



## The Diversity of Alleles at the *hsd* Locus in Natural Populations of *Escherichia coli*

Victoria A. Barcus,<sup>1</sup> Annette J. B. Titheradge and Noreen E. Murray

*Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom*

Manuscript received January 2, 1995  
Accepted for publication May 17, 1995

### ABSTRACT

In enteric bacteria three discrete families of type I restriction and modification systems (IA, IB and ID) are encoded by alleles of the *serB*-linked *hsd* locus. Probes specific for each of the three families were used to monitor the distribution of related systems in 37 of the 72 wild-type *Escherichia coli* strains comprising the ECOR collection. All 25 members of group A in this collection were screened; 12 were probe-positive, nine have *hsd* genes in the IA family, two in the IB and one in the ID. Twelve strains, representing all groups other than A, were screened; five were probe-positive, one has *hsd* genes in the IA family, one in the IB and three in the ID. The type ID genes are the first representatives of this family in *E. coli*, the probe-negative strains could have alternative families of *hsd* genes. The type IA and IB systems added at least five new specificities to the five already identified in natural isolates of *E. coli*. The distribution of alleles is inconsistent with the dendrogram of the bacterial strains derived from other criteria. This discrepancy and the dissimilar coding sequences of allelic *hsd* genes both imply lateral transfer of *hsd* genes.

**E**XCEPTIONALLY high intraspecific allelic diversity has been described for a number of loci in both eukaryotes and prokaryotes. This extreme genetic variability often correlates with a need to differentiate "foreign" from "self". In eukaryotes, examples of such systems are the MHC class II alleles in mammals (FIGUEROA *et al.* 1988; LAWLOR *et al.* 1988), mating-type loci in fungi (KÜES and CASSLETON 1992), and the self-incompatibility loci of certain plants (IOERGER *et al.* 1990). For these systems, selection for variation has resulted in the maintenance of a large number of alleles and high intraspecific sequence divergence consistent with gene lineages that predate speciation.

In bacteria, a high degree of variation among the genes encoding a variety of surface antigens is seen as a means of improving the bacterium's chances of escaping the host's immune system. For example, the somatic O lipopolysaccharide is a polymorphic surface antigen encoded by the *rfb* gene cluster; ~60 forms of the O antigen have been identified in *Salmonella*, and >160 in *Escherichia coli* (see REEVES 1993). In different antigenic groups of *Salmonella enterica*, *rfb* genes of limited sequence similarity are flanked by well-conserved DNA sequences. Recombination may replace one set of *rfb* genes with *rfb* alleles of dissimilar sequence (WANG *et al.* 1992). Other highly polymorphic systems include the flagellin genes of *S. enterica* (SMITH *et al.* 1990) and

the genes concerned with capsular serotypes in *E. coli* (DRAKE *et al.* 1993).

It is often argued that bacteria need to defend themselves against invasion by foreign DNA. Restriction and modification (R-M) systems enable bacteria to distinguish "foreign" DNA from their own (for a recent review, see BICKLE and KRÜGER 1993). The modification component of the system monitors the methylation state of the cell's own DNA and methylates specific bases within a recognition sequence, ensuring that newly replicated, hemimethylated, DNA will be fully modified. DNA with unmodified target sequences will be recognized as foreign and cleaved by the restriction component of the system. Considerable evidence already indicates allelic diversity for the genetic determination of type I R-M systems of enteric bacteria (reviewed in BARCUS and MURRAY 1995).

The type I R-M enzymes each comprise three subunits, encoded by the *hsdR*, *M* and *S* genes. The *S* and *M* subunits together form a DNA methyltransferase that methylates adenine residues, one on each strand within an asymmetric, bipartite recognition sequence. The methyltransferase component of some type I R-M systems has a preference for hemimethylated DNA. When all three subunits are present, the alternative activities of restriction and modification are dictated by the methylation state of the target sequence; hemimethylated targets are modified, unmethylated targets elicit restriction.

In *E. coli* K-12 the chromosomal genes encoding the type I system *EcoKI* are flanked on one side by *mrr* and on the other by *mcrBC* (see RALEIGH 1992). The *mrr* and *mcr* genes encode two additional restriction systems, but

Corresponding author: Noreen E. Murray, Institute of Cell and Molecular Biology, Darwin Building, The King's Buildings, Mayfield Road, Edinburgh EH9 3JR, UK.

<sup>1</sup> Present address: School of Biological Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, UK.

TABLE 1  
Target sequences of type I R-M systems of enteric bacteria

Family	Enzyme	Target	Reference
IA	<i>EcoKI</i>	ACC (N <sub>6</sub> ) GTGC	KAN <i>et al.</i> (1979)
	<i>EcoBI</i>	TGA (N <sub>8</sub> ) TGCT	LAUTENBERGER <i>et al.</i> (1978)
	<i>EcoDI</i>	TTA (N <sub>7</sub> ) GTCY	RAVETCH <i>et al.</i> (1978)
	<i>StyLTIII</i> <sup>a</sup>	GAG (N <sub>6</sub> ) RTAYG	NAGARAJA <i>et al.</i> (1985a)
IB	<i>StySPI</i>	AAC (N <sub>6</sub> ) GTRC	NAGARAJA <i>et al.</i> (1985b)
	<i>EcoAI</i>	GAG (N <sub>7</sub> ) GTCA	SURI <i>et al.</i> (1984)
	<i>EcoEI</i>	GAG (N <sub>7</sub> ) ATGC	COWAN <i>et al.</i> (1989)
	<i>CfrAI</i>	GCA (N <sub>8</sub> ) GTGG	KANNAN <i>et al.</i> (1989)
IC	<i>EcoR124I</i>	GAA (N <sub>6</sub> ) RTCG	PRICE <i>et al.</i> (1989)
	<i>EcoDXXI</i>	TCA (N <sub>7</sub> ) RTTC	GUBLER <i>et al.</i> (1992)
	<i>EcoprrI</i>	CCA (N <sub>7</sub> ) RTGC	TYNDALL <i>et al.</i> (1994)

<sup>a</sup> *StyLTIII* previously known as *StySB*.

ones that lack a corresponding modification component and are activated, not blocked, by the presence of methylated bases. The short (~15kb) region of the bacterial chromosome that includes *mrr*, *hsdRMS* and *mcrBC* has been referred to as the immigration control region (ICR), and is hypervariable in *E. coli* (see RALEIGH 1992). Alternative *hsd* genes within the ICR confer diagnostic specificities to the laboratory strains *E. coli* K-12, B and 15T<sup>-</sup> (BOYER 1964; ARBER and WAUTERS-WILLEMS 1970). Complementation tests indicated that the *hsd* genes of *E. coli* strains B and K-12 are sufficiently similar that their polypeptide products are exchangeable and that the subunit encoded by *hsdS* confers sequence specificity to the R-M complex. *EcoKI* and *EcoBI* became founder members of a family of closely related type I R-M systems, subsequently called type IA.

A first hint that allelic genes might also encode sufficiently dissimilar type I R-M systems to warrant their separation into a different family came from hybridization screens of bacterial DNAs and serological screens of cell extracts. It was found, as expected, that the nucleotide sequences of the *hsd* genes for *EcoKI* and *EcoBI* hybridized to each other, and antibodies raised against *EcoKI* reacted with *EcoBI*. In contrast, DNA probes made from the *EcoKI* genes failed to hybridize with those of *E. coli* 15T<sup>-</sup> which encoded *EcoAI*; similarly antibodies against *EcoKI* did not cross-react with *EcoAI* (MURRAY *et al.* 1982). The *hsd* genes in these two strains are of very different nucleotide sequence, but they behave as alleles in genetic tests dependent on recombination.

Eleven naturally occurring type I R-M systems, each recognizing a unique DNA sequence have been allocated to three discrete families (see Table 1). The members of the IA and IB, but not IC, families are encoded by allelic genes. Two of the three IC representatives are specified by plasmid-borne genes. Interfamily comparisons of the predicted amino acid sequences of the polypeptides of representatives of the three families suggest that the genes are homologous (SHARP *et al.* 1992; MURRAY *et al.* 1993).

Unpublished evidence (A. J. B. TITHERADGE) shows that the *serB*-linked *hsd* genes of *S. enterica* serovar blegdam (BULLAS *et al.* 1980) encode the first member of a fourth family of type I R-M genes (ID); these *hsd* genes do not cross-hybridize with those encoding the other three families of enzymes.

In this paper, we use probes specific for the three families of allelic *hsd* genes to monitor the distribution of related type I R-M systems in *E. coli*. Representatives of the ID family were identified in *E. coli*, where previously this family was only known to occur in *Salmonella*. The distribution of *hsd* genes among different populations of *E. coli* is not consistent with the relatedness of the bacterial strains based on other criteria, including multilocus enzyme electrophoresis (MLEE). This is so for the different families of enzymes and for alternative specificities within one family.

#### MATERIALS AND METHODS

**Bacteria, bacteriophages and plasmids:** Members of the ECOR collection were generally obtained from two sources—the American Type Culture Collection (ATCC) and either Dr. T. S. WHITTAM or Dr. H. OCHMAN. Discrepancies were encountered for only two strains, ECOR18 (ATCC No. 35337) and ECOR23 (ATCC No. 35342). The lyophile ATCC No. 35337 was a mixture of two strains; one included *hsd* genes of the type IB family, the other failed to hybridize with all *hsd* probes. A negative result was obtained with the strain from T.S.W. The lyophile ATCC No. 35342 was negative with the *hsd* probes, while that from T.S.W. included *hsd* genes of the type IA family. The strain from the ATCC lyophile did not have the enzyme profile of ECOR23 (T. S. WHITTAM, personal communication).

Other bacterial strains used were the  $r_K^+m_K^+$  strain C600 (APPLEYARD 1954), an *hsdΔ* strain NM679 (KING and MURRAY 1995), a *recA hsdΔ* strain DL795 (D. R. LEACH, unpublished results), ED8654 an  $r_K^-m_K^+$  host (BORCK *et al.* 1976), and K803, which is *hsdS*<sup>-</sup> and therefore  $r_K^-m_K^-$  (WOOD 1966).

$\lambda$ vir was generally used as a test for restriction and modification. Phages grown on strains lacking a modification system are denoted by the symbol .0, e.g.,  $\lambda$ vir.0. Modifications imposed on phages grown on R-M-proficient strains are identified by the name of the R-M system, e.g.,  $\lambda$ vir.*EcoKI* and

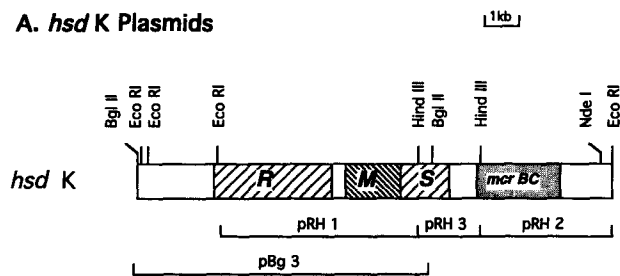
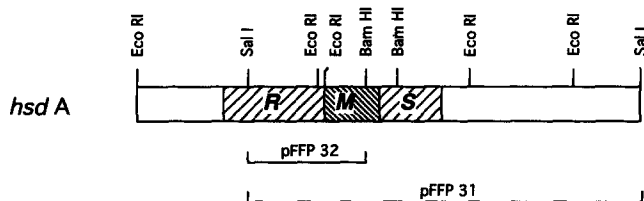
A. *hsd* K PlasmidsB. *hsd* A Plasmids

FIGURE 1.—Plasmids containing *hsd* and *mcr* genes. All are derivatives of pBR322. The type IA-specific probe, pRH1, has the *hsdM* gene, as well as parts of *hsdR* and *hsdS*, from *E. coli* K-12. The type IB-specific probe, pFFP32, contains a similar region from *E. coli* 15T<sup>-</sup>. pRH3 was used as a probe for type IA *hsdS* genes and pRH2 as a probe for *mcrBC*. The 3-kb *EcoRI* fragment downstream of *hsdS*, in pFFP31, and the *HindIII*-*NdeI* fragment spanning *mcrBC* in pRH2 were purified and used for probing cosmids. pBg3 was the source of HsdR and M polypeptides in complementation tests.

*lvir.EcoR5I*, where *EcoKI* is the type I R-M system of *E. coli* K-12 and *EcoR5I* denotes the type I R-M system identified in the ECOR5 strain. Phages P3 (BULLAS *et al.* 1980) and the P1 *darA* strain P1CmTc1 (IDA *et al.* 1987) were also used as substrates to test for restriction.

The plasmids used, derivatives of pBR322, have been described (SAIN and MURRAY 1980; FULLER-PACE *et al.* 1985) and are shown in Figure 1.

**Media and microbial techniques:** Media and general methods (MURRAY *et al.* 1977) and tests for estimating restriction and modification (FULLER-PACE *et al.* 1985) have been described. The complementation tests for a functional *hsdS* gene were done using the multicopy plasmid pBg3 (see Figure 1) as the source of HsdM and HsdR (KELLEHER *et al.* 1991).

**Enzymes and chemicals:** Restriction enzymes were purchased from Boehringer Mannheim UK except for *Sau3A*, which was obtained from Northumbria Biologicals Ltd. HK<sup>®</sup> phosphatase was purchased from Epicentre Technologies.

**Preparation, analysis and ligation of DNA:** Bacterial DNA was extracted using the miniprep procedure of REDFIELD and CAMPBELL (1987). Phage  $\lambda$  DNA was extracted according to a miniprep method devised by A. J. B. TITHERADGE (unpublished data). Vector DNA, however, was always isolated from phages that had been purified in a CsCl gradient. Restriction endonuclease digestion was done using the buffers and conditions recommended by the supplier, ligations by standard methods (SAMBROOK *et al.* 1989).

**Genomic libraries:** Libraries of *EcoRI* fragments were made in the  $\lambda$  replacement vector NM574 (BORCK *et al.* 1976). A library of fragments generated by partial digestion of ECOR9

DNA with *Sau3A* was made in  $\lambda$ NM1249, a derivative of EMBL3 in which the *dI857* gene replaces the deletion *dKH54*. The recombinant phages were recovered by *in vitro* packaging using extracts from Amersham International plc. Generally, *hsd* deletion hosts were used for the recovery, propagation and analysis of the recombinant phages.

Libraries of fragments generated by partial digestion of bacterial DNAs with *Sau3A* were made in the cosmid vector Supercos I (Stratagene Ltd) according to the supplier's instructions. Gigapack II Gold *in vitro* packaging extracts (Stratagene Ltd) were used to recover recombinant cosmids by infection of DL795.

**Detection of DNA sequences by hybridization:** Chromosomal DNA fragments from agarose gels and recombinant DNA molecules from plaques or colonies on agar plates were transferred to nylon filters (Hybond N, Amersham International plc) following standard procedures (see SAMBROOK *et al.* 1989).

The Boehringer DIG labeling and detection system was used as indicated by the suppliers, most recently using the Easy Hyb buffer recommended for the DIG system. Hybridization was carried out at 37° and washes were done at room temperature. For the analysis of bacterial DNA, *HindIII* digests of  $\lambda$  DNA labeled with DIG (Boehringer Mannheim UK) were used to provide size markers.

Plasmid probes were linearized or fragments were excised and purified (see Figure 1) before labeling. The ID-specific probe was a purified DNA fragment containing *hsd* genes from *S. enterica* serovar blegdam (A. J. B. TITHERADGE, unpublished results).

## RESULTS

The ECOR collection consists of 72 wild-type *E. coli* strains (OCHMAN and SELANDER 1984), chosen on the basis of MLEE to represent the genetic diversity of *E. coli* as a species and, in particular, of the major subspecific groups (OCHMAN *et al.* 1983; WHITTAM *et al.* 1983; SELANDER *et al.* 1986). Classification by MLEE is believed to give a reliable estimate of genetic relatedness of strains and generally agrees with classification based on restriction analyses and DNA sequence comparisons (OCHMAN *et al.* 1983; MILKMAN and BRIDGES 1990, 1993).

A dendrogram showing the relatedness of members of the ECOR collection, based on MLEE of 35 polymorphic enzymes (HERZER *et al.* 1990), provides a framework upon which to superimpose and analyze the distribution of other biological traits in evolutionary terms, in this case the *hsd* genes that encode type I R-M systems (see Figure 2). A total of 37 strains from the ECOR collection was probed for the presence of *hsd* genes from each of three families (IA, B and D). Of the 37 strains, 25 comprise group A, a section of the dendrogram including strains closely related to *E. coli* K-12 (HERZER *et al.* 1990). The others are distributed throughout the collection and include eight commonly used in other surveys of genetic diversity of *E. coli* (MILKMAN and CRAWFORD 1983; DUBOSE *et al.* 1988; DYKHUIZEN and GREEN 1991). The results of all the hybridizations are summarized in Table 2.

DNA from 12 (48%) of the strains in group A of the ECOR collection hybridized to one of the family-

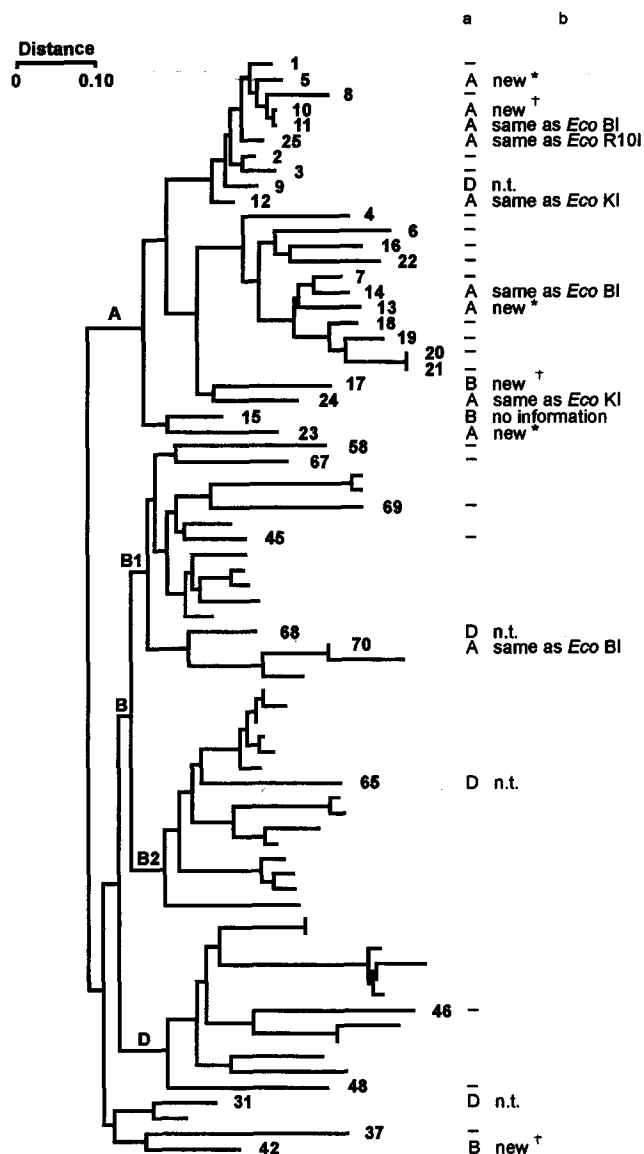


FIGURE 2.—Dendrogram of the ECOR strains. This tree was derived by HERZER *et al.* (1990) on the basis of polymorphisms at 38 enzyme-encoding loci. Only those strains tested with *hsd* probes are identified. Column a shows the results from probing genomic DNA. A, hybridization with the IA-specific probe; B, hybridization with the IB-specific probe; D, hybridization with the ID-specific probe; “—” indicates no hybridization detected with any of the three *hsd*-specific probes. Column b, summarizes the information on the specificity of the R-M systems identified. The sources of the ECOR strains are given in MATERIALS AND METHODS. \*, see Table 3 for additional information; †, see Table 4 for additional information; n.t., not tested—the type ID members are the first representatives of this family in *E. coli*.

specific probes (see Table 2); nine to the IA-specific probe (pRH1, see Figure 1), two to the IB probe (pFFP32, see Figure 1), and one to the ID probe. No strain hybridized to more than one probe. Outside of group A, five of 12 strains tested (42%) had DNA that hybridized to one of the probes; one to IA, one to IB, and three to ID.

The ECOR strains used in this survey were also screened for the presence of the genes encoding the McrBC restriction system (RALEIGH 1992). The *mcrBC* genes have been found previously only in strains encoding type IA R-M systems (DANIEL *et al.* 1988), as if the IA family *hsd* genes and the *mcrBC* genes have been acquired concomitantly. However, *mcrBC* genes were not detected in all strains with type IA *hsd* genes; both E166 (encoding *EcoDI*) and *S. enterica* serovar typhimurium (*Sty*LTHIII) lacked these genes. The results of the current survey reinforce the conclusions of DANIEL *et al.* (1988). Only strains that hybridized to the IA-specific probe hybridized to the *mcrBC* probe, and ECOR5, though positive with the type IA-specific probe, was negative with the *mcrBC*-specific probe (Table 2).

**Cloning and analysis of type IA *hsd* genes in  $\lambda$  vectors:** In *E. coli* K-12, the IA-specific probe (pRH1, see Figure 1) hybridizes to a single *EcoRI* fragment encompassing all of *hsdM* and *S* and most of *hsdR* (SAIN and MURRAY 1980). Additionally, the *mcrBC* genes are located in this fragment (ROSS *et al.* 1987; RALEIGH *et al.* 1989). Several of the ECOR strains (ECOR5, 12, 13, 23 and 24) resembled *E. coli* K-12 in that a single, large, *EcoRI* fragment hybridized to the *hsd*-specific probe. For each strain, this *EcoRI* fragment was cloned in an *att<sup>-</sup>ral<sup>-</sup>d857* replacement vector,  $\lambda$ NM574. The origin of the *hsd* region in the various recombinant phages is identified by the ECOR strain number (*e.g.*,  $\lambda$ 574-12 carries the *hsd* region of ECOR12). Each cloned fragment also hybridized to the *hsdS* probe (pRH3, see Figure 1) and, with the exception only of that in  $\lambda$ 574-5, to the *mcrBC* probe (pRH2, see Figure 1). The latter result is anticipated since the ECOR5 strain itself was negative with this probe. It is likely that all five phages include *hsdM* and *S*, though only a functional *hsdS* gene is essential to check specificity, since the M and R polypeptides can be provided in *trans*.

The specificity polypeptide encoded by each  $\lambda$  derivative was tested for its ability to form a restriction-proficient complex in the presence of the products of the *hsdR* (R) and *hsdM* (M) genes. Previous experiments (KELLEHER *et al.* 1991) have shown that a  $\lambda$ *hsd* phage encoding a functional S polypeptide has a low efficiency of plating (e.o.p.) on a nonmodifying strain carrying a multicopy plasmid (pBg3, see Figure 1) encoding the R and M polypeptides. A low e.o.p. is presumed to depend on the production of a functional endonuclease within a cell that lacks appropriately modified DNA. On the basis of this complementation test, all five phages encode a functional S polypeptide (see Table 3).

The differentiation of new specificities from those already characterized is permitted by an alternative test that also depends on the interchange of subunits between enzymes within the same family. In this test restriction-proficient lysogens are made by replacing the defective *S* gene of an *hsdR<sup>+</sup>M<sup>+</sup>S<sup>-</sup>* bacterium (K803) with the functional *S* gene of the  $\lambda$ *hsdS* phage. This is

TABLE 2  
SIZES OF *EcoRI* FRAGMENTS IDENTIFIED BY HYBRIDIZATION

ECOR No.	Group <sup>a</sup>	Probe			
		ID specific	IB specific	IA specific	<i>mcrBC</i>
5	A	—	—	12.5 kb	—
9	A	6 kb	—	—	—
10	A	—	—	5.5 kb	7 kb
11	A	—	—	6 kb	8 kb
12	A	—	—	13 kb	13 kb
13	A	—	—	11 kb	11 kb
14	A	—	—	5.5 kb	11.5 kb
15	A	—	3.5 kb	—	—
17	A	—	6 kb	—	—
23	A	—	—	11 kb	11 kb
24	A	—	—	12.5 kb	12.5 kb
25	A	—	—	6 kb	11 kb
31	* <sup>b</sup>	4 kb	—	—	—
42	* <sup>b</sup>	—	5.5, 4.5 kb	—	—
65	B2	4 kb	—	—	—
68	B1	4 kb	—	—	—
70	B1	—	—	5.5 kb	8 kb

The sizes given are approximate. DNA from the following ECOR strains did not hybridize to any of the four probes: 1, 2, 3, 4, 6, 7, 8, 16, 18, 19, 20, 21, 22, 37, 45, 46, 48, 58, 67, 69. ECOR 18, obtained from the American Type Culture Collection, was found to be a mixture of two strains. Only the contaminant strain hybridized to the IB-specific probe.

<sup>a</sup>This is the group assigned on the dendrogram of HERZER *et al.* (1990).

<sup>b</sup>ECOR31 and 42 are placed in the same group, but this was unassigned by HERZER *et al.* (1990).

readily achieved by the homology-dependent integration of the phage (GOUGH and MURRAY 1983).

The results of restriction tests are summarized in Table 3. The  $\lambda$ 574-5,  $\lambda$ 574-12, and  $\lambda$ 574-24 lysogens [K803(5), K803(12) and K803(24), respectively], restrict  $\lambda$ vir.0 with e.o.ps ranging from  $1 \times 10^{-3}$  to  $2 \times 10^{-4}$ . Neither the  $\lambda$ 574-13 nor the  $\lambda$ 574-23 lysogen restrict  $\lambda$ vir.0. The K803(23) lysogen did, however, restrict P1 *dar.0* and P3.0 ~100-fold, indicating that the ECOR23 *hsdS* gene imparts a specificity that has no target sequences in  $\lambda$ . As all known type I R-M systems

recognize targets in  $\lambda$ , the ECOR23 system has a novel specificity. On the basis of the complementation test, ECOR13 includes a functional *S* gene, but none of our tester phages was restricted by K803(13). The evidence for a new specificity in ECOR13, therefore, is circumstantial.

$\lambda$ vir modified by propagation on *E. coli* strains encoding type I R-M systems of different specificities were used to determine whether the ECOR5, 12 and 24 systems had the same specificity as any of those previously identified. The e.o.p. of various appropriately modified

TABLE 3  
ANALYSIS OF  $\lambda$  *hsd* PHAGES

Phage <sup>a</sup>	Complementation of <i>hsdR</i> <sup>+</sup> <i>M</i> <sup>+</sup> plasmid	E.o.p. of phages on lysogen <sup>b</sup>				Specificity
		$\lambda$ vir.0	$\lambda$ vir.K	P3.0	P1 <i>dar.0</i>	
$\lambda$ 574-5	+	$2 \times 10^{-4}$	$2 \times 10^{-4}$	nt	nt	New <sup>c</sup>
$\lambda$ 574-12	+	$7 \times 10^{-4}$	1	nt	nt	<i>EcoKI</i> <sup>d</sup>
$\lambda$ 574-13	+	1	nt	1	1	New <sup>c</sup>
$\lambda$ 574-23	+	1	nt	$4 \times 10^{-2}$	$1 \times 10^{-2}$	New <sup>f</sup>
$\lambda$ 574-24	+	$10^{-3}$	1	nt	nt	<i>EcoKI</i> <sup>d</sup>

nt, not tested.

<sup>a</sup>The number after the dash identifies the ECOR strain from which the donor fragment came.

<sup>b</sup>Phage ( $\lambda$  574-5, -12, etc.) integrated into the *hsdS*<sup>-</sup> derivative of *E. coli* K-12, K803, to give K803(5), etc.

<sup>c</sup> $\lambda$ vir modified by any known chromosomally-encoded type I R-M system is restricted by K803(5).

<sup>d</sup>See text for distinction between the specificities of *EcoKI* and *StyS*PI.

<sup>e</sup>Presumed to be a new specificity since it appears to restrict *E. coli* but not  $\lambda$ vir. All previously characterized type IA R-M systems have targets in  $\lambda$  DNA.

<sup>f</sup>Differs from K803(13) in that it restricts P3 and P1 *dar*.

$\lambda$  phages was checked on the restriction-proficient lysogens. Both  $\lambda vir.EcoKI$  and  $\lambda vir.StySPI$  plated with an e.o.p. of 1 on lysogens specifying either the ECOR12 or ECOR24 restriction system (Table 3). The *StySPI* R-M system recognizes a degenerate form of the *EcoKI* recognition sequence (NAGARAJA *et al.* 1985b), so that phages modified by *EcoKI* will still be restricted by *StySPI*.  $\lambda$  phages recovered after propagation on K803(12) and K803(24) were checked for modification against both *EcoKI* and *StySPI* restriction. The modification was only active against *EcoKI* restriction (Table 3), confirming that both ECOR12 and ECOR24 encode type I R-M systems with K-specificity.

K803(5) lysogens restricted  $\lambda vir$ , modified by propagation on strains encoding methyltransferases with specificities of all known type IA and IB R-M systems,  $\sim 10^3$ -fold (data not shown). This implies that ECOR5, like ECOR23, encodes a type I R-M system with a specificity previously not identified in the IA family.

**Cloning and analysis of type IA and IB *hsd* genes in cosmid vectors:**  $\lambda$  libraries have provided a reliable route for cloning *hsd* genes (BORCK *et al.* 1976; FULLER-PACE *et al.* 1985). The DNA of  $\lambda$  phages that include *hsdM* and *S* is modified with the specificity dictated by the cloned *hsdS* gene. However, the effective detection and analysis of different specificities need restriction-proficient bacteria and for new systems this will rely on the integration of a  $\lambda hsdMS$  phage into an *hsdS*<sup>-</sup> (*R*<sup>+</sup> *M*<sup>+</sup>) strain (see Table 3). The requirement for making lysogens can be obviated by a bacterium harboring a stably maintained, *hsd*<sup>+</sup> plasmid or cosmid, because this provides a direct test for both restriction and modification. The *hsd* region encoding *EcoAI* was transferred from a  $\lambda$  phage to a plasmid and the resulting transformants were restriction proficient (FULLER-PACE *et al.* 1985); no barrier to the acquisition of these type IB *hsd* genes was detected. For these reasons, the remaining type IA and IB *hsd* genes were cloned by making libraries of cosmids containing fragments generated from partial *Sau3A* digests of bacterial DNAs. Cosmids rather than plasmids were chosen to provide efficient recovery of large inserts.

The libraries were recovered in an *hsd* $\Delta recA$  host (DL795) and colonies harboring *hsd* genes were detected by their hybridization to probes specific for type IA or type IB systems. Positives from each library were tested for their ability to restrict  $\lambda vir.0$ . Restriction-proficient clones were isolated for five of the eight ECOR strains. These clones were also tested for their ability to restrict a collection of modified lysates of  $\lambda vir$  chosen to document the known repertoire of type IA and IB modification specificities. The five strains included two with previously unidentified specificities, one member of type IA and one of type IB (see Table 4).

No restriction-proficient clones were found for three ECOR strains (see Table 4), despite the expectation that the majority of cosmids selected would include the

entire *hsd* region. It is probable, though not proven, that some recombinant cosmids are not maintained sufficiently well for the host to be scored as restriction proficient. The library from ECOR70 gave the most favourable results: 10/14 clones were modification proficient and six of these were restriction proficient. However, even for these recombinants, if a culture was to be scored as restriction proficient, it was essential to grow the bacteria in medium supplemented with kanamycin (50  $\mu\text{g ml}^{-1}$ ), rather than ampicillin. Kanamycin enforces better maintenance of the cosmid than ampicillin. This observation stresses the importance of maintaining the cosmid because cultures in which only half of the cells have maintained the *hsd*<sup>+</sup> cosmid would score as restriction deficient. In the absence of a restriction-proficient strain there is no biological test for modification.

Another explanation for the absence of restriction-proficient clones is that the ECOR strains themselves have mutations that make their *hsd* genes defective. Alternatively there could be selection against the recovery of genes encoding a restriction enzyme in the absence of prior establishment of modification. This has been observed for some R-M genes encoding type II and type III systems (DEBACKER and COLSON 1991).

**The chromosomal location of three families of *hsd* genes in *E. coli*:** The archetypal members of the IA, IB and ID families of R-M systems are encoded by *serB*-linked *hsd* genes (BOYER 1964; ARBER and WAUTERS-WILLEMS 1970; BULLAS *et al.* 1980). It is likely that the *hsd* genes detected by the three family-specific probes (Table 2) share a common chromosomal location. Chromosomal segments of DNA known to be adjacent to *hsdS* in their respective *E. coli* strains, K-12 and 15T<sup>-</sup>, were used to probe the clones derived from the ECOR strains. The probe from *E. coli* K-12 spans *mcrB* and *C* (see Figure 1A and legend), that from 15T<sup>-</sup> is the 3-kb *EcoRI* fragment adjacent to *hsdS* (see Figure 1B and legend). The clones tested included those for the 13 type IA and IB *hsd* genes and one representative of the ID family. The data are summarized in Table 5.

Each clone derived from a chromosomal DNA shown to cross-hybridize with pRH2 was shown to include DNA that hybridized with the probe derived from pRH2, which includes most of *mcrB* and *C*. This result is consistent with the conservation of close linkage between the type IA *hsd* genes and the neighbouring *mcrB* and *C* genes of the immigration control region (RALEIGH 1992). All of the five clones derived from the DNA of ECOR strains that failed to cross-hybridize with pRH2 were found to hybridize with the flanking probe derived from *E. coli* 15T<sup>-</sup>. Three of the cosmids that include type IA *hsd* genes were tested with both probes and were positive with both as expected if the chromosomal location of the immigration control region is conserved in enteric bacteria.

TABLE 4  
Analysis of cosmid clones

Donor DNA	Probe	No. of restriction proficient clones <sup>a</sup>	Modification that protects against restriction	Specificity
ECOR10	IA	2 (10)	<i>EcoR25I</i>	New
ECOR11	IA	0 (13) <sup>b</sup>	nt <sup>c</sup>	Same as <i>EcoBI</i>
ECOR14	IA	1 (9)	<i>EcoBI</i>	Same as <i>EcoBI</i>
ECOR15	IB	0 (18)	nt <sup>c</sup>	No information
ECOR17	IB	0 (18)	nt <sup>c</sup>	<sup>d</sup>
ECOR25	IA	4 (7)	<i>EcoR10I</i>	Same as <i>EcoR10I</i>
ECOR42	IB	2 (6)	None	New
ECOR70	IA	6 (14)	<i>EcoBI</i>	Same as <i>EcoBI</i>

<sup>a</sup> Values in parentheses are number of clones tested.

<sup>b</sup> Two clones modify with the same specificity as *EcoBI*.

<sup>c</sup> In the absence of restriction, these tests cannot be done.

<sup>d</sup> Specificity known to be new (P. H. THORPE, personal communication)

#### DISCUSSION

**Occurrence:** The utility of restriction enzymes in molecular biology has motivated extensive searches for type II R-M enzymes with different specificities. JANULAITIS and coworkers (1988; A. JANULAITIS, personal communication) have screened the Enterobacteriaceae for sequence-specific endonucleases by incubating cell extracts with suitable DNA substrates, then looking for discrete fragments following separation by electrophoresis through agarose gels. Sequence-specific nucleases were detected in nearly 25% of 1000 strains tested. Many type II and type III R-M systems, however, are plasmid-encoded and, given the transmissibility of plas-

mids, chromosomally located genes are the preferred choice for analysis in the context of the dendrogram of ECOR strains.

The majority of known type I R-M systems are encoded by chromosomal genes, but these enzymes do not generate discrete fragments of DNA and their detection has often relied on the restriction of bacteriophages *in vivo* (BERTANI and WEIGLE 1953; BULLAS *et al.* 1980). This biological screen for R-M systems in natural isolates of bacteria is limited to strains that are sensitive to phage infection. Unfortunately, many wild-type *E. coli* strains are resistant to some aspect of productive infection by common laboratory phages, including even P1 (BARCUS 1993). Genes coding for members of known families of type I R-M systems, however, can be detected by hybridization to DNA probes derived from representative *hsd* genes (DANIEL *et al.* 1988). We have used this approach with probes that identify three discrete families of type I R-M enzymes, IA, IB and ID. The *hsd* genes for the representatives of each family have been shown to be allelic, although their nucleotide sequences are so dissimilar that no cross-hybridization has been detected between them (DANIEL *et al.* 1988; A. J. B. TITHERADGE, unpublished observations).

Of the 37 ECOR strains tested, 17 have DNA that hybridizes to one of the three family-specific probes. Because our analysis is limited by the available probes, we have no information for the remaining 20 strains, and it is quite possible that they have alternative families of *hsd* genes. Present experiments are exploring this possibility. BULLAS *et al.* (1980) screened a number of *Salmonella* serotypes for their ability to restrict phages and found evidence for eight R-M systems encoded by genes linked, like those for *EcoKI*, to *serB*. One of these is the first representative of the ID family and others could identify alternative families.

The numbers of families of type I R-M systems in the ECOR strains remains unknown and there is no indication how many, if any, of the strains will be like

TABLE 5  
Cocloning of *hsd* genes and the adjacent chromosomal DNA

ECOR strain	Vector	Family of <i>hsd</i> genes	Hybridization to chromosomal sequence	
			From K-12 <sup>a</sup>	From 15T <sup>-b</sup>
5	λ	A	- <sup>c</sup>	+
9	λ	D	- <sup>c</sup>	+
10	cosmid	A	+	+
11	cosmid	A	+	nt
12	λ	A	+	nt
13	λ	A	+	nt
14	cosmid	A	+	nt
15	cosmid	B	- <sup>c</sup>	+
17	cosmid	B	- <sup>c</sup>	+
23	λ	A	+	nt
24	λ	A	+	nt
25	cosmid	A	+	+
42	cosmid	B	- <sup>c</sup>	+
70	cosmid	A	+	+

nt, not tested.

<sup>a</sup> See Figure 1A and legend.

<sup>b</sup> See Figure 1B and legend.

<sup>c</sup> Bacterial DNA used to make clones does not hybridize with this probe.