INFECTION AND IMMUNITY, June 1999, p. 2692–2699 Vol. 67, No. 6
0019-9567/99/$04.00+0

Phylogenetic Analysis of Enteroaggregative and Diffusely Adherent *Escherichia coli*

JOHN R. CZECZULIN,[1] THOMAS S. WHITTAM,[2] IAN R. HENDERSON,[1] FERNANDO NAVARRO-GARCIA,[1,3] AND JAMES P. NATARO[1]*

Departments of Pediatrics and Microbiology and Immunology, Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, Maryland 21201[1]; Department of Biology, Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, Pennsylvania 16802[2]; and Department of Public Health, Faculty of Medicine, UNAM, 04510 Mexico DF, Mexico[3]

Received 7 January 1999/Returned for modification 3 March 1999/Accepted 16 March 1999

The phylogenetics of the various pathotypes of diarrheagenic *Escherichia coli* are not completely understood. In this study, we identified several plasmid and chromosomal genes in the pathogenic enteroaggregative *E. coli* (EAEC) prototype strain 042 and determined the prevalence of these loci among EAEC and diffusely adherent *E. coli* strains. The distribution of these genes is analyzed within an evolutionary framework provided by the characterization of allelic variation in housekeeping genes via multilocus enzyme electrophoresis. Our data reveal that EAEC strains are heterogeneous with respect to chromosomal and plasmid-borne genes but that the majority harbor a member of a conserved family of virulence plasmids. Comparison of plasmid and chromosomal relatedness of strains suggests clonality of chromosomal markers and a limited transfer model of plasmid distribution.

Diarrheagenic *Escherichia coli* strains are major pathogens associated with enteric disease in many parts of the world. Currently, five pathotypes of diarrheagenic *E. coli* have been unequivocally associated with diarrheal illness: enterotoxigenic *E. coli*, enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC) and enteroaggregative *E. coli* (EAEC) (reviewed in reference 26). A sixth *E. coli* pathotype, diffusely adherent *E. coli* (DAEC), has been proposed on the basis of epidemiologic studies; however, unlike the other five categories, in which outbreaks and volunteer studies have identified truly pathogenic isolates, there are no such data for DAEC strains and their enteric pathogenicity is in question (26).

EAEC may be an emerging diarrheal pathogen. This pathotype, defined by aggregative adherence (AA) to HEp-2 cells in culture, has been associated characteristically with persistent diarrhea among infants, particularly in the developing world (3, 4, 9, 28, 43). However, recent outbreaks and volunteer studies suggest that EAEC strains are virulent in adults (29, 40) and have a global distribution (17, 18).

Volunteer studies suggest that the virulence of EAEC strains for humans is genetically and phenotypically heterogeneous (29). The factors conferring this heterogeneity have yet to be characterized, but several candidate EAEC virulence factors have been identified; none of these factors are present in all EAEC strains. We have shown that most EAEC strains harbor a 60 to 65-MDa plasmid which may encode the AA fimbria AAF/I or AAF/II (10, 25, 30, 42) and, in some cases, the enterotoxin EAST1 (36). Recently, Eslava et al. cloned and sequenced a 104-kDa enterotoxin (called the plasmid-encoded enterotoxin, or Pet), which is also encoded on the AA plasmid of some EAEC strains (13). Yet despite the fact that the large plasmids of EAEC are heterogeneous with regard to fimbriae and toxin expression, Baudry et al. (2) isolated a plasmid-derived fragment which hybridizes with many EAEC strains; thus, the degree of conservation of the AA plasmids is an important and unresolved question. Indeed, a highly conserved plasmid could encode a large number of virulence factors that may provide important clues to the pathogenic mechanism of EAEC, as well as to its phylogenic origins. Moreover, it has been suggested that DAEC strains should be categorized within the EAEC pathotype (28); indeed, the predominant adhesins of DAEC strains are related to fimbrial adhesins encoded on EAEC plasmids (10).

We undertook this study to develop a phylogenetic framework for EAEC and DAEC strains with the aims of understanding their natural history and their pathogenesis, as well as to illuminate overall aspects of *E. coli* phylogeny. A collection of strains from various epidemiologic studies was selected and subjected to multilocus enzyme electrophoresis (MLEE). In addition, we performed colony blot hybridization using a selection of chromosomal and plasmid-borne genes which we and others have identified. This study was made possible by analysis of the prototype AA plasmid derived from one proven human pathogenic EAEC strain (strain 042). The study of EAEC AA plasmid loci demonstrates that many EAEC strains harbor a partially conserved plasmid and that this plasmid has undergone significant horizontal dissemination among EAEC clones.

MATERIALS AND METHODS

Strains. Prototype EAEC strain 042 was isolated from a child with diarrhea in the course of an epidemiologic study in Lima, Peru, in 1983 (23); this strain elicited diarrhea in adult volunteers (29). EAEC strains used in colony blot analysis were from the collection of the Center for Vaccine Development and were isolated from epidemiologic studies in various sites throughout the developing world. Strain 101-1 was implicated in a large outbreak of EAEC diarrhea among schoolchildren in Japan (18) and was obtained from Y. Itoh. Strain C1096 was implicated in an outbreak of EAEC diarrhea in a nursery in Serbia (7) and was obtained from M. Cobeljic. Strain RD8 is a Shiga toxin-producing EAEC strain that was implicated in an outbreak of hemolytic-uremic syndrome in France (21) and was obtained from A. Caprioli. *Shigella flexneri* YSH6000 (35) was obtained from S. Austin. *E. coli* HB101 (6) and DH5α (1) were used as recipient strains for genetic manipulations. All strains were stored at −70°C in Trypticase soy broth with 15% glycerol. Strains were routinely passed on Luria-Bertani broth or agar with ampicillin (200 μg/ml) where appropriate.

* Corresponding author. Mailing address: Center for Vaccine Development, 685 W. Baltimore St., Baltimore, MD 21201. Phone: (410) 706-8442. Fax: (410) 706-6205. E-mail: jnataro@umaryland.edu.

TABLE 1. Gene probes used in this study

Gene or probe	Description of target (reference)	Location of gene	Size of probe	PCR primers or restriction fragment[a] (reference)
aggA	AAF/I fimbrial subunit (37)	Plasmid	450 bp	(a) TTAGTCTTCTATCTAGGG (b) AAATTAATTCCGGCATGG
shf	Cryptic ORF (this study)	Plasmid	613 bp	(a) ACTTTCTCCCGAGACATTC (b) CTTTAGCGGGAGCATTCAT
aspU	Cryptic secreted protein (25)	Plasmid	232 bp	(a) CTTTTCTGGCATCTTGGGT (b) GTAACAACCCCTTTGGAAGT
aggR	Transcriptional activator of AAFs (24)	Plasmid	308 bp	(a) CTAATTGTACAATCGATGTA (b) ATGAAGTAATTCTTGAAT
she	Secreted protein of *S. flexneri* (33)	Chromosome	1175 bp	(a) GGGTATTGTCCGTTCCGAT (b) ACAACGATACCGTCTCCCG
sigA	Autotransporter homolog of *S. flexneri* chromosome (33)	Chromosome	2.5 kb	(a) AATGTCAGGGTTCCGGCCAG (b) CAGCCCGGGCCGCGTTTGCG
irp2	Yersiniabactin biosynthetic gene (38)	Chromosome	264 bp	(a) AAGGATTCGCTGTTACCGGAC (b) TCGTCGGGCAGCGTTTCTTCT (38)
aafA	AAF/II fimbrial subunit (10)	Plasmid	770 bp	*Eco*RI/*Pst*I fragment from pJC2 (10)
CVD432	AA probe (2)	Plasmid	0.7 kb	*Eco*RI/*Pst*I fragment from pCVD432 (2)
pet	104-kDa enterotoxin (13)	Plasmid	2.2 kb	*Pst*I/*Apa*I fragment pJPN205 (13)

[a] Primers (upstream [a] and downstream [b]) are written 5′→3′. Primers not referenced were derived in this study.

Molecular cloning. All genetic manipulations were performed by standard methods (1). Plasmid DNA was extracted by using a Plasmid Midi kit (Qiagen Inc., Chatsworth, Calif.). Purification of DNA fragments and extraction from agarose gel slices were performed with Geneclean (Bio 101, La Jolla, Calif.). Plasmid DNA was introduced into *E. coli* HB101 or DH5α by transformation of competent cells (obtained from Gibco/BRL, Gaithersburg, Md.) according to the manufacturer's instructions.

Construction of a pAA2 library. Purified plasmid DNA from strain 042 was sonicated to generate fragments 1 to 5 kb in length. After separation in a 0.7% agarose gel, 1.5 to 2.5-kb fragments were eluted from the gel, and ends were filled with *Pfu* DNA polymerase as instructed by the manufacturer (Stratagene). Fragments were ligated into *Sma*I-digested pBluescript. Ligated DNA was transformed into DH5α, selecting for ampicillin resistance. A pAA2 cosmid library in vector pCVD301 (19) has been described previously (13). The pAA2 plasmid map was derived by restriction enzyme analysis of the parent plasmid as well as insert DNA from several of the pCVD301 cosmid clones.

DNA sequence analysis. Nucleotide sequences were determined in the Biopolymer Core Laboratory (Department of Microbiology and Immunology, University of Maryland School of Medicine), with an Applied Biosystems model 373A automated sequencer via dye terminator cycle sequencing with *Taq* polymerase (Perkin-Elmer Corp., Norwalk, Conn.) according to manufacturer's instructions. Each sequence was derived by assembly of contigs with the SEQUENCHER 3.0 for Macintosh program (Gene Codes Corporation, Ann Arbor, Mich.) and was analyzed with GENEPRO sequence analysis software (version 5.00; Riverside Scientific, Bainbridge Island, Wash.) and the Wisconsin Genetics Computer Group sequence analysis package (available through the Center of Marine Biotechnology, University of Maryland). The predicted amino acid sequence of each open reading frame (ORF) was compared with protein sequences listed in EMBL/GenBank by using the Genetics Computer Group TFASTA program and the BLAST algorithms (National Center for Biotechnology Information).

Nucleic acid hybridization studies. Colony blot hybridization was performed by standard methods (1) on blots prepared by methods previously described (14); 25 ng of probe fragment was labeled by random primer extension (Prime-it kit; Stratagene Cloning Systems, La Jolla, Calif.) with 50 μCi of [^{32}P]dATP (Amersham Life Science Products, Arlington Heights, Ill.). Southern hybridization was performed by standard methods (1). PCR primers used to derive the DNA probes are described in Table 1. Restriction fragment length polymorphism (RFLP) analysis was performed with probes consisting of cosmids which carried the regions encoding AAF/I (cosmid pJPN31 from strain 17-2) (25) and AAF/II (cosmid D6 from strain 042). The CVD432 probe is the original AA probe described by Baudry et al. (2) and was derived from plasmid pCVD432.

HEp-2 adherence assay. The HEp-2 adherence assay was performed for a 3-h incubation as initially described by Cravioto et al. (8). AA and DA (diffuse adherence) patterns were interpretated as described by Nataro et al. (27).

MLEE clonal analysis. Genetic variation was analyzed by horizontal starch gel electrophoresis to detect protein polymorphism for housekeeping enzymes (39). Electromorphs were compared to standard mobility variants and assigned mobility ranks by the rate of migration. Isolates lacking enzyme activity were designated null for the particular locus. Every strain was characterized by its multilocus profile of electromorphs, or alleles, for the 20 enzyme-encoding loci (45). Each distinctive allelic array was designated an electrophoretic type (ET) (39). Phylogenetic relationships between ETs were determined based on a matrix of genetic distances between all pairs constructed by comparison of the allelic arrays. The neighbor-joining algorithm (34) was used to construct a dendrogram with the computer program MEGA (20).

Nucleotide sequence accession number. The region downstream of the *aafC* gene (Fig. 1) has been deposited in GenBank under accession no. AF134403.

RESULTS

Identification of plasmid-borne loci. We derived an *Mlu*I map of the prototype AA plasmid pAA2 (Fig. 1). By Southern hybridization, several previously described EAEC plasmid genes were localized, including a gene cluster encoding the Pet and EAST1 enterotoxins (*pet* and *astA* genes, respectively), as well as the AAF/II-related genes, *aafA* (the fimbrial subunit) and *aggR* (the fimbrial gene regulator). The sequence upstream of the fimbrial gene activator *aggR* contained a gene, which we have termed *aspU* (EAEC secreted protein U), that lies 821 bp upstream and in the same orientation as *aggR* (Fig. 1). *aspU* is 99% identical to an ORF partially sequenced from strain 17-2 (called ORF1 in reference 24), and is linked to *aggR* in that strain as well. *aspU* encodes the previously described 14-kDa secreted protein of EAEC strain 17-2 (25), but a role for this protein has not been elucidated. Notably, strain 17-2 expresses the AAF/I fimbrial antigen (the product of the *aggA* gene), whereas 042 expresses AAF/II but not AAF/I; both of these fimbriae are under *aggR* control (10, 24). AAF/II genes are organized as two distinct clusters (Fig. 1): region 1, comprising *aafA* and *aafR* as well as the chaperone *aafD*; and region 2, which contains the usher protein *aafC* (12).

Sequencing of DNA downstream of *aafC* (counterclockwise on fragment G in Fig. 1) revealed a cluster of three ORFs transcribed in the same direction and each ca. 93% identical (at the amino acid level) to a similar cluster located on the large plasmid of *S. flexneri* (32). The last gene of this cluster (and the one closest to *aafC*) is homologous to *virK* (GenBank accession no. D11025), a gene which encodes a 316-amino-acid (36.8-kDa) protein which has been suggested to be a posttranscriptional regulator of *virG* expression (22). Notably, however, hybridization of 042 DNA by using an *S. flexneri virG* fragment probe did not yield a signal, suggesting that plasmid pAA2 does not encode VirG (data not shown). The first gene of our *virK* cluster is predicted to encode a protein of 280 amino acids (predicted molecular mass of 32.8 kDa). Interestingly, the ho-

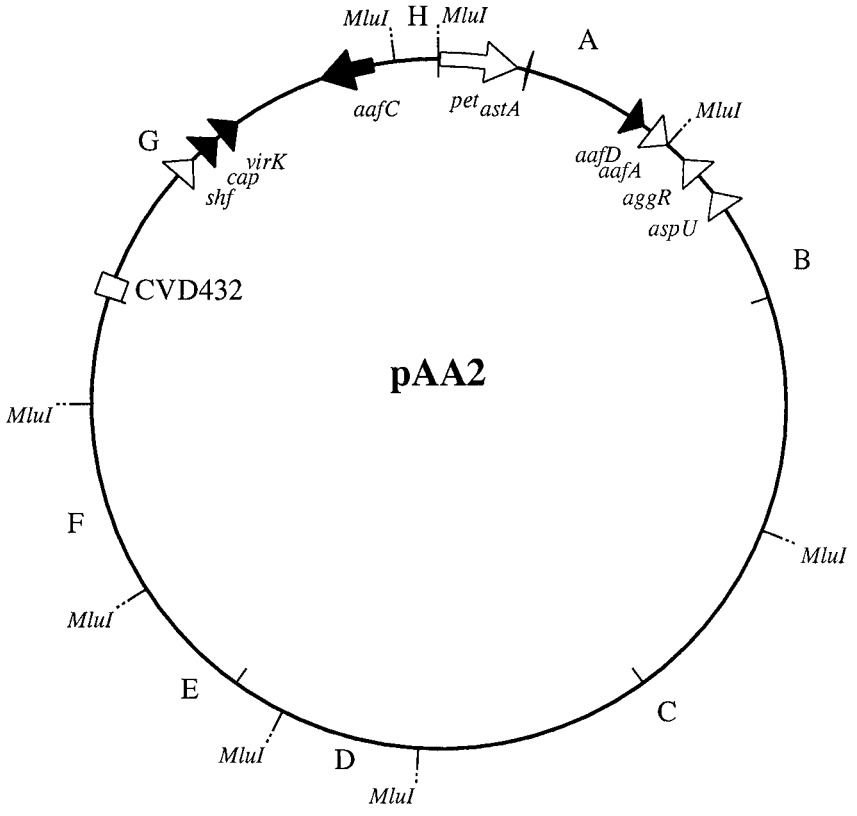

FIG. 1. *Mlu*I restriction map of the 100-kb plasmid pAA2 from strain 042. The positions of plasmid-borne probes used in this work are indicated by open symbols; the positions of other genes mentioned in the text are indicated by solid symbols. Inner tick marks are 20 kb apart. Uppercase letters represent specific *Mlu*I fragments.

mologous DNA in *S. flexneri* is frameshifted and therefore may not encode such a protein; the two ORFs in *S. flexneri*, separated by a single frameshift, are designated *shf1* and *shf2* in GenBank (accession no. AF012082), and we have therefore adopted the designation *shf* to refer to our single ORF. Notably, a homolog of this ORF is also present on the *E. coli* O157:H7 60-MDa plasmid (accession no. AF043470). No function is known for the predicted Shf proteins, but the closest well-characterized homolog (at 25% amino acid identity) is the IcaB protein of *Staphylococcus epidermidis*, a protein implicated in intercellular adhesion (15). Between the *shf* and *virK* ORFs lies another ORF which, although similar in *S. flexneri*, has not been reported previously. The predicted product of this ORF, 251 amino acids (29.0 kDa) in length in both *S. flexneri* and EAEC, would encode a protein 50% identical to an *rfbU*-related lipopolysaccharide biosynthetic gene of *E. coli* O157:H7 (accession no. AB011549). Both genes exhibit motifs which we have found to be specific for hexosyltransferase enzymes (16).

Identification of chromosomally encoded loci. Strain 042 has been shown to secrete into the supernatant two high-molecular-mass proteins of 104 and 116 kDa (13). The 104-kDa protein is the product of the *pet* gene, which encodes a plasmid-encoded autotransporter enterotoxin (13). We determined the N-terminal sequence of the 116-kDa secreted protein of strain 042 and found that it is identical to the sequence of the predicted protein product of the *she* gene of *S. flexneri* (GenBank accession no. U35656) (33). This gene is notable in that the *Shigella* enterotoxin 1 gene is encoded on the antisense strand. We derived PCR primers from the reported sequence of the *she* gene (Table 1) and amplified a 1,175-bp product from *S. flexneri* 2457T. This fragment was then used as a DNA probe in colony blot and Southern blot hybridization. Strains 2457T and 042 each yielded strong hybridization signals in both colony blot and Southern blot analyses of restricted genomic DNA. However, neither *E. coli* HB101 nor *E. coli* HB101(pAA2) yielded a signal with this *she* probe. Moreover, the *she* PCR primers yielded the same amplification product from 042 template DNA as from 2457T DNA but yielded no PCR product from HB101 or HB101(pAA2) DNA. Therefore, we concluded that a *she* homolog was encoded on the chromosome of EAEC strain 042.

In their description of the *she*-encoding pathogenicity island in the *S. flexneri* chromosome, Rajakumar et al. (33) also reported the partial sequence of another autotransporter-encoding gene located approximately 4 kb downstream from the *she* gene; this gene was designated *sigA*. To determine whether the complete chromosomal island of *S. flexneri* is also present in EAEC, we derived primers from the available partial sequence of *sigA* (accession no. U97487) and amplified a fragment from *Shigella* strain YSH6000. This fragment hybridized with parent *S. flexneri* strain YSH6000 but did not hybridize with 042 DNA. These data suggested that whereas both *S. flexneri* and EAEC strain 042 carried the *she* gene, the overall organizations of the purported pathogenicity islands are dissimilar.

Schubert et al. (38) have recently reported that 93% of EAEC hybridized with a probe derived from the *irp2* gene, which encodes a protein necessary for yersiniabactin biosynthesis in *Yersinia* species and which is part of the high-pathogenicity island. Using published primers derived from the *irp2* sequence, we generated a PCR product from *Yersinia enterocolitica* and found that genomic DNA of 042, but not HB101

(pAA2), hybridized with this *irp2* fragment, suggesting that this gene is indeed located on the chromosome of strain 042.

Chromosomal phylogeny studies. Having established the presence of several plasmid-borne and chromosomal loci among EAEC strains, we examined the prevalence of these genes on a defined phylogenic map. These gene probes used are listed in Table 1. MLEE for 20 chromosomally encoded enzymes was performed on a preselected collection of EAEC and DAEC strains. These strains were chosen to represent well-characterized strains from diverse locations and serotypes. Roughly three-fourths of the EAEC strains that we selected were CVD432 probe positive, approximately the same proportion of probe positivity for all of the strains in our collection. Interestingly, five originally selected CVD432 probe-negative strains were found not to be *E. coli* (by MLEE and biochemical analysis), and two were found not to be HEp-2 adherent upon follow-up testing after the MLEE analysis. These strains were excluded from analysis. Results are shown in Fig. 2.

The phylogenetic analysis of EAEC and DAEC strains reveals five large clusters of strains (Fig. 2). Two of the clusters were found to comprise largely EAEC strains (designated EAEC1 and EAEC2), two contained predominantly DAEC strains (DAEC1 and DAEC2), and one cluster contained strains of both pathotypes. Although the strains were found in characteristic clusters, the overlapping nature of the phylogeny was notable: in addition to the combined AA/DA cluster, both EAEC clusters contained DAEC strains, and DAEC2 contained one EAEC strain. The presence of multiple discrete clusters of strains from a single pathotype is reminiscent of EPEC and EHEC strains; these clusters are illustrated as shaded boxes in Fig. 2. Several previously characterized diarrheagenic *E. coli* clusters (45) are also superimposed on our phylogram.

Despite the overlapping nature of DAEC and EAEC strains, hybridization with the chromosomal *she* locus was restricted to EAEC strains. Not only was hybridization with the *she* probe characteristic of strains located only in EAEC1, EAEC2, and AA/DA, but within those clusters, *she*-homologous sequence were restricted to EAEC strains. In contrast, hybridization with the *irp2* probe was characteristic of both EAEC and DAEC clusters.

Analysis of EAEC plasmids. The presence of EAEC plasmid-specific sequences was highly correlated with EAEC strains; in only one case was a DAEC strain positive for any of the plasmid probes (DA E1058B in EHEC1). In addition to being specific for EAEC strains, we found that AA plasmid-borne genes were highly linked to each other. Of the 44 strains displaying the AA phenotype (and therefore defined as EAEC), only 2 failed to hybridize with any of the AA plasmid probes (Table 2). The high correlation of CVD432, *aggR* and *aspU* (widely separated on the pAA2 restriction map) is the best evidence for the existence of a conserved EAEC plasmid family.

Various plasmid-borne markers were found to segregate with each other (Table 2). *pet* and AAF/II, for example, although each present in only a minority of strains, correlated well: of the eight strains that hybridized with the AAF/II probe, seven were also positive with the *pet* probe. Also of interest were the observations that AAF/I and AAF/II were mutually exclusive (as previously suggested [10]) and that although AAF/I and AAF/II were nearly always present along with *aagR*, the latter was present in several cases in the absence of the fimbrial subunit genes, suggesting the presence of as yet undiscovered AAFs.

The placement of plasmid genes on the dendrogram presented evidence of vertical transmission of plasmid-borne genes. Most prominently, AAF/I was found mainly in cluster EAEC1, whereas AAF/II predominated in the AA/DA cluster. Also, as seen in Table 2, particular patterns of plasmid genes were commonly conserved within a single ET. For example, ET9 and ET52 each comprised three strains from different areas with nearly identical plasmid gene profiles. However, we also observed evidence of horizontal transmission of plasmids throughout the dendrogram. This may be best exemplified by the presence of both AAF/I- and AAF/II-encoding plasmids within the same clusters (e.g., ET43). To assess further whether or not such horizontal transfer may have occurred, the plasmids of strains representing two specific gene combinations that were found in strains from widely separated clusters were subjected to RFLP analysis. The combinations selected were CVD432-AAFI-*aggR-aspU* (probe a) and CVD432-*shf*-AAFII-*pet-aggR-aspU* (probe b). As probes in these analyses we used the AAF-encoding regions derived from previously constructed AAF/I- and AAF/II-encoding cosmid clones (13, 25). As shown in Fig. 3, the plasmids with probe b were largely conserved with regard to the RFLP profiles; and in fact, three of the strains revealed identical patterns despite their derivation from widely dispersed MLEE clusters. Conservation of plasmids hybridizing with probe a was not as great; nonetheless, the profiles revealed at least partial conservation of both nucleotide sequence (i.e., multiple fragments hybridized on Southern blot) and of restriction fragment patterns. These observations substantiated the hypothesis that AA plasmids are inherited by a combination of horizontal and vertical transmission.

DISCUSSION

EAEC is an emerging enteric pathogen, yet neither virulence factors nor their phylogenetics have been completely characterized. Significant EAEC strain-to-strain heterogeneity has been suggested by the prevalence of various virulence factors (10, 13, 28, 36, 37) and by the fact that only a minority of EAEC strains tested have caused diarrhea in volunteers (29). Here, we suggest that EAEC strains comprise a heterogeneous set of pathogens that share certain chromosomal and plasmid-borne genes.

Using MLEE analyses, we found that EAEC and DAEC strains segregated into defined chromosomal clusters. However, we noted significant overlap among these clusters, and specific EAEC or DAEC strains were occasionally found within clusters characteristic of the other pathotype. Overall, this situation is similar to that observed among other *E. coli* pathotypes, such as EPEC and EHEC (also shown in Fig. 2), yet the precise relationship of EAEC and DAEC to each other is still not clear. Presumably, the presence of AAFs confers the AA phenotype regardless of the chromosomal background. But our data also suggest that, like EPEC and EHEC strains, EAEC strains may comprise a package of genes (i.e., the *she* island and AA plasmid genes) that travel together in an evolutionary sense, exhibiting horizontal spread and conservation as a result of the conference of a particularly adaptive pathogenic profile. Indeed, as for EHEC and EPEC strains, it appears that the characteristic EAEC plasmid and chromosomal packages have arisen on multiple independent occasions (45).

Our phylogenetic analysis has also allowed us to place several previously described diarrheagenic *E. coli* clusters (45) (DEC 6, 7, 13, 14 and 15) on our dendrogram. Our analyses suggest that these diarrheagenic *E. coli* may belong to EAEC and DAEC pathotypes, and further analyses of these strains are ongoing. Recent work in our laboratory suggests that mem-

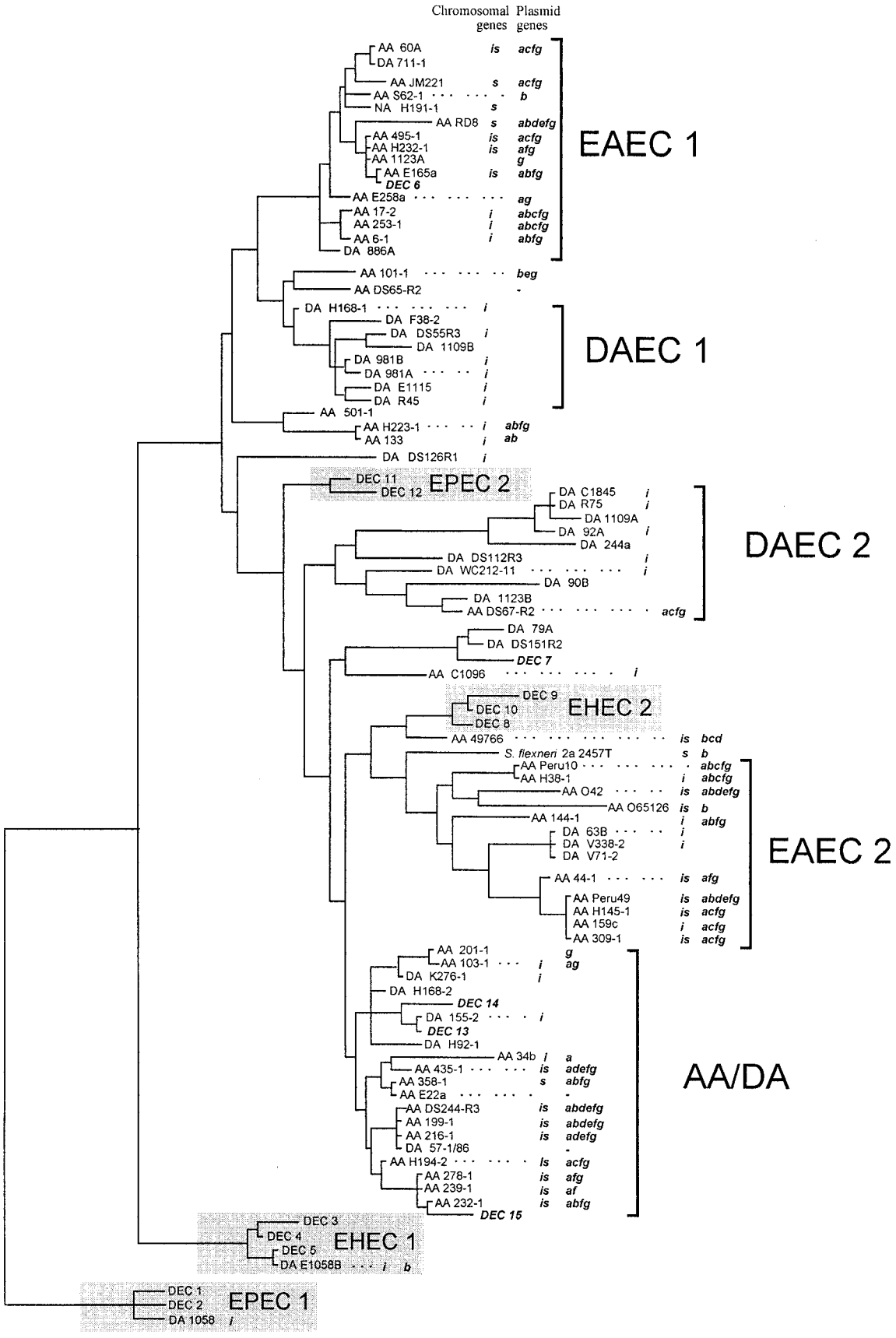

FIG. 2. Phylogram of EAEC and DAEC strains. EAEC and DAEC strains were subjected to MLEE for 20 enzymes as previously described (39, 45). Results were analyzed by the neighbor-joining method. Bold italic letters refer to positive hybridization reactions with the probes listed in Table 2: ***a***, CVD432; ***b***, *shf*; ***c***, AAF/I; ***d***, *pet*; ***e***, AAF/II; ***f***, *aggR*; ***g***, *aspU*; ***s***, *she* homolog; ***i***, *irp2*. DA, DAEC; NA, non-adherent strain; AA, EAEC (as defined by HEp-2 adherence phenotype). Dotted lines are added to assist in visual alignments. DEC clones represent diarrheagenic *E. coli* clusters previously described by Whittam et al. (45).

TABLE 2. Distribution of plasmid and chromosomal genes among EAEC strains

Strain	ET no.[a]	Location	Serotype	Probe result[b]								
				Plasmid gene							Chromosomal gene	
				CVD432	*shf*	AAF/I	*pet*	AAF/II	*aggR*	*aspU*	*she*	*irp2*
60A	1	Mexico	ND[c]	+		+			+	+	+	+
JM221	2	Mexico	O92:H33	+		+			+	+	+	
S62-1	3	China	ND		+							+
RD8	5	France[d]	O111:H2	+	+		+	+	+	+	+	
E165a	6	China	ND	+	+				+	+	+	+
H232-1	6	Peru	ND	+	+				+	+	+	+
1123A	6	China	ND							+		
495-1	6	Thailand	O130:H27	+		+			+	+	+	+
E258a	8	China	ND	+						+		
006-1	9	Thailand	Rough:H2	+	+				+	+		+
253-1	9	Thailand	O3:H2	+	+	+			+	+		+
17-2	9	Chile	O3:H2	+	+	+			+	+		+
101-1	11	Japan[d]	O?:H10		+			+		+		
501-1	20	Thailand	Rough:H53									
H223-1	21	Peru	ND	+	+				+	+	+	+
133	21	Peru	ND	+	+						+	+
C1096	34	Serbia[d]	O4:H?									+
49766	35	Mexico	O?:H10		+	+	+				+	+
PERU10	37	Peru	ND	+	+	+			+	+		
H38-1	37	Peru	ND	+	+	+			+	+		+
O42	38	Peru	O44:H18	+	+		+	+	+	+	+	+
65126	39	Mexico	ND		+						+	+
144-1	40	Thailand	O77:H−	+	+				+	+		+
44-1	42	Thailand	O36:H18	+					+	+	+	+
H145-1	43	Peru	ND	+		+			+	+	+	+
PERU49	43	Peru	ND	+	+		+	+	+	+	+	+
309-1	43	Thailand	O130:H27	+		+			+	+	+	+
159c	43	Mexico	ND	+		+			+	+		+
103-1	44	Thailand	O148:H28	+						+		+
201-1	44	Thailand	O81:H−							+		
K276-1	45	Israel	ND		+							
34b	49	India	O?:H11	+								+
435-1	50	Thailand	O33:H16	+			+	+	+	+	+	+
E22a	51	China	ND									
358-1	51	Thailand	O125ac:H12	+	+				+	+	+	
199-1	52	Thailand	Rough:H1	+	+		+	+	+	+	+	+
DS244-R3	52	Philippines	ND	+	+		+	+	+	+	+	+
216-1	52	Thailand	Rough:H27	+			+	+	+	+	+	
H194-2	53	Peru	ND	+		+			+	+	+	+
278-1	54	Thailand	O125ac:H21	+					+	+	+	+
232-1	54	Thailand	O127:H21	+					+	+	+	+
239-1	54	Thailand	Rough:H21	+						+	+	+
H32-1	ND	Peru	ND	+	+	+			+	+		+
H46-2	ND	Peru	ND	+		+			+	+	+	+
No. (%) of strains hybridizing with gene probe				34 (77)	22 (50)	14 (32)	8 (18)	8 (18)	29 (66)	35 (80)	25 (57)	32 (73)

[a] Arbitrarily assigned from the dendrogram in Fig. 2.
[b] For AA strains only. Blank spaces signify negative probe result.
[c] ND, not done.
[d] Outbreak strain.

bers of DEC 6, 11, and 15 are generally EAEC and are *she* positive, while some members of DEC 7, 9, 13, and 14 are EAEC but are *she* negative.

The high prevalence of the *Shigella* chromosomal gene *she* within EAEC clusters is notable. This gene, found predominantly in *S. flexneri* 2a strains, has been suggested to be part of a larger chromosomal pathogenicity island (33). Our data suggest that at least some *S. flexneri* strains may be descended from recent ancestors shared with EAEC strains. However, in one *S. flexneri* strain, *she* is closely linked to the *sigA* autotransporter gene (33), which we did not find by hybridization in any of our EAEC strains. Thus, the chromosomal islands of *S. flexneri* and EAEC are apparently not identical throughout their lengths.

A further inference drawn from our data is that EAEC strains, like EPEC and EHEC strains (26), carry related pathotype-specific plasmids harboring several highly conserved loci. The previously described AA probe (CVD432) has been localized to the AA plasmid of strains 042 and 17-2 and has been used as a diagnostic reagent for detection of EAEC (2). However, the analyses presented here suggest that some CVD432 probe-negative strains also carry an AA plasmid, as evidenced by the presence in such strains of several other AA plasmid genes. The specificity of these genes for EAEC strains is sup-

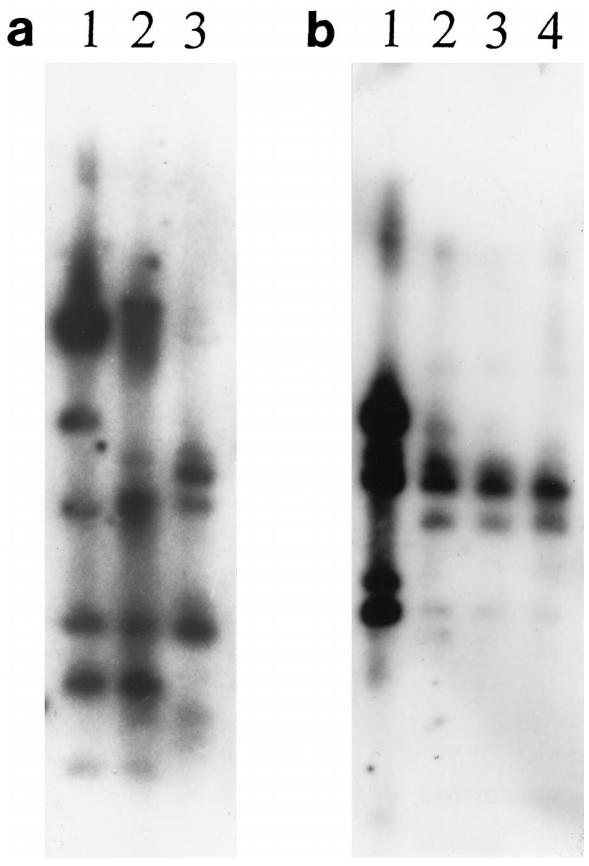

FIG. 3. Southern blot hybridization of AA plasmids by using AAF/I or AAF/II-expressing cosmid clones as probes. Total plasmid DNA was extracted from the strains indicated below; the DNA was digested with *Eco*RV and separated by agarose gel electrophoresis. DNA was then blotted and hybridized using cosmid pJPN31 (which expresses AAF/I) (a) or cosmid D6 (which expresses AAF/II) (b). (a) Lane 1, JM221; lane 2, H145-1; lane 3, H194-2. (b) Lane 1, 042; lane 2, Peru49; lane 3, DS244R3; lane 4, 199-1.

ported by the complete lack of hybridization of all the DAEC strains and the high degree of correlation with the CVD432 probe. However, our data suggest that *aspU* and *shf* may each be more sensitive (although the latter is less specific) than the CVD432 probe and that *aggR* is present in the majority of EAEC strains, albeit with a slightly lower frequency than CVD432. We acknowledge that our collection of EAEC strains was selected from various studies and that the prevalence of the AA plasmid within EAEC worldwide may not be accurately reflected. Prospective epidemiologic studies in several areas will be required before a more complete picture can be generated.

On a broader scale, our data allow us to draw inferences regarding the natural history of EAEC plasmids. *Rhizobium* species have been shown to display features of combined horizontal and vertical plasmid transmission (41, 44). A similar model is suggested for the large plasmids of EHEC (5). Here, we show that EAEC exhibits a similar pattern of transmission, in which a virulence plasmid is acquired and then stably maintained among clonal descendants. In addition, our analyses reveal that horizontal transmission of plasmids between distinct chromosomal clusters also occurs. The mechanism for horizontal transmission is presumably conjugation, as we and others have shown that AA plasmids are transmissible (31). Of note, the studies presented here represent characterization of only one side of the plasmid; however, studies ongoing in our laboratory suggest that the remainder of the plasmid is devoted mostly to transfer and replication functions (11).

The full phylogenetic and pathogenic character of EAEC is unknown, and the roles of the various factors described herein warrant further investigation. In addition, the value of the various loci as EAEC diagnostic probes is being evaluated.

ACKNOWLEDGMENTS

This work was funded by Public Health Service grants AI33096 (to J.P.N.) and AI/GM42391 (to T.S.W.)

We thank Stuart Austin, Alfredo Caprioli, Miloje Cobeljic, Peter Echeverria, Ana Gil, Carlos Eslava, John Mathewson, and Y. Itoh for bacterial strains.

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.).** 1989. Current protocols in molecular biology. John Wiley & Sons, New York, N.Y.
2. **Baudry, B., S. J. Savarino, P. Vial, J. B. Kaper, and M. M. Levine.** 1990. A sensitive and specific DNA probe to identify enteroaggregative *Escherichia coli*, a recently discovered diarrheal pathogen. J. Infect. Dis. **161:**1249–1251.
3. **Bhan, M. K., V. Khoshoo, H. Sommerfelt, P. Raj, S. Sazawal, and R. Srivastava.** 1989. Enteroaggregative *Escherichia coli* and Salmonella associated with nondysenteric persistent diarrhea. Pediatr. Infect. Dis. J. **8:**499–502.
4. **Bhan, M. K., P. Raj, M. M. Levine, J. B. Kaper, N. Bhandari, R. Srivastava, R. Kumar, and S. Sazawal.** 1989. Enteroaggregative *Escherichia coli* associated with persistent diarrhea in a cohort of rural children in India. J. Infect. Dis. **159:**1061–1064.
5. **Boerlin, P., S. Chen, J. K. Colbourne, R. Johnson, S. De Grandis, and C. Gyles.** 1998. Evolution of enterohemorrhagic *Escherichia coli* hemolysin plasmids and the locus for enterocyte effacement in shiga toxin-producing *E. coli*. Infect. Immun. **66:**2553–2561.
6. **Boyer, H. W., and R. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41:**459–468.
7. **Cobeljic, M., B. Miljkovic-Selimovic, D. Paunovic-Todosijevic, Z. Velickovic, Z. Lepsanovic, D. Savic, R. Ilic, S. Konstantinovic, B. Jovanovic, and V. Kostic.** 1996. Enteroaggregative *Escherichia coli* associated with an outbreak of diarrhoea in a neonatal nursery ward. Epidemiol. Infect. **117:**11–16.
8. **Cravioto, A., R. J. Gross, S. Scotland, and B. Rowe.** 1979. An adhesive factor found in strains of *Escherichia coli* belonging to the traditional enteropathogenic serotypes. Curr. Microbiol. **3:**95–99.
9. **Cravioto, A., A. Tello, A. Navarro, J. Ruiz, H. Villafan, F. Uribe, and C. Eslava.** 1991. Association of *Escherichia coli* HEp-2 adherence patterns with type and duration of diarrhoea. Lancet **337:**262–264.
10. **Czeczulin, J. R., S. Balepur, S. Hicks, A. Phillips, R. Hall, M. H. Kothary, F. Navarro-Garcia, and J. P. Nataro.** 1997. Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia coli*. Infect. Immun. **65:**4135–4145.
11. **Dubovsky, F., and J. P. Nataro.** Unpublished observations.
12. **Elias, W. P., J. R. Czeczulin, I. R. Henderson, L. R. Trabulsi, and J. P. Nataro.** 1999. Organization of biogenesis genes for aggregative adherence fimbria II defines a virulence gene cluster in enteroaggregative *Escherichia coli*. J. Bacteriol. **181:**1779–1785.
13. **Eslava, C., F. Navarro-Garcia, J. R. Czeczulin, I. R. Henderson, A. Cravioto, and J. P. Nataro.** 1998. Pet, an autotransporter enterotoxin from enteroaggregative *Escherichia coli*. Infect. Immun. **66:**3155–3163.
14. **Gicquelais, K. G., M. M. Baldini, J. Martinez, L. Maggi, W. C. Martin, V. Prado, J. B. Kaper, and M. M. Levine.** 1990. Practical and economical method for using biotinylated DNA probes with bacterial colony blots to identify diarrhea-causing *Escherichia coli*. Infect. Immun. **28:**2485–2490.
15. **Heilmann, C., O. Schweitzer, C. Gerke, N. Vanittanakom, D. Mack, and F. Goetz.** 1996. Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis* Mol. Microbiol. **20:**1083–1091.
16. **Henderson, I., C. E. Eslava, and J. P. Nataro.** Unpublished observations.
17. **Huppertz, H. I., S. Rutkowski, S. Aleksic, and H. Karch.** 1997. Acute and chronic diarrhoea and abdominal colic associated with enteroaggregative Escherichia coli in young children living in western Europe. Lancet **349:**1660–1662.
18. **Itoh, Y., I. Nagano, M. Kunishima, and T. Ezaki.** 1997. Laboratory investigation of enteroaggregative *Escherichia coli* associated with a massive outbreak of gastrointestinal illness. J. Clin. Microbiol. **35:**2546–2550.
19. **Keen, N. T., S. Tamaki, D. Kobayashi, and D. Trollinger.** 1988. Improved broad-host-range plasmids for DNA cloning in Gram-negative bacteria. Gene **70:**191–197.
20. **Kumar, S., K. Tamura, and M. Nei.** 1993. MEGA: molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park, Pa.

21. **Morabito, S., H. Karch, P. Mariani-Kurkdjian, H. Schmidt, F. Minelli, E. Bingen, and A. Caprioli.** 1998. Enteroaggregative, Shiga toxin-producing *Escherichia coli* O111:H2 associated with an outbreak of hemolytic-uremic syndrome. J. Clin. Microbiol. **36:**840–842.
22. **Nakata, N., C. Sasakawa, N. Okada, T. Tobe, I. Fukuda, T. Suzuki, K. Komatsu, and M. Yoshikawa.** 1992. Identification and characterization of virK, a virulence-associated large plasmid gene essential for intercellular spreading of *Shigella flexneri*. Mol. Microbiol. **6:**2387–2395.
23. **Nataro, J. P., M. M. Baldini, J. B. Kaper, R. E. Black, N. Bravo, and M. M. Levine.** 1985. Detection of an adherence factor of enteropathogenic *Escherichia coli* with a DNA probe. J. Infect. Dis. **152:**560–565.
24. **Nataro, J. P., Y. Deng, Y. Deng, and K. Walker.** 1994. AggR, a transcriptional activator of aggregative adherence factor I expression. J. Bacteriol. **176:** 4691–4699.
25. **Nataro, J. P., Y. Deng, D. R. Maneval, A. L. German, W. C. Martin, and M. M. Levine.** 1992. Aggregative adherence fimbriae I of enteroaggregative *Escherichia coli* mediate adherence to HEp-2 cells and hemagglutination of human erythrocytes. Infect. Immun. **60:**2297–2304.
26. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142–201.
27. **Nataro, J. P., J. B. Kaper, R. Robins-Browne, V. Prado, and M. M. Levine.** 1987. Patterns of adherence of diarrheagenic *Escherichia coli* to HEp-2 cells. Pediatr. Infect. Dis. J. **6:**829–831.
28. **Nataro, J. P., T. Steiner, and R. L. Guerrant.** 1998. Enteroaggregative *Escherichia coli*. Emerg. Infect. Dis. **4:**251–261.
29. **Nataro, J. P., D. Yikang, S. Cookson, A. Cravioto, S. J. Savarino, L. D. Guers, M. M. Levine, and C. O. Tacket.** 1995. Heterogeneity of enteroaggregative *E. coli* virulence demonstrated in volunteers. J. Infect. Dis. **171:** 465–468.
30. **Nataro, J. P., D. Yikang, J. A. Giron, S. J. Savarino, M. H. Kothary, and R. Hall.** 1993. Aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli* requires two unlinked plasmid regions. Infect. Immun. **61:** 1126–1131.
31. **Okeke, I. N., A. Lamikanra, J. Czeczulin, J. B. Kaper, and J. P. Nataro.** Unpublished observations.
32. **Rajakumar, K., F. Luo, C. Sasakawa, and B. Adler.** 1996. Evolutionary perspective on a composite *Shigella flexneri* 2a virulence plasmid-borne locus comprising three distinct genetic elements. FEMS Microbiol. Lett. **144:**13–20.
33. **Rajakumar, K., C. Sasakawa, and B. Adler.** 1997. Use of a novel approach, termed island probing, identifies the *Shigella flexneri she* pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. Infect. Immun. **65:**4606–4614.
34. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method of reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
35. **Sasakawa, C., K. Kamata, T. Sakai, S. Y. Murayama, S. Makino, and M. Yoshikawa.** 1986. Molecular alteration of the 140-megadalton plasmid associated with loss of virulence and Congo red binding activity in *Shigella flexneri*. Infect. Immun. **51:**470–475.
36. **Savarino, S. J., A. Fasano, D. C. Robertson, and M. M. Levine.** 1991. Enteroaggregative *Escherichia coli* elaborate a heat-stable enterotoxin demonstrable in an in vitro rabbit intestinal model. J. Clin. Investig. **87:**1450–1455.
37. **Savarino, S. J., P. Fox, Y. Deng, and J. P. Nataro.** 1994. Identification and characterization of a gene cluster mediating enteroaggregative *Escherichia coli* aggregative adherence fimbria I biogenesis. J. Bacteriol. **176:**4949–4957.
38. **Schubert, S., A. Rakin, H. Karch, E. Carniel, and J. Heeseman.** 1998. Prevalence of the "high-pathogenicity island" of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. Infect. Immun. **66:** 480–485.
39. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of mutlilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:**873–884.
40. **Smith, H. R., T. Cheasty, and B. Rowe.** 1997. Enteroaggregative *Escherichia coli* and outbreaks of gastroenteritis in UK. Lancet **350:**814–815.
41. **Souza, V., and L. E. Eguiarte.** 1997. Bacteria gone native vs. bacteria gone awry?: plasmidic transfer and bacterial evolution. Proc. Natl. Acad. Sci. USA **94:**5501–5503.
42. **Vial, P. A., R. Robins-Browne, H. Lior, V. Prado, J. B. Kaper, J. P. Nataro, D. Maneval, A. Elsayed, and M. M. Levine.** 1988. Characterization of enteroadherent-aggregative *Escherichia coli*, a putative agent of diarrheal disease. J. Infect. Dis. **158:**70–79.
43. **Wanke, C. A., J. B. Schorling, L. J. Barrett, M. A. Desouza, and R. L. Guerrant.** 1991. Potential role of adherence traits of *Escherichia coli* in persistent diarrhea in an urban Brazilian slum. Pediatr. Infect. Dis. J. **10:** 746–751.
44. **Wernegreen, J. J., E. E. Harding, and M. A. Riley.** 1997. *Rhizobium* gone native: unexpected plasmid stability of indigenous *Rhizobium leguminosum*. Proc. Natl. Acad. Sci. USA **94:**5483–5488.
45. **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: P. E. Orndorff