Mosaic Structure of Plasmids From Natural Populations of *Escherichia coli*

E. Fidelma Boyd, Charles W. Hill,[1] Stephen M. Rich[2] and Daniel L. Hartl

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

Manuscript received December 10, 1995
Accepted for publication March 27, 1996

ABSTRACT

The distribution of plasmids related to the fertility factor F was examined in the ECOR reference collection of *Escherichia coli.* Probes specific for four F-related genes were isolated and used to survey the collection by DNA hybridization. To estimate the genetic diversity of genes in F-like plasmids, DNA sequences were obtained for four plasmid genes. The phylogenetic relationships among the plasmids in the ECOR strains is very different from that of the strains themselves. This finding supports the view that plasmid transfer has been frequent within and between the major groups of ECOR. Furthermore, the sequences indicate that recombination between genes in plasmids takes place at a considerably higher frequency than that observed for chromosomal genes. The plasmid genes, and by inference the plasmids themselves, are mosaic in structure with different regions acquired from different sources. Comparison of gene sequences from a variety of naturally occurring plasmids suggested a plausible donor of some of the recombinant regions as well as implicating a chi site in the mechanism of genetic exchange. The relatively high rate of recombination in F-plasmid genes suggests that conjugational gene transfer may play a greater role in bacterial population structure than previously appreciated.

NATURAL isolates of *Escherichia coli* typically harbor some one to five small plasmids and one to two large plasmids (Hartl *et al.* 1986; Selander *et al.* 1987). The small plasmids are usually smaller than 7.5 kb, and the large ones range from 40 to 200 kb. Among the large plasmids are the fertility factor F and the related R plasmids, which are ~100 kb. The laboratory version of F is notable among plasmids for to its ability, when integrated into the chromosome, to support the conjugational transfer of chromosomal genes (reviewed in Willetts and Skurray 1987). Soon after bacterial recombination was first described, natural isolates were examined for their ability to transfer chromosomal genes into the laboratory strain K-12 (Cavalli and Heslot 1949; Lederberg *et al.* 1952; Lederberg and Tatum 1953). In these experiments, between 3 and 14% of tested isolates were found to yield recombinants. In retrospect, the results of the recombination experiments are very difficult to interpret because of complications discovered since the tests were performed. For example, F is not the only plasmid capable of chromosomal integration and conjugation. In addition, most naturally occurring F plasmids are not as efficient in conjugational transfer as the laboratory F. The naturally occurring F plasmids have their transfer functions repressed owing to the action of the fertility inhibition gene *finO,* which in laboratory F is interrupted and inactivated by an insertion of IS*3.* Furthermore, the classical experiments detected recombination from all sources including transduction and sexduction. Still another potential bias comes from DNA restriction-modification systems, the incompatibility of which can greatly reduce or eliminate the recovery of recombinants between otherwise fertile strains.

Hence, the true prevalence of F-related plasmids among natural isolates remains a matter of speculation. In this paper, we have used current methods of DNA hybridization along with PCR and DNA sequencing to detect and study F-related plasmids among natural isolates comprising the ECOR reference collection of *E. coli* (Ochman and Selander 1984). Previous studies of plasmids in natural isolates have focused on the diverse set of plasmids that produce colicins and, in particular, on the colicin and immunity genes themselves (Riley 1993a,b; Ayala *et al.* 1994; Riley *et al.* 1994). In surveys of natural isolates of *E. coli,* 35% of the ECOR collection of 72 reference strains (Riley and Gordon 1992) and 51% of 234 pathogenic isolates (Achman *et al.* 1983) were found to possess colicin plasmids.

In the present study, the ECOR collection was examined for the presence of DNA sequences found in the F-related plasmids F and R1. The genes were *finO, traY* (F type), *traY* (R1 type), *traD,* and *repA.* The *finO* gene is the fertility inhibition gene that, when interrupted, derepresses the transfer functions. The transfer genes *traY* (F) and *traY* (R1) code for a component of conjugational DNA metabolism and differ substantially in sequence in F and R1 plasmids. The transfer function *traD* is also implicated in conjugational DNA metabo-

Corresponding author: Daniel L. Hartl, Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Ave., Cambridge, MA 02138. E-mail: d_hartl@harvard.edu

[1] *Permanent address:* Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey PA 17033-0850.

[2] *Present address:* Department of Ecology and Evolutionary Biology, University of California, Irvine CA 92717.

TABLE 1

Plasmid genes and PCR primers

Gene	Map[a]	Function	Primer sequence[b]
finO	99.5–100/0	Repression of transfer	F: 5′-GAAGCGACCGGTACTGACACTG-3′ R: 5′-GCCTGAAGTTCTGCCTTTATCCG-3′
traD	90.9–92.1	Conjugational DNA metabolism	F: 5′-CAGATTGCGTCCATGCGTATCC-3′ R: 5′-ATCACCACaCATATCACCGCGC-3′
traY (F)	68.2–68.8	Conjugational DNA metabolism	F: 5′-AAGATTTGGTACACGTTCTGC-3′ R: 5′-CTTCCTCTTTATCTGCCTCCC-3′
traY (R1)	53.8–53.1	Conjugational DNA metabolism	F: 5′-GTGAGGAGGCGTAACGCGAG-3′ R: 5′-GTTGACTCGTTCTCTTCGATC-3′
repA	38.0–39.9	RepFIB replicon	F: 5′-TCGCTGCAAACCTTGTCACT-3′ R: 5′-GGAGATCCTGCGTACACTGCCT-3′

[a] Except for *traY* (R1), the position of the gene in the laboratory F plasmid; for *traY* (R1), the position refers to the R1 plasmid.
[b] F, forward primer; R, reverse primer.

lism but has a similar sequence in F and R plasmids. Finally, *repA* is a component of the RepFIB region that serves as a secondary replicon of F, functionally distinct from the primary replicon RepFIA. RepFIB is present widely among multireplicon plasmids in the IncF groups (Bergquist *et al.* 1986; Gibbs *et al.* 1993). For the probes examined, the proportion of ECOR strains showing hybridization ranged from 7% with *traY* (F) to 38% with *finO*. A total of 11 strains (15%) harbored an F-related plasmid as judged from hybridization with probes for *finO, traD, repA,* and either *traY* (F) or *traY* (R1). One strain contained two distinct F-related plasmids.

DNA sequence variation was also examined in portions of the *finO, traY, traD,* and *repA* regions. Comparison of the relationships among the plasmid genes with the relationships among the host bacteria, as estimated from electrophoresis of 35 enzymes (Herzer *et al.* 1990), indicates relatively frequent horizontal transfer of F and R among natural isolates. The sequence data also indicate that the plasmid genes are mosaics formed by multiple recombination events between diverse ancestral genes. The number of recombination events detected among the plasmid genes is considerably greater than that observed in chromosomal genes, hence conjugational transfer and recombination is an important determinant of the disposition of genetic variation among plasmids.

MATERIALS AND METHODS

Bacterial strains: The ECOR strains (Ochman and Selander 1984) were from the set originally provided to C.W.H. by R. K. Selander and T. S. Whittam. The wild-type *E. coli* K-12 strain, designated CGSC4401, contains an F factor and is lysogenic for phage lambda.

PCR amplification: Primers for PCR (Table 1) and DNA sequencing were designed from the known sequences of *finO* (Mcintire and Dempsey 1987), *traD* (Jalajakumaril and Manning 1989), *traY* (R1) (Finlay *et al.* 1986a), *traY* (F) (GenBank U01159), and the *repA* gene of the RepFIB replicon (Gibbs *et al.* 1993). Following amplification, the PCR products were purified using the Qiaquick PCR purification kit.

Southern blot analysis: DNA probes for each of the genes, *finO, traD, traY* (R1), *traY* (F) and *repA,* were obtained by PCR and used to screen the ECOR collection by DNA hybridization. The *finO* and *traY* (R1) probes were prepared using DNA from ECOR50 as template, whereas the others were prepared using DNA from K-12 as template. The probes were labeled with fluorescin-conjugated nucleotides and after hybridization were detected by the ECL system of Amersham (Arlington Heights, IL). Genomic DNA was extracted using G-Nome DNA isolation kits from Bio101 (Vista, CA).

DNA from the 72 ECOR strains was digested with *Mlu*I and the fragments separated by agarose gel electrophoresis. The DNA fragments were transferred to Hybond-N nylon membranes (Amersham) for hybridization at 60° in 5× SSC, 0.1% SDS and 5% dextran sulfate.

DNA sequences: From the *finO* gene, 480 bp were sequenced for each of 16 ECOR strains; from *traD,* 540 bp for each of 14 strains; from *traY* (F), 326 bp for each of five strains; from *traY* (R1), 171 bp for each of 12 strains; and, from *repA,* the first 381 bp as well as 249 bp of the upstream region were sequenced for each of 12 strains. Sequencing of the PCR products was performed with an Applied Biosystem model 373A automated DNA sequencing system using the DyeDeoxy terminator cycle sequencing kit. For the all genes, both strands were sequenced.

Computer analysis: DNA sequence data were assembled and edited with the Sequencher program. The phylogenetic analysis was conducted using MEGA (Kumar *et al.* 1993), and additional programs were written and provided by T. S. Whittam. The gene sequences for *finO, traD, repA, traY* (R1) and *traY* (F) are available through Gen Bank accession numbers U50650–U50706.

RESULTS

DNA hybridization results: The ECOR collection of 72 strains was initially screened by DNA hybridization for the presence of three genes characteristic of plasmids related to the fertility factor F, namely, *finO, traD,* and *traY* (F). Among DNA samples from the ECOR strains, 27 (38%) hybridized with the *finO* probe (Figure 1). ECOR37 yielded two strong bands after the DNA was digested with either *Mlu*I (Figure 1) or *Eco*RV (data not shown), suggesting the possible existence of two F-related plasmids. Hybridization with the *traD* probe yielded 20 positives, all of which were also positive with *finO,* but among these there were only five samples that scored positive for *traY* (F). This result suggested that

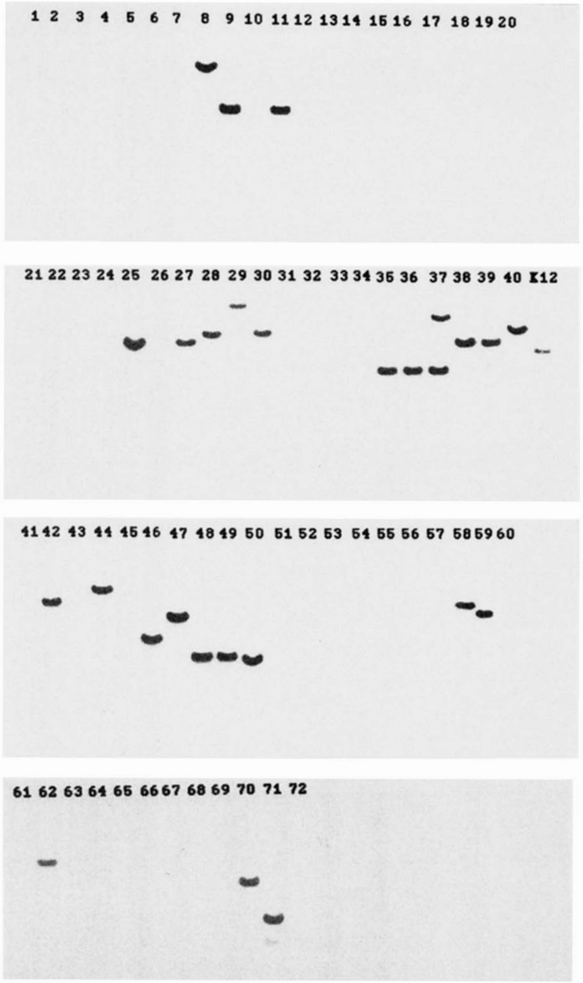


FIGURE 1.—DNA hybridization of DNA from 72 ECOR strains digested with *Mlu*I and probed with a *finO* PCR product. Lanes 1–72 contain digests of DNA from the ECOR strains (EC1–EC72). A sample from *E. coli* K-12 is also included.

plasmids bearing genes similar to those of F are common among the ECOR strains but that the plasmids are not identical.

Because certain resistance transfer factors, such as R1, have a *finO* gene closely related to that of F but have a *traY* gene that is not only much more divergent but also of a different size, PCR primers were used to amplify a segment of the *traY* (R1) gene and used to probe the ECOR DNA samples. In a further effort to determine the relationship of the putative conjugational plasmids present in these strains, hybridization with a probe for the *repA* gene of the RepFIB secondary replicon was also carried out. The *traY* (R1) probe gave a positive signal with 13 (18%) of the ECOR strains, all of which had also tested positive for *finO*. The strain ECOR37, which had given evidence of two *finO* genes, tested positive for both *traY* (F) and *traY* (R1); this strain apparently has two F-related plasmids. No other strains tested positive for both *traY* (F) and *traY* (R1). Relative to the *repA* gene of RepFIB, 20 strains (28%) were positive, and all of these were also positive for *finO*.

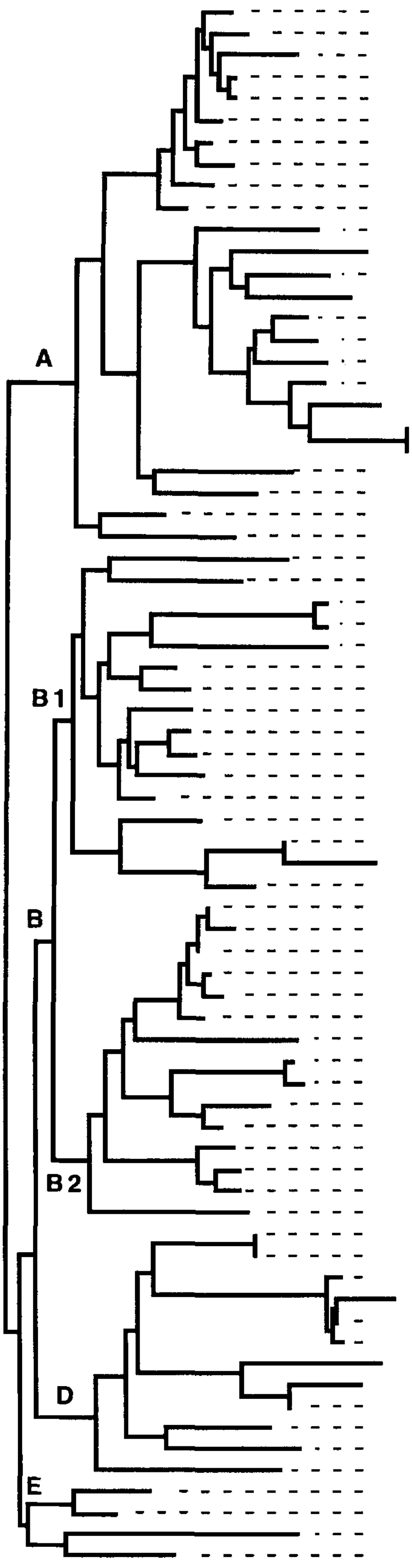
The hybridization data are summarized in Figure 2. As additional evidence for the presence of the plasmid sequences detected by hybridization, DNA samples from the ECOR strains were used as templates in PCR for each of the probes. The results were completely concordant with the hybridization results in Figure 2 except that DNA from strain ECOR70 did not support amplification with the *finO* primers, possibly because the *finO* gene in this strain has excessive mismatches with the primers.

For the sake of concreteness, we will designate plasmids that are positive for the probes *finO, traY* (F), *traD,* and *repA* as "F" and those that are positive for the probes *finO, traY* (R1), *traD,* and *repA* as "R." These designations are not intended to be definitive for each type of plasmid but serve only to identify which type of *traY* gene it contains, either *traY* (F) or *traY* R1). In terms of this restricted operational definition of F and R plasmids, the ECOR collection contains four F and eight R plasmids in 11 strains (ECOR37 contains one representative of each). The prevalence of F and R plasmids is therefore 6 and 11%, respectively, in the ECOR collection. Four strains contained plasmids that scored positive for *finO, traD,* and *repA* but were negative for both *traY* probes. The plasmids in these strains may be F-related plasmids with divergent *traY* genes and, if they are included, the overall prevalence of F-related plasmids is 15/72 = 21%.

The distribution of F and R plasmids in Figure 2 is not completely random. The plasmids are underrepresented in strains of the groups A and B2; the majority of the F and R plasmids are found in the groups B1, D, and E. Although more than half of the ECOR collection is made up of strains in the A and B2 groups (25 and 15 strains, respectively), these groups account for only about one-fourth of the strains scoring positive for *finO, traD,* or *repA* ($P < 0.05$).

Association of plasmid genes with plasmids: To verify that the hybridizing bands were plasmid determined, hybridization experiments were carried out using uncleaved genomic DNA in which electrophoresis in 0.6% agarose for an extended time separates the sheared high molecular weight chromosomal DNA from the circular plasmid DNA. After transfer, the DNA was blotted and probed for *finO*. In all 25 strains probed with *finO,* the hybridization was clearly associated with a large plasmid that separated from the main band formed by the bacterial chromosome. In the case of ECOR37, two different sized plasmids were observed. Of 18 strains also probed with *repA,* 17 also showed hybridization with this probe, and in all 17 cases the *finO* and *repA* signals coincided, indicating that they were associated with the same plasmid. In one strain, the position of the *finO* hybridization in the gel could not be resolved clearly from the chromosomal material, and so the presence of the gene in a free plasmid could not be rigorously demonstrated.

Nucleotide polymorphism: The four F-related plas-



strain	finO	traY(F)	traY(R1)	traD	repFIB
K-12	++	++		++	++
1					
5					
8	++				++
10					
11	++		++	++	++
25	++		++		++
2					
3					
9	++		++	++	
12					
4					
6					
16					
22					
7					
14					
13					
18					
19					
20					
21					
17					
24					
15					
23					
58	++	++		++	
67					
26					
27	++			++	++
69					
28	++		+	++	++
45					
29	++			++	++
32					
33					
34					
30	++	++		++	++
68					
70	++		++	++	++
71	++		++	++	++
72					
51					
52					
54					
56					
57					
55					
65					
61					
62	++	++		++	++
63					
64					
53					
59	++	++		++	++
60					
66					
35	++			++	++
36	++			++	++
38	++				++
39	++				++
40	++				++
41					
46	++		+	++	
49	++		++	++	++
50	++		++	++	++
44	++				++
47	++		++	++	
48	++		++	++	
31	+				
43					
37	++(2)	++	++	++	++
42	++		++	++	++

FIGURE 2.—Distribution of sequences that hybridize with probes for *finO, traD, traY* (R1), *traY* (F), and *repA* among strains in the ECOR reference collection. The EC numbers refer to ECOR strain designations. The tree was derived by HERZER *et al.* (1990) on the basis of electrophoretic polymorphisms among 38 enzymes. Major phylogenetic subgroups are labeled with the letters A–E. ++ and + represent strong and weak hybridization signals. DNA from strain ECOR37 yields two hybridizing bands.

TABLE 2

Sequence polymorphism and diversity

Gene	Percentage G + C	Sequenced region		No. of polymorphic sites	
		Base pairs[a]	Amino acids	Base pairs[b]	Amino acids[b]
finO	56	441 (79)	147	71 (16)	20 (14)
traD	49	540 (25)	180	54 (10)	11 (6)
traY (R1)	41	171 (75)	57	14 (8)	1 (2)
traY (F)	38	324 (90)	108	2 (0.6)	0
repA	51	381 (39)	127	22 (6)	2 (2)
5′ *repA*	39	249		25 (10)	

[a] Values in parentheses are the percentage of the total coding region sequenced.

[b] Values in parentheses are the percentage of the total number of polymorphic sites.

mid genes were sequenced from a representative sample of the major groups of the ECOR collection, including A, B1, B2, D, and E. The sequences include a region of 480 bp of *finO,* 540 bp of *traD,* 326 bp of *traY* (F), 171 bp of *traY* (R1), and 630 bp of *repA* were sequenced in 16, 14, six, 12, and 12 ECOR strains, respectively. The results are summarized in Table 2. The percentage G + C of *finO, traD,* and *repA* do not differ significantly from the average of the *E. coli* chromosome as a whole nor from the overall average of the F plasmid. However the *traY* genes of F and R1 have a lower G + C content of 38 and 41%, respectively, as does the 5′ noncoding region of *repA* at 39%.

Figure 3 shows the position and identity of the polymorphic nucleotide sites observed in *finO, traD, traY* (R1), and *repA* in the ECOR strains (EC designations). Homologous genes from other plasmids are included for comparison. Among the 16 *finO* sequences, there were 71 polymorphic sites, including 20 amino acids replacements. Among the 12 *traD* genes, there were 54 polymorphic sites, including 11 amino acid replacements. The five *traY* (F) sequences included only two polymorphisms, both silent, but the 12 *traY* (R1) sequences included 14 polymorphic sites with one amino acid replacement. The 12 sequences from RepFIB each comprised a 381-bp region of the *repA* open reading frame along with 249 bp from the 5′ flanking region; these yielded 47 polymorphic sites of which 22 were in the coding region and resulted in two amino acid replacements.

Level of synonymous and nonsynonymous polymorphism: For the four plasmid genes *finO, traD, traY* (R1), and *repA,* we estimated the genetic diversity in all pairwise comparisons using the methods of NEI and GOJOBORI (1986) and NEI and LIN (1989). The results are summarized in Table 3 along with some chromosomal genes for comparison. The value of π is the average number of differences per 100 nucleotide sites in all pairwise comparisons; d_s and d_n are the average number of nucleotide differences per 100 synonymous sites and per 100 nonsynonymous sites, respectively, among all pairwise comparisons. (For *repA,* the π value includes both coding and noncoding nucleotides.) Table 3 shows that the estimates of plasmid gene diversity are somewhat greater than those calculated for chromosomal genes.

Evolutionary relationships among plasmids and their host strains: Evolutionary trees constructed from polymorphic nucleotide sites by the neighbor-joining method (SAITOU and NEI 1987) are shown in Figure 4. The most notable feature of this analysis is the lack of congruence of the four plasmid gene trees with each other and with the inferred relationships among their host strains. Where possible, trees were also constructed with putative recombined segments removed; this also resulted in a lack of concordance of the MLEE tree of ECOR and plasmid gene trees. The lack of congruence with the host strains implies horizontal transmission, which is expected among conjugational plasmids. More surprisingly, sequences from plasmids present in strains of each of the five major ECOR groups (A, B1, B2, D, and E) are widely distributed on the various plasmid-gene trees, suggesting frequent intragenic recombination among the plasmid genes. For example, the *finO* and *traD* trees have 12 strains in common, and the relationships among these strains is different for the two genes. Furthermore, in each of the gene trees, strains of the *E. coli* subgroups do not cluster together. One common feature among all the trees is the clustering of group A strain ECOR11 with the group D strains ECOR49 and ECOR50, all of which are isolates from humans in Sweden. Generally, the four gene trees cluster the group D strains together, with the exception of ECOR47, which clusters with the group B1 strains. Another notable feature of Figure 4 is that, in the *finO* and *traD* gene trees, the sequences from the ECOR59 plasmid are very divergent from the plasmid genes found in all other isolates.

Evidence for recombination among plasmid genes: Among the four plasmid-encoded genes, there were 188 polymorphic sites; the positions of which are shown in Figure 3. Statistical tests for recombination or gene conversion based on the polymorphic synonymous sites were carried out with the cluster-detection methods of STEPHENS (1985) and SAWYER (1989) and the maximum chi-squared method of MAYNARD SMITH (1992). All tests

finO

```
                     11111111 1111111112 2222222222 2222233333 3333333333 3344444444 4
         133366779 9900224445 5666778990 0112244455 5668900333 5666778899 9900112222 2
        4105906254 6928070180 9278170351 5062803625 9012803679 9039284706 7958231235 9

EC9     CCGCAAGGTA AAAGGTACTG GCACCACTAC GAAGCCGAAC GCGGAGGGTT CCGCGGGGGT TTGAGTTCAA T
EC11    .......... .......... .......... .......... .......... .G........ ......G... .
EC25    ......CAGG G....C..A. ..G.AG...T .CG....... .......... GT........ ...G.CG..C .
EC27    .....GCAAG G...A...A. .T...G.... .GG....... ....C..... G......... ...G.CG..C .
EC30    T....GCAAG G...A...A. .T...G.... .GG....... ....C..... T...T.A... ..A..C.... .
EC71    ....GGCAAG G.....TA.. T...GA.... .GG....... .......... G......... ...G.CG... .
EC59    .........G GGGA..CG.. .G..T...G. AGGTTAT.TG T.TT..ATGC G.CTAC.AAC AAA.CACAGC C
EC62    .....GCAAG G...A...A. .T...G.... .GG....... ....C..... G....A.... ..AG.CG... .
EC35    .......... ........AA A......... ..G...G... .......... .......... .....G.... .
EC39    .......... .........A ...T.G.C.. .CG....... .T..CT.... G......... .....CG... .
EC47    .T..GGCAAG G.......AA ...T...... .GG....... .......... G......... ...G.CG..C .
EC48    .......... .......... .......... .......... .......... G......... .....G.... .
EC49    .......... .......... .......... .......... .......... .......... .....G.... .
EC50    .......... .......... .......... .......... .......... G......... .....G.... .
EC37    ..AA.G...G ......CT.A ...T.G.... .C........ .T..CT.... G......... .....G.... .
EC42    ......CAGG G........A ...T.G.... ..G....... .......... G......... ...G.C.... .
R100    .......... .......... .......... .......... .......... G......... .....G.... .
R6-5    .......... ........AA A......... ..G....... .......... G......... .....G.... .
ColB2   ....G.CACG G........A ...T.G.... .GG....... .......... G......... ...G.CG... .
         * *     *      ** * * *       *    *  ** **     * *             ***  *
```

traD

```
                   1111111111 2222222222 2222223333 3334444444 5555
           3345699 1244557899 0133345566 6777895788 8990223457 0234
        5690922106 1334367958 1645742517 8039878514 7693098701 1280

EC9     CCGTACGGCT TAAGGCTGCT TTCGACCATC AATATCGGAC TCTCCCCCCC GACA
EC11    ......A... ....AA.... C......... .......A.. .......... .G..
EC27    T......... .....A..T. .......... .....T.A.. ....T.A... ....
EC30    T......A.. ...A...C.. .C.ATT.CGT GT...T.A.. .......A.. ....
EC58    T..C...... .....AC... ..T......T .....T.... .........T ....
EC70    T......... .....A..T. .......... .....TA.T. .GGATT..TT TG..
EC71    .......... .....A..T. .......... .....T.... ......A... ..GC
EC59    T.A.CA..TC C.G...C..G ....TTG..G TTCCC.A.T. .GGATT..TT TG..
EC62    T......... .G....C... .........T .....T.AGG C.....A... ....
EC35    T.....A... ....AA.... C......... .......... .......... .G..
EC47    T.......T. C.G....... .....T.... .....T.A.. .......... ....
EC49    T.....A... ....AA.... C......... .......A.. .......... .GGC
EC50    T.....A... ....AA.... C......... .......... .......... .GGC
EC42    TA........ ..G..AC... .........T .....T.AGG C......... ....
        * *   **     *           **      *     *  *  *           *
```

***traY* R1**

```
                      111
         112266789 9035
        6281412813 9220

EC9     TCGCGAATAG ACAC
EC11    G.ATACGCTT GTGT
EC25    .......... ....
EC70    .......... ..G.
EC71    .......... ..G.
EC72    .....C.... G...
EC47    .T........ ..G.
EC48    G.A.ACGCTT GTGT
EC49    G.ATACGCTT GTGT
EC50    G.ATACGCTT GTGT
EC37    G.A.ACGCTT GTGT
EC42    .......... ...T
R1      G.A.ACG.TT GTG.
ColB4   .......... ....
             *
```

repA

```
                     5' noncoding region   coding region
                           11111 111112  22333 33344444444 45555566666
            111233456 6699922244 556690  67034 59901233567 82346900112
           5123802855 6817826978 242380  10665 40384358021 64737409587

EC11       CATGGTCCGA ATTCGACGGA GTCAA.  GGACT GCCAGTAAG.G CTGCCCAA...
EC25       ..CTTCA.C. .....GGAAG ..TC..  ....A ATTGTC.GA.C .CAT.TCC...
EC27       ..CTTCA.C. ..GT.GG.AG A...T.  C..TG ATTGACTG..C .C.T.TCC...
EC30       ..CTA.AAAC .C........ ......  ...TG ATTGACTG..C TC.TTTCC...
EC70       ..CTTCA.C. ..GT.G..AG A...T.  C..TG ATTGAC.G..C TC.TTT.C...
EC71       ..CTA.AAAC .......... ......  ....G ........... .C.T..C....
EC59       .......... .......... ......  ..G.G ....AC..... .C.TTTCC...
EC39       .......... .......... ......  ..... ........... .C.T..C....
EC40       .......... .......... ......  ..... ........... ...T..C....
EC49       T......... .......... ......  ..... ........... ......C....
EC50       .......... .......... ......  ..G.. ........... .C.T..C....
EC37       .GATTCG.A. C...A..... .G....  C...G ATT.AC.G... .C.T..C....
F          ..C.TCA.C. ...T.G..AG A...T.  ...TG ATTGACTG..C .C.TTTCC...
ColV       ..CTTCA.C. ...T.G..AG A...T.  C..TG ATTGACTG..C TC.TTTCC...
ColV3K30   ..CTACA.C. ...T.G..AG A...T.  C..TG ATTGACTG..C TC.TTTCC...
ColVtrp    .......... ...T...... ....T.  .A... ..........C .C.TTTCC...
R124       ..CTTCA.C. ...T.G..AG A...T.  C..TG ATTGACTG..C TC.TTTCC..T
R386       ..CTTCA.C. ...T.G..AG A....A  ...TG ATTGTCTG..C .C.TTTCCCC.
pHH502     .......... .......... .....A  ...TG ATTGTCTG..C .C.TTTCC...
pHH507     .......... .......... ......  ...TG ATTGACTG.TC .C.TTTCC...
                                                             *    *
```

FIGURE 3.—Distribution of polymorphic sites in four plasmid genes. The strain designations are on the left, where EC denotes an ECOR strain. The numbering of the nucleotides is given above the sequences. An asterisk at the bottom indicates an amino acid replacement site. Homologous sequences are also shown for *finO* genes from the plasmids R100 (MCINTIRE and DEMPSEY 1987), R6–5 (CRAM *et al.* 1991), and ColB2 (VAN BIESEN and FROST 1992); for *traY* (R1) from the plasmids ColB4 (FINLAY *et al.* 1986) and R1 (KORAIMANN and HOGENAUER 1989); and for *repA* from the plasmids ColV, ColV3-K30, ColVB*trp,* R124, R386, pHH502, pHH507 (GIBBS *et al.* 1993) and the laboratory F factor (SAUL *et al.* 1989).

yielded statistically significant evidence of recombination. The STEPHENS test identified the 15 statistically significant partitions ($P < 0.05$) shown in Table 4. The maximum chi-squared method confirmed the mosaic structure of the same genes identified by the STEPHENS test. Also, when one constructed trees based on the segregating sites 5′ and 3′ of an exchange event inferred from the STEPHENS test in the *repA* sequences, the resultant trees gave different topologies.

For the 71 polymorphic sites among the 16 sequences of *finO,* three significant partitions were identified (Table 4). The first partition separated *finO* in ECOR59 from all other sequences and was supported by a total of 30 polymorphic sites in a 330-bp segment that included an 87-bp segment of consecutive nonpolymorphic sites. The second partition united seven *finO* sequences from ECOR strains in groups A, B1, B2, D, and E; this partition was supported by only two sites in a 6-bp segment, but $P \approx 0.02$. The third partition separated the ECOR37 plasmid from all other sequences and was also supported by two sites ($P \approx 0.02$).

Among the 14 *traD* sequences, there were 54 polymorphic sites. The STEPHENS test detected six significant partitions (Table 4). The first separated *traD* in ECOR59 from all other sequences and was supported by nine sites in a 279-bp segment. The second partition united the *traD* sequences from ECOR11, ECOR35, ECOR49, and ECOR50 owing to their sharing three polymorphic sites in a 149-bp segment as well as a 101-bp segment of consecutive nonpolymorphic sites ($P \approx$

TABLE 3

Nucleotide diversity in chromosomal and plasmid encoded genes

Gene	π	d_s	d_n	d_n/d_s
Plasmid genes				
finO	4.00 ± 2.09	12.2 ± 1.8	1.7 ± 0.4	0.14
traD	2.80 ± 1.50	10.1 ± 1.4	0.8 ± 0.2	0.07
traY (R1)	4.10 ± 2.23	19.4 ± 5.9	0.7 ± 0.5	0.03
repA	2.80 ± 1.51	9.9 ± 2.2	0.2 ± 0.2	0.03
Chromosomal genes				
gapA	0.02 ± 0.03	0.8 ± 0.3	0.1 ± 0.1	0.12
mdh	1.10 ± 0.59	3.7 ± 0.7	0.2 ± 0.1	0.04
putP	2.40 ± 1.27	9.0 ± 1.5	0.2 ± 0.1	0.02

Data for *gapA* from NELSON *et al.* (1991), for *mdh* from BOYD *et al.* (1994), and for *putP* from NELSON and SELANDER (1992).

0.007). The third partition separated *traD* in ECOR30 from all others and was supported by eight sites in a 386-bp segment and a 186-bp segment of consecutive nonpolymorphic sites. The fourth partition grouped ECOR70 and ECOR59 by virtue of seven polymorphic sites in a 143-bp segment. The *traD* genes grouped by the two other significant partitions were ECOR62/ECOR42 and ECOR71/ECOR49/ECOR50, both of which groups were supported by two shared polymorphisms.

Application of the STEPHENS test to the 12 *repA* sequences identified six statistically significant partitions (Table 4). The first separated *repA* from ECOR37 from all other sequences and was supported by four sites in a 143-bp segment. Three significant partitions grouped the *repA* genes from ECOR25, ECOR27, ECOR30, and ECOR70 together either with ECOR71 and ECOR37 (supported by two shared sites in a 20-bp segment), with ECOR37 (four sites in an 84-bp segment) or with ECOR59 (two sites in a 15-bp segment). Other partitions grouped *repA* in ECOR27/ECOR71 by virtue of four sites in a 107-bp segment and *repA* in ECOR11/ECOR25/ECOR30 supported by three sites in a 26-bp segment.

Altogether, the *finO, traD,* and *repA* sequences yielded 15 significant partitions, each a putative recombination event. For the 12 *traY* (R1) sequences, there were only 14 polymorphic sites, and both the STEPHENS test and the SAWYER test failed to detect significant clusters. However the maximum chi-squared method identified a shared mosaic structure in ECOR11, ECOR48, ECOR49, ECOR50, and ECOR37 involving a total of 12 sites with a crossover point of nucleotide 102; on the other hand, the observed pattern of similarity between the sequences more likely results from shared common ancestry than recombination.

DISCUSSION

The ECOR collection was selected from among ~2600 *E. coli* isolates with the intention of encompassing the range of genetic diversity found within the species (OCHMAN and SELANDER 1984). If the plasmids present in the ECOR strains are representative, then our data suggest that F-related plasmids are quite common. In the ECOR strains, the R1 type of *traY* was found somewhat more frequently than the F type of *traY* (eight plasmids *vs.* four). Overall, 15% of the ECOR strains show hybridization with probes for *finO, traD, repA,* and either *traY* (F) or *traY* (R1), indicating the presence of an F-related plasmid. On the other hand, the plasmids are not distributed randomly among the strains: there appears to be somewhat of an under representation of F-related plasmids among strains of the A and B2 subgroups.

The gene diversity observed in the F-related plasmids may be compared with that found among housekeeping genes present in the chromosome. Table 3 indicates diversities (π) among plasmid genes ranging from 2.80 to 4.10%. Among chromosomal genes, *gapA* is the least variable, averaging only 0.02%. However, *gapA* has a high codon usage bias and its rate of evolution is relatively slow (LAWRENCE *et al.* 1991; NELSON *et al.* 1991). Average diversity estimates are 1.1% for *mdh* (BOYD *et al.* 1994), 1.7% for *phoA* (DUBOSE *et al.* 1988), 2.4% for *putP* (NELSON and SELANDER 1992), and 7.2% for *gnd* (NELSON and SELANDER 1994). Only *gnd* has a level of diversity greater than that observed for the genes in the F-related plasmids, and the level of variation in the *gnd* gene is believed to be a consequence of its linkage to the *rfb* locus at which diversifying selection seems to act (BARCAK and WOLF 1988; BISERCIC *et al.* 1991; DYKHUIZEN and GREEN 1991).

Figure 2 shows an inferred phylogeny of the ECOR strains based on enzyme electrophoresis. It is therefore an inferred phylogeny of the host strains that harbor the F-related plasmids. Evidence for horizontal transmission of the plasmids derives from comparison of Figure 2 with the branching patterns of the plasmid gene trees in Figure 4. From analysis of enzyme electrophoresis, chromosomal nucleotide sequences, and Rhs elements, it has been shown that the group A strains of the ECOR collection are all closely related (BOYD *et al.* 1994; HILL *et al.* 1995); however, in the four plasmid-gene trees, sequences from plasmids in strains of the A group do not cluster together. The plasmids in strains ECOR11, ECOR48, and ECOR50 have identical *finO* genes, but the chromosomes of the host bacteria are highly divergent. Similarly, the plasmids in ECOR11, ECOR48, and ECOR50 have identical *traY* (R1) sequences, but the chromosomes of the host strains are again very divergent. These results suggest that there have been many events of horizontal transmission of F-related plasmids among the isolates of the ECOR collection.

Comparison of the gene trees for *finO, traD, traY* (R1), and *repA* reveals many contradictions. The gene-tree topologies are not congruent with one another. The discrepancies are caused by plasmids that have

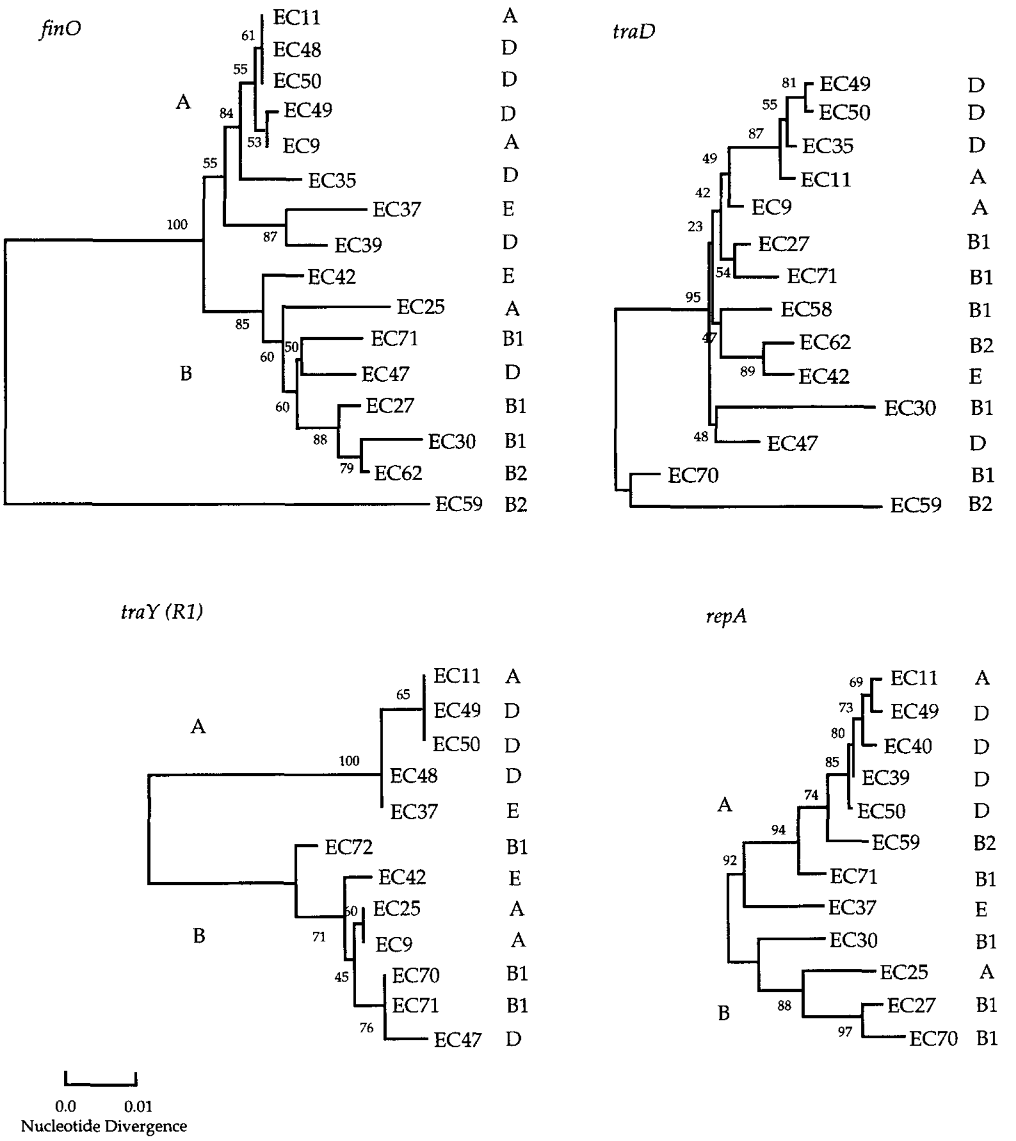


FIGURE 4.—Inferred gene trees for the genes *finO* (16 alleles), *traD* (14 alleles), *traY* (R1) (12 alleles) and *repA* (12 alleles). The trees were generated by the neighbor-joining method (SAITOU and NEI 1986) from a matrix of pairwise distances based on all nucleotide sites. Each EC number is an ECOR strain designation; the subgroup (A–E) to which the strain belongs (see Figure 2) is also indicated. Each number adjacent to a node indicates the percentage of 1000 bootstrap trees that contain the node.

closely related sequences for one gene but highly divergent sequences for another gene. Such a pattern may result from recombination among plasmid genes. Recombination in chromosomal genes of *E. coli* is well known: examples include the genes *phoA* (DUBOSE *et al.* 1988), *gnd* (BISERCIC *et al.* 1991; DYKHUIZEN and GREEN 1991; NELSON and SELANDER 1994), and the *trp* region (MILKMAN and BRIDGES 1993). However, the frequency of recombination is low. The test of STEPHENS identified only three recombination events among sequences for *gapA* (NELSON *et al.* 1991), *putP* (NELSON and SELANDER 1992), and *mdh* (BOYD *et al.* 1994). Furthermore, HALL and SHARP (1992) could find no evidence of recombination at *celC, crr,* and *gutB* in sequences from the ECOR strains.

In contrast with the low frequency of recombination detected in chromosomal genes, recombination is readily detected within the plasmid genes analyzed in this study. A total of 15 statistically significant recombination events have been detected by the STEPHENS test for clusters of polymorphic sites. An opportunity for recombination is exemplified by the coexistence of an "F plasmid" [*traY* (F)] and an "R plasmid" [*traY* (R1)] in the strain ECOR37. Additional insight into the recombination process can be gained by aligning the *finO, repA,* and *traY* genes from a range of naturally occurring IncF-group plasmids with the sequences determined from plasmids in the ECOR strains (Figure 3). With respect to *finO,* the ECOR strains ECOR11, ECOR48, and ECOR50 are identical with the sequence from plasmid R100. Furthermore, in the gene tree for *finO* (Figure 4), the sequences in the A cluster are similar to that in plasmid R6-5, whereas those in the B cluster are similar to that from plasmid ColB2. Concerning the

TABLE 4

Recombination events detected in genes in F-related plasmids

	Significant partitions		
Gene	No.	Partition	Inferred recombination event
finO	3	ECOR59 *vs.* others	330-bp segment (30 sites) from unknown donor
		ECOR25, 27, 30, 71, 62, 47, 42 *vs.* others	6-bp segment (two sites) shared by listed strains
		ECOR37 *vs.* others	5-bp segment (two unique polymorphic sites)
traD	6	ECOR59 *vs.* others	279-bp segment (nine sites) from unknown donor
		ECOR11, 35, 49, 50 *vs.* others	149-bp segment (three sites) shared
		ECOR30 *vs.* others	386-bp segment (eight sites) from unknown donor
		ECOR70, 59 *vs.* others	143-bp segment (seven sites) shared
		ECOR62, 42 *vs.* others	3-bp segment (two sites) shared
		ECOR71, 49, 50 *vs.* others	2-bp segment (two sites) shared
repA	6	ECOR37 *vs.* others	143-bp segment (four sites) from unknown donor
		ECOR25, 27, 30, 70, 71, 37 *vs.* others	20-bp segment (two sites) shared
		ECOR25, 27, 30, 70, 37 *vs.* others	84-bp segment (four sites) shared
		ECOR25, 27, 30, 70, 59 *vs.* others	15-bp segment (two sites) shared
		ECOR27, 71 *vs.* others	107-bp segment (four sites) shared
		ECOR11, 25, 30 *vs.* others	26-bp segment (three sites) shared

traY (R1) gene, the sequences from plasmids ColB4 and R1 are 95% identical at the nucleotide level (FINLAY *et al.* 1986a,b); Figure 3 indicates some of the sites at which they differ. Yet, in the gene tree for *traY* (R1) in Figure 4, the group of sequences designated A are clearly most closely related to the *traY* from plasmid ColB4, whereas those designated B are clearly most closely related to the *traY* from plasmid R1. In the gene tree for *repA* (Figure 4), the cluster of sequences denoted A is very similar to *repA* from the plasmid ColV*trp*, whereas the sequences denoted B are affiliated with *repA* from the plasmids ColV, ColV3-K30, R124, and R386 as well as the laboratory F factor (Figure 3). Conclusive evidence for recombination comes from the *repA* gene in the R-type plasmids pHH502 and pHH507 (Figure 3), in which the 5′ half strongly resembles the group A sequences in the *repA* gene tree (Figure 4) but in which the 3′ half strongly resembles the group B sequences.

In the *repA* gene, there is a recombination-stimulating sequence in the region containing the putative recombination point at which the sequences in plasmids pHH502 and pHH507 switch their close resemblance from the A cluster of ECOR strains in Figure 4 to the B cluster. The recombination-stimulating sequence is the chi sequence (5′-GCTGGTGG-3′), which is recognized by the *E. coli* enzyme RecBCD and promotes homologous recombination via the RecBCD pathway (SMITH 1987; WEST 1992). The *repA* gene includes a single-base variant of the chi sequence (5′-GCTGGTGA-3′) located at positions 270–277, which suggests that the *repA* recombination event in the common ancestor of plasmids pHH502 and pHH507 might well have been mediated by the chi site. Good matches to the chi sequence were also found at positions 179–186 in the *finO* gene (5′-CCTGGTGG-3′) and at positions 101–108 (5′-ACTGGTGG-3′) in the *traD* gene. No chi sites were observed in either of the *traY* genes.

The results of our study of F-related plasmids also bear on the history of the F factor present in laboratory strains. It has often been remarked that LEDERBERG was extremely lucky to use a strain in which the transfer functions of the F factor were derepressed owing to an insertion of IS*3* in the *finO* gene. However, the frequency of F-related plasmids in the ECOR strains shows that there is a reasonable chance that any randomly selected strain might have such a plasmid: 6% of the ECOR strains have an F plasmid [*traY* (F)] and 11% have an R plasmid [*traY* (R1)]. What is remarkable in the laboratory F factor is the insertion mutation in *finO*. In our PCR amplifications of 26 *finO* genes in plasmids in the ECOR collection, not one yielded a product other than the expected size. When did the insertion in the laboratory F factor take place? Probably not in nature. Nor is it likely to have happened in LEDERBERG's laboratory because he did not select for crossing ability among K-12 subcultures (J. LEDERBERG, personal communication). Very possibly, the insertion of IS*3* into the *finO* gene in the laboratory F arose from unconscious selection for plasmid transfer functions in the ~25 years in which the K-12 strain was stored in nutrient agar, with occasional subculturing, between the time of its isolation in 1922 to its use in genetic crosses in the late 1940s. Transposable insertion sequences are known to be active in strains stored under these conditions (GREEN *et al.* 1984; NAAS *et al.* 1994).

We are very grateful to JOSHUA LEDERBERG for the letter about his early bacterial crosses and to BARBARA BACHMANN for information about the early history of *E. coli* K-12. This work was supported by National Institutes of Health (NIH) grant number GM-40322 to D.L.H. and salary support from NIH grant GM-16329 to C.W.H.

LITERATURE CITED

ACHMAN, M., A. MERCER, B. KUSECEK, A. POHL, M. HEUZENROEDER *et al.*, 1983 Six widespread bacterial clones among *Escherichia coli* K1 isolates. Infect. Immun. **39:** 315–335.

Ayala, F. J., D. E. Krane and D. L. Hartl, 1994 Genetic variation in IncI1-ColIb plasmids. J. Mol. Evol. **39:** 129–133.

Barcak, G. J., and R. E. Wolf, 1988 Comparative nucleotide sequence analysis of growth-rate-regulated *gnd* alleles from natural isolates of *Escherichia coli* and *Salmonella typhimurium* LT-2. J. Bacteriol. **170:** 372–379.

Bergquist, P. L., S. Saadi and W. K. Maas, 1986 Distribution of basic replicons having homology with RepFIA, RepFIB, and RepFIC among IncF plasmids. Plasmid **15:** 19–34.

Boyd, E. F., K. Nelson, F.-S. Wang, T. S. Whittam and R. K. Selander, 1994 Molecular genetic basis of allelic polymorphism in malate dehydrogenase *(mdh)* in natural populations of *Escherichia coli* and *Salmonella enterica.* Proc. Natl. Acad. Sci. USA **91:** 1280–1284.

Bisercic, M., J. Y. Feutrier and P. R. Reeves, 1991 Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli:* evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. J. Bacteriol. **173:** 3894–3900.

Cavalli, L. L., and H. Heslot, 1949 Recombination in bacteria: outcrossing *Escherichia coli* K12. Nature **164:** 1057–1058.

Cram, D. S., S. M. Loh, K.-C. Cheah and R. A. Skurray, 1991 Sequence and conservation of genes at the distal end of the transfer region on plasmid F and R6–5. Gene **104:** 85–90.

DuBose, F., D. E. Dykhuizen and D. L. Hartl, 1988 Genetic exchange among natural isolates of bacteria: recombination within the *phoA* gene of *Escherichia coli.* Proc. Natl. Acad. Sci. USA **85:** 7036–7040.

Dykhuizen, D. E., and L. Green, 1991 Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. **173:** 7257–7268.

Finlay, B. B, L. S. Frost and W. Paranchych, 1986a Nucleotide sequences of the R1–19 plasmid transfer genes *traM, finP, traJ,* and *traY* and the *traYZ* promoter. J. Bacteriol. **166:** 368–374.

Finlay, B. B, L. S. Frost and W. Paranchych, 1986b Origin of transfer of incF plasmids and nucleotide sequences of the type II *oriT, traM,* and *traY* alleles from ColB4-K98 and the type IV *traY* allele from R100–1. J. Bacteriol. **168:** 132–139.

Gibbs, M. D., A. J. Spiers and P. L. Bergquist, 1993 RepFIB: a basic replicon of large plasmids. Plasmid **29:** 165–179.

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.

Hall, B. G., and P. M. Sharp, 1992 Molecular population genetics of *Escherichia coli:* DNA sequence diversity at the *celC, crr,* and *gutB* loci of natural isolates. Mol. Biol. Evol. **9:** 654–665.

Hartl, D. L., M. Medhora, L. Green and D. E. Dykhuizen, 1986 The evolution of DNA sequences in *Escherichia coli.* Phil. Trans. R. Soc. London B 312: 191–204.

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.

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.

Koraimann, G., and G. Hogenauer, 1989 A stable core region of the tra operon mRNA of plasmid R1–19. Nucleic Acids Res. **17:** 1283–1298.

Jalajakumari, M. B., and P. A. Manning, 1989 Nucleotide sequence of the *traD* region in the *Escherichia coli F* sex factor. Gene **81:** 195–202.

Kumar, S., K. Tamura and M. Nei, 1993 MEGA: molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park, PA.

Lawerence, J. G., H. Ochman and D. L. Hartl, 1991 Molecular and evolutionary relationships among enteric bacteria. J. Gen. Microbiol. **137:** 1911–1921.

Lederberg, J., and E. L. Tatum, 1953 Sex in bacteria: genetic studies, 1945–1952. Science **118:** 169–175.

Lederberg, J., L. L. Cavalli and E. M. Lederberg, 1952 Sex compatibility in *Escherichia coli.* Genetics **37:** 720–730.

Maynard Smith, J., 1992 Analyzing the mosaic structure of genes. J. Mol. Evol. **34:** 126–129.

McIntire, S. A., and W. B. Dempsey, 1987 Fertility inhibition gene of plasmid R100. Nucleic Acids Res. **15:** 2029–2042.

Milkman, R., and M. M. Bridges, 1990 Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics **126:** 505–517.

Naas, T., M. Blot, W. M. Fitch and W. Arber, 1994 Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. Genetics **136:** 721–730.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

Nei, M., and L. Jin, 1989 Variances of the average numbers of nucleotide substitutions within and between populations. Mol. Biol. Evol. **6:** 290–300.

Nelson, K., and R. K. Selander, 1992 Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli.* J. Bacteriol. **174:** 6886–6895.

Nelson, K., and R. K. Selander, 1994 Intergenic transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. Proc. Natl. Acad. Sci. USA **91:** 10227–10231.

Nelson, K., T. S. Whittam and R. K. Selander, 1991 Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli.* Proc. Natl. Acad. Sci. USA **88:** 6667–6671.

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

Riley, M. A., 1993a Molecular mechanisms of colicin evolution. Mol. Biol. Evol. **10:** 1380–1395.

Riley, M. A., 1993b Positive selection for colicin diversity in bacteria. Mol. Biol. Evol. **10:** 1048–1059.

Riley, M. A., and D. M. Gordon, 1992 A survey of Col plasmids in natural isolates of *Escherichia coli* and an investigation into the stability of Col-plasmid lineages. J. Gen. Microbiol. **138:** 1345–1352.

Riley, M. A., Y. Tan and J. Wang, 1994 Nucleotide polymorphism in colicin E1 and Ia plasmids from natural isolates of *Escherichia coli.* Proc. Natl. Acad. Sci. USA **91:** 11276–11280.

Saul, D., A. J. Spiers, J. McNulty, M. G. Gibbs, P. L. Bergquist *et al.,* 1989 Nucleotide sequence and replication characteristics of RepFIB, a basic replicon of IncF plasmids. J. Bacteriol. **171:** 2697–2707.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Sawyer, S., 1989 Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6:** 526–538.

Sawyer, S. A., D. E. Dykhuizen and D. L. Hartl, 1987 Confidence interval for the number of selectively neutral amino acid polymorphisms. Proc. Natl. Acad. Sci. USA **84:** 6225–6228.

Selander R. K., D. A. Caugant and T. S. Whittam, 1987 Genetic structure and variation in natural populations of Escherichia coli, pp. 1625–1654 in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology,* edited by F. C. Neidhardt, J. L. Ingraham, K. Brooks Low, B. Magasanik, M. Schaechter and H. E. Umbarger. American Society for Microbiology Press, Washington, DC.

Smith, G., 1987 Mechanism and control of homologous recombination in *Escherichia coli.* Annu. Rev. Genet. **21:** 179–201.

Stephens, J. C., 1985 Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. Mol. Biol. Evol. **2:** 539–556.

Van Biesen T., and L. S. Frost, 1992 Differential levels of fertility inhibition among F-like plasmids are related to the cellular concentration of *finO* mRNA. Mol. Microbiol. **6:** 771–780.

West, S. C., 1992 Enzymes and molecular mechanisms of genetic recombination. Annu. Rev. Biochem. **61:** 603–640.

Willetts, N., and R. Skurray, 1987 Structure and function of the F factor and mechanism of conjugation, pp. 1110–1133 in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology,* edited by F. C. Neidhardt, J. L. Ingraham, K. Brooks Low, B. Magasanik, M. Schaechter and H. E. Umbarger. American Society for Microbiology Press, Washington, DC.

Communicating editor: A. G. Clark