Correlation of *Rhs* Elements with *Escherichia coli* Population Structure

Charles W. Hill, Gregory Feulner, Margaret S. Brody, Sheng Zhao, Alesia B. Sadosky and Carol H. Sandt

Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033

Manuscript received November 4, 1994
Accepted for publication May 31, 1995

ABSTRACT

The *Rhs* family of composite genetic elements was assessed for variation among independent *Escherichia coli* strains of the ECOR reference collection. The location and content of the *RhsA-B-C-F* subfamily correlates highly with the clonal structure of the ECOR collection. This correlation exists at several levels: the presence of *Rhs* core homology in the strain, the location of the *Rhs* elements present, and the identity of the *Rhs* core-extensions associated with each element. A provocative finding was that an identical 1518-bp segment, covering core-extension-b1 and its associated downstream open reading frame, is present in two distinct clonal groups, but in association with different *Rhs* elements. The sequence identity of this segment when contrasted with the divergence of other chromosomal segments suggests that shuffling of *Rhs* core extensions has been a relatively recent variation. Nevertheless the copies of core-extension-b1 were placed within the respective *Rhs* elements before the emergence of the clonal groups. In the course of this analysis, two new *Rhs* elements absent from *E. coli* K-12 were discovered: *RhsF*, a fourth member of the *RhsA-B-C-F* subfamily, and *RhsG*, the prototype of a third *Rhs* subfamily.

THE *Rhs* elements represent a novel category of genetic elements (Hill *et al.* 1994). There are five *Rhs* elements in *Escherichia coli* K-12 (Feulner *et al.* 1990; Sadosky *et al.* 1991; Zhao *et al.* 1993), and a sixth has been isolated from strain ECOR-50 (Zhao and Hill 1995). Together, they comprise a family of complex genetic composites. Some of their components are highly conserved throughout the family, while other components show little or no sequence homology. The nonhomologous components, however, do share analogous features. A prototypical *Rhs* element is depicted in Figure 1A. Each known element varies somewhat from the structure shown even though all of the features represented are seen in most elements. The most prominent feature of each element is a large open reading frame (ORF) that can potentially encode a protein in excess of 150 kD. The function of this protein has yet to be established, but properties predicted from its primary sequence suggest that it may be a cell surface, ligand-binding protein (Hill *et al.* 1994). This large ORF is itself a mosaic, being comprised of two distinct components, a 3.7-kb conserved core and a highly divergent core extension. The G+C contents of the core and its extension are radically different, the cores being close to 62% G+C, while the core extensions are typically <40% G+C. Both of these values are distinctly different from the typical 51% G+C observed for *E. coli* sequences. This and other considerations led to the idea that these *Rhs* components evolved separately in backgrounds that were respectively GC rich and GC poor, and that they later joined together and entered the *E. coli* species (Feulner *et al.* 1990; Hill *et al.* 1994). A second, much smaller ORF lies in the AT-rich region immediately downstream from the core extension (Figure 1A). These downstream ORFs have dissimilar sequences, but most have N termini with characteristics expected of a signal peptide for export from the cytoplasm. This capability has been demonstrated in two cases, *RhsC* and *RhsD* (Zhao *et al.* 1993; Hill *et al.* 1994). The positions of six *Rhs* elements with respect to the *E. coli* K-12 map (Bachmann 1990) are shown in Figure 2.

Two additional features of the *Rhs* elements were of special importance for this study. The first is that the *Rhs* elements are divided into subfamilies, according to sequence divergence of the core components. The cores of the *RhsA-B-C-F* subfamily differ from the *RhsD-E* subfamily by 22% at the nucleotide level, while divergence within subfamilies is limited to about 4% (Sadosky *et al.* 1991; Zhao *et al.* 1993; Zhao and Hill 1995). Second, several *Rhs* elements exhibit a curious feature in that one or more partial repetitions of the core occur downstream from the primary core (Figure 1B). These core fragments vary in size, but always include the 3′ end of the core. Furthermore, they are always bordered by AT-rich extensions of the remnant core ORF. The fragments do not appear to be part of intact reading frames in that they do not maintain the start codon that defines the 5′ end of the core nor do they have alternative start codons. The significance of the 3′-core fragments is not clear, but a plausible hy-

Corresponding author: Charles W. Hill, Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, PA 17033.
E-mail: chill@cor-mail.biochem.hmc.psu.edu

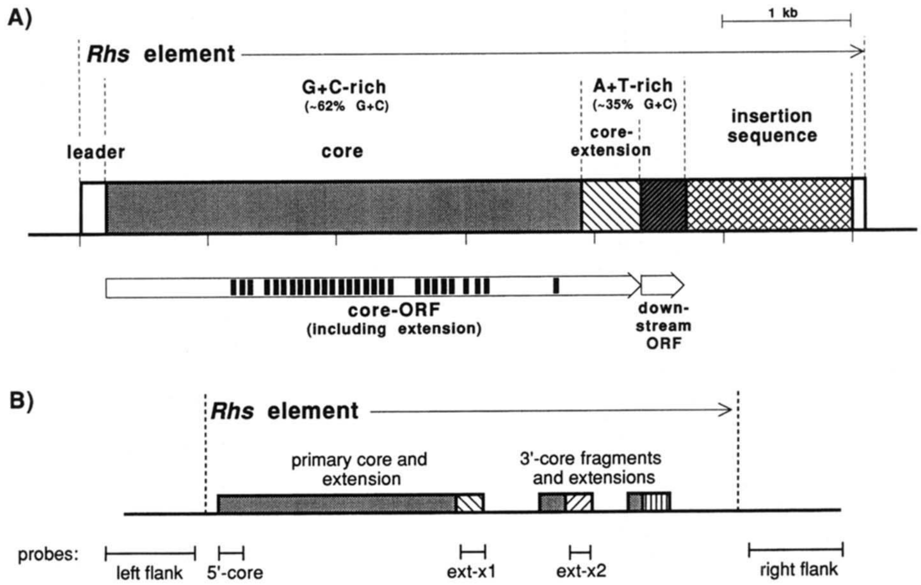


FIGURE 1.—Structural components of an *Rhs* element. (A) Schematic representation of an idealized *Rhs* element. The various components are identified above the diagram, while the extended core ORF and downstream ORF are shown below. Positions of a repeated peptide motif present in the core ORF are shown as black bars. Several different IS-like sequences have been found as parts of *Rhs* elements (ZHAO *et al.* 1993; ZHAO and HILL 1995). Of the various *Rhs* components, only the cores are clearly homologous for all elements. No known *Rhs* element has precisely the arrangement shown. (B) Schematic representation of an *Rhs* element containing 3′-core fragments and associated core extensions. Strategy for the selection of probes used in the analysis is illustrated: left flank, genomic DNA common to both Rhs^o and Rhs$^+$ strains upstream from the element; right flank, genomic DNA common to both Rhs^o and Rhs$^+$ strains downstream from the element; 5′ core; ext-1 and ext-2, DNA encoding different core extensions.

pothesis is that they and their extensions were once part of a primary core unit and that subsequently a new extension attached to the primary core by a mechanism that left a piece of the core and the old extension nearby. In strain K-12, there are a total of nine core extensions, five associated with the primary cores and four with a secondary 3′-core fragment. All nine are distinctly different in sequence. We have adopted a nomenclature for these extensions based on the following conventions: extensions present in K-12 are named according to the *Rhs* element with which they are associated and according to their position within the element. Thus ext-a1 is the extension associated with the primary core of K-12 *RhsA*, whereas ext-a2 is the extension associated with the first partial core repetition of *RhsA*. If the extension is subsequently found at a different location in a different strain, the original designation is nevertheless maintained.

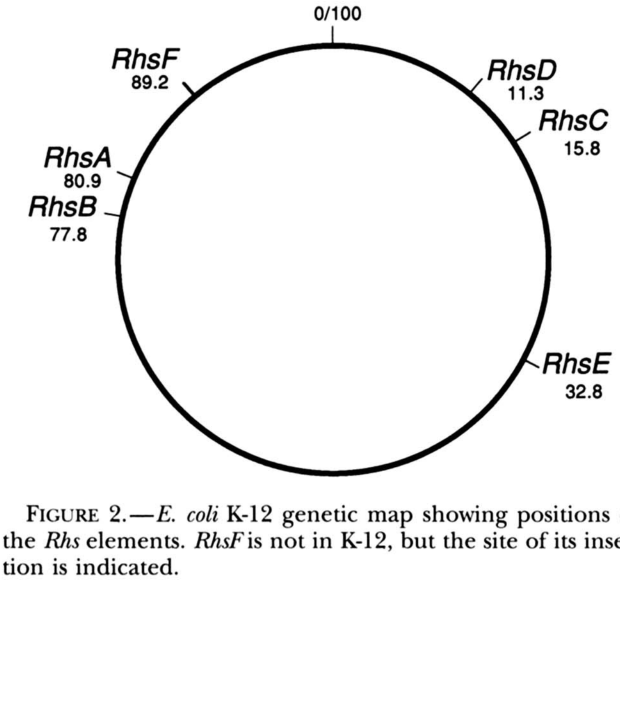


FIGURE 2.—*E. coli* K-12 genetic map showing positions of the *Rhs* elements. *RhsF* is not in K-12, but the site of its insertion is indicated.

Many kinds of accessory genetic elements have been described for *E. coli* species, including plasmids, prophages, insertion sequences and transposons (CAMPBELL 1981). The content of accessory elements tends to be different for each independent strain of *E. coli.* Of the various types of accessory elements, insertion sequences may be the most common and widespread in *E. coli* populations. Their numbers and locations, however, vary considerably from strain to strain and may vary between subcultures of individual clones (GREEN *et al.* 1984; LAWRENCE *et al.* 1989, 1992; NAAS *et al.* 1994; UMEDA and OHTSUBO 1990). In fact, insertion sequence profiles may be sufficiently variable that they distinguish strains so closely related as to appear identical to sensitive techniques such as multilocus enzyme electrophoresis (MLEE) (SAWYER *et al.* 1987).

Two considerations led us to question whether the *Rhs* elements of natural *E. coli* populations would be highly variable. First, by analogy to other accessory elements, we might expect *Rhs* elements to vary in number and position. Second, the presence of a different core extension adjacent to each of the nine 3′-core ends in K-12 might be an indication that there are many different extensions in the population. To assess these aspects of the *Rhs* elements, we surveyed a set of independent natural *E. coli* for their *Rhs* element content. The ECOR reference collection was chosen because it contains 72 natural *E. coli* isolates selected to embrace the genetic diversity of the species (OCHMAN and SELANDER 1984). MLEE has shown that clonal relationships exist among the ECOR strains (SELANDER *et al.* 1987), and four major clonal groups have been designated (HERZER *et al.* 1990). In addition to the four groups, the collection contains four outlying strains. This work demonstrates that the *Rhs* elements correlate closely with the clonality of the *E. coli* population, and that placement of core extension, ext-b1, must have occurred in at least one lineage between the time of clone divergence and the selective sweep that brought that clonal group to prominence.

MATERIALS AND METHODS

The ECOR strain collection was obtained from ROBERT SELANDER and THOMAS WHITTAM. The *E. coli* K-12 strain used was CGSC No. 4401. It contains an *F* factor and is lysogenic for lambda (BACHMANN 1973). Methods for DNA isolation and analysis were as described previously (ZHAO *et al.* 1993), with the exception that the more recent *E. coli* genomic DNA preparations utilized G-Nome DNA isolation kits (Bio101), and DNA transfers for Southern analysis used HyBond N membranes (Amersham Corp.). For hybridization studies, DNA fragments were separated by agarose gel electrophoresis, and labeled with (^{32}P)-nucleotide using a Boehringer-Mannheim random primed DNA kit. The DNA probes are listed in Table 1. Hybridization was done in 50% formamide, 0.75 M NaCl, 0.075 M sodium citrate, 0.02 M sodium phosphate, pH 6.5, 10% polyethylene glycol 8000, denatured calf thymus DNA (0.1 mg/ml) and Denhardt's reagent. The hybridization temperature was 42°, except 37° was used when reduced stringency was desired to enhance cross hybridization between *Rhs* core-specific probes. Aqueous washes were done at 42°. The *RhsA* element was cloned from ECOR-32 on a 20-kb *Sal*I genomic DNA fragment using pUC-19 as the cloning vector, resulting in pGF3510 (FEULNER 1990). The sequence of a 6980 bp segment, commencing at the *Mlu*I site 762 bp upstream from the left boundary of *RhsA* was determined and submitted to the GenBank database (accession number U16247). A portion of *RhsG* from ECOR-50 was cloned as a 7-kb *Bgl*II fragment resulting in pSZ3952 (ZHAO 1992). The sequence of the first 471 bp of its core was determined (accession number U25142). PCR amplification of genomic DNA was used to facilitate sequencing of the ext-b1 homologies from ECOR-2, 16, 24, 27 and 28. Specifically, two primers (5′-TGAATGAAGAGAACCCGCATCAGC-3′ and 5′-CTGGAGTCCTCCACCTAC-3′) were selected from the sequence of K-12 *RhsB* and used to amplify a 1121-bp segment that includes ext-b1 and the overlapping downstream ORF. The PCR primers were used along with a set of primers from within the fragment to sequence a 950-bp segment of each; this sequence included the last 105 bp of the core as well as the entire ext-b1 and downstream ORF. Direct sequencing of the PCR products utilized the CircumVent Thermal Cycle (New England Biolabs) sequencing system.

TABLE 1

DNA hybridization probes

Probe source[a]	Plasmid[b]	Identification[c]
RhsC 5′ core	pJG1637	0.4-kb *Eco*RV-*Eco*RV
RhsD 5′ core	pJG1997	0.2-kb *Bss*HII-*Eco*RV
RhsA, left flank	pJG1634 exoC	0.56-kb *Sal*I-exoIII
RhsA, right flank	pGF3504	0.4-kb *Pvu*II-*Pvu*II
RhsB, left flank	pJG1634exo10L	0.3-kb *Taq*I-exoIII
RhsB, right flank	pJG1890	0.5-kb *Hin*dIII-*Bgl*II
RhsC, left flank	pJG1637exo12C	0.6-kb *Eco*RI-exoIII
RhsC, right flank	pJZ3712	0.4-kb *Acc*I-*Pvu*II
RhsD, left flank	pAS3188	0.6-kb *Bst*EII-*Eco*RI
RhsD, right flank	pJG1883	0.3-kb *Nru*I-*Mlu*I
RhsF, left flank	pSZ3848	0.4-kb *Acc*I-*Sma*I
RhsF, right flank	pSZ3844	1.1-kb *Sma*I–*Bgl* II
ext-a1	pDV1868	0.2-kb *Ssp*I-exoIII
ext-a2	pJG1702	0.4-kb *Kpn*I-*Sph*I
ext-b1	pGF3545	0.5-kb *Eco*RV-*Pst*I
ext-c2	pSZ3966	0.2-kb exoIII-*Msc*I
ext-f3	pSZ3846	0.4-kb *Eco*RI-*Nde* I
iso-IS*1* (*RhsF*)	pSZ3847	0.5-kb *Hinc*II-*Hpa*I

[a] Conventions are as specified in Figure 1B. The various left flank, right flank and extension probes are all nonhomologous.

[b] All plasmids were derived from strain K-12 except pGF3545, which was from ECOR-2, and pSZ3836 and pSZ3847, which were from ECOR-50.

[c] exoIII refers to fact that the segment end was generated by controlled exonuclease III digestion.

RESULTS

Distribution of *Rhs* elements: The first questions addressed were whether the *Rhs* elements are widespread and whether the same elements found in *E. coli* K-12 are found in other strains. It should be emphasized that the *Rhs* elements are defined by their location, not by their specific structure or content. Therefore, we screened the ECOR strains for sequences homologous to the *Rhs* core, then attempted to locate their position within the bacterial genome.

Genomic DNA was prepared from each ECOR strain and from K-12, and the DNA was digested with *Mlu*I and with a combination of *Hin*dIII and *Sal*I. The resulting fragments were tested for homology to various *Rhs* components. After fractionation by agarose gel electrophoresis and transfer of the genomic DNA to membranes, the Southern blots were probed with a succession of probes. The initial probings were performed with DNA fragments derived from the 5′ ends of the *RhsC* and *RhsD* cores. Selection of these probes was based on two considerations. First, probes specific for the 5′ ends were used so that only the primary cores and not the 3′-core fragments would be detected (see Figure 1B). Second, the degree of homology between the subfamilies is such that a probe from a particular element gives a strong signal with other members of the same subfamily, but a weak signal with members of other subfamilies.

TABLE 2

***Rhs* core homology in the ECOR reference collection**

ECOR Group	*Rhs*$^+$ ECOR strains	*Rhs*^{o} ECOR strains
A	1, 5, 8, 10, 11, 25, 2, 3, 9, 12, 4, 6, 16, 22, 7, 14, 13, 18, 19, 20, 2, 17, 24, 15, 23	
B1	58, 67, 26, 27, 69, 28, 45, 29, 32, 33, 34, 30, 68, 70, 71, 72	
B2		51, 52, 54, 56, 57, 55, 65, 61, 62, 63, 64, 53, 59, 60, 66
D	46, 49, 50, 48	35, 36, 38, 39, 40, 41, 44, 47
Outlying	31, 43, 37, 42	

Use of a probe from each subfamily allows the subfamilies to be distinguished.

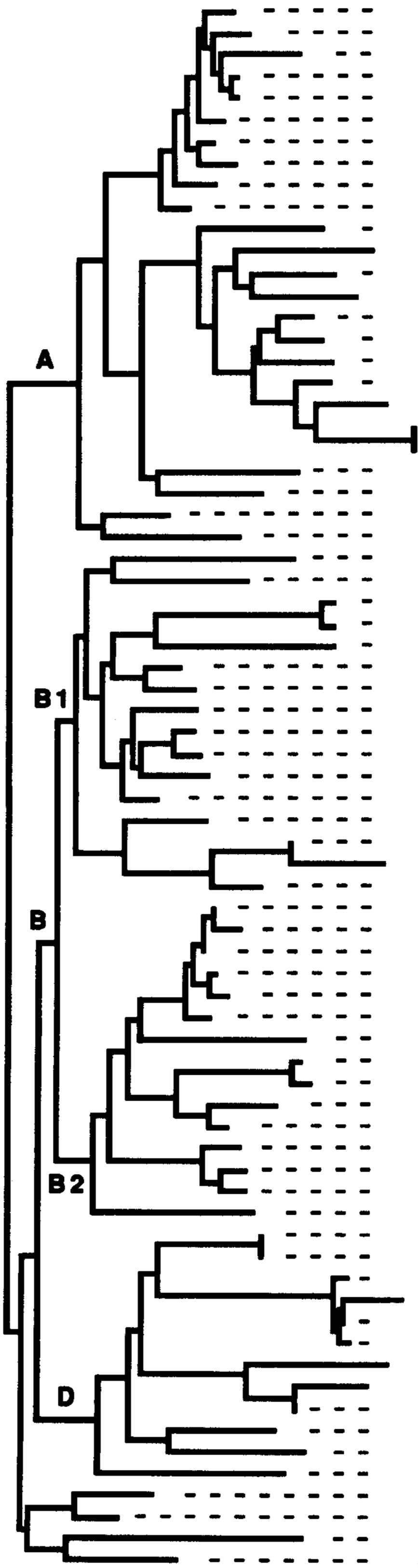
Initial tests revealed that DNA from 23 ECOR strains had no detectable homology to either core probe. These 23 strains were classified as Rhs^o (Table 2). The other 49 ECOR strains displayed multiple bands and were classified as Rhs$^+$. Inspection of Table 2 reveals that the Rhs$^+$ and Rhs^o strains were not randomly distributed among the four ECOR groups. All 15 group B2 strains were Rhs^o, while all group A and all group B1 were Rhs$^+$. In addition, the four outlying strains of the collection were Rhs$^+$. Only group D contained both Rhs$^+$ and Rhs^o strains, with four of the 12 strains testing positive. It should be noted that K-12 is closely related to the group A strains, especially to ECOR-2 and its nearest neighbors (Herzer *et al.* 1990).

In the early stages of this project, the absence of *Rhs* elements from some strains was used to define the chromosomal boundaries of individual elements. Specifically the corresponding region from a negative strain was cloned, and its sequence compared with that of a strain containing an *Rhs* element. The element boundaries were defined as the points of divergence between the Rhs$^+$ and Rhs^o strains. For *RhsA* and *RhsC*, ECOR-55 (Rhs^o) was compared to K-12. In contrast to *RhsA* and *RhsC* in K-12, 32 and 10 bp, respectively, of unrelated DNA were found at the same location in the ECOR-55 chromosome (Feulner *et al.* 1990). For *RhsD*, comparison of ECOR-39 with K-12 showed that *RhsD* was found in place of a 224-bp sequence (Sadosky *et al.* 1991). Finally, an 807-bp alternative to *RhsB* was detected when ECOR-32 was compared to K-12 (Zhao *et al.* 1993). This knowledge was used to develop a set of probes that permitted the assignment of the core homologies to specific locations. For each element, probes specific for the common regions on the left flank and on the right flank were selected (Figure 1B). The resulting patterns were compared. Coincidence of left flank, right flank and core signals was evidence that the particular *Rhs* element was present. In some instances where the left and right flank probe signals did not themselves coincide, the left flank signal nevertheless coincided with a core probe signal. Those strains were also scored as positive for that element. In the *Mlu*I analysis, this situation occurred only nine of 143 times for the *RhsA-B-C-F* subfamily. If the left flank and right flank signals coincided with one another but not with a core signal, that particular element was scored as being absent.

The results are tabulated in Figure 3. The elements hybridizing strongly with the *RhsC*-core probe, but weakly with the *RhsD*-core probe will be considered first. Of the 49 strains originally scored as Rhs$^+$, 46 had *RhsA* with an overlapping set of 46 having *RhsC*. However, only 27 of the 49 Rhs$^+$ strains had *RhsB*. A number of strains had an additional element whose core hybridized strongly to the *RhsC*-core probe and weakly to the *RhsD*-core probe, but which was not *RhsA*, *RhsB* or *RhsC*. This indicated the presence of at least one additional member of the subfamily. The additional element was characterized from ECOR-50, a group D strain (Zhao and Hill 1995). The new element was named *RhsF*, and it was found to replace 12 bp of the genome at a position corresponding to 89.2 min on the K-12 map (Figure 2). Based on this characterization, left flank and right flank probes for *RhsF* were prepared and *RhsF* was found in 24 of the 49 Rhs$^+$ strains. All *Rhs* elements in the ECOR collection giving strong signals with the *RhsC*-core probe were shown to be either *RhsA*, *RhsB*, *RhsC* or *RhsF*. Consequently, the subfamily has been redesignated *RhsA-B-C-F*. Inspection of Figure 3 reveals that there is a strong inverse correlation between the presence of *RhsB* and *RhsF* within the ECOR clonal groupings. *RhsB* is common in group A, but almost absent from group B1, whereas the opposite is the case for *RhsF*.

The results for elements hybridizing strongly with the *RhsD*-core probe, but weakly with the *RhsC*-core probe were more difficult to interpret. This was due in part to the presence of internal *Hin*dIII and/or *Mlu*I sites in many *RhsD* and *RhsE* elements and to the much greater complexity within the *RhsE* structure. Overall, *RhsD* is present in all but 11 of the 49 Rhs$^+$ strains (Figure 3). Investigation of *RhsE* continues, and results concerning *RhsE* are not reported here.

Prototype of a new core subfamily: In addition to



strain	*Rhs* elements					core-extension location				
	A	*C*	*B*	*F*	*D*	xa1	xb1	xa2	xc2	xf3
K-12	+	+	+		+	A	Ⓑ	A	C	
1	+	+	+		+	A	B	A	C	
5	+	+	+		+	A	B	A	C	
8	+	+	+		+		B	A	C	
10	+	+	+		+	A	B	A	C	
11	+	+	+		+	A	B	A	C	
25	+	+	+		+		B	A	C	
2	+	+	+		+	A	Ⓑ	A	C	
3	+	+	+		+	A	B	A	C	
9	+	+	+		+	A	B	A	C	
12	+	+	+		+	A	B	A	C	wk
4	+	+	+	+		A	B	A	C	F
6	+	+	+		+	A	B	A		
16	+	+	+			A	Ⓑ	A	C	
22	+	+	+	+		A	B	A	C	F
7	+	+	+		+	A	B	A	C	
14	+	+	+		+	A	B	A	C	
13	+	+			+	A		A	C	
18	+	+	+		+	A	B	A	C	
19	+	+	+		+	A	B	A	C	
20	+	+	+		+		B	A	C	
21	+	+	+		+		B	A	C	
17	+	+	+			A	B	A	C	
24	+	+	+			A	Ⓑ	A		
15	+	+	+	+		A	B	A	C	F
23	+	+		+		A		A	C	F
58	+	+		+	+		A	A	C	F
67	+			+			A	A		F
26	+	+		+	+		A	A	C	F
27	+	+		+	+		Ⓐ	A	C	F
69	+	+		+	+			A	C	F
28	+	+		+	+		Ⓐ	A	C	F
45	+	+		+	+		A	A	C	F
29	+	+		+	+		A	A	C	F
32	+	+		+	+		Ⓐ	A	C	F
33	+	+		+	+		A	A	C	F
34	+	+		+			A	A	C	F
30	+	+		+	+		A	A	C	F
68	+	+		+	+		A	A	C	F
70	+	+		+		A		A	C	F
71	+	+		+		A		A	C	F
72	+	+	+	+	+	A	B	A	C	F
51										
52										
54										
56										
57										
55										
65										
61										
62										
63										
64										
53										
59										
60										
66										
35										
36										
38										
39										
40										
41										
46	+	+	+	+	+			F	B	F
49		+	+	+	+			F	B	F
50		+	+	+	+			F	B	F
44										
47										
48	+	+			+				A	
31	+				+					
43	+				+		A			
37	+	+		+	+			A	C	F
42		+			+				C	
	A	*C*	*B*	*F*	*D*	xa1	xb1	xa2	xc2	xf3

FIGURE 3.—Distribution of *Rhs* elements and core extensions. The ECOR phylogeny as determined by MLEE is redrawn from HERZER *et al.* (1990). The presence of *Rhs* elements *A, B, C, D* and *F* were scored as described in the text; the presence of an element is indicated by + . Core-extensions ext-a1 (xa1), ext-b1 (xb1), ext-a2 (xa2), ext-c2 (xc2) and ext-f3 (xf3) were scored as described in the text; a capitalized letter indicates both presence and location of a core-extension. For the *RhsABCF* subfamily, signals generated by left flank and right flank probes for each location coincided in the *Mlu*I digests in all cases except *RhsA* of ECOR-20, 21 and 23, *RhsB* of ECOR-6, 14 and 46 and *RhsC* of ECOR-6, 13 and 50. A weak signal of unknown origin observed with the ext-f3 probe is indicated by wk.

hybridization signals that could be assigned to one of the six known *Rhs* elements, one or more additional bands hybridizing very weakly to both core probes were seen in some strains. One of these elements, from ECOR-50, was named *RhsG*, and has been partially characterized. Based on the 471-bp sequence nearest the 5′ end of its core, *RhsG* appears to be the prototype of a third core subfamily. Its sequence diverges from *RhsD* by 23% and *RhsA* by 27% at the nucleotide level. Its position relative to the K-12 map has yet to be established. No *RhsG*-core homology was detected in the Rhs^0 strains listed in Table 2.

Distribution of core extensions: We next determined the extent of variation among the core extensions. DNA probes were prepared to detect specific *Rhs* core-extensions among the Rhs$^+$ strains. Because *RhsA* is widely distributed among the ECOR strains, we first tested for ext-a1, which is the primary core-extension associated with K-12 *RhsA*. Ext-a1 homology was found in only 24 of the 46 *RhsA* strains (Figure 3). If ext-a1 was present in a particular strain, it was always associated with *RhsA*. Extension location was judged by coincidence of its signal with those produced by probes specific for the left flank and right flank of particular elements. Simple inspection of Figure 3 shows that the presence of ext-a1 homology correlates strongly with the ECOR grouping. Ext-a1 homology was most generally associated with strains of group A. Despite the fact that all 16 group B1 strains had *RhsA*, only three (the ECOR-70, 71 and 72 cluster) had ext-a1. None of the group D or outlying strains had ext-a1 homology even though several had *RhsA*.

We next tested for ext-b1, which is the primary core extension associated with K-12 *RhsB*. Considering that *RhsB* has a much more limited distribution than *RhsA*, we were surprised to find that ext-b1 homology occurred in more strains than had ext-a1 (37 strains compared with 24; Figure 3). Even more interesting was its location, where 12 group B1 strains that did not even have *RhsB* nevertheless had ext-b1. In these group B1 strains ext-b1 was associated with *RhsA*. In group A, ext-b1 was associated with *RhsB* (save for two strains that lacked *RhsB* and ext-b1 altogether). Besides the group A and group B1 strains, the ext-b1 homology was present in only one other strain, ECOR-43. In this outlying strain, it was associated with *RhsA*.

Some of the core-extensions found in K-12 *Rhs* elements were associated with secondary 3′-core fragments rather than primary cores (Figure 1B). It has been postulated that these were once associated with primary cores, but were displaced by rearrangements that introduced other extensions. Ext-a2 is the more proximal of two secondary core extensions associated with K-12 *RhsA*. Ext-a2 homology was found in 45 of the 49 Rhs$^+$ strains. It was associated with *RhsA* in all 25 members of group A and all 16 members of group B1 and the outlier ECOR-37. However, as described elsewhere (Zhao and Hill 1995), it was associated with *RhsF* in the group D cluster of ECOR-46, 49 and 50. For at least ECOR-50, it serves as the primary extension of *RhsF*. *RhsF* of ECOR-50 also has a secondary core extension, ext-f3. Ext-f3 has no homology in K-12. When the Rhs$^+$ strains of the ECOR collection were tested for ext-f3, however, there was a complete correlation between its presence and the presence of *RhsF* in the strain. Only strains with *RhsF* had homology to ext-f3, and it was always associated with *RhsF* in those strains. Adjacent to ext-f3 in *RhsF* of ECOR-50, there is an IS sequence that is a distant relative of IS*1*. Using a DNA probe from the interior of this IS, we found that every *Mlu*I band that had shown ext-f3 homology also hybridized with the iso-IS*1* probe. Furthermore, no bands were observed with the iso-IS*1* probe that had not been observed with the ext-f3 probe, save a weak band in ECOR-21. Thus this particular IS may have an exclusive relationship with *RhsF*.

Finally, another secondary core extension, ext-c2, was found to be associated with more than one element. Ext-c2 is associated with a 3′-core fragment of K-12 *RhsC* (Figure 3). A homologous sequence was detected in 44 strains, generally in association with *RhsC*. The four exceptional strains were all from group D. Ext-c2 was associated with *RhsB* in ECOR-46, 49 and 50, and with *RhsA* in ECOR-48.

Sequence conservation of the ext-b1 core extension: The preceding has shown a striking correlation between the location of ext-b1 and the ECOR groupings deduced from MLEE. In some cases, ext-b1 was associated with *RhsB* (as in K-12), and in others it was associated with *RhsA*. To better understand this relationship, *RhsA* was cloned from ECOR-32 and characterized by restriction mapping and nucleotide sequencing (Figure 4). The sequence data included a continuous 6980-bp segment that covered the 609-bp ORF immediately to the left of *RhsA* through the first 3′-core fragment (see materials and methods).

ECOR-32 *RhsA* showed many conserved features when compared with K-12 *RhsA* (Figure 4). The left and right boundaries were the same, confirming more precisely the correspondence of the two *RhsA* elements. The leaders separating the cores from the left boundaries were similar, differing by 9 of 190 bp. These leader regions are believed to contain the promoters for core expression, but the leaders of different *Rhs* elements are quite different (Zhao *et al.* 1993). ECOR-32 *RhsA* contained a primary core that differed from that of K-12 *RhsA* by 92 of 3714 bp (2.5%). This is comparable to the 2.9% divergence that was found between the K-12 *RhsA* and *RhsC* cores (Zhao *et al.* 1993). The core reading frame in ECOR-32 *RhsA* remained uninterrupted. The two 3′-core fragments of K-12 *RhsA* were preserved in ECOR-32 along with adjacent sequences.

The prominent difference between the *RhsA* elements of ECOR-32 and K-12 was that the primary core

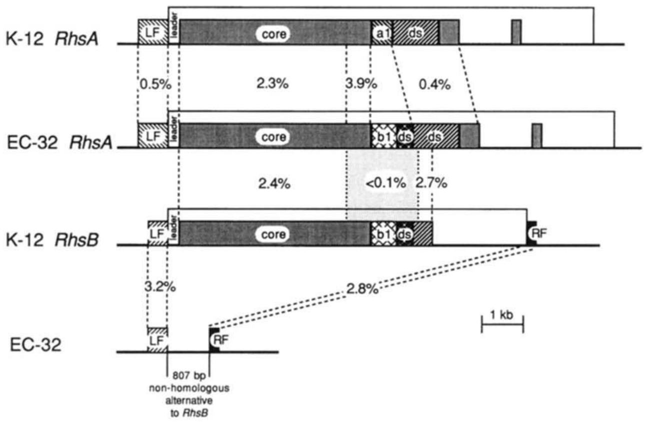


FIGURE 4.—Comparison of ECOR-32 *RhsA* to K-12 *Rhs* elements. The *RhsA* element of ECOR-32 is aligned with *RhsA* and *RhsB* of K-12. K-12 *RhsB* is in turn aligned with the analogous region of the ECOR-32 chromosome. The large white rectangles indicate the extent and positions of the whole *Rhs* elements. Relevant *Rhs* components and flanking ORFs are shown by smaller rectangles, with homologous structures signified by identical fillings. LF, left flanking ORF; RF, right flanking ORF; ds, downstream ORF; a1, core-extension ext-a1; b1, core-extension ext-b1. The nucleotide sequence divergence of homologous structures is expressed as a percentage, and the 1518-bp identity between ECOR-32 *RhsA* and K-12 *RhsB* is emphasized by shading. The structure of ECOR-32 *RhsA* beyond the first 3′-core fragment is based primarily on restriction site mapping and limited sequencing to verify key structures.

extension in ECOR-32 was ext-b1 instead of ext-a1. Also the downstream ORF immediately following ext-b1 was the same in both ECOR-32 *RhsA* and K-12 *RhsB*. The percentage divergence of *RhsA* from ECOR-32 with respect to homologous regions of *RhsA* and *RhsB* from K-12 is shown in Figure 4. One notable finding was that the sequence of a 1518-bp segment beginning in the core and covering the ext-b1 core extension and downstream ORF was absolutely identical in ECOR-32 *RhsA* and K-12 *RhsB*. Other homologies showed more divergence.

To determine whether the ext-b1 sequence identity seen in ECOR-32 and K-12 held true generally for groups A and B1, a 950-bp segment containing ext-b1 and the adjacent downstream ORF was sequenced from three group A strains (ECOR-2, 16 and 24) and from two group B1 strains (ECOR-27 and 28; MATERIALS AND METHODS). These five examples were found to be identical in sequence to the corresponding 950-bp sequence of both K-12 and ECOR-32. The seven strains for which ext-b1 was sequenced are indicated by circles in Figure 3. Because they represent different clusters from within group A and group B1, we conclude that the high sequence conservation can be generalized to be typical of ext-b1 within these groups.

DISCUSSION

The *Rhs* elements share, along with other accessory elements, the property of substantial variation among independent *E. coli* strains. However, *Rhs* element variation is restricted and correlates with established genetic groupings to a degree unprecedented for other accessory elements. This correlation is observed at three levels: the simple presence or absence of *Rhs* core homology, the presence of specific *Rhs* elements, and the association of specific core extensions with different *Rhs* elements. Our current analysis focuses on the *RhsA-B-C-F* subfamily; however, it should be emphasized that the Rhs° strains (Table 2) had no detectable homology to any of the three core subfamilies. The Rhs° class included all ECOR group B2 and some group D strains, while groups A and B1 were all Rhs$^+$. Of the four groups, group D was the only one that contained both Rhs$^+$ and Rhs° strains. According to the MLEE data (HERZER *et al.* 1990), the phylogeny of group D is consistent with its subdivision into four clusters (Figure 3): a) ECOR-35, 36, 38, 39, 40 and 41; b) ECOR-46, 49 and 50; c) ECOR-44 and 47; d) ECOR-48. The first and third of these clonal subgroups were Rhs°, while the second and fourth were Rhs$^+$ (Table 2); thus the presence of *Rhs* elements correlates within group D with the clonal relationships.

Assignment of core homologies to specific *Rhs* elements led to finer resolution of the Rhs$^+$ strains. Both *RhsA* and *RhsC* were widely distributed among Rhs$^+$ strains providing limited opportunity for discrimination, if the matter of core extensions is left aside. There was one interesting point with regard to *RhsA*; namely, of the three Rhs$^+$ strains that lacked *RhsA*, two are closest neighbors in group D (ECOR-49 and 50). In ECOR-50, a 32-bp alternative appears instead of *RhsA* (FEULNER 1990; FEULNER *et al.* 1990), and this 32-bp sequence differs by two bases from the otherwise identical sequence found in the Rhs° strains, ECOR-55 and 39.

The occurrence of *RhsB* and *RhsF* was more restricted among the ECOR strains, and their distribution correlated well with the MLEE data. *RhsF* was present and *RhsB* was absent throughout group B1, with the exception of ECOR-72. In contrast, *RhsB* was present in 23 of 25 group A strains, whereas *RhsF* was present in only four of the 25 strains. The only group A strain to have *RhsF* but not *RhsB* was ECOR-23, which is outlying among group A strains. For group D, the interrelated subgroup of ECOR-46, 49 and 50 had both *RhsB* and *RhsF*, while the group D outlier, ECOR-48, had neither.

A key question at the beginning of this study was whether the number and diversity of core extensions observed in K-12 reflects a reservoir of enormous variation in the *E. coli* population. In the extreme, the same extension might never be observed in two independent strains. The results show that the variation is not nearly that great. The same extensions are seen in many strains (Figure 3). However, the survey does establish an important type of variation, namely that the same extension can be associated with different elements in different strains. Several examples of this were seen. For

TABLE 3

Group specific *Rhs* profiles

Rhs profile	Example	*Rhs* profile characteristics[a]
I	K-12, ECOR-1, 5, 10, 11, 2, 3, 9, 12, 6, 16, 7, 14, 18, 19, 17, 24	A^{+} (ext-a1 and a2) C^{+} (ext-c2) B^{+} (ext-b1) F°
II	ECOR-58, 26, 27, 28, 45, 29, 32, 33, 34, 30, 68	A^{+} (ext-b1 and a2) C^{+} (ext-c2) B° F^{+} (ext-f3)
III	ECOR-70, 71, 23	A^{+} (ext-a1 and a2) C^{+} (ext-c2) B° F^{+} (ext-f3)
IV	ECOR-49, 50	A° C^{+} (ext-?) B^{+} (ext-c2) F^{+} (ext-a2 and f3)

[a] Core extensions associated with various *Rhs* elements are listed, but there is no implication as to the association with primary cores or with 3′-core fragments.

instance, ext-b1 is the primary extension of K-12 *RhsB*, and its association with *RhsB* was conserved in 23 ECOR strains. However, it was associated with *RhsA* in 13 other ECOR strains. Ext-a2 is a secondary extension of K-12 *RhsA*, and its association with *RhsA* was seen in 41 ECOR strains. However, it was associated with *RhsF* in three strains. Ext-c2, a secondary extension of K-12 *RhsC* was also associated with *RhsC* in 40 ECOR strains. However, it too varied in location, being associated with *RhsB* in three strains and *RhsA* in one. In contrast, ext-a1 and ext-f3 had exclusive associations with *RhsA* and *RhsF*, whenever these core extensions were present.

The presence and location of core extensions correlated with the ECOR groupings in much the same way as the complete elements, but there were some refinements. For example, the cluster of nearest neighbors within group D, ECOR-46, 49 and 50, had ext-a2 associated with *RhsF* and ext-c2 associated with *RhsB*. These were the only three strains where either of these relationships held. Within group B1, only three strains had ext-a1 associated with *RhsA*. These were the cluster of ECOR-70, 71 and 72, which, along with ECOR-68, form an outlying cluster of group B1 strains. Based on the *RhsA-B-C-F* subfamily results, we propose that there are group-specific *Rhs* profiles (Table 3): *Rhs* profile I shared by 15 group A strains including K-12; *Rhs* profile II shared by 11 group B1 strains, including ECOR-32; Rhs profile III shared by ECOR-70 and 71, both from group B1; *Rhs* profile IV shared by ECOR-49 and 50, both from group D.

Recombinants of these four profiles were also evident. ECOR-72, which is a nearest neighbor of ECOR-70 and ECOR-71 in the phylogeny determined by MLEE, provides such an example. These three strains all have *Rhs* profile III, except that ECOR-72 has an *RhsB* element similar to the one found in K-12. Because *RhsB* is otherwise absent in group B1, the simplest explanation is that a progenitor of ECOR-72 acquired *RhsB* through recombination. The same reasoning can be extended to ECOR-22 and ECOR-4. In their case, the presence of *RhsF* distinguishes their profiles from the *Rhs* profile I of their nearest neighbors. It is probable that progenitors of ECOR-22 and ECOR-4 acquired *RhsF* by recombination.

The genetic distance between ECOR groups should be related to the divergence of their homologous DNA sequences, and a considerable amount of sequence data has accumulated that allows such a comparison (DuBose *et al.* 1988; Dykhuizen and Green 1991; Milkman and Bridges 1993; Hall and Sharp 1992; Nelson and Selander 1992). Most comparisons have found that genes from group A and group B1 strains are more closely related to each other than they are to genes from groups B2 and D, although there are exceptions. Individual genes from group A strains appear to diverge from group B1 between 0.8 and 5.6%, depending on the gene, but values in the 1–2% range seem most typical. Divergence of genes within a clonal group tends to be less, on the order of 0.0–0.3%, although there are exceptions involving the *gnd* locus where divergence of 15% is seen (Dykhuizen and Green 1991). Recombination can both reduce variation of sequences between strains of different groups (DuBose *et al.* 1988) and increase variation between strains within groups (Guttman and Dykhuizen 1994a).

In this context, the sequence conservation of ext-b1 and its adjacent downstream ORF in *RhsB* of K-12, ECOR-2, 16 and 24 and *RhsA* of ECOR-32, 27 and 28 is exceedingly provocative. The identity of the 1518-bp segment of K-12 and ECOR-32 (Figure 4) contrasts starkly with the divergences of adjacent structures. Specifically, the 609-bp ORF immediately to the left of *RhsA* diverges 0.5% between K-12 and ECOR-32, and the sequences of ORFs flanking *RhsB* diverge somewhat more, ∼3%. Consequently, the chromosomal framework bordering the elements seems typical of the respective strains in terms of mutual divergence. Furthermore, the core of ECOR-32 *RhsA* diverges ∼2% from the cores of either *RhsA* or *RhsB* of K-12 over the first 3200 bp. The probability of finding a particular 1518-bp identity if the expectation of divergence is 1% per nucleotide is $<10^{-6}$. The absolute sequence conservation of ext-b1 and the adjacent downstream ORF applies to all group A and group B1 strains that possess it, judging from the identity of the 950-bp subsegment in three additional group A strains and two additional group B1 strains.

These considerations produce an interesting enigma. While the general chromosomal frames of group A and group B1 strains diverged long enough ago for their various genes to become ≥1% divergent, the ext-b1/downstream ORF structures are so recently related that they remain identical. Nevertheless, ext-b1 is present throughout the respective groups. The fact that the ext-b1 structures are at different chromosomal locations in the groups adds to the complexity.

The enigma resolves partially if we keep in mind that the divergence of the clonal groups does not necessarily coincide with their emergence as prevalent in the *E. coli* population. The linkage disequilibrium observed for the *E. coli* population implies that a limited number of highly successful clones comprise the bulk of *E. coli* in the world today (SELANDER *et al.* 1987). The factor(s) responsible for the success of these clones are unknown, but apparently the emergence of the successful clones is recent compared to the time required to obscure traces of clonality through recombination (GUTTMAN and DYKHUIZEN 1994a,b). The conservation of the ext-b1 structure can be explained if it was transferred from a member of one group to a member of the other group long after the groups diverged, but before the emergence of the second group as prevalent. If this recipient happened to be a progenitor of the founder cell responsible for the blossoming of the second clonal group, then ext-b1 would hitchhike on the periodic selection event and be observed as wide-spread. The great conservation of the ext-b1 structure would indicate that this happened in a cell much closer to the recent founder than to the more ancient common progenitor of groups A and B1. It is also possible that the ext-b1 structure was introduced from a more exotic source into the lineages of both clonal groups before the final periodic selection events. If this happened, the founders responsible for the blossoming of their respective clonal groups must be recent descendants of cells that received ext-b1. The absolute sequence conservation of the 950-bp segment among multiple strains of both group A and group B1 is consistent with this latter course.

A similar conserved relationship may exist for ext-a2, although available data is less extensive. Ext-a2 occurs as the primary extension of the *RhsF* core in ECOR-50, but as a secondary one in *RhsA* of K-12 (ZHAO and HILL 1995). Comparative sequence data is only available for 173 bp, yet the two versions of ext-a2 are identical within this limited region.

Additional questions concerning the establishment of *Rhs* profiles remain. The evidence suggests that the *Rhs* elements were in place long before the placement of the ext-b1 sequences at their current locations. First, there are only four locations for the *RhsA-B-C-F* subfamily, yet each element occurs in two or more forms when the core extensions are considered (Figure 3). This is most readily understood by assuming that elements inserted only once at a location, then diverged by acquisition/loss of their core extensions. The relationship between the respective segments separating the left boundary and the core of K-12 and ECOR-32 *RhsA* supports this idea. On the one hand, these leaders (which contain the presumed core promoter) are much more similar in sequence than either is to the leaders of *RhsB*, *RhsC* or *RhsF*. Those leaders are mutually highly divergent (ZHAO *et al.* 1993; ZHAO and HILL 1995). On the other hand, the two *RhsA* leaders have nine differences in 190 base pairs. Together, these facts suggest a common origin for *RhsA* in K-12 and ECOR-32. Multiple independent insertions at a given location cannot be ruled out absolutely, but the absence of any tell-tale homologies at the boundaries of the *Rhs* elements makes it seem unlikely.

These results and considerations discussed elsewhere regarding G+C content, sequence divergence and codon selection of *Rhs* components (FEULNER *et al.* 1990; HILL *et al.* 1994), suggest to us the following account of their development. The cores evolved in a GC-rich background, the three core subfamilies diverging long enough ago to permit 20–25% sequence divergence. Separately, the core extensions and the adjacent downstream ORFs evolved in an AT-rich background. Although limited homology can be seen between certain pairs, most extensions are so dissimilar as to leave no traces of common origin. At some point, cores and extensions became joined and at least one member of each core subfamily entered the *E. coli* species. We see nothing compelling that causes us to favor one of these events as preceding the other. Once in the *E. coli* species, the *RhsA-B-C-F* subfamily established itself (minimally) at the four known locations, the *RhsD-E* subfamily at the two known locations and the *RhsG* subfamily at two or more unknown locations. The divergences of the cores of the *RhsA-B-C-F* subfamily are small enough (<5%) that the divergence may have occurred after introduction of a single prototype into *E. coli.* For each element (*i.e.*, location), both *Rhs*$^+$ and *Rhs*^{o} alternatives exist, and different combinations were presumably produced by general recombination. The elements were in place before the divergence of the ECOR clonal groups. A much later event consisted of the shuffling of core extensions, and possibly the acquisition of core extensions from exotic sources. Finally, periodic selection resulted in the emergence of clonal groups with specific *Rhs* profiles brought along by hitchhiking.

The development of *Rhs* profiles seems to have involved four different kinds of recombination. Insertion of the elements into the host chromosome required a nonhomologous mechanism. There seem to be no circumstances suggesting site specificity of insertion. Indeed, the events were ancient and possibly rare. The original attachment of core extensions to cores certainly required a site-specific mechanism (HILL *et al.* 1994), although the details of the mechanism are obscure. One could envision that chromosomes with different combinations of elements (*i.e.*, *Rhs* profiles, Table 3) could readily be produced by homologous recombination that excluded participation of the elements themselves. Finally, there is the matter of shuffling of extensions between elements. Many mechanisms could be imagined, and the results summarized in Figure 4 may offer some insight. The 1518-bp identity segment covering ext-b1 also included the last 483 bp of the core. To the left of this point, considerable divergence was observed for the core sequences. This suggests that a homology-based cross-over occurred in the 3′ end of the core during the movement of ext-b1. The nature of the event defining the right end is less clear, however. The insertion sequences generally found within *Rhs* elements (Figure 1) may contribute to core-extension shuffling.

Despite evidence discussed for the shuffling of core extensions between *Rhs* elements and the occurrence of recombinant *Rhs* profiles, a curious circumstance exists. To date, no core extension has been observed to occur twice in the same strain. This could be due to chance, but it raises the question of whether a cell with such a profile would be at a selective disadvantage.

The functions, let alone the nature of selective pressures exerted on the *Rhs* elements, are completely unknown, yet the selective pressures must be very strong. As discussed elsewhere (Hill *et al.* 1994), the maintenance of such large ORFs as seen in the *Rhs* cores would demand strong, independent selection for each given the extent of mutual sequence divergence. The conservation of complex *Rhs*-profiles (Table 3) is also suggestive of strong selection. Understanding their function has been hampered by the fact that core ORFs are not expressed under conditions of routine laboratory cultivation, at least in K-12 (Hill *et al.* 1994). We suspect that they respond to environmental signals and play a role in the success of the cell in natural settings. A legitimate question is whether this function might relate in a key way to the periodic selection event(s) that influence *E. coli* population structure.

We thank Jane Gray and Du Chungen for expert technical assistance. We thank Robert Selander and Thomas Whittam for bacterial stocks. C.W.H. acknowledges the hospitality of Dan Hartl during the writing of this manuscript. This work was supported by Public Health Service grant GM-16329 from the National Institute of General Medical Sciences.

LITERATURE CITED

Bachmann, B. J., 1973 Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol. Rev. **36:** 525–557.

Bachmann, B. J., 1990 Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:** 130–197.

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

DuBose, R. 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.

Feulner, G. J., 1991 The structural characterization of *RhsA* and *RhsB*: exchange of a core extension. Ph.D. Thesis, The Pennsyvania State University, Hershey.

Feulner, G., J. A. Gray, J. A. Kirschman, A. F. Lehner, A. B. Sadosky *et al.*, 1990 Structure of the *rhsA* locus from *Escherichia coli* K-12 and comparison of *rhsA* with other members of the *rhs* multigene family. J. Bacteriol. **172:** 446–456.

Green, L., R. D. Miller, D. E. Dykhuizen and D. L. Hartl, 1984 Distribution of DNA insertion element IS*5* in natural isolates of *Escherichia coli.* Proc. Natl. Acad. Sci. USA **81:** 4500–4504.

Guttman, D. S., and D. E. Dykhuizen, 1994a Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science **266:** 1380–1383.

Guttman, D. S. and, D. E. Dykhuizen, 1994b Detecting selective sweeps in naturally occurring *Escherichia coli.* Genetics **138:** 993–1003.

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.

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

Lawrence, J. G., D. E. Dykhuizen, R. F. DuBose and D. L. Hartl, 1989 Phylogenetic analysis using insertion sequence fingerprinting in *Escherichia coli.* Mol. Biol. Evol. **6:** 1–14.

Lawrence, J. G., H. Ochman and D. L. Hartl, 1992 The evolution of insertion sequences within enteric bacteria. Genetics **131:** 9–20.

Milkman, R., and M. M. Bridges, 1993 Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons. Genetics **133:** 455–468.

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.

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.

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

Sadosky, A. B., J. A. Gray and C. W. Hill, 1991 The *RhsD-E* subfamily of *Escherichia coli* K-12. Nucleic Acids Res. **19:** 7177–7183.

Sawyer, S. A., D. E. Dykhuizen, R. F. DuBose, L. Green, T. Mutanga-dura-Mhlanga *et al.*, 1987 Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli.* Genetics **115:** 51–63.

Selander, R. K., D. A. Caugant and T. S. Whittam, 1987 Genetic structure and variation in natural populations of Escherichia coli, pp. 1625–1648 in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology,* edited by F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter and H. E. Umbarger. American Society for Microbiology, Washington, DC.

Umeda, M., and E. Ohtsubo, 1990 Mapping of insertion element IS*5* in the *Escherichia coli* K-12 chromosome. Chromosomal rearrangements mediated by IS*5*. J. Mol. Biol. **213:** 229–237.

Zhao, S., 1992 *Rhs* elements of *Escherichia coli*: complex composites of shared and unique sequences. Ph.D. Thesis, The Pennsylvania State University, Hershey, PA.

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

Zhao, S., C. H. Sandt, G. Feulner, D. A. Vlazny, J. A. Gray *et al.*, 1993 *Rhs* elements of *Escherichia coli* K-12: complex composites of shared and unique components that have different evolutionary histories. J. Bacteriol. **175:** 2799–2808.

Communicating editor: W. F. Eanes