INFECTION AND IMMUNITY, Aug. 1985, p. 407–413
0019-9567/85/080407-07$02.00/0

Vol. 49, No. 2

Genetic Diversity in Relation to Serotype in *Escherichia coli*

DOMINIQUE A. CAUGANT,[1,2] BRUCE R. LEVIN,[3]* IDA ØRSKOV,[4] FRITS ØRSKOV,[4] CATHARINA SVANBORG EDEN,[2] AND ROBERT K. SELANDER[1]

Department of Biology, University of Rochester, Rochester, New York 14627[1]; Department of Clinical Immunology, University of Göteborg, Göteborg, Sweden[2]; Department of Zoology, University of Massachusetts, Amherst, Massachusetts 01003[3]; and International Escherichia and Klebsiella Center (WHO), Statens Seruminstitut, Copenhagen, Denmark[4]

Received 22 January 1985/Accepted 13 May 1985

The extent of chromosomal-gene diversity among 261 isolates of *Escherichia coli* sharing single O, K, or H antigens and various combinations thereof was estimated by multilocus enzyme electrophoresis, which detects allelic variation in structural genes. The results of this study indicate that the genetic diversity among isolates sharing single antigenic determinants can approach or equal that observed among randomly chosen strains; that the magnitude of the diversity varies among antigens; and that the genetic diversity is reduced, but not eliminated, among isolates sharing two antigenic determinants. With one exception, isolates of the same O:K:H serotype were of identical or closely related electrophoretic types (ETs). Isolates of the same ET generally shared the same combination of antigenic determinants, but some ETs included isolates of different serotypes. The implications of these findings for epidemiological research and the clone hypothesis of population structure are discussed, and possible evolutionary mechanisms causing antigenic divergence and convergence are considered.

Since its introduction by Kauffmann in the 1940s (12, 13), serotyping has been the most widely used method of identifying strains of *Escherichia coli* for epidemiological and other purposes. The potential of serological techniques for strain discrimination is considerable. Some 170 O (lipopolysaccharide), 71 K (capsular), and 56 H (flagellar) antigens have been distinguished, and several thousand combinations of these antigens have been identified among isolates of *E. coli* from natural sources (26, 29, 30). Three epidemiologically important observations have been made concerning this extensive antigenic diversity. First, the relative frequencies of many of the antigens vary with the source of the isolates; for example, some O and K antigens are more common among strains associated with extraintestinal infections than among those recovered from the normal intestinal flora (19, 31). Second, the O, K, and H antigens are nonrandomly associated with one another, and specific O:K:H serotypes are, in turn, nonrandomly associated with certain other phenotypic traits, such as fermentation markers and virulence determinants (28). Third, some multiple-antigen serotypes associated with pathogenicity are worldwide in distribution and temporally stable (27, 28). Together, these observations were the basis for formulation of the "clone concept," a hypothesis postulating that the genetic structure of natural populations of *E. coli* is that of an array of stable cell lineages (clones), among which there is very little recombination of chromosomal genes (27).

Recently, electrophoretic analyses of outer-membrane proteins (1, 16) and water-soluble enzymes (8, 9, 21–23, 33, 35, 36) have been used to study genetic diversity and structure in natural populations of *E. coli*. Although several of these investigations (1, 22) have provided strong support for the clone hypothesis, they have also demonstrated that serotypes often do not mark genetically homogeneous or even similar groups of isolates. Hence, the effectiveness of conventional serological methods in the identification of clones has been questioned.

* Corresponding author.

To further examine the nature and extent of chromosomal-gene diversity within serotypes of *E. coli*, we analyzed electrophoretically demonstrable variations in enzymes among isolates recovered from humans and typed for O, K, and H antigens. Our results indicate that there is, in general, a considerable degree of similarity, if not identity, in electrophoretic type (ET) among isolates of the same O:K:H serotype, although genetically very different isolates may in some cases share the same three-antigen serotype. In contrast, isolates with the same one- or two-antigen determinants are generally quite diverse in chromosomal genotype. We also found evidence of antigenic diversity among isolates of the same multilocus enzyme ET.

MATERIALS AND METHODS

Bacterial isolates. A total of 261 isolates of *E. coli* recovered between 1970 and 1974 from fecal or urine samples from girls in Göteborg, Sweden, were analyzed. The collection included 93 isolates from the normal fecal flora and 168 isolates from urine. Among the latter, 47 isolates were from schoolgirls with asymptomatic bacteriuria, 70 isolates were from patients with acute cystitis, and 51 isolates were from patients with acute pyelonephritis. Each individual contributed only one isolate. A detailed analysis of genetic diversity in this collection of isolates has been reported elsewhere (7).

Enzyme electrophoresis. Each isolate was grown overnight in 100 ml of nutrient broth (Difco Laboratories) supplemented with 0.24 g of isopropyl-β-D-thiogalactopyranoside per liter to induce synthesis of β-galactosidase. After centrifugation, the bacterial pellet was suspended in 2 ml of 0.01 M Tris–0.001 M EDTA buffer (pH 6.8) and sonicated for 1 min with ice cooling. This suspension was centrifuged at 37,000 × *g* for 30 min, and the supernatant was stored at −70°C until used for electrophoresis.

Procedures employed for starch-gel electrophoresis and the demonstration of activity of specific enzymes were generally similar to those described by Selander et al. (34). The following 13 enzymes were assayed: malate dehydroge nase, 6-phosphogluconate dehydrogenase, β-galactosidase,

TABLE 1. Numbers of isolates assayed for ET and for O, K, and H antigens and combinations of the three

Character and category	No. of isolates		No. of types identified
	Assayed	Typed (%)	
ET	261	261 (100)	116
Single antigen			
O antigen	261		
Specifically determined		219 (84)	53
Rough (OR)		31 (12)	1
Nontypeable (ON)		11 (4)	1
K antigen	261		
Specifically determined		167 (64)	26
K−		45 (17)	1
K+		7 (3)	1
Nontypeable (KN)		42 (16)	1
H antigen	224		
Specifically determined		138 (62)	27
Rough (HR)		1 (1)	1
H−		71 (32)	1
Nontypeable (HN)		14 (6)	1
Antigen combination			
O:K determined	261	144 (55)	47
O:H determined	224	131 (58)	63
K:H determined	224	78 (35)	29
O:K:H determined	224	73 (33)	35

adenylate kinase, phenylalanyl-leucine peptidase, leucyl-leucyl-glycine peptidase, isocitrate dehydrogenase, phosphoglucose isomerase, aconitase, mannose-6-phosphate isomerase, glucose-6-phosphate dehydrogenase, alcohol dehydrogenase, and glutamic oxaloacetic transaminase.

The electrophoretic mobilities of the electromorphs (allozymes) were recorded, and electromorphs of an enzyme were equated with alleles at the corresponding structural gene locus. The ET of an isolate was defined by the combination of alleles at the 13 loci. Two lines of evidence suggest that the enzymes assayed are encoded by chromosomal rather than plasmid-borne genes. All 13 enzymes showed activity in several isolates in which plasmid could not be detected (8), and the genes for 11 of the enzymes have been mapped on the chromosome of *E. coli* K-12 (3).

Serotyping. The O, K, and H antigens of the isolates were determined by standard serological methods, as described by Ørskov and Ørskov (26). Briefly, O antigens were identified by agglutination in 164 standard O antisera of broth cultures boiled overnight. The presence of polysaccharide capsular K antigen was determined by the Cetavlon method (25), and K antigens were identified by countercurrent immunoelectrophoresis against 71 standard K antisera. H antigens were determined by agglutination tests of motile cultures with 56 standard H antisera. All isolates were tested for O and K antigens, but only 224 of the 261 isolates were analyzed for H antigen.

Statistical analyses. We calculated the number of electromorph differences between all pairs of isolates of specific serotypes and the mean ($\overline{D}$) and variance (V) of these pairwise differences. $\overline{D}$ is a measure of allelic diversity in structural genes among isolates of a given serotype. Since 13 enzymes were analyzed, $\overline{D}$ could range from 0 (no diversity) to 13 (no identity). Statistical inference was based on a "bootstrapping" procedure. By using the Monte Carlo sampling technique on the collection of 261 isolates, we generated the distributions of $\overline{D}$ values that would be expected in random samples of different numbers of isolates. These distributions were estimates of the probability of obtaining the observed mean number of pairwise electromorph differences among isolates by chance alone.

RESULTS

Overall serotypic and electrophoretic diversity. Results of the serotypic analyses of the 261 isolates are summarized in Table 1. Among the three single-antigen serotypes, serotype O was the most discriminating, with the highest frequency of typeable isolates (84%) and the largest number (53) of types. The number of isolates that could be serotyped for multiple specific antigens was smaller than that for single antigens. Among the antigenic pairs, O:H was the most discriminating, with 63 distinct antigenic combinations among the 131 typeable isolates. Of the 224 isolates tested for all three antigens, only 33% could be assigned a complete O:K:H serotype.

All 13 enzyme loci were polymorphic, and the mean number of alleles per locus was 5.3 (range, from 2 alleles for glutamic oxaloacetic transaminase to 9 alleles for β-galactosidase). All 261 isolates were typeable by this procedure, and 116 distinct ETs were identified. For more detailed analyses of enzyme variation in these isolates and the distribution of ETs among samples from fecal and urinary tract sources, see reference 7.

Among the 116 ETs, 17 differed from one another by having a "null" phenotype (no enzyme activity) for one or two enzymes. Most of these nulls involved alcohol dehydrogenase, but null phenotypes were also recorded for β-galactosidase, 6-phosphogluconate dehydrogenase, phenylalanyl-leucine peptidase, isocitrate dehydrogenase, and mannose-6-phosphate isomerase. Many of the nulls may represent mutations selected for during the 8 or so years that the isolates were stored in culture in the laboratory (7). ETs that differed by single null alleles were grouped in the same ET class. There were 99 such ET classes among the 116 ETs. In estimating $\overline{D}$, null alleles were included, but in Tables 2, 3, and 4, cases in which ETs differed solely in the occurrence of nulls are noted.

Electromorph variation within serotypes. ET variation was analyzed within each single-antigen serotype represented by five or more isolates (Table 2). The numbers of ETs and of electromorph differences between paired ETs demonstrated the existence of extensive genetic variation among isolates sharing the same single antigenic determinant. However, all single-antigen serotypes were not equally variable. For example, there was relatively little variation among O16, K13, and H1 isolates, whereas $\overline{D}$ values for O8, K2, and H4 isolates were not significantly smaller than those for groups of randomly chosen isolates. $\overline{D}$ values for isolates assigned to nonspecific serotypes (OR, ON, K−, KN, K+, and H−) were as large as those for randomly chosen isolates. In some cases, the genetic differences between groups of isolates characterized by the same single antigenic determinant were very large. This was indicated by a large variance in D relative to the $\overline{D}$ value. For example, 13 of the 16 isolates of serotype O7 belonged to a single ET class. The three remaining O7 isolates were identical to each other but were of an ET class that differed from the first one in as many as 8 of the 13 enzymes.

In general, the amount of genetic variation among isolates typed for multiple antigenic determinants was less than the variation for those classified by single antigens (Tables 3 and 4). For all two- and three-antigen serotypes, there was less electromorph variation than for randomly chosen isolates. In fact, there was no variation among isolates of the O16:K1 or O16:K1:H6 serotypes examined or among isolates of

serotypes O25:K5, O25:H1, or O25:K5:H1. For several other two- and three-antigen serotypes, the only variation was the absence of activity of one or two enzymes in some isolates. On the other hand, there was a substantial amount of enzyme variation among some isolates characterized by the same two- or three-antigen serotypes. Differences at more than four enzyme loci were noted for isolates of the O1:K1, O2:K1, O18:K5, O2:H4, and K1:H4 serotypes and also for isolates of the O2:K1:H4 serotype. The electromorph profiles of isolates sharing two- and three-antigen determinants are shown in Table 5.

Serotypic variation within ET classes. The O, K, and (when tested) H antigenic determinants of isolates within each ET class represented by five or more isolates are listed in Table 6. Some of the ET classes were largely associated with a single serotype. For example, most isolates of ET classes 1/32, 15/76/90, 19/34/44/102/105/113, and 55/92/99 were of serotypes O2:K1:H4, O7:K1:H−, O1:K1:H−, and O18:K1:H7, respectively. If the rough and other nontypeable isolates are pooled with those having specifically determined antigens, even more ET classes are dominated by single serotypes.

In contrast, some ET classes included isolates of several different serotypes. For example, among the 18 isolates of ET class 6/60/66, five O, five K, and four H antigens were identified. The 35 isolates of ET class 4/68 were characterized by three different O and five different K antigens, but only a single flagellar antigen, H1, was identified among them.

TABLE 2. Electromorph variation among isolates of the same single-antigen serotype[a]

Serotype	No. of:		$\bar{D}$	V	D_{max}
	Isolates	ETs			
O antigen					
O1	16	10	4.27[b]	8.25	8
O2	28	10	3.99[c]	9.26	9
O4	15	6	1.16[c]	1.83	5
O6	25	8	2.56[c]	6.92	9
O7	16	4	2.88[c]	12.91	8
O8	6	6	5.73	5.35	9
O16	10	2	1.80[c]	13.25	9
O18	13	8	3.01[c]	2.51	5
O25	15	5	3.15[c]	12.59	9
O75	9	3	0.83[c]	1.40	3
O83	7	4	4.48	7.86	8
O112	5	3	3.80	8.40	7
OR	31	26	6.54	5.27	11
ON	11	11	6.42	3.80	10
K antigen					
K1	74	29	4.68[c]	6.68	10
K2	13	8	4.65	12.33	9
K5	29	12	4.86[b]	8.81	10
K12	8	4	2.00[c]	6.67	7
K13	6	2	0.53[c]	0.27	1[d]
K53	7	4	6.27	16.91	10
K−	45	40	5.66	3.51	10
K+	7	6	5.43	4.06	8
KN	42	32	5.99	5.45	11
H antigen					
H1	26	3	0.29[c]	0.22	2
H4	26	13	5.71	9.79	11
H5	13	5	2.38[c]	3.64	5
H6	6	3	2.80[b]	6.17	6
H7	16	11	3.56[c]	5.12	8
H10	5	5	6.30	2.90	9
H18	5	4	4.20	5.96	7
H−	71	39	5.77	5.70	10
HN	14	7	2.64[b]	2.83	6

[a] $\bar{D}$ is the average number of electromorph differences between all pairs of strains within a serotype class, and V is the variance. D_{max} is the largest number of electromorph differences observed between pairs of isolates.
[b] The random probability of a $\bar{D}$ value as low as or lower than that observed is less than 5%.
[c] The random probability of a $\bar{D}$ value as low as or lower than that observed is less than 1%.
[d] The only electrophoretic difference involved "null" alleles.

TABLE 3. Electromorph variation among isolates of the same two-antigen serotype[a]

Serotype	No. of:		$\bar{D}^b$	V	D_{max}
	Isolates	ETs			
O:K serotype					
O1:K1	15	9	3.81	7.66	7
O2:K1	14	4	1.76	3.65	6
O2:K2	5	3	1.00	0.44	2
O2:K5	6	3	0.67	0.38	2
O4:K12	7	3	0.57	0.36	2
O6:K5	5	2	0.40	0.27	1[c]
O6:K13	6	2	0.53	0.27	1[c]
O7:K1	11	3	0.51	0.33	2
O16:K1	9	1	0.00		0
O18:K1	5	3	1.00	0.67	2
O18:K5	7	4	3.14	3.33	5
O25:K5	5	1	0.00		0
O:H serotype					
O2:H4	13	4	3.13	10.99	8
O4:H5	9	4	0.83	0.43	2
O6:H1	11	2	0.33	0.22	1[c]
O25:H1	10	1	0.00		0
K:H serotype					
K1:H4	14	4	2.37	10.48	9
K1:H6	5	2	1.20	1.07	2
K1:H7	10	7	1.71	0.98	4
K2:H1	5	2	0.60	0.27	1[c]
K5:H1	10	2	0.20	0.16	1[c]

[a] A See Table 2, footnote *a*, for explanation of abbreviations.
[b] See Table 2, footnote *c*.
[c] See Table 2, footnote *d*.

DISCUSSION

The present analysis confirms and extends the results of other electrophoretic studies of genetic variation in *E. coli* serotypes (1, 22, 24) in demonstrating that there often is substantial chromosomal-gene variation among strains ex-

TABLE 4. Electromorph variation among isolates of the same O:K:H serotype[a]

Serotype	No. of:		$\bar{D}^b$	V	D_{max}
	Isolates	ETs			
O1:K1:H7	4	3	1.00	0.40	2
O2:K1:H4	11	3	1.42	4.40	6
O4:K12:H5	4	2	0.50	0.30	1[c]
O6:K2:H1	3	2	0.67	0.33	1[c]
O6:K5:H1	5	2	0.40	0.27	1[c]
O16:K1:H6	3	1	0.00		0
O18:K1:H7	4	2	1.00	1.20	2[c]
O25:K5:H1	4	1	0.00		0
O25:K7:H1	4	1	0.00		0

[a] See Table 2, footnote *a*, for explanation of abbreviations.
[b] See Table 2, footnote *c*.
[c] See Table 2, footnote *d*.

TABLE 5. Electromorph profiles of isolates sharing two or three antigens

Antigen combination	ET	Electrophoretic relative mobility at enzyme locus[a]: MDH	βGA	6PG	AK	PE2	PE3	IDH	PGI	ACO	MPI	G6P	ADH	GOT	Third antigen	No. of isolates
O:K																
O1:K1	19	6	12	1	4	5	3	5	7	7	7	2	6	3	H−	5
	34	6	0	1	4	5	3	5	7	7	7	2	6	3	H−	1
	44	6	12	1	4	5	3	5	7	7	7	2	0	3	H−	1
	102	6	12	1	4	0	3	5	7	7	7	2	0	3	H−	1
	113	6	12	1	4	0	3	5	7	7	7	2	6	3	H−	1
	62	2	4	1	5	4	3	5	7	6	3	2	0	3	H−	1
	91	2	7	1	5	4	3	5	7	5	3	2	6	3	H7	1
	97	2	7	1	5	4	3	5	7	6	3	2	0	3	H7	1
	101	2	7	1	5	4	3	5	7	6	3	2	6	3	H−, H7	1, 2
O2:K1	6	2	7	6	4	4	3	5	7	6	3	2	6	3	H6	2
	1	2	7	14	5	4	3	5	7	6	3	2	0	3	H4	7
	32	2	7	14	5	4	3	5	7	6	3	2	6	3	H4, H7	3, 1
	15	6	7	4	5	5	3	5	7	7	7	2	6	3	H4	1
O2:K2	4	2	7	6	5	2	3	5	7	6	3	2	6	3	H1, HN	1, 1
	68	2	7	6	5	2	3	5	7	6	3	2	0	3	H1, HN	1, 1
	94	2	11	6	5	2	3	5	7	6	3	2	6	3	H−	1
O2:K5	25	5	6	14	4	0	3	5	9	5	7	2	6	3	H−	1
	57	5	6	14	4	5	3	5	9	5	7	2	6	3	H−, H4	2, 2
	115	5	6	14	4	5	3	5	9	5	7	2	0	3	H−	1
O4:K12	66	2	0	6	4	4	3	5	7	6	3	2	6	3	H5	1
	6	2	7	6	4	4	3	5	7	6	3	2	6	3	H5, HN	3, 2
	95	2	7	6	4	4	6	5	7	6	3	2	6	3	H−	1
O6:K5	4	2	7	6	5	2	3	5	7	6	3	2	6	3	H1	4
	68	2	7	6	5	2	3	5	7	6	3	2	0	3	H1	1
O6:K13	4	2	7	6	5	2	3	5	7	6	3	2	6	3	H1, HN	2, 2
	68	2	7	6	5	2	3	5	7	6	3	2	0	3	HN	2
O7:K1	15	6	7	4	5	5	3	5	7	7	7	2	6	3	H−	8
	76	6	7	4	5	5	3	5	7	7	7	2	0	3	H−	1
	90	6	7	4	5	5	3	5	7	7	0	2	6	3	H−	2
O16:K1	89	2	7	6	5	4	3	5	7	6	3	3	6	3	H−, H6, H−	4, 3, 2
O18:K1	55	2	7	0	5	4	3	5	7	6	3	2	0	3	H7	1
	92	2	7	9	5	4	3	5	7	6	3	2	6	3	H7	3
	99	2	7	9	5	4	3	5	7	6	3	2	0	3	HN	1
O18:K5	2	2	11	9	5	4	3	5	7	6	7	2	0	3	H−	1
	69	2	11	9	5	4	3	5	7	5	7	2	6	3	H−	3
	1	2	7	14	5	4	3	5	7	6	3	2	0	3	H−	1
	39	2	7	9	5	2	3	6	7	5	7	2	0	3	H−	2
O25:K5	4	2	7	6	5	2	3	5	7	6	3	2	6	3	H1, H−	4, 1
O:H																
O2:H4	1	2	7	14	5	4	3	5	7	6	3	2	0	3	K1	7
	32	2	7	14	5	4	3	5	7	6	3	2	6	3	K1	3
	15	6	7	4	5	5	3	5	7	7	7	2	6	3	K1	1
	57	5	6	14	4	5	3	5	9	5	7	2	6	3	K5	2
O4:H5	60	2	7	6	4	4	3	5	7	6	3	2	0	3	K−	1
	66	2	0	6	4	4	3	5	7	6	3	2	6	3	K12	1
	6	2	7	6	4	4	3	5	7	6	3	2	6	3	K3, K12, KN	1, 3, 1
	83	2	7	6	4	2	3	5	7	6	3	2	6	3	K−, KN	1, 1
O6:H1	4	2	7	6	5	2	3	5	7	6	3	2	6	3	K5, K2, K13, K53	4, 2, 2, 1
	68	2	7	6	5	2	3	5	7	6	3	2	0	3	K5, K2	1, 1
O25:H1	4	2	7	6	5	2	3	5	7	6	3	2	6	3	K5, K7, KN, K53	4, 4, 1, 1
K:H																
K1:H4	1	2	7	14	5	4	3	5	7	6	3	2	0	3	O2, OR	7, 1
	10	2	15	4	4	5	6	2	4	7	3	2	1	3	O12	1
	32	2	7	14	5	4	3	5	7	6	3	2	6	3	O2, OR	3, 1
	15	6	7	4	5	5	3	5	7	7	7	2	6	3	O2	1
K1:H6	6	2	7	6	4	4	3	5	7	6	3	2	6	3	O2	2
	89	2	7	6	5	4	3	5	7	6	3	3	6	3	O16	3
K1:H7	55	2	7	0	5	4	3	5	7	6	3	2	0	3	O18	1
	32	2	7	14	5	4	3	5	7	6	3	2	6	3	O2	1
	91	2	7	1	5	4	3	5	7	5	3	2	6	3	O1	1
	92	2	7	9	5	4	3	5	7	6	3	2	6	3	O18	3
	93	2	11	6	5	4	3	5	7	5	3	2	6	3	O6	1
	97	2	7	1	5	4	3	5	7	6	3	2	0	3	O1	1
	101	2	7	1	5	4	3	5	7	6	3	2	6	3	O1	2

Continued on following page

TABLE 5—*Continued*

Antigen combination	ET	Electrophoretic relative mobility at enzyme locus[a]: MDH	βGA	6PG	AK	PE2	PE3	IDH	PGI	ACO	MPI	G6P	ADH	GOT	Third antigen	No. of isolates
K2:H1	4	2	7	6	5	2	3	5	7	6	3	2	6	3	O2, O6	1, 2
	68	2	7	6	5	2	3	5	7	6	3	2	0	3	O2, O6	1, 1
K5:H1	4	2	7	6	5	2	3	5	7	6	3	2	6	3	O6, O25, OR	4, 4, 1
	68	2	7	6	5	2	3	5	7	6	3	2	0	3	O6	1
O:K:H																
O1:K1:H7	91	2	7	1	5	4	3	5	7	5	3	2	6	3		1
	97	2	7	1	5	4	3	5	7	6	3	2	0	3		1
	101	2	7	1	5	4	3	5	7	6	3	2	6	3		2
O2:K1:H4	1	2	7	14	5	4	3	5	7	6	3	2	0	3		7
	32	2	7	14	5	4	3	5	7	6	3	2	6	3		3
	15	6	7	4	5	5	3	5	7	7	7	2	6	3		1
O4:K12:H5	66	2	0	6	4	4	3	5	7	6	3	2	6	3		1
	6	2	7	6	4	4	3	5	7	6	3	2	6	3		2
O6:K2:H1	4	2	7	6	5	2	3	5	7	6	3	2	6	3		2
	68	2	7	6	5	2	3	5	7	6	3	2	0	3		1
O6:K5:H1	4	2	7	6	5	2	3	5	7	6	3	2	6	3		4
	68	2	7	6	5	2	3	5	7	6	3	2	0	3		1
O16:K1:H6	89	2	7	6	5	4	3	5	7	6	3	3	6	3		3
O18:K1:H7	55	2	7	0	5	4	3	5	7	6	3	2	0	3		1
	92	2	7	9	5	4	3	5	7	6	3	2	6	3		3
O25:K5:H1	4	2	7	6	5	2	3	5	7	6	3	2	6	3		4
O25:K7:H1	4	2	7	6	5	2	3	5	7	6	3	2	6	3		4

[a] MDH, Malate dehydrogenase; βGA, β-galactosidase; 6PG, 6-phosphogluconate dehydrogenase; AK, adenylate kinase; PE2, phenylalanyl-leucine peptidase; PE3, leucyl-leucyl-glycine peptidase; IDH, isocitrate dehydrogenase; PGI, phosphoglucose isomerase; ACO, aconitase; MPI, mannose-6-phosphate isomerase; G6P, glucose-6-phosphate dehydrogenase; ADH, alcohol dehydrogenase; GOT, glutamic oxaloacetic transaminase.

pressing the same antigenic determinant(s). The magnitude of this genetic diversity varies among antigenic determinants and with the number of antigenic determinants tested for strain identification. For some isolates sharing the same single-antigen serotype, the diversity approaches or equals that of randomly chosen isolates. This result could have been anticipated from the fact that DNA hybridization experiments did not demonstrate preferentially high relatedness between isolates of *E. coli* and *Shigella* spp. with identical or cross-reacting antigens (5, 6). Among isolates sharing two antigenic determinants, chromosomal-gene variation is reduced, but there is still a good possibility that very different strains will be serologically indistinguishable. Isolates sharing the same O, K, and H antigens tend to be of identical or similar ET, but we have shown that, in some cases, genetically distinctive isolates can have the same O:K:H serotype. Our analysis and that of Ochman et al. (24) also revealed variations in serotype among isolates of the same ET class.

Despite the relatively modest number of isolates of each serotype examined and the geographically limited source of the collection analyzed, the results of our investigation are fully consistent with the major findings of Ochman and Selander (22), who examined only four serotypes (O1:K1, O7:K1, O16:K1, and O18:K1), but a larger number of isolates from more diverse sources. In accordance with their results, the O1:K1 isolates analyzed here represented two major ET classes that were genetically very distinct (differing in alleles at 6 of the 13 enzyme loci); there was virtually no enzyme variation within the O7:K1, O16:K1 and O18:K1 isolates; and the O18:K1 isolates were closely related to one of the ET classes identified among the O1:K1 isolates (differing by only one or two electromorphs). However, the present study suggested that the two genetically distinct groups of O1:K1 isolates may be partially distinguished by the flagellar serotype. All nine isolates belonging to one of the O1:K1 genetic groups were nonflagellated (H−), whereas four of the six isolates belonging to the other group were H7, the two remaining isolates being H− (Table 5). The O18:K1 isolates were related to the O1:K1:H7 isolates, and they also were characterized by the H7 antigen. The present analysis also identified a number of additional cases in which isolates of the same two-antigen serotype (O2:K1, O2:H4, O18:K5, and K1:H7) differed by three or more electromorphs.

In interpreting the results of our study and evaluating their implications, the character of the collection of isolates examined should be considered. The 261 isolates were chosen for study because they had been extensively analyzed both for enzyme variation and for serotype. They were neither a random sample of *E. coli* nor a collection of specific serotypes, such as those studied by Achtman et al. (1) and Ochman and Selander (22). The restricted nature of the collection may limit the quantitative conclusions drawn from our analysis, but probably not the qualitative ones. We suspect that the magnitude of the ET variation within a serotype is even greater than that demonstrated here.

From both epidemiological and evolutionary perspectives, it is important to understand the mechanisms responsible for antigenic diversity among isolates of the same ET class (antigenic divergence) and enzyme diversity among isolates sharing the same multiple antigenic determinants (antigenic convergence). The intensity of natural selection for cell surface antigens must surely be substantially greater than that for the electromorph variants of enzymes, the majority of which probably are physiologically equivalent or nearly so. The surface antigens of *E. coli* are unlikely to be cryptic to mammalian hosts; and, moreover, as components of membranes, capsules, and flagella, these surface antigens may affect the capacity of the bacteria to invade and maintain populations in hosts. In contrast, the enzymes assayed are primarily intracellular components for which minor

TABLE 6. Serotypic variation within ET classes

ET class	Serotype	No. of isolates	ET class	Serotype	No. of isolates
ET-1/ET-32	O2:K1:H4	10	ET-15/ET-76/ET-90	O7:K1:H−	11
	OR:K1:H4	2		O7:KN:H−	1
	O2:KN:H−	1		OR:K1:H−	1
	O2:KN	1		OR:K1	2
	O2:K1:H7	1		O7:K12:H−	1
	O14:K1	1		O2:K1:H4	1
	O18:K5:H−	1		O23:KN:H6	1
ET-4/ET-68	O2:K2:H1	2	ET-19/ET-34/ET-44/ET-102/ET-105/ET-113	O1:K1:H−	9
	O2:K2:HN	2		OR:K1:H−	1
	O2:KN:H1	1		OR:K1	1
	O6:K2:H1	3		OR:K23:H7	1
	O6:K5:H1	5			
	O6:K13:H1	2	ET-7/ET-109	O83:K1:H33	1
	O6:K13:HN	4		O83:K1:H−	1
	O6:K53:H1	1		O83:K1:HN	1
	O6:K−:H−	1		OR:K1:H−	1
	O25:K5:H1	4		OR:K1	1
	O25:K7:H1	4			
	O25:K53:H1	1	ET-25/ET-57/ET-115	O2:K5:H4	2
	O25:K5:H−	1		O2:K5:H−	4
	O25:KN:H1	1		OR:KN	1
	OR:K5:H1	1			
	OR:KN	2	ET-55/ET-92/ET-99	O18:K1:H7	4
				O18:K1:HN	1
ET-6/ET-60/ET-66	O2:K1:H6	2	ET-64/ET-67	O75:K5:H−	4
	O4:K3:H5	1		O75:KN:H5	4
	O4:K3:H−	1		OR:K95:H−	1
	O4:KN:H5	1		OR:K3	1
	O4:K−:H5	1			
	O4:K12:H5	4	ET-69	O18:K5:H−	3
	O4:K12:HN	2		OR:KN	2
	O4:KN:HN	1			
	O6:K15:HN	1	ET-89	O16:K1:H6	3
	O9:K+:H10	1		O16:K1:H−	4
	O83:K24:H31	1		O16:K1:HN	2
	O83:KN:H31	2			

structural variations are unlikely to have significant effects on the ecological performance of the bacteria (14). This hypothesis is supported by the results of experimental studies of selection on allozyme variants of several enzymes in chemostats (10, 11). When bacteria of the same or closely related clones encounter different habitats (e.g., host species or tissues), selection could lead to divergence in antigenic structure, with little or no change in genes that do not contribute to the determination of antigenic type. Similarly, genetically unrelated clones in the same habitat could evolve similar or identical adaptative antigenic structures.

In attempting to account for the apparent identity of antigenic determinants among different ET classes, it should be remembered that different antigenic structures can be recognized by the same antiserum. For example, Kusecek et al. (16) showed that the antigens of K1 isolates typed as O1 represent two different lipopolysaccharides, each of which is associated with a different outer-membrane protein pattern. If different molecular configurations frequently are immunologically indistinguishable, antigenic convergence would be more readily achieved than if all molecular configurations were serologically unique.

Although the results of this and earlier investigations demonstrate that there is much genetic variation within serotypes and ETs, these results are not inconsistent with the hypothesis of clonal population structure. Even with periodic selection and random extinction purging variants (2, 15, 17, 20), a modest amount of genetic variation would be expected among independent isolates of the same lineage. But when differences between isolates of the same serotype involve many genes, it seems reasonable to assume that separate lineages (clones) are included within the same serotype.

As a method for the identification of isolates of *E. coli*, multilocus electrophoresis has a number of advantages over serotyping. An ET can be assigned to every isolate analyzed. In comparison, a complete O:K:H serotype could be determined for only one-third of the isolates tested in the present study, and the proportion is reported to be even less for isolates from animal sources (4). Furthermore, electrophoretic data provide estimates of the degree of genetic relatedness or distance among isolates, whereas serological data indicate only the existence of differences.

While we see these advantages as good reason for using multilocus electrophoresis in epidemiological and systematic studies of *E. coli* and other bacteria (18, 32; J. M. Musser, D. M. Granoff, P. E. Pattison, and R. K. Selander, Proc. Natl. Acad. Sci. U.S.A., in press), we do not consider them arguments for abandoning serological procedures. Because antigenic determinants are highly variable characters, they are useful epidemiological markers, even though serotypic identity of strains cannot be taken as evidence of clonal

identity or even, in many cases, similarity. Extensive collections of serotyped isolates of known origin are available, and, most importantly, established relationships exist between serotype and other biological and clinical properties of strains of *E. coli*. Because of the likelihood that serotypes are adaptive characters, they may be more diagnostic of the clinical properties of strains than is the ET. In any event, the demonstration of extensive genetic variation within serotypes argues for a pluralistic approach to the identification of *E. coli* isolates that includes serotype, ET, and outer-membrane protein pattern.

ACKNOWLEDGMENTS

Neal Bogdanovich wrote the computer program used in this study.

This work was supported by grant 215 from the Swedish Medical Council; by grants from the Ellen, Walter and Lennart Hesselmann Foundation for Scientific Research, the Medical Faculty of the University of Göteborg (D.A.C. and C.S.E.), and the Volkswagen Foundation (I.Ø. and F.Ø.); and by U.S. Public Health Service grants GM19848, GM33782 (B.R.L.), and GM22126 (R.K.S.) from the National Institutes of Health.

LITERATURE CITED

1. **Achtman, M., A. Mercer, B. Kusecek, A. Pohl, M. Heuzenroeder, W. Aaronson, A. Sutton, and R. P. Silver.** 1983. Six widespread bacterial clones among *Escherichia coli* K1 isolates. Infect. Immun. **39**:315–335.
2. **Atwood, K. C., L. K. Schneider, and F. J. Ryan.** 1951. Periodic selection in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **37**:146–155.
3. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.
4. **Bettelheim, K. A.** 1978. The sources of 'OH' serotypes of *Escherichia coli*. J. Hyg. **80**:83–113.
5. **Brenner, D. J., G. R. Fanning, G. V. Miklos, and A. G. Steigerwalt.** 1973. Polynucleotide sequence relatedness among *Shigella* species. Int. J. Syst. Bacteriol. **23**:1–7.
6. **Brenner, D. J., G. R. Fanning, A. G. Steigerwalt, I. Ørskov, and F. Ørskov.** 1972. Polynucleotide sequence relatedness among three groups of pathogenic *Escherichia coli* strains. Infect. Immun. **6**:308–315.
7. **Caugant, D. A., B. R. Levin, G. Lidin-Janson, T. S. Whittam, C. Svanborg Eden, and R. K. Selander.** 1983. Genetic diversity and relationships among strains of *Escherichia coli* in the intestine and those causing urinary tract infections. Prog. Allergy **33**:203–227.
8. **Caugant, D. A., B. R. Levin, and R. K. Selander.** 1981. Genetic diversity and temporal variation in the *E. coli* population of a human host. Genetics **98**:467–490.
9. **Caugant, D. A., B. R. Levin, and R. K. Selander.** 1984. Distribution of multilocus genotypes of *Escherichia coli* within and between host families. J. Hyg. **92**:377–384.
10. **Dykhuizen, D. E., and D. L. Hartl.** 1980. Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. Genetics **96**:801–817.
11. **Dykhuizen, D. E., and D. L. Hartl.** 1983. Functional effects of PGI allozymes in *Escherichia coli*. Genetics **105**:1–18.
12. **Kauffmann, F.** 1943. Uber neue thermolabile Korperantigene der Coli-Bakterien. Acta Pathol. Microbiol. Scand. **20**:21–44.
13. **Kauffmann, F.** 1944. Zur Serologie der Coli-Gruppe. Acta Pathol. Microbiol. Scand. **21**:20–45.
14. **Kimura, M.** 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.
15. **Koch, A. L.** 1974. The pertinence of the periodic selection phenomenon to prokaryote evolution. Genetics **77**:127–142.
16. **Kusecek, B., H. Wloch, A. Mercer, V. Vaisänen, G. Pluschke, T. Korhonen, and M. Achtman.** 1984. Lipopolysaccharide, capsule, and fimbriae as virulence factors among O1, O7, O16, O18, or O75 and K1, K5, or K100 *Escherichia coli*. Infect. Immun. **43**:368–379.
17. **Levin, B. R.** 1981. Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. Genetics **99**:1–23.
18. **Levin, M. H., R. A. Weinstein, C. Nathan, R. K. Selander, H. Ochman, and S. A. Kabins.** 1984. Association of infection caused by *Pseudomonas aeruginosa* serotype O11 with intravenous abuse of pentazocine mixed with tripelennamine. J. Clin. Microbiol. **20**:758–762.
19. **Lidin-Janson, G., L. A. Hanson, B. Kaijser, K. Lincoln, U. Lindberg, S. Olling, and H. Wedel.** 1977. Comparison of *Escherichia coli* from bacteriuric patients with those from feces of healthy schoolchildren. J. Infect. Dis. **136**:346–352.
20. **Maruyama, T., and M. Kimura.** 1980. Genetic variablity and effective population size when local extinction and recolonization of subpopulations are frequent. Proc. Natl. Acad. Sci. U.S.A. **77**:6710–6714.
21. **Milkman, R.** 1973. Electrophoretic variation in *Escherichia coli* from natural sources. Science **182**:1024–1026.
22. **Ochman, H., and R. K. Selander.** 1984. Evidence for clonal population structure in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **81**:198–201.
23. **Ochman, H., T. S. Whittam, D. A. Caugant, and R. K. Selander.** 1983. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J. Gen. Microbiol. **129**:2715–2726.
24. **Ochman, H., R. A. Wilson, T. S. Whittam, and R. K. Selander.** 1983. Genetic diversity within serotypes of *Escherichia coli*, p. 202–224. *In* Proceedings of the Fourth International Symposium on Neonatal Diarrhea. VIDO Publications, University of Saskatchewan, Saskatoon, Canada.
25. **Ørskov, F.** 1976. Agarose electrophoresis combined with second dimensional cetavlon precipitation. A new method for demonstration of acidic polysaccharide K antigens. Acta Pathol. Microbiol. Scand Sect. B **84**:319–320.
26. **Ørskov, F., and I. Ørskov.** 1984. Serotyping of *Escherichia coli*, p. 43–112. *In* T. Bergan (ed.), Methods in microbiology, vol. 14. Academic Press, Ltd., London.
27. **Ørskov, F., I. Ørskov, D. J. Evans, Jr., R. B. Sack, D. A. Sack, and T. Wadström.** 1976. Special *Escherichia coli* serotypes among enterotoxigenic strains from diarrhoea in adults and children. Med. Microbiol. Immunol. **162**:73–80.
28. **Ørskov, I., and F. Ørskov.** 1977. Special O:K:H serotypes among enterotoxigenic *E. coli* strains from diarrhea in adults and children. Occurrence of the CF (colonization factor) antigen and of hemagglutinating abilities. Med. Microbiol. Immunol. **163**:99–110.
29. **Ørskov, I., F. Ørskov, B. Jann, and K. Jann.** 1977. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. Bacteriol. Rev. **41**:667–710.
30. **Ørskov, I., F. Ørskov, and B. Rowe.** 1984. Six new *E. coli* O groups: O165, O166, O167, O168, O169, and O170. Acta Pathol. Microbiol. Scand. Sect. B **92**:189–194.
31. **Robbins, J. B., G. H. McCracken, E. C. Gotschlich, F. Ørskov, I. Ørskov, and L. A. Hanson.** 1974. *Escherichia coli* K1 capsular polysaccharide associated with neonatal meningitis. N. Engl. J. Med. **290**:1216–1220.
32. **Schill, W. B., S. R. Phelps, and S. W. Pyle.** 1984. Multilocus electrophoretic assessment of the genetic structure and diversity of *Yersinia ruckeri*. Appl. Environ. Microbiol. **48**:975–979.
33. **Selander, R. K., and B. R. Levin.** 1980. Genetic diversity and structure in *Escherichia coli* populations. Science **210**:545–547.
34. **Selander, R. K., M. H. Smith, S. Y. Yang, W. E. Johnson, and J. B. Gentry.** 1971. Biochemical polymorphism and systematics in the genus *Peromyscus*. I. Variation in the old-field mouse (*Peromyscus polionotus*), p. 49–90. *In* Studies in genetics VI. University of Texas publication no. 7103. University of Texas Press, Austin.
35. **Whittam, T. S., H. Ochman, and R. K. Selander.** 1983. Multilocus genetic structure in natural populations of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **80**:1751–1755.
36. **Whittam, T. S., H. Ochman, and R. K. Selander.** 1983. Geographic components of linkage disequilibrium in natural populations of *Escherichia coli*. Mol. Biol. Evol. **1**:67–83.