1988).

**Note added in revision:** While this manuscript was being reviewed a paper describing the sequence of IS*150* appeared (SCHWARTZ, KRÖGER and RAK 1988). The sequence of IS*150* is, with minor discrepancies, identical to the sequence reported here for IS*103*. In an earlier paper SCHWARTZ, HERBERGER and RAK (1988) reported that IS*150* was observed to transpose to a single target site: bp 387–388 in IS*1*. The major differences between our sequences are that they report a sequence of 1443 bp and a target duplication of 3 bp (SCHWARTZ, HERBERGER and RAK 1988), while we report a sequence of 1441-bp and a 6-bp target duplication (Figure 5). The border sequences that we find in the *bgl* operon are remarkably similar to those reported in IS*1* (Figure 5). Because the sequences of the target genes are known both for IS*150* inserted into IS*1*, and for IS*103* inserted into *bglF*, the length of the target duplication should be unambiguous.

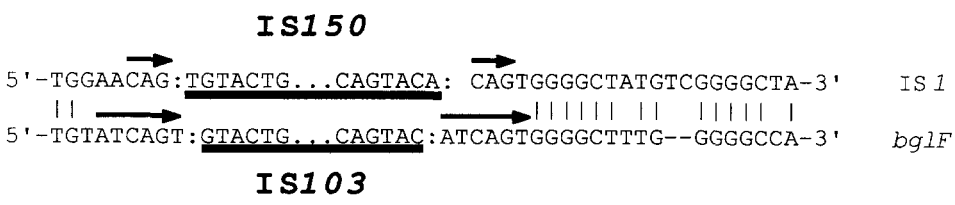


FIGURE 5.—Alignment of border sequence of IS*150* inserted into IS*1* with IS*103* inserted into *bglF*. A bar is drawn beneath each insertion sequence, and arrows are drawn above the target site direct repeat duplications. Vertical lines connect identical bases that lie *outside* of the insertion sequences and of the duplicated portions of the target sites.

SCHWARTZ, HERBERGER and RAK (1988) report that, depending upon the precise site of insertion of IS*150* into IS*1*, the target duplication may be 1, 3 or 5 bp. We can not reconcile our border sequences with any of those possibilities. Aside from the trivial explanation of sequencing errors, we are left with two interesting possibilities to explain these different results: IS*150* and IS*103* are slightly different variants that actually generate different target site repeat length, or IS*103/50* produces a target duplication of variable length.

The observation that the sequences surrounding the insertions sequences are so similar supports the notion (SCHWARTZ, HERBERGER and RAK 1988) that IS*150/103* possesses a strong target specificity for insertion. On the other hand, IS*1* is clearly not the only target for transposition of IS*150/103*. Although *E. coli* K12 carries 6–8 copies of IS*1* (DEONIER 1987), one of the two additional copies of IS*103* present in strain χ342 is inserted into *bglF* rather than into a resident IS*1*. Furthermore, ECOR strain 46 which carries no copies of IS*1* (SAWYER *et al.* 1987) carries three copies of IS*103* in its chromosome. Several other ECOR strains have IS*103* copy numbers that exceed IS*1* copy numbers in either chromosomes or plasmids.

The similarity of the two target sites brings up the interesting possibility that one of the assumptions common to all of the SAWYER *et al.* (1987) models, essentially unlimited targets, may be invalid. It is not yet clear how this possibility should be evaluated. For instance, in the "A group" strains (Figure 4), where IS*103* is most prevalent, the copy number of IS*1* is usually several times higher than the copy number of IS*103*, *i.e.*, in those cases the number of available target sites *is* in excess of the number of elements. It remains completely valid to say that the observed distribution of IS*103* is consistent with the distribution that is expected if transposition is strongly regulated and the assumptions of the model are met. Whether or not that distribution is, in fact, generated by the distribution of available target sites in members of the *E. coli* population will remain uncertain until more is known about the specific regulation of IS*103*.

Readers are referred to SCHWARTZ, HERBERGER and RAK (1988) and to SCHWARTZ, KRÖGER and RAK (1988) for details of transposition frequency, open reading frames, and detailed comparisons with other insertion sequences. We have retained the name IS*103* in this paper both because the work was done independently and because that designation was used in another paper (HALL 1988). Because the name IS*150* appeared first we recommend that, to avoid confusion, the designation IS*150* should be used in future reports and that the name IS*103* not be assigned in the future.

We are very grateful to N. KLECKNER for comments, discussion, and for correction of some misinterpretations. We are grateful to L. DEBESI for assistance in subcloning some of the fragments used for sequencing. This work was supported by U.S. Public Health Service grants GM37110 to B.G.H. and GM40322 to D.L.H.

## LITERATURE CITED

AKROYD, J., B. BARTON, P. LUND, S. M. SMITH, K. SULTANA and N. SYMONDS, 1984 Mapping and properties of the *gam* and *sot* genes of phage Mu: their possible roles in recombination. Cold Spring Harbor Symp. Quant. Biol. **49:** 261–266.

CRAIG, N. L., and N. KLECKNER, 1987 Transposition and site-specific recombination, pp. 1054–1070 in *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.

DEONIER, R. C., 1987 Locations of native insertion sequence elements, pp. 982–989 in *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Biology*, edited by F. C. Neidhardt. American Society for Microbiology, Washington, D.C.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DYKHIUZEN, D. E., S. A. SAWYER, L. GREEN, R. D. MILLER and D. L. HARTL, 1985 Joint distribution of insertion elements IS4 and IS5 in natural isolates of *Escherichia coli.* Genetics **111:** 219–231.

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. Nat. Acad. Sci. USA **81:** 4500–4504.

GROISMAN, E. A., and M. J. CASADABAN, 1986 Mini-Mu bacteriophage with plasmid replicons for in vivo cloning and *lac* gene fusing. J. Bacteriol. **168:** 357–364.

GROISMAN, E. A., and M. J. CASADABAN, 1987 Cloning of genes from members of the family Enterobacteriaceae with mini-mu bacteriophage containing plasmid replicons. J. Bacteriol. **16:** 3687–3691.

HALL, B. G., 1988 Adaptive evolution that requires multiple spontaneous mutations. I. Mutations involving an insertion sequence. Genetics **120:** 887–897.

HALL, B. G., and P. W. BETTS, 1987 Cryptic genes for cellobiose utilization in natural isolates of *Escherichia coli.* Genetics **115:** 431–439.

HALL, B. G., P. W. BETTS and M. KRICKER, 1986 Maintainance of the cellobiose utilization genes of *Escherichia coli* in a cryptic state. Mol. Biol. Evol. **3:** 389–402.

HARTL, D. L., M. MEDHORA, L. GREEN and D. E. DYKHUIZEN, 1986 The evolution of DNA sequences in *Escherichia coli.* Philos. Trans. R. Soc. Lond. B **312:** 191–204.

MAHADEVAN, S., A. E. REYNOLDS and A. WRIGHT, 1987 Positive and negative regulation of the *bgl* operon of *Escherichia coli.* J. Bacteriol. **169:** 2570–2578.

MATSUTANI, S., H. OHTSUBO, Y. MAEDA and E. OHTSUBO, 1987 Isolation and characterization of IS elements repeated in the bacterial chromosome. J. Mol. Biol. **196:** 445–455.

MILLER, J. H., 1972 *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

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

PARKER, L. L., P. W. BETTS and B. G. HALL, 1988 Activation of a cryptic gene by excision of a DNA fragment. J. Bacteriol. **170:** 218–222.

PRASAD, I., and S. SCHAEFLER, 1974 Regulation of the β-glucoside system in *Escherichia coli* K-12. J. Bacteriol. **120:** 638–650.

RALEIGH, E., and N. KLECKNER, 1984 Multiple IS*10* rearrangements in *Escherichia coli.* J. Mol. Biol. **173:** 437–461.

REYNOLDS, A. E., J. FELTON and A. WRIGHT, 1981 Insertion of DNA activates the cryptic *bgl* operon of *E. coli* K-12. Nature **293:** 625–629.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SAWYER, S., and D. L. HARTL, 1986 Distribution of transposable elements in prokaryotes. Theor. Popul. Biol. **30:** 1–16.

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

SCHNETZ, K., C. TOLOCZYKI and B. RAK, 1987 β-Glucoside (*Bgl*) operon of *Escherichia coli* K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. J. Bacteriol. **169:** 2579–2590.

SCHWARTZ, E., C. HERBERGER and B. RAK, 1988 Second-element turn-on of gene expression in an IS*1* insertion mutant. Mol. Gen. Genet. **211:** 282–289.

SCHWARTZ, E., M. KRÖGER and B. RAK, 1988 IS*150*: distribution, nucleotide sequence and phylogenetic relationships of a new *E. coli* insertion element. Nucleic Acids Res. **16:** 6789–6802.

SELANDER, R. K., D. A. CAUGENT 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 F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.

Communicating editor: N. KLECKNER