INFECTION AND IMMUNITY, May. 1993, p. 1619–1629
0019-9567/93/051619-11$02.00/0

Vol. 61, No. 5

Clonal Relationships among *Escherichia coli* Strains That Cause Hemorrhagic Colitis and Infantile Diarrhea

THOMAS S. WHITTAM,[1]* MARIE L. WOLFE,[1] I. KAYE WACHSMUTH,[2] FRITS ØRSKOV,[3] IDA ØRSKOV,[3] AND RICHARD A. WILSON[1,4]

Institute of Molecular Evolutionary Genetics, Department of Biology,[1] and Department of Veterinary Sciences,[4] Pennsylvania State University, University Park, Pennsylvania 16802; Division of Bacterial Diseases, Center for Infectious Disease, Centers for Disease Control, Atlanta, Georgia 30333[2]; and International Escherichia and Klebsiella Centre, Statens Seruminstitut, DK-3000 Copenhagen S, Denmark[3]

Received 18 November 1992/Accepted 2 February 1993

The genetic relationships among 1,300 isolates of *Escherichia coli* representing 16 serotypes associated with enteric disease, including O157:H7 strains recovered from patients with hemorrhagic colitis and hemolytic uremic syndrome and O26:H11, O55:H6, O55:H7, O111:H2, and O128:H2 strains, many of which were isolated originally from infants with diarrhea, were estimated from allelic variation among 20 enzyme-encoding genes detected by multilocus enzyme electrophoresis. Multiple electrophoretic types were observed among isolates of each serotype, with isolates of the same O serogroup differing on average at 28% of the enzyme loci. Comparisons of the multilocus enzyme profiles revealed that 72% of the isolates belong to 15 major electrophoretic types, each of which corresponds to a bacterial clone with a wide geographic distribution. Genetically, the O157:H7 clone is most closely related to a clone of O55:H7 strains that has long been associated with worldwide outbreaks of infantile diarrhea. We propose that the new pathogen emerged when an O55:H7-like progenitor, already possessing a mechanism for adherence to intestinal cells, acquired secondary virulence factors (Shiga-like cytotoxins and plasmid-encoded adhesins) via horizontal transfer and recombination.

Enterohemorrhagic *Escherichia coli* (EHEC), a newly recognized class of enteric pathogen (10, 23, 48), was first linked to a clinically distinct syndrome of diarrheal disease called hemorrhagic colitis (HC) following two outbreaks of HC in North America in 1982 (36). The EHEC strains recovered from these outbreaks were serotype O157:H7 and did not possess the virulence determinants typical of other *E. coli* that cause infectious enteric disease (36): they failed to produce the classical toxins characteristic of enterotoxigenic *E. coli* (ETEC), lacked the invasive abilities of enteroinvasive *E. coli*, and were serotypically distinct from strains that have long been associated with worldwide outbreaks of infantile diarrhea, the enteropathogenic *E. coli* (EPEC). Recent studies have shown that O157:H7 EHEC strains express potent Shiga-like cytotoxins (30, 32), carry plasmids that encode adhesins mediating bacterial adherence to intestinal cells (19, 25), and have a chromosomal gene (*eae* locus) that is essential for the production of intimin (11) and the attaching and effacing lesions characteristic of some EPEC strains (11, 17, 58). Illnesses caused by *E. coli* O157:H7, including HC and hemolytic uremic syndrome, as well as diseases caused by other Shiga-like cytotoxin-producing strains (also referred to as verocytotoxigenic *E. coli*), have emerged as a major health problem in North America and Europe (5, 10, 20, 27, 44).

To determine the genetic relationships of O157:H7 strains to other pathogenic forms of *E. coli*, we used multilocus enzyme electrophoresis (39) to study the genetic diversity and clonal relationships among O157:H7 isolates and strains of other serotypes implicated in diarrheal disease. By detecting allelic variation at polymorphic enzyme loci, the multilocus approach provides a sensitive system of genetic markers for characterizing the chromosomal genotypes of strains and for estimating the overall genomic relatedness of isolates (39). Previously, we found that O157:H7 isolates from recent epidemics of HC in North America were not closely allied to Shiga-like cytotoxin-producing strains of other *E. coli* serotypes (54), many of which produce a clinically similar form of bloody diarrhea (49, 50). We also demonstrated that O157:H7 strains are only distantly related to other strains of the O157 group associated with enteric infections in animals (55). These findings suggested that if the O157:H7 strains have evolved recently from an extant *E. coli* lineage, strains closely related to O157:H7 strains would be found in an O serogroup other than O157.

Here we assess the genetic relatedness among *E. coli* isolates obtained from patients and animals in diverse localities on five continents, with special emphasis on strains of 16 specific O:H serotypes that have been associated with infectious diarrheal disease. We focused our study on strains of the following serotypes, which generally have one or more properties in common with O157:H7 strains, as summarized in Table 1: EHEC O26:H11 and O111:H8, strains which may produce Shiga-like cytotoxin and cause HC; EPEC O55:H6, O111:H2, O111:H12, O111:H21, O128:H2, and O128:H12, strains which may attach to intestinal epithelial cells by a mechanism similar to that of O157:H7 strains; and EPEC O55:H7 and ETEC O128:H7, bacteria which express H7 flagellar antigen. In addition, we examined isolates of several other H types and nonmotile strains of the O26, O55, O111, O128, and O157 serogroups.

MATERIALS AND METHODS

Bacterial strains. A total of 1,300 *E. coli* strains of five O serogroups, the bulk of which represent 16 O:H serotypes associated with diarrhea and other intestinal diseases in

* Corresponding author.

TABLE 1. Distinctive attributes[a] of *E. coli* O157:H7 strains and strains of other serotypes recovered from cases of diarrheal disease

Serotype[b]	Disease class	H7 expressed	Sorbitol negative[c]	SLT[d]	A/E[e]	Associated disease[f]	
						HC	HUS
O157:H7	EHEC	+	+	+	+	+	+
O26:H11	EHEC	−		+	+	+	+
O26:NM	EHEC	−		+	+	+	+
O55:H6	EPEC	−	+	−	+		
O55:H7	EPEC	+		+	+		
O55:NM	EPEC	−		+	+		+
O111:H2	EPEC	−	+	−	+		
O111:H8	EHEC	−		+	+		
O111:H12	EPEC	−		−	−		
O111:H21	EPEC	−		−	−		
O111:NM	EPEC	−		+	−	+	
O128:H2	EPEC	−	+	+	+		+
O128:H7	ETEC	+		−	−		
O128:H12	EPEC	−		+	−		
O128:NM	EPEC	−		+	+	+	

[a] A + indicates that an attribute has been reported for some strains of a serotype; it does not mean that all strains of a serotype have the attribute.
[b] Classification of serotypes is from reference 23.
[c] Fermentation of sorbitol requires more than 24 h. Data are from references 33 and 46.
[d] SLT, expression of Shiga-like toxin (or verocytotoxin). Data are from reference 20.
[e] A/E, attaching and effacing lesions. Data are from references 22 and 33.
[f] HUS, hemolytic uremic syndrome; data for both HC and HUS are from references 20 and 44.

humans, were characterized for allelic variation at 20 enzyme-encoding loci by multilocus enzyme electrophoresis (39). Most of the isolates were obtained from the collections of the Centers for Disease Control (CDC; Atlanta, Ga.), the *E. coli* Reference Center (University Park, Pa.), and the Statens Seruminstitut (Copenhagen, Denmark). The samples were originally collected in the period from 1934 to 1991 from North America (Canada, Guatemala, Greenland, Jamaica, Mexico, Panama, and the United States), South America (Argentina, Brazil, Colombia, Suriname, and Peru), Africa (Egypt, Kenya, South Africa, and Zaire), Europe (Belgium, Bulgaria, Denmark, England, France, Germany, Sweden, Switzerland, and Wales), Asia (Afghanistan, Iran, Israel, Japan, Malaysia, Sri Lanka, Thailand, and Vietnam), and Australia. Additional isolates were kindly supplied by L. Beutin, A. Cravioto, H. Lior, W. Maas, J. Mathewson, A. O'Brien, and P. Tarr.

The sample set contains cultures studied previously by multilocus enzyme electrophoresis, including 76 O157:H7 isolates originally obtained from patients with HC and hemolytic uremic syndrome in North America (7, 54), 194 O157 strains of a variety of H types originally isolated from domesticated animals with colibacillosis (55), and 24 EPEC strains of serotypes O55:H6, O111:H2, and O128:H2, originally recovered from cases of infantile diarrhea (34).

For comparative purposes, we also included five O-antigen type strains that are used for the preparation of standard O antisera (34), as follows: H311b (O26:H—), Su3912-41 (O55:H—), Stoke W (O111:H—), Cigleris (O128:H2), and A2 (O157:H19).

Enzyme electrophoresis. To analyze enzyme electrophoretic variation, extracts of water-soluble proteins from disrupted cells were prepared and subjected to horizontal starch gel electrophoresis and selective enzyme staining as described previously (39). For each enzyme, electromorphs (allozymes) were distinguished by differences in the migration rate of specifically stained protein bands. The 20 enzymes analyzed individually for mobility variants are listed in Table 3. Electromorphs for each enzyme were determined by comparison with standard mobility variants and assigned numbers by their rate of anodal migration. Isolates that lacked detectable enzyme activity were assigned a null allelic state at the locus in question.

Electromorphs of an enzyme were equated with alleles at the corresponding structural gene loci, so that each bacterial strain was characterized by its multilocus genotype (allele combination) for the enzyme-encoding loci assayed (39). Distinctive multilocus genotypes were designated electrophoretic types (ETs), which were numbered by their inferred relationships from a cluster analysis (see below).

Shiga-like toxin gene probes. Gene probes specific for *slt-I* and *slt-II* were prepared by amplifying DNA by the polymerase chain reaction (PCR). For *slt-I*, template DNA for the PCR was prepared from *E. coli* 3143-85B (54) and amplified with primers 5′-TCGTATGGTGCTCAAGGA GTA-3′ and 5′-TCTGAGTCAACGAAAAATAA-3′. For *slt-II*, DNA was prepared from strain 871717 and amplified with primers 5′-TCAGCCAAAAGGAACACCTG-3′ and 5′-CTGCACTTCAGCAAATCCGG-3′. The PCR products were spin filtered and labeled with [α-^{32}P]dCTP by nick translation prior to hybridization, according to the procedure of the manufacturer (GIBCO Laboratories, Grand Island, N.Y.).

DNA hybridizations. Total chromosomal DNA from 50 ml of overnight cultures was isolated by the protocol of Silhavy et al. (43). For each isolate, 10 μg of DNA was immobilized on a nylon membrane (Zeta-Probe; Bio-Rad Laboratories, Richmond, Calif.) with a vacuum apparatus by the alkaline blotting protocol of the manufacturer (Bio-Dot; Bio-Rad), cross-linked on a transilluminator for 1 min, and baked at 80°C for 30 min.

Hybridizations were performed by the standard protocol for Zeta-Probe membranes (Bio-Rad) with bottles (300 by 35 mm) in a mini-hybridization oven (Hybaid; National Labnet Co., Woodbridge, N.J.). Membranes were prehybridized at 62°C for 30 min in 15 ml of heated prehybridization solution. The prehybridization solution was removed, replaced with 15 ml of heated hybridization solution containing 0.75 μg of labeled probe DNA per ml, and incubated at 62°C with rotation for 20 to 24 h. Membranes were washed twice with 80 to 100 ml of 5% sodium dodecyl sulfate (SDS) and twice with 1% SDS. Washes were performed in rotating bottles at 62°C for 30 min. Membranes were sealed into plastic bags and exposed to X-ray film overnight at −70°C.

Autoradiographs were scored for the amount of exposure for DNA from the negative (strain K-12) and positive (*slt-I*, 3143-85B; *slt-II*, 871717) controls that were included on every membrane. Each bacterial isolate was tested on at least two filters for each gene probe.

Phylogenetic analysis. The phylogenetic relationships among ETs were inferred by distance methods (12) with the average-linkage (AL) method (45). For the analysis, a matrix of genetic distances between all pairs of ETs was computed from comparisons of electrophoretic profiles. Distance (d) was calculated as $d_{ij} = -\ln(1 - D)$, where D equals the proportion of loci with different alleles between the ith and jth ETs (15, 29). The principle underlying this formula is that any allelic difference in electrophoretic mobility results from at least one codon difference at the nucleotide level, and under the assumption that codon changes occur indepen-

TABLE 2. Genetic diversity among *E. coli* isolates of serotypes commonly associated with enteric disease

Serotype	No. of isolates	No. of ETs	%P[a]	Avg no. of alleles	Avg genetic diversity[b] (SE)	
					H_I	H_{ET}
O26:H11	47	6	75	2.0	0.069 (0.014)	0.373 (0.058)
O26:NM	13	5	50	1.8	0.194 (0.052)	0.300 (0.075)
O26 group	93	20	85	3.1	0.208 (0.038)	0.382 (0.059)
O55:H6	60	12	75	3.0	0.183 (0.035)	0.465 (0.075)
O55:H7	81	14	90	3.0	0.123 (0.021)	0.434 (0.061)
O55:NM	69	16	80	3.3	0.269 (0.046)	0.437 (0.071)
O55 group	298	57	90	5.0	0.376 (0.062)	0.451 (0.068)
O111:H2	31	3	15	1.2	0.013 (0.007)	0.100 (0.055)
O111:H8	13	2	40	1.4	0.113 (0.032)	0.400 (0.112)
O111:H12	17	4	55	1.8	0.112 (0.026)	0.383 (0.084)
O111:H21	16	5	70	2.2	0.172 (0.034)	0.420 (0.073)
O111:NM	58	10	60	2.4	0.197 (0.056)	0.321 (0.071)
O111 group	203	36	90	4.3	0.281 (0.058)	0.407 (0.065)
O128:H2	40	11	85	2.7	0.157 (0.036)	0.331 (0.055)
O128:H7	19	2	5	1.1	0.005 (0.005)	0.050 (0.050)
O128:H12	10	5	45	1.6	0.222 (0.060)	0.230 (0.062)
O128:NM	5	3	60	1.7	0.360 (0.070)	0.433 (0.084)
O128 group	103	35	95	3.8	0.275 (0.054)	0.340 (0.059)
O157:H7	369	6	30	1.4	0.008 (0.006)	0.123 (0.048)
O157:NM	46	11	90	3.3	0.371 (0.056)	0.475 (0.064)
O157 group	604	50	100	5.3	0.275 (0.049)	0.445 (0.063)
Combined	1,300	191	100	8.5	0.399 (0.063)	0.445 (0.064)

[a] %P, percent polymorphic loci, based on 20 enzyme loci assayed.
[b] Average genetic diversity calculated from the frequency of alleles in isolates (H_I) and from the frequency of alleles in ETs (H_{ET}).

dently, *d* gives an estimate of the mean number of electrophoretically detectable codon differences per genetic locus (29). Using the AL algorithm (45), we constructed dendrograms which summarize the genetic relationships among ETs and are a source of hypotheses about the history of phylogenetic divergence of strains.

To assess the size of the error involved in the inferred genetic relationships among ETs, a bootstrap method was used to create a series of random samples from the original data (13). A computer program (ETBOOT) was designed that samples enzyme loci with replacement from the original data to generate new genotypic profiles for the original ETs, calculates a genetic distance matrix, and constructs a tree (dendrogram). This process was repeated for 1,000 phylogenetic trees, and the frequency of occurrence of different clusters in these trees was tabulated. For particular clusters of interest, the percentage of bootstrapped trees with the same nodes of the observed tree were recorded.

The bootstrap approach of repeated sampling from the original data provided a method for obtaining confidence limits on clusters in dendrograms. In addition to the AL algorithm, we also used the neighbor-joining (NJ) method (37) for comparison in the bootstrap analysis. The NJ method has been shown in computer simulations to work better than the AL method in recovering a phylogeny when the rate of divergence varies across lineages (18).

RESULTS

Genetic diversity within serotypes. The analysis of protein polymorphisms among 1,300 *E. coli* isolates of five serogroups revealed extensive genetic variation in the structural genes encoding soluble enzymes; all 20 enzymes had electrophoretically detectable mobility variants, with an average of 8.5 alleles per enzyme locus (Table 2). Within each of the 16 O:H serotypes associated with diarrheal disease, there were multiple ETs, with the number of ETs ranging from 2 (O111:H8 and O128:H7 strains) to 16 (O55:NM strains) (Table 2). For the 15 serotypes other than O157:H7, there is a significant linear relationship between the number of ETs and the number of isolates (Fig. 1). Essentially, this correlation means that, for isolates of the same serotype, a new ET is found for each six strains examined—at least for sample sizes of fewer than 100 isolates. Among the 369 O157:H7 strains, however, only six ETs were resolved, which reflects the low level of genetic variation among strains of this serotype. In contrast to the effect of sample size on the number of ETs of a serotype, the average genetic diversity per locus (H) is not correlated with sample size, regardless of whether it is calculated for isolates (H_I) or ETs (H_{ET}).

O26 strains. Among the 93 isolates of the O26 serogroup, there were 20 ETs which differed, on average, at 38% (H_{ET} = 0.382) of the enzyme loci (Table 2). The 47 isolates of EHEC serotype O26:H11 represented a total of six ETs, although most isolates belong to one of two closely related ETs. The genetic diversity among isolates of this serotype (H_I = 0.069) accounts for 33% of the genetic diversity among isolates of the O26 group as a whole (Table 2).

Genetic relationships among the 20 ETs of the O26 serogroup, inferred from cluster analysis by the AL method (Fig. 2A) and the NJ algorithm (not shown), revealed two major groups of strains: the top cluster includes ETs 1 to 8 and was observed in 88% of the bootstrapped AL trees (92% of the NJ trees), and the bottom cluster of ETs 9 to 17 was

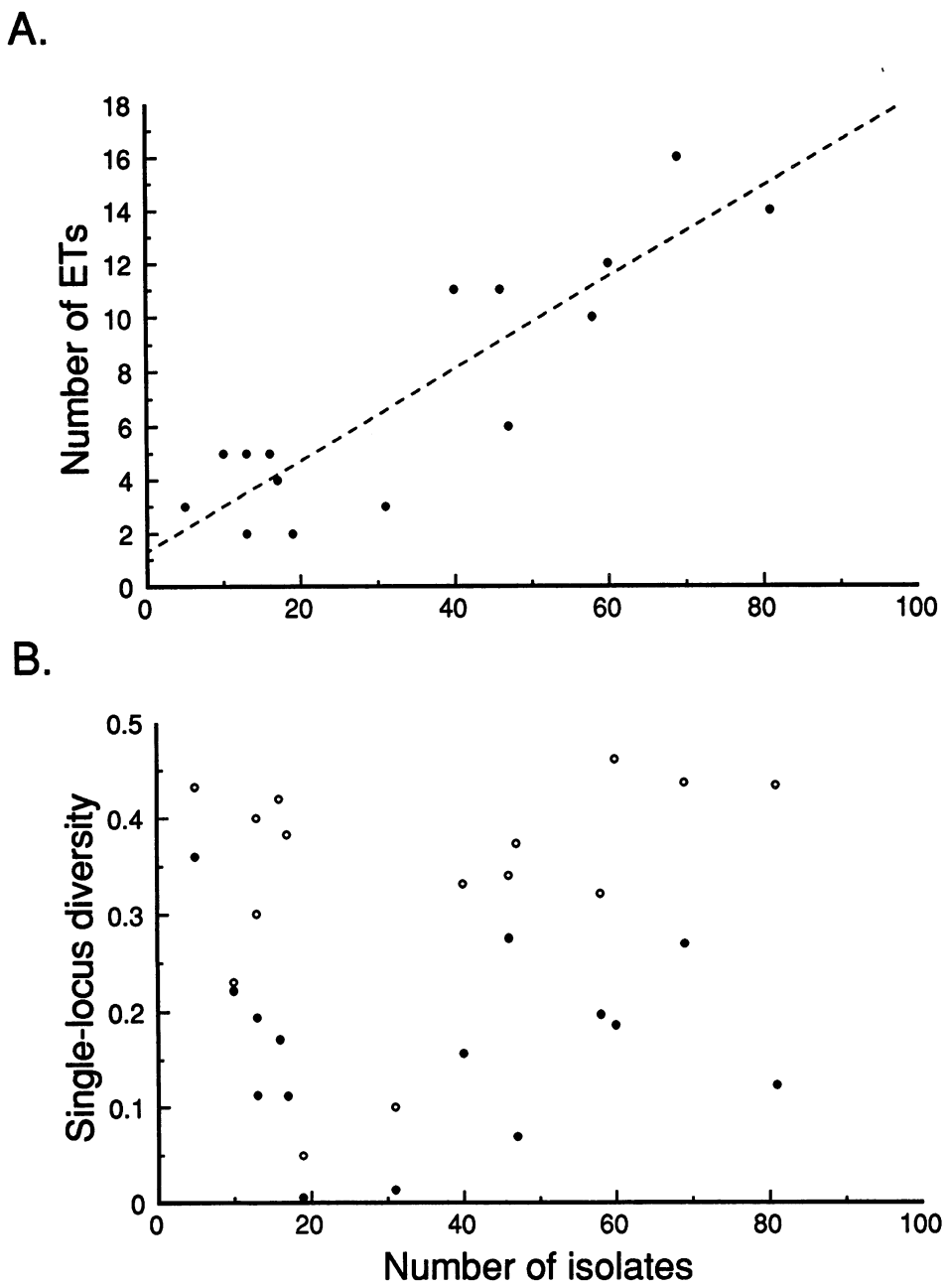

FIG. 1. Level of genetic variation as a function of sample size for 15 serotypes of *E. coli*. (A) Number of multilocus enzyme genotypes (ETs) as a function of the number of isolates. The dashed line is the linear regression equation $y = 1.326 + 0.170x$ ($R^2 = 0.79$, $P < 0.01$). (B) Single-locus diversity calculated across isolates (●) and ETs (○). Because of the disproportionately large sample size, values for the 369 O157:H7 strains are not given.

recovered in 76% of the AL trees (36% of the NJ trees). The major clusters differ in the frequency of two flagellar antigen types; most isolates in the top cluster have H32 flagellar antigen, whereas a majority of the strains in the bottom cluster have H11 antigen. The association between cluster and H type is not complete, because an O26:H11 strain (ET 7) also occurs in the top cluster (Fig. 2A). The two ETs most frequently isolated, ET 13 and ET 15, are closely related and together account for 61% of the isolates of the O26 serogroup.

O55 strains. We examined nearly equal numbers of the three EPEC serotypes among 298 O55 strains: 20% of the isolates were O55:H6, 27% were O55:H7, and 23% were nonmotile (Table 2). Among the O55:H6 and O55:H7 strains, we identified 12 and 14 ETs, respectively, with the diversity among isolates within serotypes accounting for 49% (O55:H6) and 33% (O55:H7) of the total diversity among isolates of the O55 serogroup ($H_I = 0.376$). The nonmotile O55 strains were more variable than either the H6 or H7 strains, with a diversity among isolates that is 72% of that for the whole serogroup.

Comparison of the allele profiles of O55 strains disclosed a

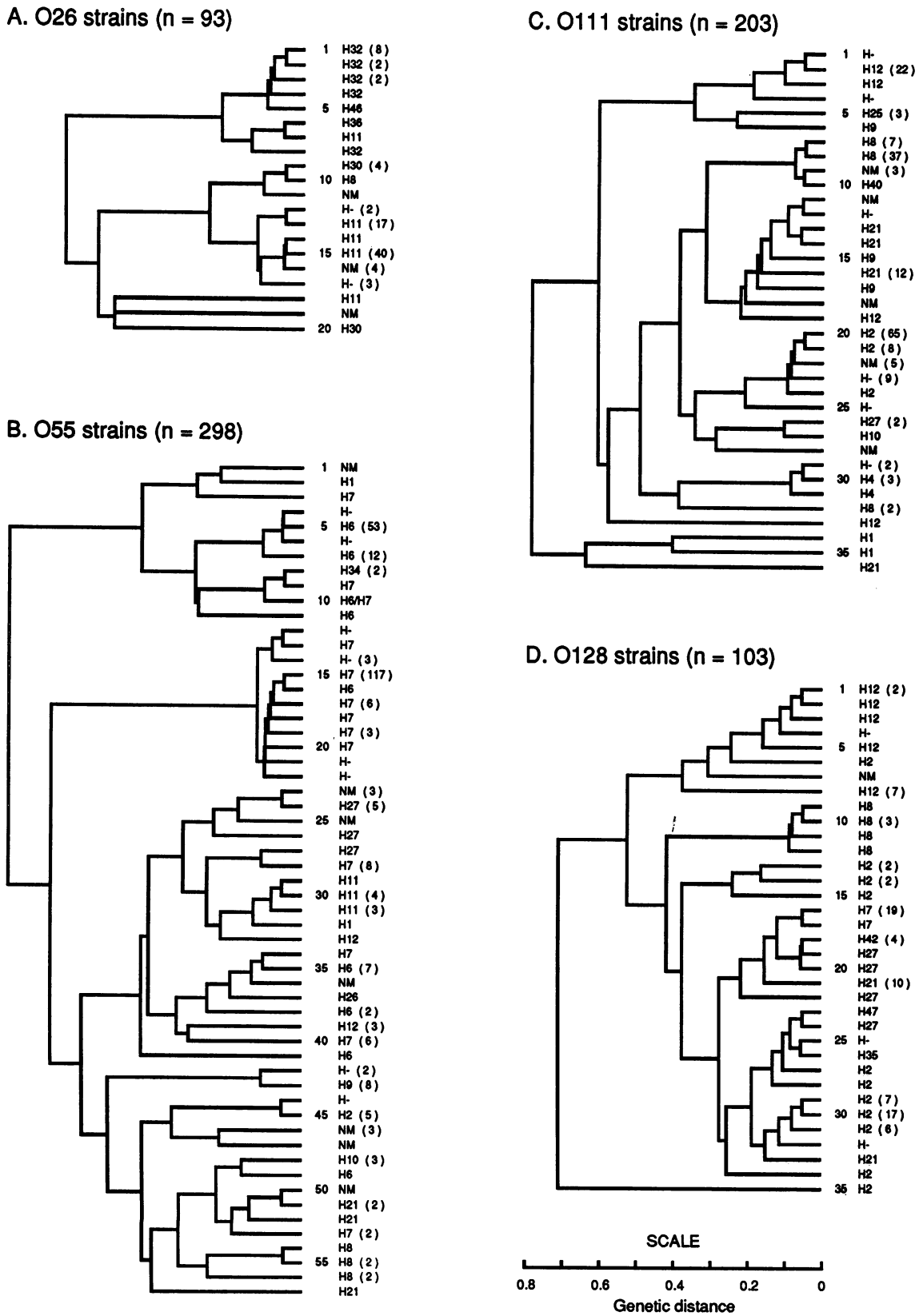

FIG. 2. Genetic relationships of multilocus enzyme genotypes (ETs) of *E. coli* strains of four serogroups, O26 (A), O55 (B), O111 (C), and O128 (D). Dendrograms were produced by the AL method applied to the matrix of standard genetic distances between multilocus genotypes. The predominant flagellar antigen type (H type) of strains is marked at the right. The number in parentheses is the number of isolates of each ET that was isolated more than once.

total of 57 ETs which differed, on average, at 45% (H_{ET} = 0.451) of the enzyme loci (Table 2). Twenty-five of the ETs (43%) were represented by more than one isolate. The two most frequently isolated ETs fell into two distinct clusters that were observed in a majority of the bootstrapped trees (Fig. 2B). The top cluster in Fig. 2B comprises ETs 1 to 11 and was found in 71% of the AL trees (86% of the NJ trees). This cluster includes the second most common multilocus genotype (ET 5) in the O55 serogroup, which is represented by 53 isolates, mostly of serotype O55:H6. The second cluster of 11 ETs (ETs 12 to 22), which was observed in 90% of both the AL and NJ trees, contains the most frequently isolated enzyme genotype (ET 15), which was represented by 117 isolates, most of which were either O55:H7 or O55:NM. Together, the 22 ETs of these two clusters account for 71% of the isolates of the O55 serogroup. Although H6 and H7 are the predominant serotypes in the two clusters, it is noteworthy that isolates with these H antigens are found throughout the dendrogram, reflecting the substantial genetic diversity (H_{ET}) among ETs of these serotypes (Table 2) and indicating that these antigens can be expressed by strains with very different chromosomal backgrounds.

O111 strains. More than half of the 203 isolates of the O111 serogroup examined belonged to EPEC serotype H2, H12, H21, or NM (Table 2). The average H_I for the O111 strains with H2, H8, H12, or H21 antigen was 0.103, which is about one-third of the total diversity among all isolates of the serogroup. In contrast, the average H_{ET} for the four serotypes was 80% of the diversity for the O111 serogroup. The difference between the estimates of diversity reflects the fact that most of the isolates of a serotype are of a single ET. The nonmotile O111 strains were more diverse genetically (H_I = 0.198) than isolates of other O111 H types. Although there were multiple ETs among isolates of each serotype, there was, in general, a relatively low level of genetic variation among isolates with specific H types.

The most common enzyme genotype (ET 20) in the O111 serogroup is represented by 65 isolates and occurs in a cluster of four ETs (Fig. 2C). This cluster (ETs 20 to 23), which was recovered in 52% of the AL trees (42% of the NJ trees), includes 43% of the isolates of the O111 serogroup, most of which are EPEC serotype O111:H2. The second most common enzyme genotype, ET 8, belongs to a tight cluster (ETs 7 to 10) that was observed in 79% of the AL trees (85% of the NJ trees). A majority of the strains of this cluster are EHEC serotype O111:H8. Of the remaining EPEC serotypes, O111:H12 and O111:H21 strains largely fall into ET 2 and ET 16, respectively, although isolates of these serotypes also occur in other branches of the dendrogram (Fig. 2C). For example, an isolate with H21 antigen (ET 36) is highly divergent from other H21 strains and differs by 10 alleles from its closest relative at the bottom of the dendrogram, again indicating that the same O:H serotype can be associated with distantly related strains.

O128 strains. There were 11 ETs among the 40 O128:H2 EPEC strains, which accounted for 50% of the genetic diversity among strains of the O128 group (Table 2). The 19 O128:H7 ETEC strains were nearly identical in electrophoretic profile and were classified into two closely related ETs. The 5 ETs with ETEC serotype O128:H12 were genetically diverse, as reflected by the H_I and H_{ET} values, both of which exceeded 0.20. The nonmotile O128 strains separated into three highly divergent ETs; the genetic diversity among these isolates exceeded that of the serogroup as a whole.

The 35 ETs identified among the O128 strains were organized into three major clusters (Fig. 2D). Strains with H12 flagellar antigen occur exclusively in the top cluster (ETs 1 to 8), which was found in only 16% of the AL trees (48% of the NJ trees). ET 8 was the most common O128:H12 serotype and was represented by seven isolates. Six O128:H8 strains of four closely related ETs formed a second cluster (ETs 9 to 12) recovered in 82% of the AL trees (81% of the NJ trees). With the exception of ET 35, the remaining isolates of the O128 serogroup belong in the large cluster of 22 ETs, within which no consensus order of branching or significant groupings resulted from the bootstrap analysis. The most common ETs are as follows: ET 16, which represents 19 O128:H7 strains; ET 21, which includes 10 O128:H21 strains; and ETs 29 to 31, which together represent 30 O128:H2 strains. Interestingly, the H2 flagellar antigen was found in O128 strains of diverse genetic backgrounds.

O157 strains. The occurrence of extensive genetic variation among O157 strains is reflected in the following observations (Table 2). All 20 enzymes assayed were polymorphic, with an average of 5.3 alleles per enzyme locus; the mean genetic diversity per locus among isolates (H_I) was 0.275, indicating that O157 strains have different alleles, on average, at 28% of the enzyme loci; and there are 50 ETs which differ, on average, at 45% of their enzyme loci (H_{ET} = 0.445).

Only 3% (0.008/0.275) of the genetic diversity of the O157 serogroup is accounted for by variation among strains with the H7 flagellar antigen (Table 2). Most of the O157:H7 isolates (95%) belong to a single ET, designated ET 11, and the electrophoretic profiles of the closely related variant strains differ only by single alleles from that of ET 11 (Fig. 3). The tight cluster of O157:H7 strains was frequently recovered in the bootstrap analysis (93% of the AL trees and 91% of the NJ trees) and is highly divergent (d = 0.67) from the remaining clusters of the O157 serogroup. The other major cluster, accounting for a large proportion of the O157 isolates, is composed of three ETs (ETs 15 to 17) and was observed in the majority of the bootstrapped trees (92% of the AL trees and 88% of the NJ trees). There were 142 isolates of ET 16, the most frequently recovered enzyme genotype of this cluster. Most of the isolates of this ET (previously referred to as ET 14 in reference 55) were originally collected from cases of porcine colibacillosis, express K88 fimbriae, and carry H43 flagellar antigen (55).

Most frequently isolated ETs. Although the strains of 16 serotypes associated with outbreaks of diarrheal disease represent a large variety of chromosomal genotypes, as reflected in the numerous ETs of each serotype, the comparison of electrophoretic profiles revealed common ETs that together accounted for a majority of the isolates of each serogroup. Table 3 presents the allele profiles of the 15 most frequently recovered ETs, which generally are genetically distant from one another, differing on average at 55% of the 20 enzyme loci. Together, these 15 ETs, which represent 8% of a total of 191 ETs resolved in the combined data (Table 2), account for more than 70% of all the isolates examined here. Because most of the strains of these ETs were originally recovered from patients with diarrheal disease in widely separated geographic areas, we infer that they individually mark widespread clones, which we will refer to as diarrheagenic *E. coli* (DEC) clones. The serotypic characteristics of the isolates of the DEC clones are listed in Table 4.

Examination of the allele profiles presented in Table 3 discloses several close genetic relationships of DEC clones of different serotypes and disease classes. For example, there is only a single allele difference in multilocus genotype

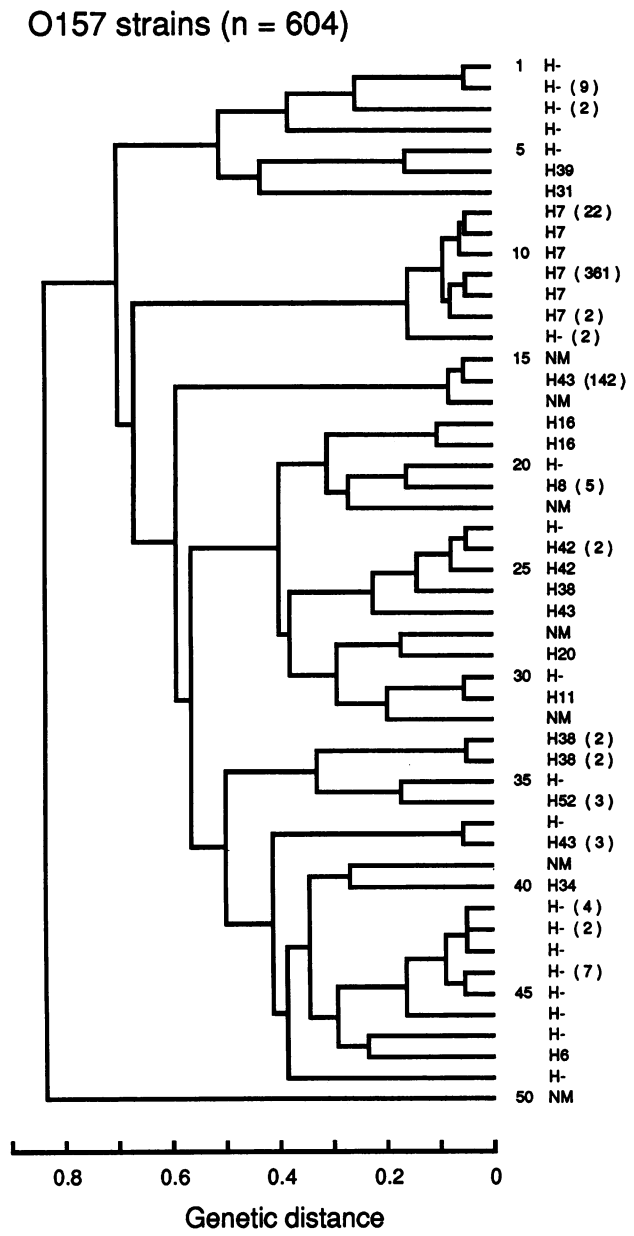

FIG. 3. Genetic relationships of multilocus enzyme genotypes (ETs) of *E. coli* strains of the O157 serogroup. The dendrogram was produced by the AL method applied to the matrix of standard genetic distances between multilocus genotypes. The predominant flagellar antigen type (H type) of strains is marked at the right. The number in parentheses is the number of isolates of each ET that was isolated more than once.

between DEC 3 and DEC 4, representing the O157:H7 clones, and DEC 5, the major O55:H7 clone (Table 3). This similarity in electrophoretic profile indicates an overall close relatedness in the genomes of these strains. These ETs are very distinct from those of the other common clones, differing at nine enzyme loci from DEC 6, the most similar DEC clone, isolates of which can differ in serotype but are usually O111:H12 (Table 4).

Other close genetic relationships among the major clones include DEC 8, 9, and 10, which show single-allele differences at the glyceraldehyde-phosphate dehydrogenase and gluconate-6-phosphate dehydrogenase loci but otherwise have the same alleles at 18 of the enzyme loci (Table 3). The strains of these clones are typically O26:H11 or O111:H8, although some other serotypes are occasionally encountered (Table 4). There is also a close relationship between DEC 11 and 12 strains (Table 3), all of which express H2 flagellar antigen but have either O111 or O128 somatic antigen (Table 4).

O standard strains. In four of the five cases examined, the standard strain for an O antigen belonged to a common ET of a serogroup. Comparison of the allele profiles showed that H311b (O26:H—) matched DEC 10, Stoke W (O111:H—) matched DEC 12, Cigleris (O128:H2) matched DEC 11, and A2 (O157:H19) matched DEC 7. The exception was Su3912-41 of the O55 serogroup, a strain originally isolated from human pus (40). This strain was ET 42 among the O55 strains (Fig. 2B) and was only distantly related to the clusters containing the common diarrheagenic O55 ETs.

***slt* genes.** Our finding that O157:H7 strains, which characteristically carry *slt* genes and produce one or two Shiga-like cytotoxins (2, 16, 31), belong to a distinct cluster in which the common ETs (DEC 3 and 4) are 95% similar in enzyme allele profile to the common O55:H7 ET (DEC 5), suggesting that these O55 strains also harbor *slt* genes. To test this idea, we extracted and purified total DNA from 93 DEC 5 isolates and 54 DEC 4 (O157:H7) isolates and tested for the presence of *slt* genes by dot-blot analysis. Hybridization with specific gene probes for *slt-I* and *slt-II* synthesized by PCR revealed that none of the DEC 5 isolates was positive for the *slt* genes, whereas all O157:H7 isolates were positive for at least one of the *slt* genes. These results indicate that, in contrast to wild-type O157:H7 strains, the O55 strains of the DEC 5 clonal lineage do not carry either *slt-I* or *slt-II*.

DISCUSSION

The subdivision of *E. coli* strains associated with diarrheal disease into categories based on distinct clinical features, discrete virulence and adherence properties, and into various O:H serotypes has been a useful and convenient system for studying the pathogenesis and epidemiology of these enteric disease agents (23). However, because many serotypes of *E. coli* are genetically heterogenous (1, 3, 8, 34, 38, 40, 55, 57), the identification and classification of pathogenic strains from epidemiological surveys based solely on O:H serotyping are not always indicative of genetic relatedness. Here, we have shown that the number of distinct multilocus genotypes observed within a collection of strains of a single serotype increases with sample size and that, consequently, sample sizes of several hundred isolates may be required to see the full extent of genotypic variation. Average genetic diversity, or the proportion of loci with different alleles between pairs of isolates, typically falls between 0.10 and 0.20 for strains of the same O:H serotype and between 0.30 and 0.40 for strains of the same O serogroup. Nonmotile isolates of an O serogroup are more variable than those of the same H type.

Evolutionary radiation of the major DEC clones. Although our analysis of strains has shown substantial genotypic variation among isolates of a serotype, the comparison of electrophoretic profiles revealed common ETs that together account for a majority of the isolates of each serogroup. Because isolates of identical enzyme genotype were originally recovered at different times from patients in separate geographic areas, we infer that these multilocus genotypes distinguish pathogenic clones of wide geographic distribu-

TABLE 3. Fifteen common ETs defined by distinct allele combinations for 20 enzyme loci

DEC clone	Serogroup-ET no.[a]	Allele at enzyme locus[b]:																		
		PGI	IDH	ACO	G3P	PE2	AK	MDH	PGD	M1P	GOT	BGA	ADH	MPI	G6P	IPO	CAK	NSP	TDH	SKD	GLU
1	O55-7	6	3	7	4	2.3	4	4	12	4	7	6	3	6	4	4	4	2	4	12	2
2	O55-5	6	3	7	4	2.3	4	4	12	3.8	7	6	2	6	4	4	4	2	4	12	2
3	O157-8	5	4	6	4	4.2	2	4	6	4	2	7	6	6	4	4	4	2	2	2	2
4	O157-11	5	4	6	4	5	2	4	6	4	2	7	6	6	4	4	4	2	2	2	2
5	O55-15	5	4	6	4	5	2	4	12	4	2	7	6	6	4	4	4	2	2	2	2
6	O111-2	5	2	7	4	5	2	4	8	6	6	8	2	4	4	4	4	2	2	3	2
7	O157-16	4	5	4	4	7	2	4	6	8	6	4	0	4	4	4	4	2	4	2	4
8	O111-8	8	5	7	4	5	2	4	8	8	6	5	6	8	4	4	4	2	4	8	4
9	O26-13	8	5	7	6	5	2	4	6	8	6	5	6	8	4	4	4	2	4	8	4
10	O26-15	8	5	7	4	5	2	4	6	8	6	5	6	8	4	4	4	2	4	8	4
11	O128-30	6	5	7	4	5	2	4	6	6	6	2.5	6	10	4	4	4	2	4	4	2
12	O111-20	6	5	7	4	5	2	4	8	7	6	2.5	6	10	4	4	4	2	4	4	2
13	O128-16	6	5	6	4	5	1	4	6	8	6	8	6	10	4	4	4	2	4	8	2
14	O128-21	6	5	6	4	5	1	4	6	8	6	4	6	11	4	4	4	2	4	8	2
15	O111-16	6	5	6	4	5	2	4	8	8	6	8	4	8	4	4	6	2	4	8	2

[a] Serogroup and ET number; see Fig. 2 and 3.
[b] Abbreviations: PGI, glucosephosphate isomerase; IDH, isocitrate dehydrogenase; ACO, aconitase; G3P, glyceraldehyde-phosphate dehydrogenase; PE2, phenylalanyl-leucine peptidase; AK, adenylate kinase; MDH, malate dehydrogenase; PGD, gluconate-6-phosphate dehydrogenase; M1P, mannitol-1-phosphate dehydrogenase; GOT, aspartate aminotransferase; BGA, β-galactosidase; ADH, alcohol dehydrogenase; MPI, mannosephosphate isomerase; G6P, glucose-6-phosphate dehydrogenase; IPO, indophenol oxidase; CAK, carbamate kinase; NSP, nucleoside phosphorylase; TDH, threonine dehydrogenase; SKD, shikimate dehydrogenase; GLU, glutamate dehydrogenase.

tions. Thus, as observed in other genetically variable pathogenic species of bacteria (41), the majority of the outbreaks and sporadic cases of diarrheal disease are caused by a small proportion of the total number of extant *E. coli* clones.

To elucidate the history and origin of the major DEC clones, we have further characterized five isolates of each DEC clone for allelic variation at an additional 13 enzyme loci (56) and used the NJ method to derive a phylogenetic tree (Fig. 4). The tree represents a phylogeny that predicts a pattern of ancestry and levels of genomic divergence among clones. Such clonal phylogenies serve as hypotheses that can be tested by examining genetic variation at additional loci. Hypothetical phylogenies also provide a framework for drawing inferences about past horizontal gene transfer and recombination events. In general, the rate of recombination in natural populations of *E. coli* appears to be low (51–53); however, individual genes, such as *gnd* (4), may have obscured the histories because they have experienced high rates of intragenic recombination.

The clonal phylogeny (Fig. 4) suggests the following features concerning the evolutionary radiation of the DEC clones. First, the genetic distances along major branches of the tree reflect substantial levels of genetic divergence between the chromosomal genomes. Here, branch lengths are measured in terms of genetic distance, which, for electrophoretically detected protein polymorphisms, estimates the net number of detectable codon differences per enzyme locus (29). From the observed accumulation of nucleotide substitutions that result in charged amino acid replacements among genes sequenced in *E. coli* K-12 and *Salmonella typhimurium* LT-2, Whittam and Ake (51) estimated a mutation rate per locus of 3.2×10^{-8} per year. (The calculation

TABLE 4. Characteristics of major DEC clones

DEC clone	No. of isolates	Geographic source[a]	Time period	Predominant serotype	Other serotypes	% Non-motile	Virulence factors[b]	
							Adherence	Toxins
1	13	NA, SA, EU, AF	1951–1977	O55:H6	O55:H7, O111:H1	0	AE, EAF	
2	53	NA, SA, EU	1951–1983	O55:H6	O55:H7/H19/H21	12	AE, EAF	
3	22	NA	1985–1988	O157:H7		0	AE	SLT
4	361	NA, SA, EU, AF, AS	1975–1990	O157:H7		2	AE	SLT
5	117	NA, SA, AF, AS	1950–1988	O55:H7	O55:H1/H11	39	AE	
6	23	NA, SA, EU	1951–1983	O111:H12	O111:H1/H4/H21, O55:H7	4		
7	142	NA	1975–1990	O157:H43	O157:H20	15	K88	LT, ST
8	37	NA, EU	1965–1988	O111:H8	O111:H11/H13	51	AE	SLT
9	17	NA, SA	1952–1979	O26:H11	O26:H46	6	AE	SLT
10	43	NA, EU, AU	1952–1989	O26:H11	O128:H8	0	AE	SLT
11	17	NA	1971–1977	O128:H2		0	AE, EAF	
12	43	NA, SA, EU, AF	1950–1986	O111:H2	O111:H28	32	AE, EAF	
13	19	NA	1971–1981	O128:H7	O128:H47	0		LT
14	10	NA, SA	1970–1983	O128:H21		20		
15	12	NA, EU	1967–1984	O111:H21		0		

[a] NA, North America; SA, South America; EU, Europe; AF, Africa; AS, Asia; AU, Australia.
[b] Based on results in references 20, 28, 34, 55, and 56. Abbreviations: AE, presence of *eae* gene, whose product is involved in attaching and effacing lesions; EAF, EPEC adherence factor; SLT, Shiga-like toxin genes; LT, heat-labile enterotoxin; ST, heat-stable enterotoxin; K88, porcine fimbrial antigen.

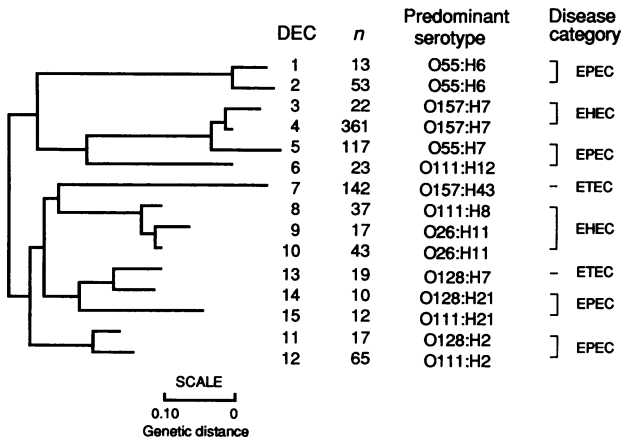

FIG. 4. Phylogenetic tree inferred from allelic variation for 33 enzyme loci (56) for 15 common ETs associated with diarrhea. The tree was constructed by the NJ method (37). Distance is defined as the net number of detectable codon differences per locus. The numbers of isolates (*n*) and serotypes of each clone are given in the right-hand columns. The 13 enzymes examined in addition to the 20 listed in Table 3 are hexokinase, propanediol dehydrogenase, formaldehyde dehydrogenase, leucine aminopeptidase, diaphorase, glycerate dehydrogenase, acid phosphatase 1 and 2, phosphoglucomutase, glutamine synthetase, peroxidase, glycerol kinase, and tyrosine aminotransferase.

assumes that protein electrophoresis detects all codon changes between alleles that involve charged amino acids in the protein.) At this rate, a genetic distance of 0.10 (net number of detectable codon differences per locus) is expected to accumulate along a branch of the tree, on average, in 3.2 million years. Hence, the length of time back to a common ancestor for the DEC clones is on the order of 10 to 15 million years, indicating that there has been sufficient time for DEC clones to accumulate many differences in the genes involved in pathogenesis. Second, clones of the same disease category are not monophyletic; that is, strains of the same disease class (e.g., EHEC) are not restricted to a single branch of the tree. This observation is not surprising, given that many of the factors implicated in virulence and used to define disease classes are encoded on plasmids or carried by bacteriophages (23, 24, 47). Finally, the tree predicts that clones of certain different disease categories or serotypes are genomically closely related; for example, EPEC DEC 5 (O55:H7) lies in the same lineage as EHEC DEC 3 and 4 (O157:H7), and ETEC DEC 13 (O128:H7) is closely related to EPEC clones DEC 14 (O128:H21) and DEC 15 (O111:H21). Such close genetic relationships among clones of a lineage suggest that these strains have in common traits involved in pathogenesis.

Virulence properties of DEC clones. Representative strains of several of the DEC clones described here have been examined for virulence factors in previous studies (28, 34). Because isolates of the same ET or clone are very closely related, as indicated by the identity in electrophoretic profile, we expect that certain properties, especially those encoded by chromosomal genes, will be virtually uniform in all members of a clone. (Indeed, Böhm and Karch [6] have demonstrated that epidemiologically unrelated isolates of O157:H7 were virtually indistinguishable in DNA fragments resolved by pulsed-field gel electrophoresis.) In contrast, the virulence traits encoded by plasmids or associated with bacteriophages may be more variable among independent isolates of a clone because of loss of these genetic elements in nature or during laboratory culture.

Two distinct complexes of clones have the localized pattern of HEp-2 cell adherence (9) that defines class I EPEC (23). The first complex, which includes DEC 1 and DEC 2, is part of a cluster of ETs composed of strains with the classic EPEC serotypes O55:H6, O125:H6, and O142:H6 (34). These clones were previously referred to as ET 4 and 5 (34), based on the analysis of isolates from infantile diarrhea that were shown to have the EPEC adherence factor (EAF) and localized adherence (LA) to HEp-2 cells (41). Similar properties (EAF^+ and LA^+) were described by Moyenuddin and colleagues (28) for two isolates of DEC 2 (CDC strain numbers 607-54 and 5513-56) that were originally recovered from outbreaks of diarrheal illness in the United States. They showed also that neither of these strains was positive for Shiga-like toxin, as measured by a Vero cell assay and DNA probes, and none was probe positive for *E. coli* heat-labile or heat-stable enterotoxins (28). The second complex includes DEC 11 and DEC 12, which belong to the cluster (ETs 12 to 19; see Fig. 2 in reference 34) comprising several EPEC serotypes with H2 antigen (34). In addition to the five isolates examined by Ørskov and coworkers (34), DEC 12 includes six O111 strains isolated from U.S. outbreaks of infantile diarrhea in the 1950s (CDC numbers 806-50, 2956-53, 4874-54, 124-55, 380-56, and 1852-59; see Table 1 in reference 28), all of which were EAF^+. These strains also did not produce verotoxin or carry gene sequences for Shiga-like toxin, or *E. coli* heat-labile or heat-stable enterotoxins (28).

Although these two complexes apparently have in common the EAF^+ and concomitant LA^+ phenotypes and both lack cytotoxin and enterotoxin genes, they are only distantly related (Fig. 4).

Reconstructing the origin of the O157:H7 clone complex. Our analysis has shown that the O157:H7 (DEC 4) clone, associated with outbreaks of HC and hemolytic uremic syndrome, and the major O55:H7 (DEC 5) clone, which is associated with infantile diarrhea, are similar genetically and thus were recently derived from a common ancestral cell. From the close relationship between the genomes of these distinct pathogens, as reflected by the 95% identity in multilocus enzyme genotype, we would expect that these clones also have many common genetic attributes, including virulence factors, inherited from their most recent ancestor. Hence, it is noteworthy that both O55:H7 (26) and O157:H7 (56) attach intimately to the surfaces of intestinal epithelial cells in the initial stages of infection, efface the microvilli, and induce characteristic histological and ultrastructural lesions in animal models (26, 50). Knutton et al. (21) have shown that these attaching and effacing lesions are composed of dense concentrations of actin microfilaments in the apical cytoplasm beneath the attached bacteria and have demonstrated that both O55:H7 and O157:H7 strains produce similar lesions in human tissue cultures. Furthermore, the production of attaching and effacing lesions requires neither plasmid-encoded products nor expression of cytotoxins (21, 49), and the mechanism is determined, in part, by the chromosomal *eae* gene (17), which presumably was present in the most recent ancestor of the O55:H7 and O157:H7 strains.

We propose a two-step scenario for the evolution of the O157:H7 clone. First, an ancestral *E. coli* evolved the chromosomally encoded gene products that mediate attaching and effacing adherence. This attribute alone may be sufficient for bacteria to cause diarrheal disease in infants, as

is the case with the contemporary O55:H7 clone. Second, an O55:H7-like progenitor cell, already able to cause disease by the attaching-effacing mechanism, acquired secondary virulence factors, such as the genes encoding Shiga-like toxins and adhesins, via horizontal genetic transfer from other strains. With these genes expressed in the attaching-effacing chromosomal background, a new pathogen causing a new type of disease emerged—the O157:H7 clone.

One prediction of the above scenario is that the O55:H7 lineage (DEC 5) should have an allele of the *eae* gene that is very similar in sequence to that of the O157:H7 lineage. Yu and Kaper (58) have shown that there is 86% similarity in nucleotide sequence at the *eae* locus between O157:H7 and a derivative of EPEC strain E2348/69 (O127, EAF$^+$). However, strain E2348/69 is only distantly related to the O55:H7/O157:H7 lineage, with different electrophoretic alleles at about 50% of the enzyme loci (56). If the *eae* gene has diverged along with the rest of the chromosomal backgrounds of these strains, we predict that the close genetic relationship of the O55:H7 and O157:H7 clones should be reflected in the relative divergence of their *eae* alleles, which should be at most 1 to 2% different. These alleles, however, could show greater levels of differentiation if, for example, natural selection has favored certain amino acid replacements in intimin molecules or if intragenic recombination has generated novel mosaic alleles (4, 51).

The hypothesis that the O157:H7 clone originated from an O55:H7-like ancestor through horizontal transfer and recombination also gains support from the observation that both Shiga-like toxins, encoded by prophage, and EPEC-associated adhesins, encoded on plasmids, occur in diverse lineages of the *E. coli* population (32); presumably, these genes can be transferred between strains in nature. Using specific gene probes for the Shiga-like toxins, we found that none of the O55 isolates examined here harbored *slt* sequences, a result suggesting that the recent ancestor of the O157:H7 and O55:H7 clones did not produce these cytotoxins. The pattern of codon usage in the *slt* genes differs from that of most *E. coli* protein-coding genes (16); indeed, the codon adaptation index (42), a measure of the degree of synonymous codon usage bias, for *slt* genes falls below the values reported for 165 *E. coli* genes (42). Such a low degree of codon adaptation strongly argues that the *slt* genes are foreign to *E. coli* and were relatively recently acquired via horizontal transfer.

Conclusions. In sum, *E. coli* strains that cause infectious diarrhea belong to a small number of geographically widespread clones. Phylogenetic analysis based on variation in multilocus enzyme genotype demonstrates that the O157:H7 clone is closely related to a group of O55 strains implicated in infantile diarrhea, a result suggesting that this new pathogen was recently derived from an O55:H7-like ancestral clone by addition of Shiga-like toxins and adhesin genes to a genome already adapted for causing diarrheal disease. These findings underscore the complex and multifactorial nature of microbial pathogenesis (14), in which the acquisition and expression of multiple virulence factors by strains with specific chromosomal backgrounds may foster the emergence of new pathogenic clones.

ACKNOWLEDGMENTS

We thank A. O'Brien, P. Tarr, A. Cravioto, H. Lior, L. Beutin, W. Maas, and J. Mathewson for supplying bacterial cultures.

This work was supported by Public Health Service grants AI 24566 and AI 00964 from the National Institutes of Health.

REFERENCES

1. **Achtman, M., and G. Pluschke.** 1986. Clonal analysis of descent of virulence among selected *Escherichia coli*. Annu. Rev. Microbiol. **40:**185–210.
2. **Beutin, L., M. A. Montenegro, I. Ørskov, F. Ørskov, J. Prada, S. Zimmermann, and R. Stephan.** 1989. Close association of verotoxin (Shiga-like toxin) production with enterohemolysin production in strains of *Escherichia coli*. J. Clin. Microbiol. **27:**2559–2564.
3. **Beutin, L., I. Ørskov, F. Ørskov, S. Zimmermann, J. Praedia, H. Gelderblom, R. Stephan, and T. S. Whittam.** 1990. Clonal diversity and virulence factors in strains of *Escherichia coli* of the classic enteropathogenic serogroup O114. J. Infect. Dis. **162:**1329–1334.
4. **Biserčić, M., J. Y. Feutrier, and P. R. Reeves.** 1991. Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. J. Bacteriol. **173:**3894–3900.
5. **Bitzan, M., H. Karch, M. G. Maas, T. Meter, H. Rüssman, S. Alkeksic, and J. Bockemühl.** 1991. Clinical and genetic aspects of Shiga-like toxin introduction in traditional enteropathogenic *E. coli*. Zentralbl. Bakteriol. **274:**496–506.
6. **Böhm, H., and H. Karch.** 1992. DNA fingerprinting of *Escherichia coli* O157:H7 strains by pulsed-field electrophoresis. J. Clin. Microbiol. **30:**2169–2172.
7. **Bopp, C. A., K. D. Greene, F. P. Downes, E. G. Sowers, J. G. Wells, and I. K. Wachsmuth.** 1987. Unusual verotoxin-producing *Escherichia coli* associated with hemorrhagic colitis. J. Clin. Microbiol. **25:**1486–1489.
8. **Caugant, D. A., B. R. Levin, I. Ørskov, F. Ørskov, C. Svanborg Eden, and R. K. Selander.** 1985. Genetic diversity in relation to serotype in *Escherichia coli*. Infect. Immun. **49:**407–413.
9. **Cravioto, A., R. J. Gross, S. M. Scotland, and B. Rowe.** 1979. An adhesive factor found in strains of *Escherichia coli* belonging to the traditional infantile enteropathogenic serotypes. Curr. Microbiol. **3:**95–99.
10. **Cryan, B.** 1990. Enterohaemorrhagic Escherichia coli. Scand. J. Infect. Dis. **22:**1–4.
11. **Donnenberg, M. S., and J. B. Kaper.** 1992. Enteropathogenic *Escherichia coli*. Infect. Immun. **60:**3953–3961.
12. **Felsenstein, J.** 1984. Distance methods for inferring phylogenies: a justification. Evolution **38:**16–34.
13. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
14. **Finlay, B. B., and S. Falkow.** 1989. Common themes in microbial pathogenesis. Microbiol. Rev. **53:**210–230.
15. **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.
16. **Jackson, M. P., R. J. Neill, A. D. O'Brien, R. K. Holmes, and J. W. Newland.** 1987. Nucleotide sequence analysis and comparison of the structural genes for Shiga-like toxin I and Shiga-like toxin II encoded by bacteriophages from *Escherichia coli* 933. FEMS Lett. **44:**109–114.
17. **Jerse, A. E., J. Yu, B. D. Tall, and J. B. Kaper.** 1990. A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. Proc. Natl. Acad. Sci. USA **87:**7839–7943.
18. **Jin, L., and M. Nei.** 1991. Relative efficiencies of the maximum-parsimony and distance-matrix methods of phylogeny construction for restriction data. Mol. Biol. Evol. **8:**356–365.
19. **Karch, H., J. Heesemann, R. Laufs, A. D. O'Brien, C. O. Tacket, and M. M. Levine.** 1987. A plasmid of enterohemorrhagic *Escherichia coli* O157:H7 is required for expression of a new fimbrial antigen and for adhesion to epithelial cells. Infect. Immun. **55:**455–461.
20. **Karmali, M.** 1989. Infection by verocytotoxin-producing *Escherichia coli*. Clin. Microbiol. Rev. **2:**15–38.
21. **Knutton, S., T. Baldwin, P. H. Williams, and A. S. MacNeish.** 1989. Actin accumulation at sites of bacterial adhesion to tissue culture cells: basis of a new diagnostic test for enteropathogenic

and enterohemorrhagic *Escherichia coli*. Infect. Immun. **57:** 1290–1298.
22. **Knutton, S., A. D. Phillips, H. R. Smith, R. J. Gross, R. Shaw, P. Watson, and E. Price.** 1991. Screening for enteropathogenic *Escherichia coli* in infants with diarrhea by the fluorescent-actin staining test. Infect. Immun. **59:**365–371.
23. **Levine, M. M.** 1987. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. J. Infect. Dis. **155:**377–389.
24. **Levine, M. M., J. P. Nataro, H. Karch, M. M. Baldini, J. P. Kaper, R. E. Black, M. L. Clements, and A. D. O'Brien.** 1985. The diarrheal response of humans to some classic serotypes of enteropathogenic *Escherichia coli* is dependent on a plasmid encoding an enteroadhesiveness factor. J. Infect. Dis. **152:**550–559.
25. **Levine, M. M., J.-G. Xu, J. B. Kaper, H. Lior, V. Prado, B. Tall, J. Nataro, H. Karch, and K. Wachsmuth.** 1987. A DNA probe to identify enterohemorrhagic *Escherichia coli* of O157:H7 and other serotypes that cause hemorrhagic colitis and hemolytic uremic syndrome. J. Infect. Dis. **156:**175–182.
26. **Moon, H. W., S. C. Whipp, R. A. Argenzio, M. M. Levine, and R. A. Giannella.** 1983. Attaching and effacing activities of rabbit and human enteropathogenic *Escherichia coli* in pig and rabbit intestines. Infect. Immun. **41:**1340–1351.
27. **Morgan, G. M., C. Newman, S. R. Palmer, J. B. Allen, W. Shepard, A. Rampling, R. E. Warren, R. J. Gross, S. M. Scotland, and H. R. Smith.** 1988. First recognized community outbreak of hemorrhagic colitis due to verotoxin-producing *Escherichia coli* O157:H7 in the UK. Epidemiol. Infect. **101:**83–91.
28. **Moyenuddin, M., I. K. Wachsmuth, S. L. Moseley, C. A. Bopp, and P. A. Blake.** 1989. Serotype, antimicrobial resistance, and adherence properties of *Escherichia coli* strains associated with outbreaks of diarrheal illness in children in the United States. J. Clin. Microbiol. **27:**2234–2239.
29. **Nei, M.** 1975. Molecular population genetics and evolution. American Elsevier Publishing Company, Inc., New York.
30. **Newland, J. W., N. A. Strockbine, S. F. Miller, A. D. O'Brien, and R. K. Holmes.** 1985. Cloning of Shiga-like toxin structural genes from a toxin-converting phage of *Escherichia coli*. Science **230:**179–181.
31. **O'Brien, A. D., and R. K. Holmes.** 1987. Shiga and Shiga-like toxins. Microbiol. Rev. **51:**206–220.
32. **O'Brien, A. D., J. W. Newland, S. F. Miller, R. K. Holmes, H. W. Smith, and S. B. Formal.** 1984. Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. Science **226:**694–696.
33. **Ørskov, F., and I. Ørskov.** Unpublished data.
34. **Ørskov, F., T. S. Whittam, A. Cravioto, and I. Ørskov.** 1990. Clonal relationships among classic enteropathogenic *Escherichia coli* (EPEC) belonging to different O groups. J. Infect. Dis. **162:**76–81.
35. **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.
36. **Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, E. S. Olcott, L. M. Johnson, N. T. Hargrett, P. A. Blake, and M. L. Cohen.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N. Engl. J. Med. **308:**681–685.
37. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
38. **Scotland, S. M., G. A. Willshaw, H. R. Smith, and B. Rowe.** 1990. Properties of strains of *Escherichia coli* O26:H11 in relation to their enteropathogenic or enterohemorrhagic classification. J. Infect. Dis. **162:**1069–1074.
39. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. H. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:**873–884.
40. **Selander, R. K., D. A. Caugant, and T. S. Whittam.** 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
41. **Selander, R. K., and J. M. Musser.** 1990. The population genetics of bacterial pathogenesis, p. 11–36. *In* B. H. Iglewski and V. L. Clark (ed.), Molecular basis of bacterial pathogenesis. Academic Press, Orlando, Fla.
42. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaptation index—a measure of the directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:**1281–1295.
43. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
44. **Smith, H. R., and S. M. Scotland.** 1988. Vero cytotoxin-producing strains of *Escherichia coli*. J. Med. Microbiol. **26:**77–85.
45. **Sneath, P. H. A., and R. R. Sokal.** 1973. Numerical taxonomy. W. H. Freeman & Co., San Francisco.
46. **Stenderup, J., and F. Ørskov.** 1983. The clonal nature of enteropathogenic *Escherichia coli* strains. J. Infect. Dis. **148:** 1019–1024.
47. **Strockbine, N. A., L. R. M. Marques, J. W. Newland, H. W. Smith, R. K. Holmes, and A. D. O'Brien.** 1986. Two toxin-converting phages from *Escherichia coli* O157:H7 strain 933 encode antigenically distinct toxins with similar biologic activities. Infect. Immun. **53:**135–140.
48. **Tesh, V. L., and A. D. O'Brien.** 1992. Adherence and colonization mechanisms of enteropathogenic and enterohemorrhagic *Escherichia coli*. Microb. Pathog. **12:**245–254.
49. **Tzipori, S., H. Karch, K. I. Wachsmuth, R. M. Robins-Browne, A. D. O'Brien, H. Lior, M. L. Cohen, J. Smithers, and M. M. Levine.** 1987. Role of a 60-megadalton plasmid and Shiga-like toxins in the pathogenesis of infection caused by enterohemorrhagic *Escherichia coli* O157:H7 in gnotobiotic piglets. Infect. Immun. **55:**3117–3125.
50. **Tzipori, S., I. K. Wachsmuth, C. Chapman, R. Birner, J. Brittingham, C. Jackson, and J. Hogg.** 1986. The pathogenesis of hemorrhagic colitis caused by *Escherichia coli* O157:H7 in gnotobiotic piglets. J. Infect. Dis. **154:**712–716.
51. **Whittam, T. S., and S. E. Ake.** 1993. Genetic polymorphisms and recombination in natural populations of *Escherichia coli*, p. 223–246. *In* N. Takahata and A. G. Clark (ed.), Mechanisms of molecular evolution. Sinauer Associates, Sunderland, Mass.
52. **Whittam, T. S., H. Ochman, and R. K. Selander.** 1983. Multilocus genetic structure in natural populations of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80:**1751–1755.
53. **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.
54. **Whittam, T. S., I. K. Wachsmuth, and R. A. Wilson.** 1988. Genetic evidence of clonal descent of *Escherichia coli* O157:H7 associated with hemorrhagic colitis and hemolytic uremic syndrome. J. Infect. Dis. **157:**1124–1133.
55. **Whittam, T. S., and R. A. Wilson.** 1988. Genetic relationships among pathogenic *Escherichia coli* of serogroup O157. Infect. Immun. **56:**2467–2473.
56. **Whittam, T. S., M. L. Wolfe, and B. Miller.** Unpublished data.
57. **Whittam, T. S., M. L. Wolfe, and R. A. Wilson.** 1989. Genetic relationships among *Escherichia coli* isolates causing urinary tract infections in humans and animals. Epidemiol. Infect. **102:**37–46.
58. **Yu, J., and J. B. Kaper.** 1992. Cloning and characterization of the *eae* gene of enterohaemorrhagic *Escherichia coli* O157:H7. Mol. Microbiol. **6:**411–417.