



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

Victoria A. Barcus,¹ 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.

EXCEPTIONALLY 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.

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