INFECTION AND IMMUNITY, Sept. 2001, p. 5553–5564 Vol. 69, No. 9
0019-9567/01/$04.00+0 DOI: 10.1128/IAI.69.9.5553–5564.2001

Comparative Sequence Analysis of the Plasmid-Encoded Regulator of Enteropathogenic *Escherichia coli* Strains

IRUKA N. OKEKE,[1]† JADE A. BORNEMAN,[2] SOOAN SHIN,[1] JAY L. MELLIES,[3] LAURA E. QUINN,[1] AND JAMES B. KAPER[1]*

Department of Microbiology and Immunology and Center for Vaccine Development[1] and Department of Medical Research and Technology,[2] University of Maryland School of Medicine, Baltimore, Maryland 21201, and Biology Department, Reed College, Portland, Oregon 97202[3]

Received 11 January 2001/Returned for modification 9 April 2001/Accepted 18 June 2001

Enteropathogenic *Escherichia coli* (EPEC) strains that carry the EPEC adherence factor (EAF) plasmid were screened for the presence of different EAF sequences, including those of the plasmid-encoded regulator (*per*). Considerable variation in gene content of EAF plasmids from different strains was seen. However, *bfpA*, the gene encoding the structural subunit for the bundle-forming pilus, bundlin, and *per* genes were found in 96.8% of strains. Sequence analysis of the *per* operon and its promoter region from 15 representative strains revealed that it is highly conserved. Most of the variation occurs in the 5′ two-thirds of the *perA* gene. In contrast, the C-terminal portion of the predicted PerA protein that contains the DNA-binding helix-turn-helix motif is 100% conserved in all strains that possess a full-length gene. In a minority of strains including the O119:H2 and canine isolates and in a subset of O128:H2 and O142:H6 strains, frameshift mutations in *perA* leading to premature truncation and consequent inactivation of the gene were identified. Cloned *perA*, *-B*, and *-C* genes from these strains, unlike those from strains with a functional operon, failed to activate the *LEE1* operon and *bfpA* transcriptional fusions or to complement a *per* mutant in reference strain E2348/69. Furthermore, O119, O128, and canine strains that carry inactive *per* operons were deficient in virulence protein expression. The context in which the *perABC* operon occurs on the EAF plasmid varies. The sequence upstream of the *per* promoter region in EPEC reference strains E2348/69 and B171-8 was present in strains belonging to most serogroups. In a subset of O119:H2, O128:H2, and O142:H6 strains and in the canine isolate, this sequence was replaced by an IS*1294*-homologous sequence.

Enteropathogenic *Escherichia coli* (EPEC) is a leading cause of infantile diarrhea, particularly in developing countries (36). Although EPEC was one of the first classes of enterovirulent *E. coli* to be identified, the precise mechanism by which EPEC strains cause diarrhea is still under investigation. Two definitive features that are observed when EPEC infects epithelial cells are the attaching-and-effacing (A/E) lesion and localized adherence (LA) (36). The A/E histopathology is characterized by bacteria attached intimately to the host cell (33). Following signal transduction, host cytoskeletal proteins, including actin, accumulate at the site of attachment. Eventually, the host cell undergoes structural modifications resulting in the elevation and cupping of the bacterial cell on a pedestal-like protrusion supported by host cell-polymerized actin (33). In the intestine, this leads to effacement of microvilli. EPEC strains possess a chromosomal pathogenicity island, called the locus of enterocyte effacement (LEE), which contains the genes required to produce A/E lesions (28). The LEE encodes a 94-kDa outer membrane protein required for intimate attachment called intimin and a type III secretion system as well as proteins that are translocated via this system (14).

Typical EPEC strains adhere to host epithelial cells in tight, three-dimensional microcolonies. Expression of a type IV pilus, the bundle-forming pilus (BFP), is required for this LA phenotype (17). The genes required for biogenesis of the BFP are carried on a large 50- to 70-MDa virulence plasmid, the EPEC adherence factor (EAF) plasmid (51), which varies in sequence among different EPEC strains but is somewhat conserved (37). The presence of this plasmid has been shown to greatly enhance the virulence of EPEC, and hence strains carrying the EAF plasmid are strongly associated with diarrhea in epidemiological studies and produce diarrhea in volunteers (5, 24, 36). In addition to the *bfp* genes, the plasmid also carries genes encoding a regulator (Per) (18). The *per* operon consists of three genes, the first of which, *perA*, encodes a protein that belongs to the AraC family of transcriptional activators (18). The other two genes, *perB* and *perC*, encode proteins that enhance the activity of PerA by an unknown mechanism (18) (53). Per has been shown, to date, to activate genes within and outside the LEE via the LEE-encoded regulator (Ler) (13, 30) and the *bfp* operon (53) as well as a chromosomal gene involved in microcolony formation, *trcA* (54). Per also activates its own promoter (26, 34). Volunteer studies conducted by Bieber et al. demonstrated that a *perA* mutant (referred to as a *bfpT* mutant in that study) is less virulent than its isogenic wild-type strain (5). Per appears, therefore, to play a pivotal role in the regulation of essential virulence genes in EPEC.

Most EPEC strains are genetically related and belong to a limited number of O:H serotypes. Classic EPEC O serogroups include O55, O86, O111, O114, O119, O128, and O142. The most common H antigens associated with EPEC are the H6

* Corresponding author. Mailing address: Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, MD 21201. Phone: (410) 706-5328. Fax: (410) 706-0182. E-mail: jkaper@umaryland.edu.

† Present address: Department of Biomedical Sciences, University of Bradford, Bradford, West Yorkshire, BD7 1DP United Kingdom.

TABLE 1. Strains employed in this study

Strain	Serotype	Yr of isolation	Source	Country	Reference
C54-58 (DEC 1a)	O55:H6	1958	Human	Dutch Guyana	40
Beta	O55:H6	1947	Human	Scotland	45
5513-56 (DEC 2b)	O55:NM	1956	Human	United States	56
E990	O86:NM	1950	Human	England	45
2966-56 (DEC 12b)	O111:H2	1956	Human	United States	56
DIF043256	O111:H2	1986	Human	Mexico	11
2309-77	O111ab:H2	1977	Human	United States	37
B171-8	O111:NM		Human		17
StokeW	O111:NM	1947	Human	Scotland	45
009-271082	O111	1982	Human	Peru	35
065542	O114:H2		Human	Brazil	4
065900	O114:H2		Human	Brazil	4
MB80	O119:H2		Human	Brazil	8
O119#6	O119:H6		Human	Brazil	19
O119#5	O119:H6		Human	Brazil	19
C35	O119:H6	1995	Human	Nigeria	39
0659-79	O119:H6	1979	Human	United States	24
E. coli 10	O119:H6	1986	Human	Chile	35
023-220982	O126		Human	Peru	27
E2348/69	O127:H6	1969	Human	England	23
1092-80	O127:NM		Human		47
MB21	O128:H2		Human	Brazil	8
E56/54	O128ab:H2	1953	Human	England	45
010-311082	O128	1982	Human	Peru	35
O142#19	O142:H6		Human	Brazil	50
O142#15	O142:H6		Human	Brazil	50
G51	O142:H6	1995	Human	Nigeria	39
C771	O142:H6	Before 1960	Human	Indonesia	45
E851/71	O142:H6	1971	Human	Scotland	23
012-050982	O142:H6	1982	Human	Peru	35
CA 4225		1989	Canine	Canada	3

and H2 antigens. EPEC strains have been subtyped into two major genetic lineages by multilocus enzyme electrophoresis. The EPEC1 lineage includes predominantly H6 strains, and the EPEC2 lineage includes predominantly H2 strains (40, 56). A less common EPEC H type is H34, and a number of EPEC strains are nonmotile in conventional tests and classified as H^-. EPEC strains belonging to nonclassic serotypes have been reported, but these rarely possess the EAF plasmid. EPEC strains from classic serogroups appear to have arisen from a few clones that acquired virulence loci, such as the LEE and the EAF plasmid, in a stepwise manner (43). Studies with other EPEC virulence genes, including *eae* (which encodes intimin), have shown that horizontally acquired genes often, but not always, show allelic variation that correlates with clonality (1, 6, 43, 56). The *per* operon is part of a plasmid and itself has a lower G+C content than most other regions of the plasmid. In strains E2348/69 and B171-8, the *per* operon is known to be flanked by insertion sequences (18, 52). We sequenced the *perA*, *-B*, and *-C* genes from multiple EPEC strains belonging to different serotypes and investigated potential effects on virulence gene activation resulting from sequence differences. We also sequenced the region immediately upstream of the *perABC* operon to study the relationship between the *per* sequence, EAF structure, and genomic background with a view towards better understanding the evolution of EPEC.

MATERIALS AND METHODS

Strains. Thirty-one EPEC strains that carry an EAF plasmid and were isolated from patients with diarrhea in different countries were employed in the study (Table 1). Thirty strains, representing nine EPEC O groups and 13 serotypes, were from human with diarrhea and had been characterized to various degrees in previous studies. One EAF-positive canine isolate was also studied. PCR-restriction fragment length polymorphism (PCR-RFLP) typing based on the *fliC* gene was performed on all strains in order to verify H type and identify H types of nonmotile strains as an EAF-independent subclassification (15). *E. coli* K-12 strain MG1655 (7), enterohemorrhagic *E. coli* strain EDL933 (44), and JPN15 (20), the plasmid-cured derivative of EPEC reference strain E2348/69, were employed in PCR and hybridization experiments as controls.

Screening for EAF plasmid genes. The EAF probe (an empirically derived probe from the EAF plasmid of EPEC strain E2348/69) (35) was used to screen all the strains by colony hybridization. The presence of other, previously described, EAF plasmid-encoded genes was determined by colony hybridization and PCR. PCR was carried out under high-stringency conditions, employing primers designed to anneal to the sequence of the target gene from plasmid B171-8 (accession number AB024946). Colony hybridizations were conducted under high-stringency conditions at 65°C employing probes that were labeled by random priming with [α-^{32}P]dCTP as described previously (39). Fragment probes for *eae, bfpA*, the EAF, and *perABC* were prepared from plasmid clones described in Table 2. Probes for *bfpG, trcP, orf35-36* (truncated EHEC *toxB* homolog), *orf61-62* (truncated *gadB* and *gadC* homologs), and *orf67-68* (*stbAB* homolog) were prepared by PCR with primers listed in Table 3, using strain B171-8 as the template, and verified by restriction analysis.

LA and A/E tests. Fluorescent actin staining of HEp-2 monolayers infected with each of the *E. coli* isolates was carried out as described by Knutton et al. (22), with modifications. HEp-2 cells were grown overnight to 50% confluence in Dulbecco's minimum Eagle medium (DMEM) (Gibco BRL, Gaithersburg, Md.) containing penicillin, streptomycin, and 10% fetal bovine serum on coverslips in 24-well plates. Bacteria were grown for 16 h in Luria broth (LB) without shaking. The HEp-2 cells were washed three times with phosphate-buffered saline (PBS), and the medium was replaced with DMEM containing 10% fetal bovine serum and 1% mannose. Ten microliters of bacterial suspension was added per well, and the slides were incubated at 37°C and 5% CO_2 for 3 h. After a 3-h incubation period, the cells were washed three times with PBS, fixed with formalin, permeabilized with 0.1% Triton X-100 in PBS, and stained with fluorescein isothio-

TABLE 2. Plasmids and constructs used in this study

Plasmid or construct	Description	Reference or source
pCVD443	1-kb *Hin*dIII *eae* probe fragment in PUC19	20
pMSD207	852-bp *Eco*RI *bfpA* probe fragment in pUC18	16
pJPN16	1-kb *Bam*HI-*Sal*I EAF probe fragment in pUC8	35
pJLM171	1.8-kb *Bam*HI *perABC* minimal clone in pBR322	30
pBR322	Cloning vector	46
p2R	1.6-kb *Sal*I fragment from cosmid clone	31
pBluescript	Cloning vector	Stratagene
JLM183	Single-copy fusion of the region −211 to +221 relative to the transcriptional start site of the LEE1 gene on the chromosome of *E. coli* K-12 strain MC4100	Quinn et al.[a]
SS630	Single-copy fusion of the region −688 to +124 relative to the transcriptional start site of *bfpA* on the chromosome of *E. coli* K-12 strain MC4100	48

[a] L. Quinn, J. Mellies, and J. Kaper, 101st Gen. Meet. Am. Soc. Microbiol. 2001, poster B-283, 2001.

cyanate-labeled phalloidin (Sigma). The monolayers were then washed with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and treated with RNase A (10 mg/ml) before being stained with propidium iodide (Molecular Probes, Eugene, Oreg.). The HEp-2 cells were observed for the adherence pattern of the red-stained bacteria as well as actin accumulation (green) at the site of bacterial attachment under incident fluorescent light with a Zeiss Axioskop routine microscope. Each strain was tested in duplicate. EPEC strain E2348/69 and *E. coli* K-12 strain MG1655 were used as positive and negative controls, respectively.

Western immunoblotting. Overnight bacterial cultures in LB were subcultured into DMEM and grown, with shaking, for 3 to 4 h to an optical density (A_{600}) of 1.0. Bacterial suspensions were pelleted and resuspended in PBS. Whole-cell lysates were prepared by boiling samples in sodium dodecyl sulfate sample buffer. Ten micrograms of total protein was loaded per lane, subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and transferred to 0.45-μm-pore-size polyvinylidene difluoride membranes (Millipore, Bedford, Mass.). Membranes were incubated with polyclonal rabbit antisera prepared against intimin, bundlin, and the LEE-encoded chaperone CesD, which had been prepared from proteins purified from EPEC strain E2348/69. The membranes were washed with PBS–0.05% Tween 20, immunostained with a 1:35,000 dilution of goat anti-rabbit horseradish peroxidase antibody conjugate (Kirkegaard and Perry Laboratories, Gaithersburg, Md.), and developed with enhanced-chemilumniscence reagents according to the instructions of the manufacturer (Amersham, Pharmacia Biotech, Piscataway, N.J.). Bands were detected by exposure to X-ray film (Kodak X-Omat blue).

Cloning and sequencing of *perABC* and the *per* upstream regions. Overlapping clones for the *per* region were prepared by amplifying with the K1547-K1693 and K182-K696 or K1694-K696 primer pairs (Table 3) with the proofreading enzyme *Pwo* polymerase (Boehringer Mannheim GmbH, Mannheim, Germany) and cloning the resulting fragments into the *Sma*I site of pBluescript (Stratagene) or into pCR-Blunt II TOPO (Invitrogen, Carlsbad, Calif.). Boiled colonies of the test strains were used as templates. Because of the tendency to select PCR-derived *per* deletion mutants when cloning *per* genes into high-copy-number vectors, clones with inserts in the inverse orientation were selected for sequencing. Clones that contained mutations were verified by carrying out another PCR with a fresh template, recloning, and resequencing the region or in a few cases by restriction analysis of the new PCR product with the enzymes *Ssp*I and *Dra*I (these enzymes gave altered restriction profiles for the *perA* alleles of O142#15 [*Ssp*I only], MB80, and CA 4225). The upstream region of *per* from strain MB80 was cloned by partially digesting genomic DNA with the frequently cutting enzyme *Taq*I (Gibco BRL) to reduce the number of nonspecific products and subjecting the resulting digest to *Taq* polymerase PCR and annealing at 50°C with primer pair K1547-K1549. The resulting multiple fragments were gel purified, and fragments that yielded an expected 350-bp product after *Bam*HI (Gibco BRL) digestion were cloned and sequenced. On the basis of this sequence, primer K1978 was designed and used with K1549 to amplify the upstream regions of strains MB80, O142#15, and CA 4225 under high-stringency conditions (annealing at 52°C with *Pwo* polymerase) and cloned for sequencing. Plasmid DNA was purified with a Wizard miniprep kit (Promega, Madison, Wis.) and subjected to automated sequencing at the University of Maryland Biopolymer Facility on an Applied Biosystems, Inc., sequencer. Primers K182, K1691, K696, K1547, and K1978 were used with vector-priming M13 oligonucleotides to obtain double-stranded sequence. Clones of the *per* operon used in functional experiments were prepared by amplifying the entire operon with the primers K1547 or K1978 and K696, which carries a *Bam*HI site in its tail, digesting with *Bam*HI, and cloning the resulting product into the *Bam*HI site of pBR322. This produced 1.8-kb

TABLE 3. PCR primers used in this study

Primer	Target	Strand	Sequence (5′–3′)
K1694	*perA*	+	GGGGTACCTTAGCGTTATTGTTAACTAATTAC
K1693	*perA*	−	CCCAAGCTTTGGCAATGTTCCTTGTGT
K182	*perA*	+	ATGAGTGATTCTTGTCTG
K1691	*perB*	−	CCCAAGCTTTGGTGCTCTTTTTGATTC
K1547	*perA* upstream region to IS*1*(νξ)	+	TGAGTCACCTCTGCCTGAG
K1549	*perA*	−	TGGATTCTATTGTGTATTCGG
K696	*perC-perD* intergenic region	−	CTCGATCCCTGAGAGATCTACTATT
K1978	*perA* upstream region to IS*1294*	+	TGTGAGAGCTTCTCAGCA
K1821	*bfpA*	+	CGGAATTCATCCTCCATGAAGCCAGTCA
K1822	*bfpA*	−	TCCCCCCGGGCGTATTATGTAGATTAC
K1823	*bfpG*	+	TCCCCCCGGGAGGTATATGTGAGGACA
K1824	*bfpG*	−	CGGGATCCCAGCTGATGTTCTGATGATA
K2019	*trcP* (pB171)	+	TGCCGACATGGATTTTGCCA
K2136	*trcP* (pB171)	−	GAATTGATAAGCATTGCTCCG
K1914	*orf35* (pB171 *toxB* homolog)	+	ATGACTTGTCTGGATACATATC
K1913	*orf36* (pB171 *toxB* homolog)	−	GAAACGACATGTCGAGTATGC
K1928	*orf61* (pB171 *gadB* homolog)	+	AGCGGTCGATATTCATAGCT
K1929	*orf62* (pB171 *gadC* homolog)	−	GTAGATAGATAGCACCAGCG
K1912	*orf67* (pB171 *stbA* homolog)	+	GAACGTATACTGCGATGATGGT
K1911	*orf68* (pB171 *stbB* homolog)	−	GCACCGAATCAGTAGCAGAAG

TABLE 4. Results of PCR and hybridization for chromosomal and EAF-encoded sequences[a]

Strain	Serotype	*fliC* (H) type(s)	Presence or absence of:						IS element upstream of *per*	Presence or absence of:					GenBank accession no. for *per* sequence
			eae	*espC*	*bfpA* (probe)	*bfpA* (PCR)	*bfpG*	*trcP*		*per*	EAF	*orf35-36*	*orf61-62*	*stbAB*	
C54-58	O55:H6	6	+	+	+	+	+	−	IS*1*(νξ)	+	+	+	+	+	AF254667
Beta	O55:H6	6	+	+	+	+	+	−	IS*1*(νξ)	+	+	+	+	+	
5513-56	O55:NM	6	+	+	+	+	+	−	IS*1*(νξ)	+	+	+	+	+	AF254666
E990	O86:NM	2	+	−	+	+	+	+	IS*1*(νξ)	+	+	+	+	+	AF255773
2966-56	O111:H2	2	+	−	+	+	+	+	IS*1*(νξ)	+	+	+	+	+	AF255771
DIF043256	O111:H2	2	+	−	+	+	+	+	IS*1*(νξ)	+	+	+	+	+	
2309-77	O111:H2	2	+	−	+	+	+	+	IS*1*(νξ)	+	+	+	+	+	
B171	O111:NM	2	+	−	+	+	+	+	IS*1*(νξ)	+	+	+	+	+	L42638 (54)
StokeW	O111:NM	2	+	−	+	+	+	+	IS*1*(νξ)	+	+	+	−	+	
009-271082	O111	2	+	−	+	+	+	+	IS*1*(νξ)	+	+	+	+	+	
065542	O114:H2	2, 6	+	−	+	+	+	+	IS*1*(νξ)	+	+	+	+	+	
065900	O114:H2	2, 2A, 6	+	−	+	+	+	+	IS*1*(νξ)	+	+	+	+	+	AF255768
MB80	O119:H2	2	+	−	+	−	−	−	IS*1294*	+	+	−	−	+	AF295090
O119#6	O119:H6	6	+	−	+	+	+	−	IS*1*(νξ)	+	+	−	−	+	AF305565
O119#5	O119:H6	6	+	−	+	+	+	−	IS*1*(νξ)	−	+	−	−	+	
C35	O119:H6	6	+	−	+	+	+	−	IS*1*(νξ)	+	+	+	−	+	
0659-79	O119:H6	6	+	−	+	+	+	−	IS*1*(νξ)	+	+	+	−	+	
E. coli 10	O119:H6	6	+	−	+	+	+	−	IS*1*(νξ)	+	+	+	−	+	
023-220982	O126	2C	+	−	+	+	+	+	IS*1*(νξ)	+	+	−	−	+	AF295088
E2348/69	O127:H6	6	+	+	+	+	+	−	IS*1*(νξ)	+	+	+	+	+	Z48561 (18)
1092-80	O127:NM	6	+	+	+	+	+	−	IS*1*(νξ)	+	+	+	−	+	AF255769
MB21	O128:H2	2	+	−	+	−	−	−	IS*1294*	+	−	−	−	+	
E56/54	O128:H2	2	+	−	+	+	+	−	IS*1*(νξ)	+	+	+	−	+	AF347302
010-311082	O128	2C	+	−	+	+	+	−	IS*1*(νξ)	+	+	+	−	+	
O142#19	O142:H6	6	+	+	+	+	+	−	IS*1294*	+	+	−	−	+	
O142#15	O142:H6	6	+	+	+	+	+	−	IS*1294*	+	+	−	−	+	AF255770
G51	O142:H6	6A	+	+	+	+	+	−	IS*1*(νξ)	+	+	+	+	+	
C771	O142:H6	6	+	+	+	+	+	−	IS*1*(νξ)	+	+	+	−	+	
E851/71	O142:H6	6	+	+	+	+	+	−	IS*1*(νξ)	+	+	+	−	+	AF255772
012-050982	O142:H6	6	+	+	+	+	+	+	IS*1*(νξ)	+	+	+	−	+	
CA 4225		U	+	−	+	−	−	−	IS*1294*	+	+	+	−	+	AF295089
JPN15		6	+	+	−	−	−	−	NA	−	−	−	−	−	
MG1655		ND	−	−	−	−	−	−	NA	−	−	−	−	−	
EDL933		7	+	−	−	−	−	−	NA	−	−	+	−	−	

[a] ND, not determined; NA, not applicable.

clones of the *per* operon that were analogous to pJLM171 (30) and included the 182 bp of the region upstream of the *perA* start codon.

Sequence analysis. Sequences were assembled by the Sequencher 3.1 program (Gene Codes, Inc.) and annotated using MacDNASIS version 3.5 (Hitachi Software Engineering America, Ltd.). The sequences of *per* operons from strains E2348/69 and B171-8 were retrieved from GenBank (accession numbers Z48561 and L42638, respectively) and included in the analysis. Multiple alignments were performed with the CLUSTAL W program. Phylogenetic trees were derived with the program MEGA2β (http://www.megasoftware.net; Sudhir Kumar, Koichiro Tamura, Ingrid Jakobsen, and Masatoshi Nei, 2000, Institute of Molecular Evolutionary Genetics, The Pennsylvania State University). The significance of synonymous and nonsynonymous substitutions was determined by the method of Nei and Gojobori (38) with the MEGA2β program.

β-Galactosidase assays. Clones of the *perABC* operon from representative strains in pBR322 were assayed for their ability to activate single-copy chromosomal fusions of the upstream region of the LEE1 gene operon (nucleotides −211 to +221 relative to the transcriptional start site) and *bfpA* (nucleotides −688 to +124 relative to the transcriptional start site) to *lacZ* in an *E. coli* K-12 (MC4100) background. Fusions were constructed by the method of Simons et al. (49) as described previously (30). Overnight bacterial cultures were subcultured into LB and grown to an optical density at 600 nm of 0.4 to 0.5. The β-galactosidase activities of culture lysates on the substrate *ortho*-nitrophenyl-β-D-galactopyranoside (Sigma) was measured in Miller units as described by Miller (32). Statistical analysis of the results was performed by Student's *t* test.

***Ssp*I-based PCR-RFLP.** The region encompassing nucleotide 54 to the end of the *perA* open reading frame was amplified by *Taq* PCR with primers K1694 and K1693. The resulting 770-bp products were digested at 37°C for 90 min with *Ssp*I (Gibco BRL). Fragment length polymorphisms were distinguished by electrophoresis on a 2% agarose gel.

Nucleotide sequence accession numbers. The nucleotide sequences of the *perA*, *-B*, and *-C* genes obtained in this study were deposited in the GenBank database under the accession numbers listed in Table 4. Sequences of *per* genes from strains E2348/69 (Z48561 [18]) and B171-8 (L42638 [53]) were retrieved from previous depositions.

RESULTS

EPEC strains vary in virulence factor expression. Expression levels of two established virulence proteins of EPEC that are regulated by Per—bundlin (the BFP structural subunit) and intimin—were examined in different EPEC strains. CesD, a LEE-encoded (and therefore Per-regulated) chaperone, was also evaluated because, unlike the other two proteins, it is not surface exposed and is less likely to vary antigenically among different strains. As shown in Fig. 1, there was considerable variation in the expression levels for bundlin, intimin, and CesD as well as in the ability of different isolates to demonstrate the A/E and LA phenotypes. Although the expression of different initimin alleles may account for some of the differences in levels of intimin expression (1), similar antigenic differences have not been demonstrated for bundlin (6) and are not expected to occur with CesD, a protein that is highly conserved between widely divergent A/E pathogens (41, 57). We therefore hypothesized that differences in the levels of

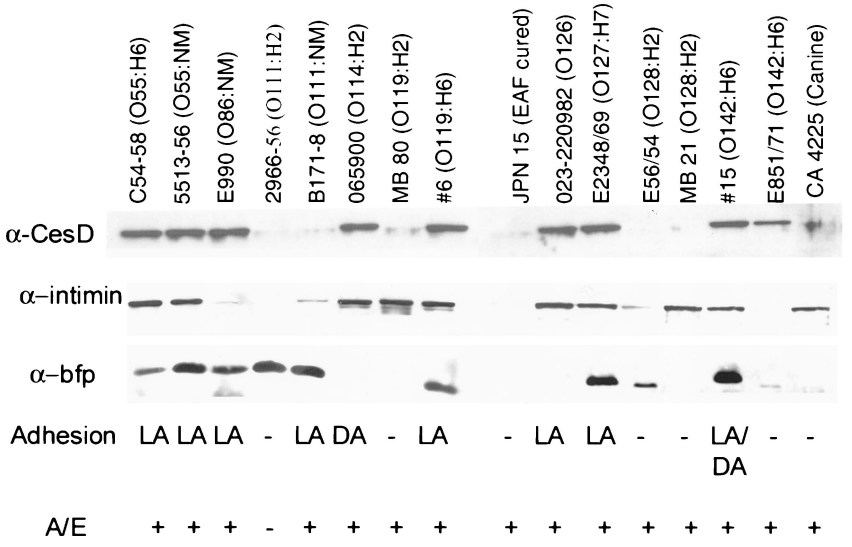

FIG. 1. Expression of intimin, bundlin, and CesD, virulence proteins known to be directly or indirectly regulated by *per*, in different EPEC isolates. α, antibody.

expression of these proteins may be related to defects in specific genes or in genes encoding key regulators. Per is known to play a central role in the regulation of these genes and phenotypes and was therefore selected for comparative sequence analysis.

EAF plasmids are partially conserved, and the *per* operon is present in almost all EAF-positive EPEC strains. In concordance with the results of previous studies (16, 18, 37), we found that, although there is variation in the gene contents of EAF plasmids of different EPEC isolates, the *bfpA* gene and the *stbAB* region are universally present in all strains examined. The *perA*, *-B*, and *-C* genes were shown, by hybridization and PCR, to be present in all but one O119:H6 isolate (O119#5). The empirically derived EAF probe was shown in previous studies to hybridize with most but not all strains that have the virulence plasmid (35) and hybridized with all but one of the strains employed in this study. As shown in Table 4, other regions of the EAF plasmid are less conserved. The *bfpG* gene was present in 28 (90.3%) strains. The three strains that were negative for *bfpG* were the O119:H2 (MB80) and O128:H2 (MB21) strains, which have previously been reported to lack this gene and the rest of the *bfp* operon (8), and the canine isolate CA 4225. Although all three isolates hybridized with the *bfpA* probe, *bfpA* could not be amplified by PCR, suggesting that the 3′ end of the gene is deleted as described by Bortolini et al. (8). As expected, all three isolates fail to produce bundlin detectable by Western blotting (Fig. 1). The region *orf35-36* in plasmid pB171 (52) (truncated homolog of the EHEC *toxB* gene) was found in 24 (77.4%) of the EPEC strains, while the region *orf61-62* (*gadB* [truncated] and *gadC* homologs) was present in only 13 (41.9%) of the strains examined. The *trcP* gene, which was originally found during the sequencing of the EAF plasmid of O111:NM strain B171 (52), was in this study detected in strains E990 (O86:NM), 065542, 065900 (O114:H2), 023-220982 (O126), 012-050982 (O142:H6), and all the O111 strains but was absent in other strains. In summary, the O55, O86, O111, O114, O126, and O127 strains and a subset of the O142 strains were positive for most of the pB171-8-derived probes, while the O119 and O128 strains and the canine isolate were more likely not to hybridize with one or more of the probes.

Nucleotide sequence variation in the *per* operon is minimal. *per* sequences for at least one strain belonging to each serotype were subjected to comparative analysis. *Ssp*I restriction fragment length polymorphisms revealed that two O142:H6 strains showed a unique profile, and therefore *per* from strain O142#15 was sequenced in addition to that from O142:H6 strain E851/71. Two previously determined *per* sequences (from strains E2348/69 and B171-8) were also included in the analysis, and the accession numbers of sequences determined in this and other studies are listed in Table 1. The sequences of the *per* genes from different strains show very little variation. We found 94 to 99% identity among the sequences of *perABC* and the upstream 100 bp at the nucleic acid level. Pairwise analysis by the Nei-Gojobori method revealed that most of the sequences were not significantly different from one another. The major exceptions were the O119:H2 (MB80), O142:H6 (O142#15), and canine (CA 4225) *perABC* operons, which showed significant differences from the others ($P < 0.05$). In addition to other substitutions, these strains have mutating frameshifts in the *perA* sequence. The *perA* nucleotide G114 is deleted in strains MB80 (O119:H2) and CA 4225 (canine), and TA is inserted after nucleotide 313 in strain O142#15 (O142:H6), which results in truncation of the gene (Fig. 2).

The predicted amino acid sequences of PerC are 100% identical for all but three strains, which each have only one substitution. Similarly, the C-terminal 37 amino acids of PerB were conserved in all the strains sequenced and strains C54-58, 5513-56 (O55), E990 (O86), 2966-56, B171 (O111), 023-220980 (O126), and E2348/69 (O127:H6) show 100% identity in the entire PerB amino acid sequence. All other strains have two conserved substitutions (at residues 43 and 92) and strains

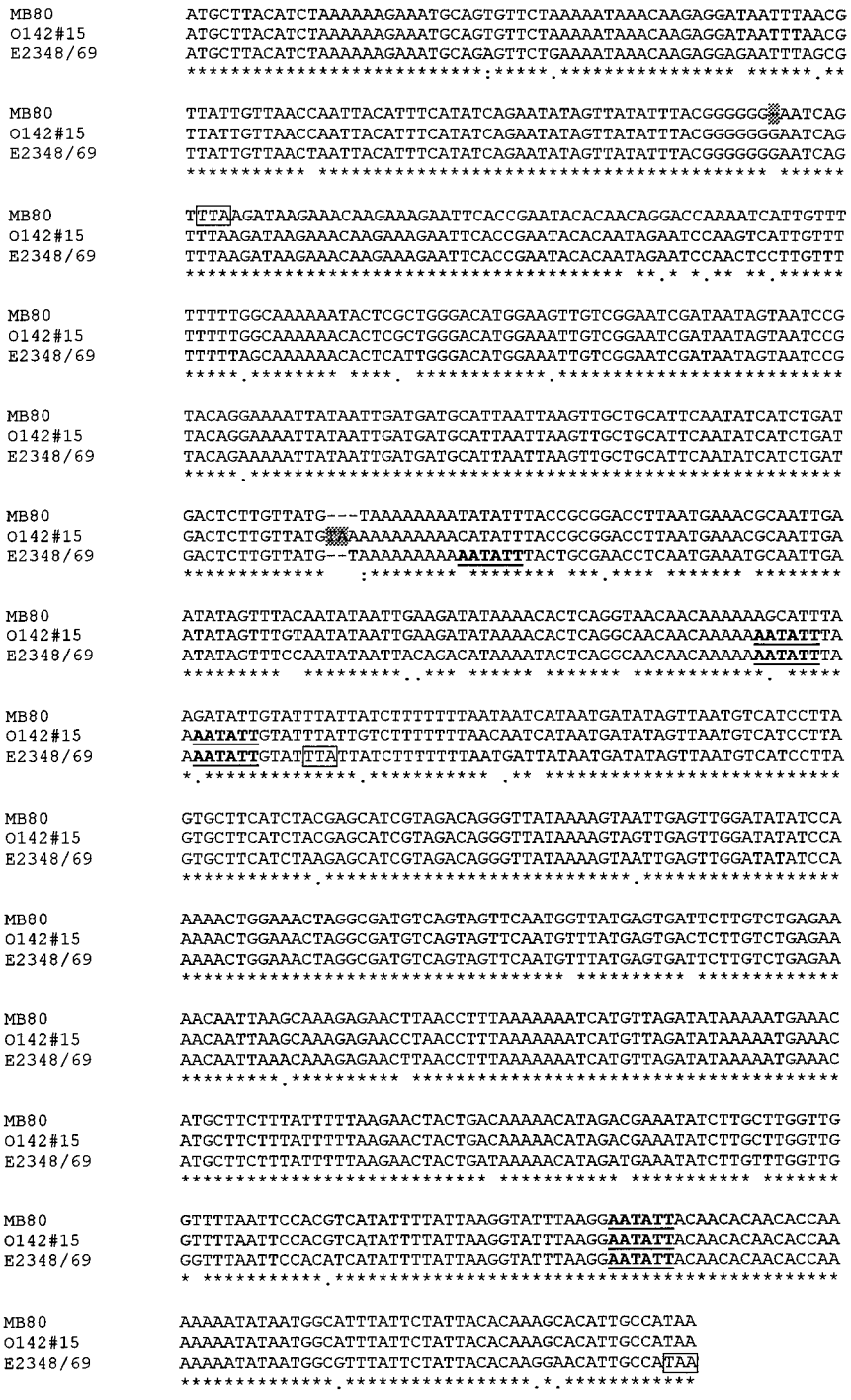

FIG. 2. Alignment of the nucleotide sequences of *perA* alleles from prototype strain E2348/69 (O127:H6) and strains MB80 (O119:H2) and O142#15 (O142:H6), which contain mutating frameshifts. Mutating frameshifts are shaded, and stop codons are boxed. *Ssp*I sites are indicated by bold underlined text.

O119#6 and O142#15 have two to six additional amino acid substitutions in PerB.

The PerA sequence is the most variable, showing 93 to 100% identity at the amino acid level in strains with an intact gene. The predicted DNA-binding domain of PerA is a double helix-turn-helix motif that lies in the C-terminal portion of the protein (18). This region is absent in the strains with frameshifts, in which cases *perA* is prematurely truncated (Fig. 2). The helix-turn-helix motif (amino acids 217 to 259) is 100% conserved at the amino acid (Fig. 3) and nucleotide levels in all *perA* alleles that were not frameshifted. The region covered by amino acids 61 to 121 marks another region of conservation,

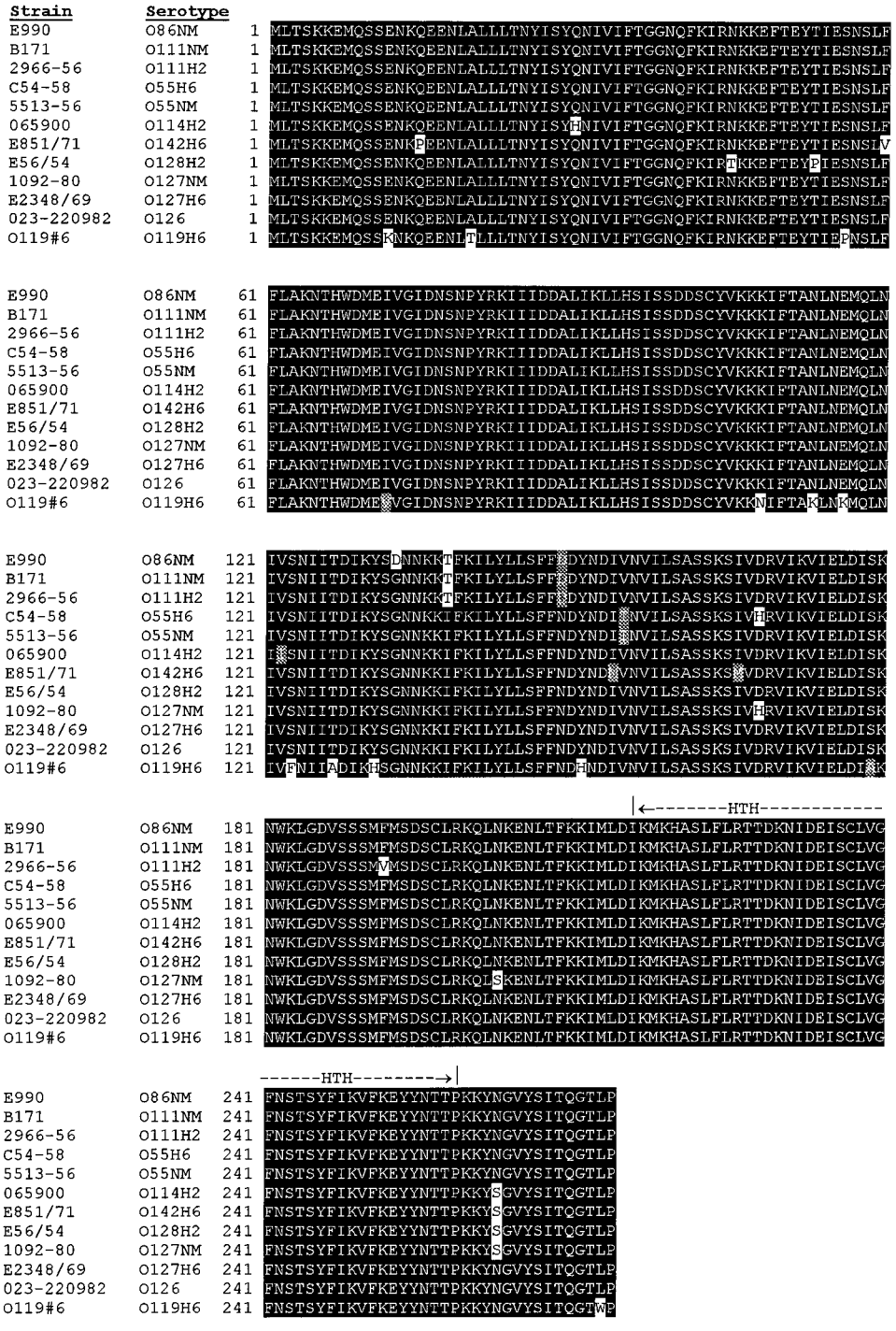

FIG. 3. Alignment of the PerA predicted amino-acid sequences of strains that have a full-length gene showing the locations of conserved (gray background) and nonconserved (white background) substitutions seen in different strains. No substitutions occurred within the DNA-binding helix-turn-helix (HTH) motif region annotated with arrows and dashed lines.

with synonymous substitution seen only in O119:H6 strain O119#6, which was the least conserved *perA* allele.

The expression of Per-regulated genes changes with environmental factors (42). The basis for this observation may be that Per itself is regulated by other factors. One such factor, the chromosomal AraC homolog GadX (YhiX), is a negative regulator of *per* transcription (48). Furthermore, PerA activates its own promoter, with the nucleotides between positions −54 and −81 relative to the transcriptional start site being critical for autoactivation (26). We hypothesized that sequence variation in the region upstream of the *perA* gene may affect the binding of regulators such as GadX as well as autoregula-

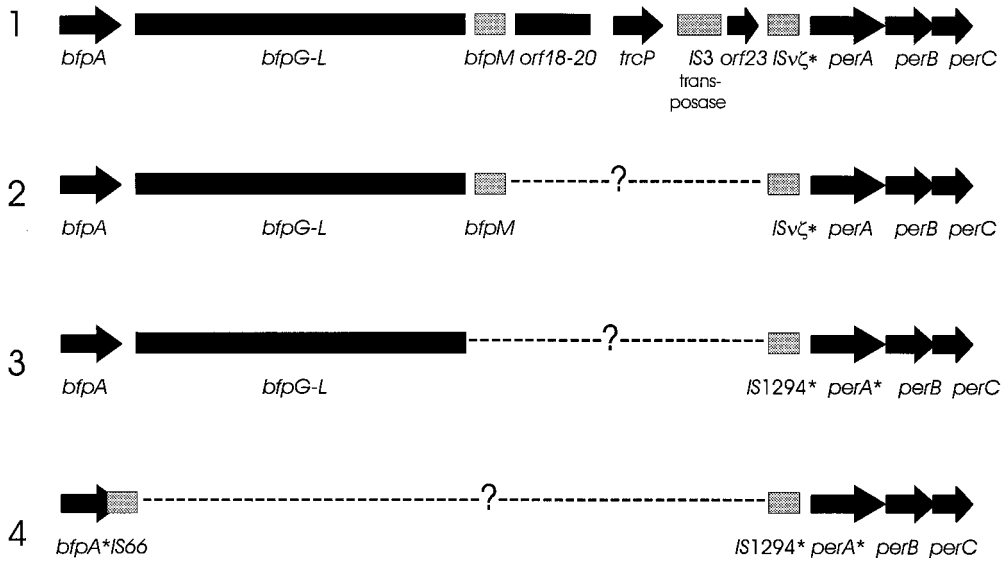

FIG. 4. The *bfp-per* region of EAF plasmids from different EPEC strains. Diagram 1, EAF plasmid of B171-8 representative of typical EPEC2 strains belonging to O111, O86, and O126 serogroups; diagram 2, EAF plasmid of E2348/69, representative of EPEC1 strains O55, O119:H6, and O127 and some O128 and O142 strains; diagram 3, EAF plasmid of O142#15, representative also of O142#19; diagram 4, EAF plasmid of MB80, representative of the O119:H2, O128:H2 (MB21), and canine strains. Asterisks indicate truncated genes, and sequences for which no data are available are represented by dashed lines.

tion and therefore determined the sequences of this region in all strains in which *per* was sequenced. We found that strains CA 4225 (canine) and MB80 (O119:H2), which have a frameshift at nucleotide 114 in *perA*, have two adenine residues deleted at positions −50 and −49. This deletion lies within the region of the promoter that is homologous with other promoters that are activatable by Per (26) and may conceivably affect the transcription of *perBC*.

The sequence upstream of the *perA* promoter region varies in different strains. The sequence of the EAF plasmid from EPEC strain B171-8 reveals that a partial IS*1*(νξ) homolog lies immediately upstream of the *per* operon (52). The region upstream of the *per* promoter (−500 bp relative to the transcriptional start site of *perA*) was shown by sequencing to be almost identical to the B171-8 sequence in O55, O86, O111, O114, O119:H6, O126, and O127 strains and subsets of O128:H2 and O142 strains. In O119:H2, O142:H6, and the canine isolate, this sequence is replaced by a partial IS*1294* element inserted 180 bp upstream of the *perA* start codon (Fig. 4). We carried out multiplex PCR with primers K1547, K1549, and K1978, which gives a 500-bp product in strains with the IS*1294* element upstream of *perA*, compared with the 650-bp product seen in strains that have an upstream IS*1*(νξ). We were able to determine that the IS*1294* element was restricted to the strains that have frameshifts in *perA*, namely, MB80 (O119:H2), CA 4225 (canine), MB21 (O128:H2), O142#15, and O142#19 (O142:H6).

Upstream of the IS*1*(νξ) homolog of EAF plasmid pB171 lies the *trcP* gene (52). This gene is located between the *bfp* and *per* operons and is bound on either side by insertion (IS) elements. The average percentages of G+C contents of the *bfp* operon (38%), *trcP* (30%), and *perABC* (30%) are considerably lower than that seen for most of the plasmid (46%), and the three regions are interspersed by transposases or IS elements (52). We could amplify the 2.5-kb region between *trcP* and *perA* by PCR in all six O111 strains as well as the O114, O126, and O86 strains but in only one other strain (O142:H6 isolate 012-050982), suggesting that this region is different in other EPEC serogroups (Fig. 4). PCR with primers K2019 and K2136 specific for *trcP* and hybridization with a DNA probe showed that most EPEC1 and a few EPEC2 (O119:H2 and O128:H2) strains lack the *trcP* gene (Table 1).

Genetic relationships among *per* sequences and correlation with multilocus enzyme electrophoresis-derived EPEC clonal lineages. A phylogenetic tree generated from alignments of the *perABC* sequence and upstream (to −106 bp) nucleic acid sequence shows the close homology of the sequences (Fig. 5). Although all the sequences are closely related, there are two major clusters within the tree. The first comprises *per* from strains MB80 and O119#6, O142#15, and CA 4225, which cluster away from all the other strains. A subcluster of the second group comprises O55, O127:H6, O126, O86, and O111 *per* genes. Strains within this subcluster include conventional EPEC1 (O55, O127) and EPEC2 (O86, O111, and O126) strains (56) and possess α-bundlin alleles (according to the classification of Blank et al. [6]). The *per* sequences from strains 0065900 (O114:H2), 1092-80 (O127:NM), and E851/71 (O142:H6) did not cluster with either of the two major groups but are closer to the EPEC1 and EPEC2 strains than to those with frameshifts. Interestingly, strains with IS*1294* elements upstream of *perA*—O142#15 (O142:H6), MB80 (O119:H2), and CA 4225 (canine)—and the majority of the *trcP*-positive strains (those belonging to the O126, O86, and O111 serogroups but not the O114:H2 strain 0065900) form distinct clusters even though these sequences were not included in the region from which the tree was generated.

Cloned frameshifted *perA*, *-B*, and *-C* alleles are unable to activate virulence gene promoters. Figure 6 shows that cloned

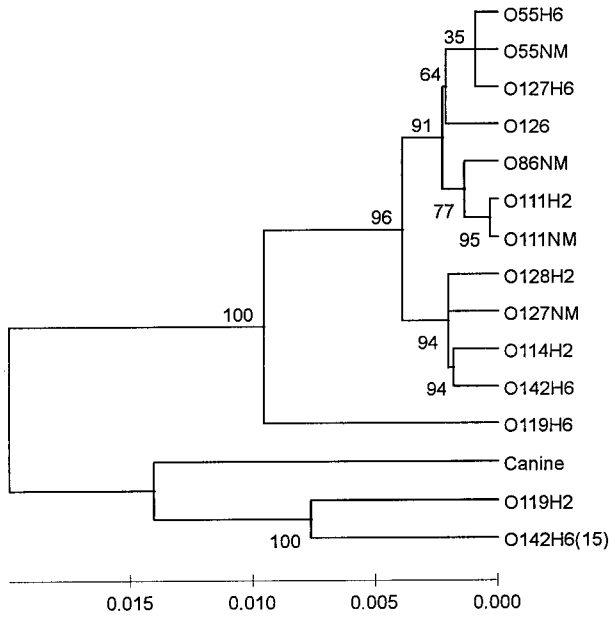

FIG. 5. Unrooted tree for the nucleotide sequence of *perABC* and the upstream 106 nucleotides derived from the neighbor-joining alogarithm based on the gamma distance with α equal to 2. Bootstrap values from 1,000 tree replicates generated using MEGA2β are given at the branch points. The scale on the *x* axis indicates the number of substitutions per 100 nucleotides. Aligned sequences used to construct the tree were obtained from strains C54-58 (O55:H6), 5513-56 (O55:NM), E990 (O86:NM), 2966-56 (O111:H2), B171-8 (O111:NM), 065900 (O114:H2), MB80 (O119:H2), O119#6 (O119:H6), 023-220982 (O126), E2348/69 (O127:H6), 1092-80 (O127:NM), E56/54 (O128:H2), 15 (O142:H6)(15), E851/71 (O142:H6), and CA 4225 (canine).

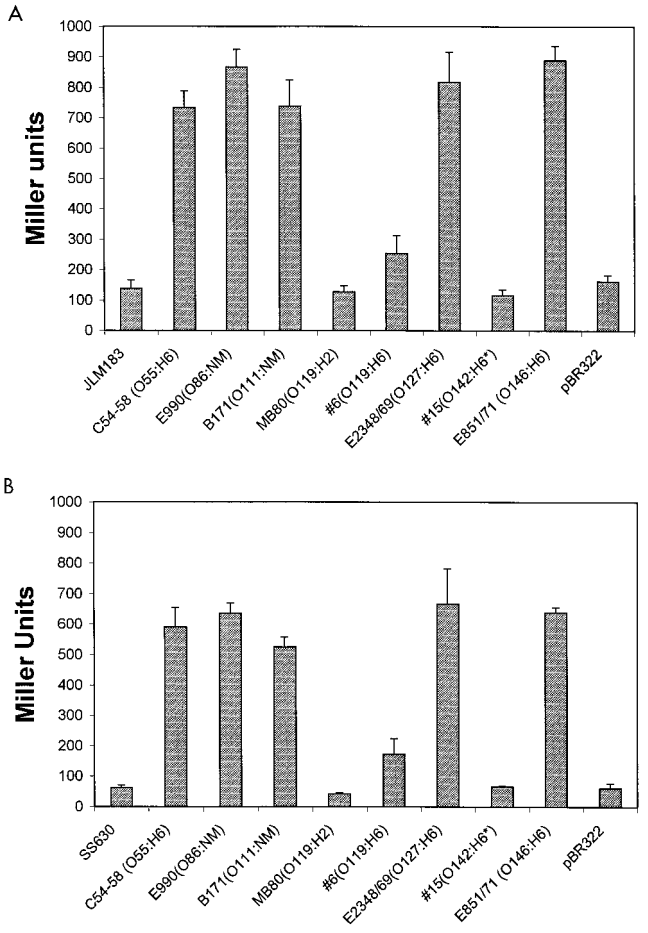

FIG. 6. β-Galactosidase activities on *LEE1* (A) and *bfpA* (B) promoter fusions of cloned *perABC* genes from different strains. (A) Activation of fusion JLM183 (−211 to +221 relative to the *LEE1* transcriptional start site fused to *lacZ*); (B) activation of fusion SS630 (−688 to +124 relative to the *bfpA* transcriptional start site fused to *lacZ*). Clones from strains C54-58 (O55:H6), E990 (O86:NM), B171 (O111:NM), and E851/71(O142:H6) were able to activate these fusions at levels comparable to those of reference strain E2348/69 (O127:H6). Clones from strains MB80 (O119:H2) and O142#15 (O142:H6), with frameshifts in *perA*, were unable to activate the fusions above background while strain O119#6 (O119:H6) showed activation that was significantly above background ($P < 0.04$) but significantly below what was seen with clones that contained untruncated *perA* genes ($P < 0.01$).

per genes from O55:H6 (C54-58), O86:NM (E990), O111:NM (B171-8), O127:H6 (E2348/69) and O142:H6 (E851/71) isolates were able to mediate 4- to 5-fold increases in the activation of the LEE1 gene ($P < 0.0005$) and 9- to 11-fold increases in the activation of *bfpA* ($P < 0.0001$) β-galactosidase promoter fusions. Conversely, clones from O119:H2 (MB80) and O142:H6 (O142#15) strains with frameshifts in *perA* were unable to activate these promoters above background level ($P > 0.05$). The cloned *perA*, *-B*, and *-C* genes from the O119:H6 strain O119#6 were able to activate both promoters but to lower levels: two- and threefold for the LEE1 gene and *bfpA* promoter fusions, respectively ($P < 0.04$). Although this level of activation was statistically significant, it was significantly lower than that seen with clones from E2348/69 and other strains with fully active *per* genes ($P < 0.01$).

Strains bearing inactive *per* genes can be identified by PCR-RFLP typing. We observed two varieties of frameshifts. The first was seen in the MB80 (O119:H2) and CA 4225 (canine) isolates. The second was seen in O142:H6 strain O142#15. In both alleles, there is an alteration in *Ssp*I restriction profile of the gene due to point mutations (Fig. 2). We therefore devised a PCR-RFLP analysis in which the *perA* gene is amplified with primers K1694 and K1693 and digested with *Ssp*I. We were able to distinguish three different patterns by agarose gel electrophoresis. Using this PCR-RFLP test, we were able to identify other strains with these alleles without sequencing (Fig. 7). O128:H2 strain MB21 produced a pattern consistent with that of the MB80 *perA* allele, and O142#19 (H6) showed a *perA* RFLP pattern similar to that of O142#15 (H6).

DISCUSSION

EPEC strains were first described as diarrheagenic *E. coli* more than 50 years ago. However, insights into their pathogenic mechanisms have been uncovered only in the last 2 decades (36). The first molecular correlate of EPEC pathogenicity to be described was the EAF plasmid, which has subsequently been shown to carry genes encoding the BFP, Per, and other putative virulence factors (12, 18, 52). All the currently known virulence genes in prototype EPEC strain E2348/69 are activated by Per, either directly or via the LEE gene-encoded regulator, Ler (13, 18, 26, 30, 53, 54). Per is known to be

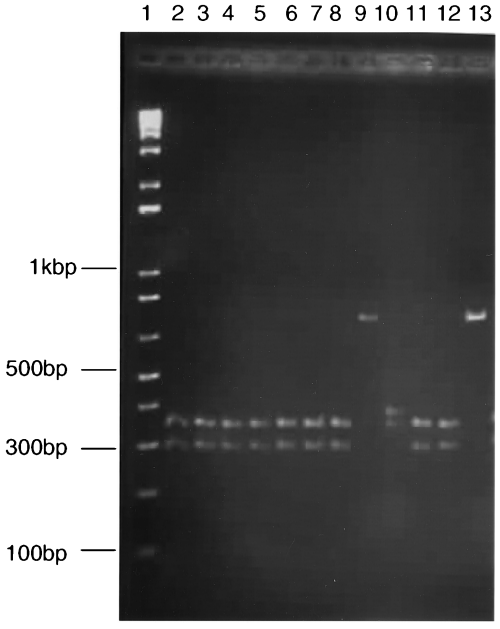

FIG. 7. *Ssp*I-based PCR-RFLP typing distinguishes normal *perA* alleles (lanes 2 to 8 and 11 to 12) from those with O119:H2, O128:H2, or canine-type frameshifts (lanes 9 and 13) and O142:H6-type frameshifts (lanes 10). Test strains shown are E2348/69 (O127:H6) (lane 2), C54-58 (O55:H6) (lane 3), 5513-56 (O55:NM) (lane 4), E990 (O86:NM) (lane 5), 2966-56 (O111:H2) (lane 6), B171-8 (O111:NM) (lane 7), 065900 (O114:H2) (lane 8), MB21 (O128:H2) (lane 9), O142#15 (O142:H6) (14) (lane 10), C771 (O142:H6) (lane 11), E851/71 (O142:H6) (lane 12), and CA 4225 (canine) (lane 13). Lane 1, 1-kb ladder plus (Gibco, BRL).

required for the full virulence of EPEC (5). We therefore sequenced the *perABC* operons from representative EPEC strains and their upstream regions to determine if sequence variation could account for differences in levels of virulence gene expression and the epidemiological prominence of some EPEC serotypes.

A study by Nataro et al. (37) showed that EAF plasmids are at least partially conserved, and Gomez-Duarte and Kaper (18) noted that most EPEC strains harbor *per* genes. By examining the prevalence of *perABC* and other EPEC sequences, we were able to corroborate both reports and infer that, although there is some heterogeneity between EAF plasmids, the presence of the *bfpA* or *per* gene may be considered definitive for EAF plasmids. This is because all EAF-positive EPEC strains possess as least vestiges of the *bfpA* gene and all but one strain possess functional or mutated *perA*, *-B*, and *-C* genes, even though the intervening sequences differ among strains. The only other ubiquitous locus was the *stbAB* region, which is found in other plasmids (52). The product of the *bfpA* gene, bundlin, is surface expressed during infection and is therefore subject to evolutionary pressure from the host cell response, in a manner similar to that of intimin, evident from the mosaic nature of the surface-exposed region of BfpA (6). In addition, a minority of EPEC strains have lost the 3′ end of their *bfpA* gene along with the rest of the operon (8). Per proteins are expressed intracellularly, and with the exception of strain O119#5, the *perABC* region was present in all the test strains. The *per* genes therefore represent a good choice for examining the phylogeny of EPEC plasmids in addition to evaluating structure-function relationships of the operon.

Most molecular and volunteer studies with EPEC have been conducted with prototype EPEC1 and EPEC2 strains, E2348/69 (O127:H7) and B171-8 (O111:NM), respectively. The *perA*, *-B*, and *-C* genes were cloned and sequenced by Gomez-Duarte and Kaper from strain E2348/69 (18). Subsequently, almost identical *per* genes were reported from strain B171-8 and designated *bfpTVW* (53). Comparison of these sequences showed that *perBC* and *bfpVW* are identical and that *bfpT* and *perA* differ very slightly (21). By sequencing the *per* operon from other strains, we demonstrate that, although these genes are highly conserved, point mutations and frameshifts in a few strains lead to reduction or abolishment of activity. We found that strains in which the *perA* gene is truncated before the C-terminal helix-turn-helix motif of the protein do not express a functional plasmid-encoded regulator. In all other strains, this helix-turn-helix region is 100% conserved. In all but the O119:H6 strain O119#6, which has multiple substitutions in *per*, all the *per* operons that were cloned demonstrated full, or in the case of those with frameshifted *perA* genes, no detectable activity. In O119#6, activity is significantly reduced but none of the many substitutions occur in the C-terminal DNA-binding domain of *perA*, suggesting that other regions are required for full activity.

Unlike the results seen with some virulence genes, notably *eae* (encoding intimin) (29), we found very little variation in the sequences of the *perABC* operons from different EPEC strains. Sequence comparison between the EPEC E2348/69 and enterohemorrhagic *E. coli* (EHEC) EDL933 LEE pathogenicity islands suggests that, while genes encoding proteins required for the regulation and delivery of virulence proteins (such as Ler and the type III secretion apparatus) are highly conserved, those encoding proteins that are exposed to the host cell (intimin and secreted proteins) are more divergent (41).

Although the sequence immediately upstream of the *per* genes is highly conserved, the sequence further upstream of this 180 bp is strikingly different. The presence of both the *bfp* and *per* operons in different strains and the homology between the *bfpA* and *perA* promotor regions (26) appears to suggest that the regions were acquired together and that the intervening DNA was acquired later. That different IS elements are found in this region may suggest that it is a hot spot for recombination or transposition. Because the *bfp* and *perABC* operons and the *trcP* gene of strain B171-8 have G+C content percentages significantly below that of most of the EAF plasmid (30 to 38% relative to 46%) and the G+C content percentage of intervening DNA is much higher (42 to 60%) (52), we cannot rule out the possibility that all three loci were acquired independently. Other IS elements have been found flanking virulence genes on large plasmids in pathogenic *E. coli*, suggesting that they play key roles in the evolution of virulence plasmids (9, 10, 25). Tobe et al. found that 29.5% of the sequence of plasmid pB171-8 was comprised of IS elements, many of which were truncated and most of which were located at truncation sites within other genes (52). This study and that of Bortoloni et al. (8) have found IS elements associated with the plasmids of other strains that are not present in the B171-8 plasmid. The variability of gene content among

strains and the abundance of IS elements suggest that these elements may have played an important role in generating the variability in EAF plasmids among strains.

Cloned *perABC* operons from MB80 and O142#15 failed to activate *bfpA* and LEE1 gene fusions, and we therefore conclude that Per in these strains is inactive. We were able to detect expressed intimin but not bundlin in MB80 (O119:H2) and MB21 (O128:H2) strains with inactivating frameshifts in *perA*. The lack of expression of bundlin is, however, a function of a conserved deletion in the *bfp* operon at the 3′ end of *bfpA* due to an IS*66* insertion, previously identified by Bortolini et al. (8); hence, their inability to express bundlin is expected. The observation that the canine strain is similarly truncated in the *bfp* operon and *perA* genes provides an explanation for the findings of Beaudry et al., who observed that although *bfpA*- and EAF-homologous sequences are present in canine EPEC strains, most of these strains (including CA 4225) were unable to demonstrate LA (3). The ability of these strains to express intimin and CesD and of strain O142#15 (O142:H6) to be unimpaired in bundlin formation can be explained by analogy to the related pathogen EHEC, which possesses the LEE but lacks the EAF plasmid and its *per* genes. It is reasonable to believe that, like EHEC, EPEC strains that lack functional *perABC* operons regulate their virulence genes by other mechanisms. The absence of activity in *perABC* clones that are truncated before the helix-turn-helix motif in PerA provides evidence that this region is essential for activity and that the reduced activity in the clone from O119#6 (an H6 strain) suggests that other residues within the PerA protein are required for full activity. The *perABC* operon is the only EAF operon not involved in plasmid maintenance that has not been found to be truncated or interrupted by IS elements in at least some EPEC strains. This may suggest that there is some pressure to maintain intact *per* genes. Interestingly, although O142#15 (H6 strain) was found to have an inactivating frameshift mutation in *perA*, it showed no deficiency in virulence gene expression or phenotype in this study. Conversely, strain 2966-56 (O111:H2) possessed an intact *per* operon but failed to elaborate detectable intimin or to elicit the LA and A/E phenotypes. These strains warrant further study.

The identification of EPEC isolates from cases of diarrhea that lack a functional plasmid-encoded regulator poses questions about the pathogenic potential of these strains. We found that most of these strains were deficient in the expression of key virulence genes and in elaborating LA and A/E phenotypes in in vitro assays. Although there is currently no animal model for EPEC infection, it is known that EPEC strains that lack the EAF plasmid are rarely associated with diarrheal disease in epidemiological studies and, although not completely attenuated, are less likely to elicit diarrhea in human volunteers (5, 24, 36, 55). Data from this study and that of Bortolini et al. (8) imply that some strains lack the two major virulence factors present on the EAF plasmid. An antecedent from this is the fact that major serotypes associated with sporadic diarrhea and outbreaks caused by EPEC are O55:H6, O111:H2/NM, O114:H2, O119:H6, and O127:H6 (2, 55), all of which were found to possess intact *perABC* operons in this study. Also, the majority of O119 strains recovered from patients with diarrhea express the H6 flagellar antigen while O119:H2 isolates are relatively rare (19). The relative success of the H6 strains may be related to their possession of intact *bfp* and *per* operons. It is desirable to determine the epidemiological significance of *per* by determining the relative frequency of strains with nonfunctional operons among the heterogeneous O128:H2 and O142:H6 serotypes. Using tools generated in this study, it should be possible to examine a large number of EPEC by PCR-RFLP typing and multiplex PCR with a view to determining which strains contain frameshifts in *perA* and possess an IS*1294* operon upstream of the *per* operon.

The results of this study have provided further evidence for the essential role of *per* in the pathogenicity of disease due to EPEC strains belonging to major enteropathogenic clones. Furthermore, in spite of the high level of conservation associated with *per*, EAF plasmids appear to be show some variation in gene arrangement and content within the *bfp-per* regions and at other loci.

ACKNOWLEDGMENTS

This work was supported by grant AI 21657 from the National Institutes of Health, and sequencing was conducted and funded by the University of Maryland Biopolymer Facility.

We thank Nicholas Ambolous and Lisa Sadzewicz of the biopolymer lab for sequences and T. Eric Blank, Michael S. Donnenberg, Jorge A. Girón, Vanessa Sperandio, and Thomas S. Whittam for strains and helpful suggestions.

REFERENCES

1. **Adu-bobie, J., G. Frankel, C. Bain, A. Goncalves, L. Trabulsi, G. Douce, S. Knutton, and G. Dougan.** 1998. Detection of intimins α, β, γ, and δ, four intimin derivatives expressed by attaching and effacing microbial pathogens. J. Clin. Microbiol. **36:**662–668.
2. **Ansaruzzaman, M., M. Albert, S. Nahar, R. Byun, M. Katouli, I. Kuhn, and R. Mollby.** 2000. Clonal groups of enteropathogenic *Escherichia coli* isolated in case-control studies of diarrhoea in Bangladesh. J. Med. Microbiol. **49:** 177–185.
3. **Beaudry, M., C. Zhu, J. M. Fairbrother, and J. Harel.** 1996. Genotypic and phenotypic characterization of *Escherichia coli* isolates from dogs manifesting attaching and effacing lesions. J. Clin. Microbiol. **34:**144–148.
4. **Beutin, L., I. Orskov, F. Orskov, S. Zimmermann, J. Prada, 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.
5. **Bieber, D., S. Ramer, C.-Y. Wu, W. Murray, T. Tobe, R. Fernandez, and G. Schoolnik.** 1998. Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. Science **280:**2114–2118.
6. **Blank, T., H. Zhong, A. Bell, T. Whittam, and M. Donnenberg.** 2000. Molecular variation among type IV pilin (*bfpA*) genes from diverse enteropathogenic *Escherichia coli* strains. Infect. Immun. **68:**7028–7038.
7. **Blattner, F., G. R. Plunkett, C. Bloch, N. Perna, V. Burland, M. Riley, J. Collado-Vides, J. Glasner, C. Rode, G. Mayhew, J. Gregor, N. Davis, H. Kirkpatrick, M. Goeden, D. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1474.
8. **Bortolini, M., L. Trabulsi, R. Keller, G. Frankel, and V. Sperandio.** 1999. Lack of expression of bundle-forming pili in some clinical isolates of enteropathogenic *Escherichia coli* (EPEC) is due to a conserved large deletion in the *bfp* operon. FEMS Microbiol. Lett. **179:**169–174.
9. **Brunder, W., H. Schmidt, M. Frosch, and H. Karch.** 1999. The large plasmids of Shiga-toxin-producing *Escherichia coli* are highly variable genetic elements. Microbiology **145:**1005–1014.
10. **Burland, V., Y. Shao, N. Perna, G. Plunkett, H. Sofia, and F. Blattner.** 1998. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. Nucleic Acids Res. **26:**4196–4204.
11. **Cravioto, A., A. Tello, A. Navarro, J. Ruiz, H. Villafán, F. Uribe, and C. Eslava.** 1991. Association of *Escherichia coli* HEp-2 adherence patterns with type and duration of diarrhoea. Lancet **337:**262–264.
12. **Donnenberg, M. S., J. A. Girón, J. P. Nataro, and J. B. Kaper.** 1992. A plasmid-encoded type IV fimbrial gene of enteropathogenic *Escherichia coli* associated with localized adherence. Mol. Microbiol. **6:**3427–3437.
13. **Elliott, S. J., V. Sperandio, J. A. Girón, S. Shin, J. L. Mellies, L. Wainwright, S. W. Hutcheson, T. K. McDaniel, and J. B. Kaper.** 2000. The locus of enterocyte effacement (LEE)-encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic *Escherichia coli*. Infect. Immun. **68:**6115–6126.
14. **Elliott, S. J., L. A. Wainwright, T. K. McDaniel, K. G. Jarvis, Y. K. Deng, L.**

C. Lai, B. P. McNamara, M. S. Donnenberg, and J. B. Kaper. 1998. The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. Mol. Microbiol. **28:**1–4.
15. **Fields, P., K. Blom, H. Hughes, L. Helsel, P. Feng, and B. Swaminathan.** 1997. Molecular characterization of the gene encoding H antigen in *Escherichia coli* and development of a PCR-restriction fragment length polymorphism test for identification of *E. coli* O157:H7 and O157:NM. J. Clin. Microbiol. **35:**1066–1070.
16. **Girón, J. A., M. S. Donnenberg, W. C. Martin, K. G. Jarvis, and J. B. Kaper.** 1993. Distribution of the bundle-forming pilus structural gene (*bfpA*) among enteropathogenic *Escherichia coli*. J. Infect. Dis. **168:**1037–1041.
17. **Girón, J. A., A. S. Ho, and G. K. Schoolnik.** 1991. An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. Science **254:**710–713.
18. **Gómez-Duarte, O. G., and J. B. Kaper.** 1995. A plasmid-encoded regulatory region activates chromosomal *eaeA* expression in enteropathogenic *Escherichia coli*. Infect. Immun. **63:**1767–1776.
19. **Gonçalves, A. G., L. C. Campos, T. A. Gomes, J. Rodrigues, V. Sperandio, T. S. Whittam, and L. R. Trabulsi.** 1997. Virulence properties and clonal structures of strains of *Escherichia coli* O119 serotypes. Infect. Immun. **65:** 2034–2040.
20. **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:**7839–7843.
21. **Kaper, J. B., and O. G. Gomez-Duarte.** 1997. The *per* regulator of enteropathogenic *Escherichia coli*. Mol. Microbiol. **23:**179–181.
22. **Knutton, S., A. Phillips, H. Smith, R. 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., E. J. Bergquist, D. R. Nalin, D. H. Waterman, R. B. Hornick, C. R. Young, and S. Sotman.** 1978. *Escherichia coli* strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. Lancet **1:**1119–1122.
24. **Levine, M. M., J. P. Nataro, H. Karch, M. M. Baldini, J. B. 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. **Makino, K., K. Ishii, T. Yasunaga, M. Hattori, K. Yokoyama, C. Yutsudo, Y. Kubota, Y. Yamaichi, T. Iada, K. Yamamoto, T. Honda, C.-G. Han, E. Ohtsubo, M. Kasamatsu, T. Hayashi, S. Kuhara, and H. Shingawa.** 1998. Complete nuclceotide sequence of 93-kb and 3.3-kb plasmids of enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. DNA Res. **5:**1–9.
26. **Martinez-Laguna, Y., C. Calva, and J. Puente.** 1999. Autoactivation and environmental regulation of *bfpT* expression, the gene coding for the transcriptional activator of *bfpA* in enteropathogenic *Escherichia coli*. Mol. Microbiol. **33:**153–166.
27. **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.
28. **McDaniel, T. K., and J. B. Kaper.** 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol. Microbiol. **23:**399–407.
29. **McGraw, E., J. Li, R. Selander, and T. Whittam.** 1999. Molecular evolution and mosaic structure of alpha, beta, and gamma intimins of pathogenic *Escherichia coli*. Mol. Biol. Evol. **16:**12–22.
30. **Mellies, J. L., S. J. Elliott, V. Sperandio, M. S. Donnenberg, and J. B. Kaper.** 1999. The Per regulon in enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). Mol. Microbiol. **33:** 296–306.
31. **Mellies, J. L., F. Navarro-Garcia, I. Okeke, J. Frederickson, J. P. Nataro, and J. B. Kaper.** 2001. *espC* pathogenicity island of enteropathogenic *Escherichia coli* encodes an enterotoxin. Infect. Immun. **69:**315–324.
32. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press Cold Spring Harbor, N.Y.
33. **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.
34. **Munson, G., L. Holcomb, and J. Scott.** 2001. Novel group of virulence activators within the AraC family that are not restricted to upstream binding sites. Infect. Immun. **69:**186–193.
35. **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.
36. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142–201.
37. **Nataro, J. P., K. O. Maher, P. Mackie, and J. B. Kaper.** 1987. Characterization of plasmids encoding the adherence factor of enteropathogenic *Escherichia coli*. Infect. Immun. **55:**2370–2377.
38. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
39. **Okeke, I., A. Lamikanra, H. Steinrück, and J. B. Kaper.** 2000. Characterization of *Escherichia coli* strains from cases of childhood diarrhea in provincial southwestern Nigeria. J. Clin. Microbiol. **38:**7–12.
40. **Órskov, F., T. S. Whittam, A. Cravioto, and I. Órskov.** 1990. Clonal relationships among classic enteropathogenic *Escherichia coli* (EPEC) belong to different O groups. J. Infect. Dis. **162:**76–81.
41. **Perna, N. T., G. F. Mayhew, G. Posfai, S. Elliott, M. S. Donnenberg, J. B. Kaper, and F. R. Blattner.** 1998. Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. Infect. Immun. **66:**3810–3817.
42. **Puente, J., D. Bieber, S. Ramer, W. Murray, and G. Schoolnik.** 1996. The bundle-forming pili of enteropathogenic *Escherichia coli*: transcriptional regulation by environmental signals. Mol. Microbiol. **20:**87–100.
43. **Reid, S., C. Herbelin, A. Bumbaugh, R. Selander, and T. Whittam.** 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature **406:** 64–67.
44. **Riley, L., R. Remis, S. Helgerson, H. McGee, J. Wells, B. Davis, R. Hebert, E. Olcott, L. Johnson, N. Hargrett, P. Blake, and M. Cohen.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N. Engl. J. Med. **308:**681–685.
45. **Robins-Browne, R. M., W. C. Yam, L. E. O'Gorman, and K. A. Bettelheim.** 1993. Examination of archetypal strains of enteropathogenic *Escherichia coli* for properties associated with bacterial virulence. J. Med. Microbiol. **38:**222–226.
46. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
47. **Savarino, S. J., A. McVeigh, J. Watson, A. Cravioto, J. Molina, P. Echeverria, M. K. Bhan, M. M. Levine, and A. Fasano.** 1996. Enteroaggregative *Escherichia coli* heat-stable enterotoxin is not restricted to enteroaggregative *E. coli*. J. Infect. Dis. **173:**1019–1022.
48. **Shin, S., M.-P. Castanie-Cornet, J. W. Foster, J. A. Crawford, and J. B. Kaper.** An activator of glutamate decarboxylase genes regulates expression of enteropathogenic *Escherichia coli* virulence genes through control of the plasmid-encoded regulator, Per. Mol. Microbiol, in press.
49. **Simons, R., F. Houman, and N. Kleckner.** 1987. Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. Gene **53:** 85–96.
50. **Sperandio, V., J. B. Kaper, M. R. Bortolini, B. C. Neves, R. Keller, and L. R. Trabulsi.** 1998. Characterization of the locus of enterocyte effacement (LEE) in different enteropathogenic *Escherichia coli* (EPEC) and Shigatoxin producing *Escherichia coli* (STEC) serotypes. FEMS Microbiol. Lett. **164:**133–139.
51. **Stone, K., H. Zhang, L. Carlson, and M. Donnenberg.** 1996. A cluster of fourteen genes from enteropathogenic *Escherichia coli* is sufficient for the biogenesis of a type IV pilus. Mol. Microbiol. **20:**325–337.
52. **Tobe, T., T. Hayashi, C. Han, G. Schoolnik, E. Ohtsubo, and C. Sasakawa.** 1999. Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. Infect. Immun. **67:**5455–5462.
53. **Tobe, T., G. Schoolnik, I. Sohel, V. Bustamante, and J. Puente.** 1996. Cloning and characterization of *bfpTVW*, genes required for the transcriptional activation of *bfpA* in enteropathogenic *Escherichia coli*. Mol. Microbiol. **21:** 963–975.
54. **Tobe, T., I. Tatsuno, E. Katayama, C. Wu, G. Schoolnik, and C. Sasakawa.** 1999. A novel chromosomal locus of enteropathogenic *Escherichia coli* (EPEC), which encodes a *bfpT*-regulated chaperone-like protein, TrcA, involved in microcolony formation by EPEC. Mol. Microbiol. **33:**741–752.
55. **Whittam, T., and E. McGraw.** 1995. Clonal analysis of EPEC serogroups. Rev. Microbiol. **27:**7–16.
56. **Whittam, T. S., M. L. Wolfe, I. K. Wachsmuth, F. Orskov, I. Orskov, and R. A. Wilson.** 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. Infect. Immun. **61:**1619–1629.
57. **Zhu, C., T. A. Agin, S. J. Elliott, L. A. Johnson, T. E. Thate, J. B. Kaper, and E. C. Boedeker.** 2001. Complete nucleotide sequence and analysis of the locus of enterocyte effacement from rabbit diarrheagenic *Escherichia coli* RDEC-1. Infect. Immun. **69:**2107–2115.

Editor: V. J. DiRita