INFECTION AND IMMUNITY, July 1997, p. 2685–2692 Vol. 65, No. 7
0019-9567/97/$04.00+0

Evolutionary Relationships among Pathogenic and Nonpathogenic *Escherichia coli* Strains Inferred from Multilocus Enzyme Electrophoresis and *mdh* Sequence Studies

GULIETTA M. PUPO, DAVID K. R. KARAOLIS,† RUITING LAN, AND PETER R. REEVES*

Department of Microbiology, The University of Sydney, Sydney New South Wales 2006, Australia

Received 18 November 1996/Returned for modification 20 February 1997/Accepted 21 April 1997

Within the species *Escherichia coli*, there are commensal strains and a variety of pathogenic strains, including enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), and urinary tract infection (UTI) strains. The pathogenic strains are identified by serotype and by possession of specific virulence determinants (toxins and adhesions, etc.) encoded by either monocistronic genes, plasmids, or pathogenicity islands. Although there are studies on the relationships between selected pathogenic strains, the relatedness among the majority of the pathogenic forms to each other, to commensal *E. coli*, and to the genus *Shigella* (which has often been suggested to be part of *E. coli*) has not been determined. We used multilocus enzyme electrophoresis (MLEE) at 10 enzyme loci and the sequence of the *mdh* housekeeping gene to study the genetic relationships of pathogenic *E. coli* strains (including *Shigella* clones), namely, 5 EPEC strains (serotypes O111 and O55), 3 EHEC strains (serotype O157), 6 ETEC strains (serotypes O78, O159, and O148), 5 EIEC strains (serotypes O124, O28, and O112), and 13 *Shigella* strains representing clones Flexneri, Dysenteriae, Boydii, and Sonnei, to commensal *E. coli* strains. Both the MLEE and *mdh* sequence trees reveal that EPEC, EHEC, ETEC, EIEC, and UTI strains are distributed among the ECOR set groups, with no overall clustering of EPEC, ETEC, EIEC, or UTI strains. The genus *Shigella* is shown to comprise a group of closely related pathogenic *E. coli* strains. Six pathogenic strains, i.e., M502 (EIEC; O112ac:NM), M503 (EPEC; O111:H12), M526 (ETEC; O159:H4), M522 (EPEC; O111ac:H12), M524 (ETEC; O78:H11), and M506 (ETEC; O78:H11), were found to have *mdh* sequences identical to those of five ECOR group A strains (ECOR5, ECOR10, ECOR14, ECOR6, and K-12). All 11 strains are closely related by MLEE. The results indicate that pathogenic strains of *E. coli* do not have a single evolutionary origin within *E. coli* but have arisen many times. The results also suggest the possibility that any *E. coli* strain acquiring the appropriate virulence factors may give rise to a pathogenic form.

Escherichia coli is a diverse species consisting of both commensal and pathogenic strains. It was first described in 1885 as *Bacillus coli*, a commensal intestinal inhabitant (13). The first *Shigella* species was described as *Bacillus dysenteriae* in 1898 (54), but later, it and *B. coli* were placed in different genera currently named *Shigella* and *Escherichia*. The two genera have long been known to be closely related; however, for some time, pathogenic forms (all human) were placed in the genus *Shigella* and, over time, four *Shigella* species were recognized and a series of characters was found to differentiate the two genera. *Escherichia* has several species, but probably only the relatively rare *Escherichia fergusonii* should be included in the same genus as *Escherichia coli* (25).

Since the 1940s, a variety of strains have been found which were classified as *Escherichia coli* rather than *Shigella* but which are pathogenic for humans or animals. They are responsible for significant worldwide diarrheal disease and include enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), and enteroinvasive *E. coli* (EIEC). EPEC was the first *E. coli* recognized to cause diarrheal disease, especially among infants (3, 4, 16, 37, 56). EPEC produces a characteristic attaching and effacing (A/E) lesion where the bacteria adhere intimately to the enterocyte surface, resulting in localized destruction of intestinal microvilli (9, 23, 30, 33, 44). EHEC has been responsible for outbreaks of hemorrhagic colitis and hemolytic-uremic syndrome (HUS). ETEC causes diarrhea in infants in developing countries and in travelers from industrialized countries who visit these areas (26, 27). EIEC causes an invasive, dysenteric form of diarrhea. Some *E. coli* bacteria can colonize outside of the intestine, and of those, the urinary tract infection (UTI) strains are perhaps the best studied.

The various types of pathogenic *E. coli* have specific virulence determinants. For example, EPEC has a pathogenicity island termed the locus for enterocyte effacement (30) which enables the production of A/E lesions by these strains; EHEC virulence factors include cytotoxins (Shiga-like toxin I [SLT-I] and SLT-II), and like EPEC, they possess genes for A/E lesions (8, 10, 12, 30, 38, 43, 53, 57); ETEC elaborates a heat-labile toxin and or a heat-stable toxin; and EIEC, in addition to being able to invade through products of genes encoded on an invasion plasmid (49–51), has recently been shown to also elaborate an enterotoxin (15). UTI strains have a specific pilus and produce hemolysins which allow colonization and tissue damage to the urinary epithelium (11, 28, 29, 58).

The distinction between *E. coli* and *Shigella* has been difficult to define since the 1940s since pathogenic *E. coli* and *Shigella* have many properties in common, EIEC being particularly *Shigella* like in its pathogenesis. In addition, EIEC isolates cross-react with traditional *Shigella* typing antisera (26), and EHEC SLT-I and SLT-II cytotoxins resemble *Shigella dysenteriae* toxin (Shiga toxin) (24). There are now good reasons based on similarity at the DNA level for considering all *E. coli* and

* Corresponding author. Phone: (612) 9351 2536. Fax: (612) 9351 4571. E-mail: reeves@angis.su.oz.au.

† Present address: Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, MD 21201.

TABLE 1. Details of strains analyzed

Disease group or clone	Strain	Serotype	Country or region of isolation	Source[a]	Original laboratory designation
EIEC	M500	O124:NM	United States (Texas)	CDC	EDL 1284
	M519	O124:NM	Argentina	SHL	C906-91
	M501	O28ac:NM	Bulgaria	CDC	3536-89
	M520	O28ac:NM	United States	IEKC	C499-89
	M502	O112ac:NM	Bulgaria	CDC	3541-89
EPEC	M503	O111ac:H12	Brazil	CDC	3608-90
	M522	O111ac:H12	Denmark	IEKC	C156-89
	M521	O111ac:H2	England	IEKC	C1250-91
	M505	O111ac:H2	Mexico	CDC	3593-90
	M504	O55:H6	United States (Philadelphia, Pa.)	CDC	607-54
ETEC	M506	O78:H11		CDC	H10407
	M524	O78:H11	United States	IEKC	C812-92
	M507	O159:H4	Central African Republic	CDC	T1918
	M526	O159:H4	Chile	IEKC	C494-88
	M508	O148:H28	Cruise ship	CDC	E4890-C1
	M525	O148:H28	United States	IEKC	C707-91
EHEC	M509	O157:H7	United States (Oregon)	CDC	EDL 933
	M527	O157:H7	Denmark	IEKC	C664-92
	M534	O157	Australia	SHL	
Flexneri	M515	2a	Cruise ship	CDC	SSU 6335
	M530	2	Australia	SHL	870N
	M528	2	Australia	SHL	224R
	M529[b]	4	African travel	SHL	200L
	M531	4	Malaysian travel	SHL	630Q
	M532	6	Australia	SHL	159N
Boydii	M533	1	Australia	SHL	158N
	M517	2	United States (Montana)	CDC	4334-74
	M518	14	United States (California)	CDC	248-59
Dysenteriae	M513	4	Panama Canal Zone	CDC	1112-74
Sonnei	M564[b]		United States	ATCC	ATCC 11060
	M565[b]		Australia	RNSH	
	M582		Japan (Tokyo)	NIHJ	624-85

[a] CDC, Centers for Disease Control, Atlanta, Ga.; SHL, State Health Laboratory, Perth, Australia; IEKC, International Escherichia and Klebsiella Centre, Copenhagen, Denmark; NIHJ, National Institute of Health, Tokyo, Japan; ATCC, American Type Culture Collection; RNSH, Royal North Shore Hospital, Sydney, Australia.

[b] Not sequenced.

Shigella strains as being in one species, *E. coli* (5–7, 14, 17, 18, 20, 40, 55, 59). Nonetheless, *Shigella* and *Escherichia* are generally treated as separate genera, and the official nomenclature has not changed. We consider it confusing when discussing relationships between strains to retain two genera where there is abundant evidence that this is not justified. We suggest that the current specific names be used with a capital first letter and without italics as names of clones or sets of clones and that the name Shigella likewise be used to describe a group of clones, without implying that a group so defined necessarily comprises related clones. In the case of Sonnei, we have good evidence that there is one clone (20), which we referred to as the Sonnei clone of *E. coli* or *E. coli* (Sonnei).

E. coli is a clonal species as can be inferred from the coexistence of different forms and as directly shown by multilocus enzyme electrophoresis (MLEE) (18, 40, 52, 59) and sequencing studies (2, 20, 35, 36). The ECOR set (39) is a reference collection of 72 wild-type *E. coli* isolates from humans and 16 other mammalian species (many from the Seattle zoo) obtained from a larger collection of approximately 2,600 isolates (31) and is thought to broadly represent genotypic variation in *E. coli* (39). Although the ECOR set includes many isolates of commensal forms from animal or humans, Milkman and McKane (32) state that none of the ECOR set strains is pathogenic. The only ECOR strains that may be considered pathogenic are those isolated from humans with UTI since there are no isolates from patients with episodes of diarrhea or dysentery. There have also been MLEE studies reported of pathogenic forms of *E. coli* which have shown that, in some cases at least, strains with the same O antigen can comprise a variety of quite divergent clones (42, 61, 63). However, these studies have not explored to any extent the relationship between pathogenic and commensal clones of *E. coli.*

To better understand the genetic relationships of commensal and pathogenic strains and the overall genetic structure of *E. coli*, including the Shigella strains, we used MLEE at 10 enzyme loci and the sequence of the *mdh* housekeeping gene to study (i) the genetic diversity and overall population structure of *E. coli*, (ii) the relationship of pathogenic clones to commensal

TABLE 2. Electrophoretic profiles

Strain[a]	Allele type at locus									
	MIP	MDH	IDH	PGD	G6P	PGM	MPI	PGI	ADK	ME
M500-EIEC	3	2	5	2	2	5	3	7	4	1
M501-EIEC	4	2	5	6	2	4	3	7	4	1
M502-EIEC	3	2	2	4	2	4	3	4	4	1
M519-EIEC	3	2	2	13	2	4	8	4	4	1
M520-EIEC	3	2	5	13	2	4	8	4	4	1
M503-EPEC	3	2	2	13	2	4	3	4	4	1
M504-EPEC	1	2	5	15	2	4	3	7	4	1
M505-EPEC	3	2	5	13	2	5	9	7	2	1
M521-EPEC	3	2	5	13	2	5	9	7	4	1
M522-EPEC	3	2	2	13	2	4	3	4	4	1
M506-ETEC	4	2	2	6	1	4	3	4	4	1
M507-ETEC	4	2	6	6	2	4	3	7	4	1
M508-ETEC	4	2	5	6	2	4	9	4	2	1
M524-ETEC	4	2	2	6	1	4	3	4	4	1
M525-ETEC	4	2	5	6	2	4	9	7	2	1
M526-ETEC	3	2	2	4	2	4	3	4	4	1
M534-EHEC	2	3	5	6	2	5	5	4	4	3
M527-EHEC	2	2	5	6	2	5	5	4	4	1
M509-EHEC	2	2	5	6	2	5	5	4	4	1
M515-Flex	4	7	5	4	2	3	5	4	4	1
M528-Flex	4	7	5	4	2	3	5	4	4	1
M529-Flex	4	7	5	4	2	3	5	4	4	1
M530-Flex	4	7	5	6	2	3	5	4	4	1
M531-Flex	4	7	5	4	2	3	5	4	4	1
M532-Flex	4	2	2	4	2	4	8	4	4	1
M533-Boyd	4	2	2	6	2	4	8	4	4	1
M517-Boyd	4	2	2	13	2	4	8	4	4	1
M518-Boyd	4	2	2	13	2	4	8	4	4	1
M513-Dysen	4	2	2	6	2	3	5	4	4	1
M564-Sonn	3	2	2	6	2	4	8	4	4	1
M565-Sonn	3	2	2	6	2	4	8	4	4	1
M582-Sonn	3	2	2	6	2	4	8	4	4	1

[a] The strain number is followed by the *E. coli* group or Shigella clone for which the strain is representative. Flex, Flexneri; Dysen, Dysenteriae; Sonn, Sonnei; Boyd, Boydii.

clones, and (iii) the genetic relationship of the Shigella clones to commensal strains and other pathogenic *E. coli* strains.

MATERIALS AND METHODS

Bacterial strains. Details of the strains used in this study are given in Table 1. Serotyping for each pathogenic strain was performed by standard techniques as described by Ørskov et al. (41).

MLEE. Enzyme lysates were prepared for each strain by sonication and analyzed electrophoretically by high-voltage zone electrophoresis with cellulose acetate gels (Cellogel; Chemetron) as the support medium. Electrophoretic methods were carried out as described by Richardson et al. (47). The electrophoretic conditions for each of the 10 loci were as follows: glucose-6-phosphate dehydrogenase (G6P), 6-phosphogluconate dehydrogenase (PGD), and phosphoglucose isomerase (PGI) were run with 0.046 M Tris-Barbitone (pH 8.8). Isocitrate dehydrogenase (IDH), phosphoglucomutase (PGM), malic enzyme (ME), mannitol-1-phosphate dehydrogenase (M1P), malate dehydrogenase (MDH), and mannose phosphate isomerase (MPI) were run with 0.015 M Tris–0.05 M EDTA–borate with 0.01 M $MgCl_2$ (pH 7.8). Adenylate kinase (ADK) was run with 0.13 M Tris–0.022 M Na_2EDTA–0.0713 M boric acid with 0.006 M NaOH (pH 8.9). Standard controls representing the alleles of the respective locus from the ECOR set were run with each sample, thus allowing use of the same scoring system as that used for the ECOR set. Comparisons of the mobilities of the alleles were made visually on the same gel sheet.

PCR amplification and nucleotide sequencing. Chromosomal DNA was prepared as described by Bastin et al. (1). The *mdh* gene encoding malate dehydrogenase (EC 1.1.1.37) was sequenced from PCR-amplified products by using 5′ ATGAAAGTCGCAGTCCTC 3′ as the 5′ primer and 5′ GGCGATATCTTTC TTCAGCG 3′ as the 3′ primer. Primer sequences were designed by taking into account the segments conserved in the published *E. coli* K-12 (34) and ECOR set strain (2) sequences. All oligonucleotides were synthesized with either universal forward or reverse M13 primer sequences attached to facilitate automated sequencing. The DNA sequence was obtained from the 34 bp after the ATG start codon to base 860 of the *mdh* gene. Double-stranded PCR DNA product was purified by use of the Wizard PCR purification system (Promega) to remove excess PCR primers and eluted in 25 μl of sterile distilled water, and the sequence was determined by the dye-labelled primer technique with a thermal cycler (Perkin-Elmer Cetus, Norwalk, Conn.) and an automated 377 DNA sequencer (Applied Biosystems, Burwood, Victoria, Australia) as described in the manufacturer's instructions.

Computer analysis. DNA sequences were analyzed by use of programs from the Australian National Genomic Information Service at the University of Sydney (46) or the MULTICOMP package (45), which gives pairwise comparisons of DNA and derived amino acid sequences and incorporates programs such as those within the PHYLIP package (version 3.4; written by Joseph Felsenstein, Department of Genetics, University of Washington, Seattle). Molecular evolutionary relationships among *mdh* sequences were examined by the neighbor-joining method of tree construction (48), based on the distance estimated by the two-parameter method of Kimura (22).

Phylogenetic trees for MLEE data were also constructed by the neighbor-joining method. A matrix of genetic distances between all pairs of electrophoretic types was calculated 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 (19).

Nucleotide sequence accession numbers. The sequences reported here have been assigned GenBank accession numbers AF004170 to AF004209.

RESULTS

MLEE variation. We undertook a study of 32 strains of pathogenic *E. coli* (including Shigella), namely, 5 EPEC strains (serotypes O111 and O55), 3 EHEC strains (serotype O157), 6 ETEC strains (serotypes O78, O159, and O148), 5 EIEC strains (serotypes O124, O28, and O112), and 13 Shigella strains representing all four major clone groups (Flexneri, Dysenteriae, Boydii, and Sonnei). These strains were chosen to

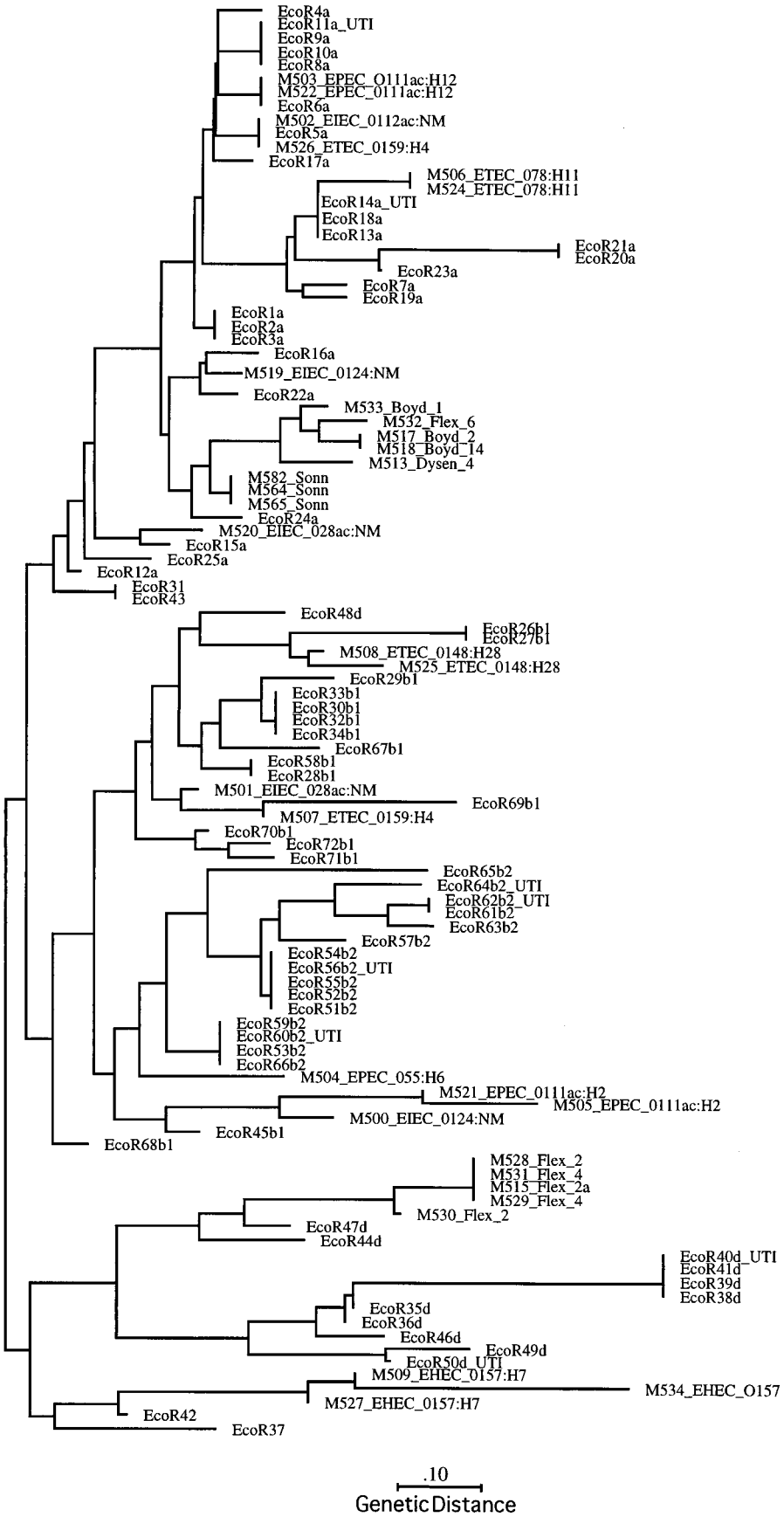

FIG. 1. Genetic relationships of commensal *E. coli* strains to pathogenic *E. coli* strains as resolved by MLEE.

```
        1122233333344456666667777777888
        2703712567756951235690125889024
        9774663192808765761605709238429
M500    TTCTCGCCCAGTCCCTGAATATTTCCCAGCC
M501    C.........A......G..............
M502    ...C......A....C..G............
M503    ...C......A....C..G............
M504    .ATC...TT........G.....C..T....
M505    ..................G.............
M506    ...C......A....C..G............
M507    C........TA......G.............
M508    .....T..........A.....ACT.A....
M509    C...T.........T..G.GTC.C...GT..
M513    ...C......A......G.............
M515    C..C..A...A......G.............
M517    ..........A......G.............
M518    ..........A......G.............
M519    ..............C..............T.
M520    .....T..........A........G.....
M521    ..................G.............
M522    ...C......A....C..G............
M524    ...C......A....C..G............
M525    .....T..........A.....AC.......
M526    ...C......A....C..G............
M527    C...T.........T..G.GTC.C...GT..
M528    C..C..A...A......G.............
M530    C..C..A...A......G.............
M531    C..C..A...A......G.............
M532    ..........A......G.............
M533    ..........A......G.............
M534    C...T........TT..G.GTC.C...GT..
M582    ...........CT....C.....C..T...T
        3333312333333321333333333233333
          *  ****   *   **********  ***
```

FIG. 2. Distribution of the polymorphic bases within the *mdh* gene of pathogenic *E. coli*. The *mdh* sequence of strain M500 is used as the master sequence, and only those nucleotides in the *mdh* gene of other pathogenic strains that differ from the M500 sequence are shown. Nucleotides identical to those of M500 are represented by dots. The position of each base within its codon is indicated by 1, 2, or 3 as shown at the bottom. Informative bases are indicated by asterisks.

represent the common serotypes involved in pathogenesis for each major category of pathogen.

The 10 enzymes used for MLEE in this study were chosen from the 35 enzymes used by the Selander group for the ECOR set (19) for which we had the MLEE data (51a). The 10 enzymes were selected because they were informative and contained no null allele. The neighbor-joining tree for the 72 ECOR set strains resulting from the use of data for the 10 enzymes was the same overall (data not shown) as that resulting from the use of 38 enzymes with the ECOR set strains (19). MLEE was therefore carried out on the 32 pathogenic *E. coli* strains by using appropriate ECOR set strains as controls to allow use of the same numbering system as that described by Ochman and Selander (39), thus allowing the data to be compared directly with published data from the ECOR set. We were then able to construct a tree for the 10 enzymes which includes all ECOR set strains and the pathogenic strains analyzed in this study.

All of the 10 loci were polymorphic in the pathogenic strains as shown in Table 2, the number of allelic states ranging from 2 at PGI, ME, ADK, and G6P to 5 at PGD, with an average of 3.0 states per locus. Most of the allele types found correspond to types already reported in the ECOR set. However, there was one new allele each for the MDH and ME loci, both present in the EHEC strain M534.

Relationship of pathogenic strains to the ECOR set by MLEE. The neighbor-joining tree derived from the MLEE data is shown in Fig. 1. The inclusion of the pathogenic strains led to only minor changes in the inferred relationships of the ECOR set strains. The A, B2, and D groups were still clearly separated, but ECOR48, the most divergent of the group D strains, was now in the group B1 cluster and two B1 strains were separated from the others by group B2 strains. As in the published tree (19), ECOR31, -37, -42, and -43 were outside of the four major groups.

The EPEC, EHEC, ETEC, EIEC, and UTI strains of *E. coli* are clearly distributed among the ECOR set strains in groups A, B1, B2, and D. There was no overall clustering of EPEC, ETEC, EIEC, or UTI strains, all of which were found in at least two of the four groups of the ECOR set; however, the limited set of EHEC strains analyzed were all associated with group D. The Shigella strains grouped together and were mostly found in a single cluster in group A, with Flexneri group 2 and 4 strains in group D.

***mdh* sequence variation.** We sequenced the *mdh* gene, which encodes the enzyme MDH, from 29 of the 32 strains of pathogenic *E. coli* included in the MLEE study. Seventeen distinctive nucleotide sequences were identified among the 29 strains analyzed. Variation was found at 31 sites distributed throughout the 860-nucleotide sites studied of the 942-bp gene. A summary of the distribution of the polymorphic sites is shown in Fig. 2. A comparison of the 29 nucleotide sequences revealed a maximum of 1.81% and an average of 0.75% pairwise difference, which is lower than the average level of 1.1% reported for the ECOR set *E. coli mdh* genes studied previously (2). Pairwise comparison of the 17 distinct nucleotide sequences revealed an average value of 0.84%.

Since 10 of the 15 ECOR set strains previously sequenced by Boyd et al. (2) belong to group D, we sequenced 11 additional ECOR set strains to represent the other groups more adequately, namely, ECOR5, -6, -27, -30, -45, -59, -61, -62, -66, -69, and -37. The sequences of the pathogenic strains, K-12, the 11 ECOR set strains sequenced in this study, and 20 additional strains (including 5 non-ECOR set strains) sequenced by Boyd et al. (2) constitute 61 *E. coli* sequences for comparison. The sequence tree for the ECOR groups A, B2, and D (Fig. 3) was similar to that derived by MLEE, but for group B1, there was only a weak relationship between strains. This is similar to the situation observed for *gapA* (36), *putP* (35), and *mdh* (2) genes. The inclusion of pathogenic strains did not greatly affect the inferred relationships of the 26 ECOR set strains.

The pathogenic strains fall among strains of the ECOR set, with six pathogenic strains, i.e., two EPEC strains (M503 and M522), 3 ETEC strains (M526, M524, and M506), and the EIEC strain M502 found to have a DNA sequence identical to that of K-12 and four group A ECOR set strains, i.e., ECOR5, -6, -10, and -14. The three EHEC strains studied were found to be outside of the four major ECOR groups, with M527 and M509 being identical to non-ECOR strains E3406 and A8190 (both also being O157:H7 EHEC strains [58a]) and M534 being closely related. All Shigella strains studied except Sonnei grouped with group B1 strains. Sonnei was associated with group B2.

DISCUSSION

Previous MLEE studies of *E. coli* have focused on a disease class or serotype (42, 60–63), with each study using an individual numbering system for enzyme mobilities, thereby preventing the inference of the evolutionary relationships between strains from different studies. There have been no studies that have investigated the relationship between groups of pathogenic *E. coli* and nonpathogenic *E. coli* strains. In this study, we

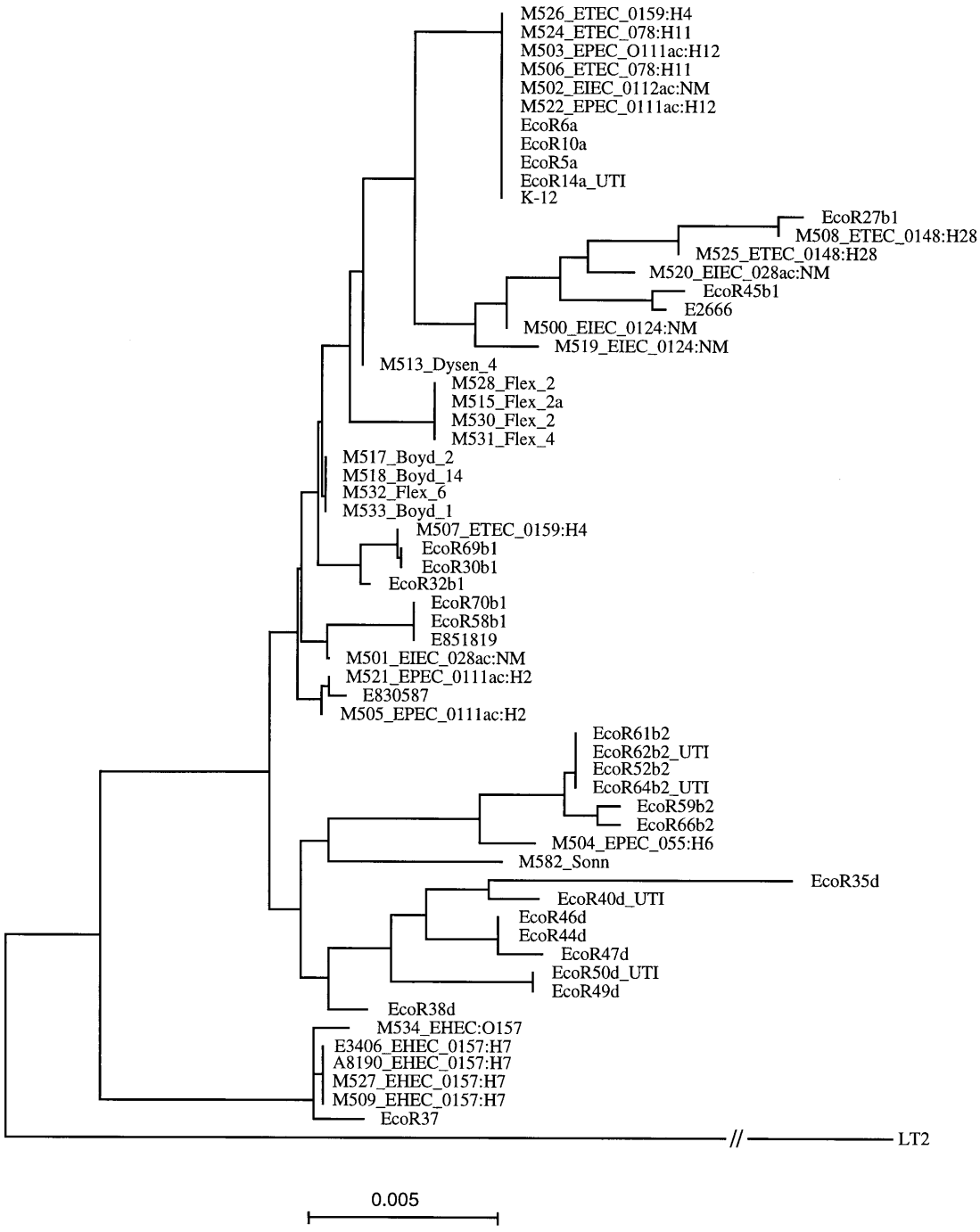

FIG. 3. Evolutionary tree for the *mdh* gene generated by the neighbor-joining method. *Salmonella enterica* LT2 was used as the outgroup.

used MLEE at 10 enzyme loci and the sequence of the chromosomal *mdh* housekeeping gene to look at the relationships between commensal and pathogenic *E. coli* strains. We found that pathogenic strains did not cluster according to their mode of pathogenesis except for Shigella and EHEC clones.

The 13 Shigella strains studied appear to be closely related but clearly within *E. coli*. With MLEE, the majority of Shigella strains (which are human-only pathogens) cluster together within ECOR group A, which comprises mainly human isolates. A second group, containing Shigella clones Flexneri 2, 2a, and 4, stands apart from the other Shigella clones and is closely related to ECOR44 and -47. Note that Flexneri 6 is in the main Shigella cluster. This association is strongly influenced by the sharing of MDH allele 7 by these Flexneri strains and ECOR47. It is interesting that the slow-moving MDH proteins with alleles 5, 6, and 7 are confined to group D. The majority of the 10 Shigella *mdh* sequences are also closely related. M531 (Flexneri 4), M528, M530 (Flexneri 2), and

M515 (Flexneri 2a) had identical *mdh* base sequences, differing from M532 (Flexneri 6) at three sites (including a nonsynonymous change). Three Boydii strains, M518 (Boydii 14), M517 (Boydii 2), and M533 (Boydii 1), had the same sequence as M532 (Flexneri 6). M582 (Sonnei) was the only Shigella sequence not in this cluster falling into group B2. As can be observed in Fig. 1 and 3, differences in mobilities of MDH observed in MLEE do not necessarily show up as changes in our DNA sequences. This is presumably because we sequenced only part of the gene. It should be noted that substitutions at codons 300 and 308 (called codons 289 and 297 by Boyd et al. [2] since they began their count at codon 12) known to give charge changes affecting MLEE were not within the segment we sequenced.

There are some unresolved aspects of the relationship of the Shigella strains of *E. coli*. Based on MLEE, the main Shigella group is with ECOR group A, whereas based on the *mdh* sequence, the main group is with some ECOR group B1 strains. Sonnei is associated with group A by MLEE but with group B2 in the *mdh* gene tree. More data are needed to determine the detailed relationships with other *E. coli* strains. With regards to the differences between Flexneri 6 and the other Flexneri strains, Flexneri 6 has a quite distinctive O antigen (21), but the O antigens of Flexneri 1 to 5 differ only in the presence or absence of specific O-acetyl or glucosyl residues determined by prophages. It would be interesting to determine whether Flexneri 3 and 5 strains group with Flexneri 2, 2a, and 4 since they share the same basic O antigen structure.

In contrast to the Shigella strains, the pathogenic strains traditionally recognized as being in *E. coli* are spread throughout the MLEE and *mdh* sequence trees, and even the EIEC strains, the strains most similar to Shigella strains in their pathogenesis and cross-reacting with traditional Shigella typing antisera (26), are present in several clusters, many including other pathogenic and nonpathogenic *E. coli*. It appears that while the Shigella strains form a related group of *E. coli* strains, the other pathogenic forms (with perhaps the exception of EHEC) do not comprise separate groups of related strains and appear to have arisen through the independent acquisition of virulence factors.

There have been a number of studies that have assessed the relationships of the EHEC O157:H7 clone and its relatedness to other pathogenic *E. coli* (60, 61, 63). The 32 pathogenic strains in our study include three O157 EHEC strains, two of which are H7 and one of which (M534) is not recorded. They are clustered by MLEE with ECOR42 and ECOR37. Strains E3406 and A8190 which were used in the Boyd et al. (2) study had *mdh* sequences identical to those of M527 and M509. The four O157:H7 strains are probably from the major HUS clone. The study of Whittam et al. (63) shows that the O157:H7 clone was not found to be closely related to other O157 strains.

The relationship of the O157:H7 clone to ECOR37 and ECOR42 strongly suggests the possibility of EHEC strains emerging from commensal *E. coli*, with pathogenic strains gaining virulence genes such as those for the A/E phenotype and SLT. It has previously been suggested by Whittam et al. (63) that the EHEC O157:H7 clone has been derived from an O55:H7-like ancestor. It would be interesting to determine the relationships of the O55:H7 strain reported by Whittam et al. (63) to the ECOR37 and ECOR42 strains. The other O157 EHEC strain, M534, although grouped with the O157:H7 strains both in MLEE and *mdh* trees, is more distant than expected for an O157:H7 derivative. It differs significantly from the others by two enzyme loci, namely, MDH and ME, both using L-malate as a substrate for reactions in the central metabolic pathways; its *mdh* sequence differs from that of the other four EHEC strains sequenced by one nonsynonymous base change. The meaning of this substantial difference is not clear.

A number of pathogenic strains were found to have electrophoretic types identical to those of ECOR set strains and identical or closely related to those of other pathogenic strains with different modes of pathogenesis, e.g., M502 (EIEC), M526 (ETEC), and ECOR5. There were six pathogenic strains, i.e., M502 (EIEC; O112ac:NM), M503 (EPEC; O111:H12), M526 (ETEC; O159:H4), M522 (EPEC; O111ac:H12), M524 (ETEC; O78:H11), and M506 (ETEC; O78:H11), which had the same *mdh* sequence as five ECOR group A strains, including K-12. The presence of pathogenic strains having the same *mdh* sequence but different pathogenic modes and O antigens indicates extensive lateral gene transfer for virulence genes and O antigen genes. Strains belonging to the same serotype did not necessarily fall into the same cluster nor were clones belonging to the same disease category monophyletic; this is in accord with the findings of Whittam et al. (63).

The major conclusion of our study is that pathogenic strains of *E. coli* do not have a single evolutionary origin within *E. coli* but have arisen several times. The presence of specific virulence factors associated with particular forms of pathogenesis may be due to lateral transfer of such properties, some of which are on plasmids and bacteriophages and which can be subject to very strong natural selection.

The genus Shigella, which has often been suggested to be part of *E. coli*, is shown to comprise a group of closely related pathogenic *E. coli* strains, whereas the EPEC, ETEC, EIEC, or UTI pathogenic strains do not form discrete groups. The EHEC strains studied probably belong to the major O157:H7 clone (61), and we cannot say if other EHEC clones will be related to them. It is interesting that the Shigella strains, traditionally divided into four species, form one cluster within *E. coli* and are far less diverse than EPEC, ETEC, EIEC, or UTI strains. These data provide compelling support for the inclusion of Shigella strains within *E. coli*; indeed, it is not possible to discuss the data sensibly on any other basis.

ACKNOWLEDGMENTS

We thank all the institutions listed in Table 1 for kindly supplying strains.

This work was supported by a grant from the Australian Research Council.

REFERENCES

1. **Bastin, D. A., L. K. Romana, and P. R. Reeves.** 1991. Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain. Mol. Microbiol. **5:**2223–2231.
2. **Boyd, E. F., K. Nelson, F.-S. Wang, T. S. Whittam, and R. K. Selander.** 1994. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc. Natl. Acad. Sci. USA **91:**1280–1284.
3. **Bray, J.** 1945. Isolation of antigenically homogeneous strains of *Bact. coli neapolitanum* from summer diarrhea of infants. J. Pathol. Bacteriol. **57:**239–247.
4. **Bray, J., and T. E. D. Beaven.** 1948. Slide agglutination of *Bact. coli var neapolitan* in summer diarrhea. J. Pathol. Bacteriol. **60:**395–401.
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, F. J. Skerman, and S. Falkow.** 1972. Polynucleotide sequence divergence among strains of *Escherichia coli* and closely related organisms. J. Bacteriol. **109:**953–965.
7. **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.
8. **Cordovez, A., A. Prado, L. Maggi, J. Cordero, J. Martinez, A. Misraji, R. Rios, G. Soza, A. Ojeda, and M. M. Levine.** 1992. Enterohemorrhagic *Escherichia coli* associated with hemolytic-uremic syndrome in Chilean children. J. Clin. Microbiol. **30:**2153–2157.
9. **Donnenberg, M. S., and J. B. Kaper.** 1992. Enteropathogenic *Escherichia coli*: minireview. Infect. Immun. **60:**3951–3961.
10. **Donnenberg, M. S., S. Tzipori, M. L. McKee, A. D. O'Brien, J. Alroy, and J. B. Kaper.** 1993. The role of the *eae* gene of enterohemorrhagic *Escherichia*

coli in intimate attachment in vitro and in a porcine model. J. Clin. Invest. **92:**1418–1424.
11. **Ekback, G., B. Morner, S. Lund, and S. Normark.** 1986. Correlation of genes in the *pap* cluster to expression of globoside-specific adhesin by uropathogenic *Escherichia coli*. FEMS Microbiol. Lett. **34:**355–360.
12. **Elliot, E., Z. Li, D. Stiel, A. Buret, J. Wallace, I. Brzusczak, and E. O'Loughlin.** 1994. Modulation of host response to *Escherichia coli* O157:H7 infection by anti-CD18 antibody in rabbits. Gastroenterology **106:**1554–1561.
13. **Escherich, T.** 1885. Die darmbakterien des neugeboren und sauglingen. Fortschr. Med. **3:**515–528.
14. **Ewing, W. H.** 1953. Serological relationships between shigella and coliform cultures. J. Bacteriol. **66:**333–340.
15. **Fasano, A., A. K. Bradford, R. B. Russel, D. R. Maneval, Jr., and M. M. Levine.** 1990. Enterotoxin and cytotoxin production by enteroinvasive *Escherichia coli*. Infect. Immun. **58:**3717–3723.
16. **Giles, C., and M. B. Sangster.** 1948. An outbreak of infantile gastroenteritis in Aberdeen. J. Hyg. **46:**1–9.
17. **Goullet, P.** 1980. Esterase electrophoretic pattern between *Shigella* species and *Escherichia coli*. J. Gen. Microbiol. **117:**493–500.
18. **Hartl, D. L., and D. E. Dykhuizen.** 1984. The population genetics of *Escherichia coli*. Annu. Rev. Genet. **18:**31–68.
19. **Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam.** 1990. Phylogenetic distribution of branch RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172:**6175–6181.
20. **Karaolis, D. K. R., R. Lan, and P. R. Reeves.** 1994. Sequence variation in *Shigella sonnei* (Sonnei), a pathogenic clone of *Escherichia coli*, over four continents and 41 years. J. Clin. Microbiol. **32:**796–802.
21. **Katzenellenbogen, E., and E. Romanowska.** 1980. Structural studies on *Shigella flexneri* serotype 6 core region. Eur. J. Biochem. **113:**205–211.
22. **Kimura, M.** 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **116:**111–120.
23. **Knutton, S., T. Baldwin, P. H. Williams, and A. S. McNeish.** 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.
24. **Konawalchuk, J., J. I. Spiers, and S. Stavric.** 1977. Vero response to a cytotoxin of *Escherichia coli*. Infect. Immun. **18:**775–779.
25. **Lawrence, J. G., H. Ochman, and D. L. Hartl.** 1991. Molecular and evolutionary relationships among enteric bacteria. J. Gen. Microbiol. **137:**1911–1921.
26. **Levine, M. M., and R. Edelman.** 1984. Enteropathogenic *Escherichia coli* of classical serotypes associated with infant diarrhea: epidemiology and pathogenesis. Epidemiol. Rev. **6:**31–51.
27. **Levine, M. M., J. B. Kaper, R. E. Black, and E. Al.** 1983. New knowledge on pathogenesis of bacterial enteric infections as applied to vaccine development. Microbiol. Rev. **47:**510–550.
28. **Lund, B., F. Lindberg, B.-I. Marklund, and S. Normark.** 1987. The papG protein is the α-D-galactopyranosyl-(1→4)-β-D-galactopyranose-binding adhesion of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA **84:** 5898–5902.
29. **Lund, B., B.-I. Marklund, N. Stromberg, F. Lindberg, K.-A. Karlsson, and S. Normark.** 1988. Uropathogenic *Escherichia coli* can express serologically identical pili of different receptor binding specificities. Mol. Microbiol. **2:**255–263.
30. **McDaniel, T. K., K. G. Jarvis, M. S. Donnenberg, and J. B. Kaper.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc. Natl. Acad. Sci. USA **92:**1664–1668.
31. **Milkman, R.** 1973. Electrophoretic variation in *E. coli* from natural sources. Science **182:**1024–1026.
32. **Milkman, R., and M. McKane.** 1995. DNA sequence variation and recombination in *E. coli*, p. 127–142. *In* S. Baumberg, J. P. W. Young, E. M. H. Wellington, and J. R. Saunders (ed.), Population genetics of bacteria. Symposium 52 of the Society for General Microbiology. Cambridge University Press, Melbourne, Australia.
33. **Moon, H. W., S. C. Whipp, M. Argenzio, M. Levine, and R. A. Gianella.** 1983. Attaching and effacing activities of rabbit and human enteropathogenic *Escherichia coli* in pig and rabbit intestines. Infect. Immun. **41:**1340–1351.
34. **Nasoff, M. S., H. V. Baker II, and R. E. Wolf, Jr.** 1984. DNA sequence of the *Escherichia coli* gene, *gnd*, for the 6-phosphogluconate dehydrogenase. Gene **27:**253–264.
35. **Nelson, K., and R. K. Selander.** 1992. Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J. Bacteriol. **174:**6886–6895.
36. **Nelson, K., T. S. Whittam, and R. K. Selander.** 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc. Natl. Acad. Sci. USA **88:**6667–6671.
37. **Neter, E., O. Westphal, O. Luderitz, and E. Al.** 1955. Demonstration of antibodies against enteropathogenic *Escherichia coli* in sera of children of various ages. Pediatrics **16:**801–807.
38. **O'Brien, A. D., A. R. Melton, C. K. Schmitt, M. L. McKee, M. L. Batts, and D. E. Griffin.** 1993. Profile of *Escherichia coli* O157:H7 pathogen responsible for hamburger-borne outbreak of hemorrhagic colitis and hemolytic-uremic syndrome in Washington. J. Clin. Microbiol. **31:**2799–2801.
39. **Ochman, H., and R. K. Selander.** 1984. Evidence for clonal population structure in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81:**198–201.
40. **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.
41. **Ørskov, F., and I. Ørskov.** 1984. Serotyping of *Escherichia coli*. Methods Microbiol. **14:**43–112.
42. **Ørskov, F., T. S. Whittam, A. Cravioto, and R. K. Selander.** 1990. Clonal relationships among classic enteropathogenic *Escherichia coli* (EPEC) belonging to different O groups. J. Infect. Dis. **162:**76–81.
43. **Ostroff, S. M., P. I. Tarr, M. A. Neill, J. H. Lewis, N. Hargrett-Bean, and J. M. Kobayashi.** 1989. Toxin genotypes and plasmid profiles as determinants of systematic sequelae in *E. coli* O157:H7 infection. J. Infect. Dis. **160:**994–998.
44. **Polotsky, Y. E., E. M. Dragunskaya, V. G. Seliverstova, T. A. Avdeeva, M. G. Chakhutinskaya, I. Ketyi, A. Vertenyi, B. Ralovich, L. Emody, I. Malovics, N. V. Safonova, E. S. Snigirevskaya, and E. I. Karayagina.** 1977. Pathogenic effect of enterotoxigenic *Escherichia coli* and *Escherichia coli* causing infantile diarrhea. Acta Microbiol. Acad. Sci. Hung. **24:**221–236.
45. **Reeves, P. R., L. Farnell, and R. Lan.** 1994. MULTICOMP: a program for preparing sequence data for phylogenetic analysis. CABIOS **10:**281–284.
46. **Reisner, A. H., C. A. Bucholtz, J. Smelt, and S. McNeil.** 1993. Australia's National Genomic Information Service, p. 595–602. *In* Proceedings of the 26th Annual Hawaii International Conferences on Systems Science.
47. **Richardson, B. J., P. R. Baverstock, and M. Adams.** 1986. Allozyme electrophoresis—a handbook for animal systematics and population studies. Academic Press, Sydney, Australia.
48. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
49. **Sansonetti, P. J., H. d'Hauteville, C. Ecobichon, and C. Pourcel.** 1983. Molecular comparison of virulence plasmids in *Shigella* and enteroinvasive *Escherichia coli*. Ann. Microbiol. (Inst. Pasteur) **134A:**295–318.
50. **Sansonetti, P. J., D. J. Kopecko, and S. B. Formal.** 1981. *Shigella sonnei* plasmids: evidence that a large plasmid is necessary for virulence. Infect. Immun. **34:**75–83.
51. **Sansonetti, P. J., D. J. Kopecko, and S. B. Formal.** 1982. Involvement of a plasmid in the invasive ability of *Shigella flexneri*. Infect. Immun. **35:**852–860.
51a. **Selander, R. K.** Personal communication.
52. **Selander, R. K., T. K. Korhonen, V. Vaisanen-Rhen, P. H. Williams, P. E. Pattison, and D. A. Caugant.** 1986. Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. Infect. Immun. **52:**213–222.
53. **Shermann, P., R. Soni, and M. Karmali.** 1988. Attaching and effacing adherence of Vero cytotoxin-producing *Escherichia coli* to rabbit intestinal epithelium in vivo. Infect. Immun. **56:**756–761.
54. **Shiga, K.** 1898. Ueber den Dysenterie-bacillus (Bacillus dysenteriae). Zentralbl. Bakteriol. (Orig.) **24:**913–918.
55. **Stevenson, G., B. Neal, D. Liu, M. Hobbs, N. H. Packer, M. Batley, J. W. Redmond, L. Lindquist, and P. R. Reeves.** 1994. Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. J. Bacteriol. **176:**4144–4156.
56. **Taylor, J., B. W. Powell, and J. Wright.** 1949. Infantile diarrhea and vomiting: a clinical and bacteriological investigation. Br. Med. J. **2:**117–141.
57. **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.
58. **Welch, R. A., R. Hull, and S. Falkow.** 1983. Molecular cloning and physical characterization of a chromosomal hemolysin from *Escherichia coli*. Infect. Immun. **42:**178–186.
58a. **Whittam, T. S.** Personal communication.
59. **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.
60. **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.
61. **Whittam, T. S., and R. A. Wilson.** 1988. Genetic relationships among pathogenic *Escherichia coli* of serogroup O157. Infect. Immun. **56:**2467–2473.
62. **Whittam, T. S., and R. A. Wilson.** 1988. Genetic relationships among pathogenic strains of avian *Escherichia coli*. Infect. Immun. **56:**2458–2466.
63. **Whittam, T. S., M. L. Wolfe, I. K. Wachsmuth, F. Ørskov, I. Ørskov, and R. A. Wilson.** 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. Infect. Immun. **61:**1619–1629.

Editor: A. O'Brien