Journal of Medical Microbiology (2012), **61**, 1114–1120 **DOI** 10.1099/jmm.0.045443-0

Genetic diversity of locus of enterocyte effacement genes of enteropathogenic *Escherichia coli* isolated from Peruvian children

C. A. Contreras,[1] T. J. Ochoa,[1,2] J. Ruiz,[3,4] D. W. Lacher,[5] D. Durand,[1] C. DebRoy,[6] C. F. Lanata[7,8] and T. G. Cleary[2]

Correspondence
T. J. Ochoa
Theresa.J.Ochoa@uth.tmc.edu

[1]Instituto de Medicina Tropical 'Alexander von Humboldt', Universidad Peruana Cayetano Heredia, Lima, Peru

[2]University of Texas School of Public Health, Houston, USA

[3]Centre de Recerca en Salut Internacional de Barcelona, Hospital Clinic/Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain

[4]CIBERESP, Barcelona, Spain

[5]Center for Food Safety and Applied Nutrition, US Food and Drug Administration, Laurel, Maryland, USA

[6]E. coli Reference Center, Department of Veterinary and Biomedical Sciences, Pennsylvania State University, Pennsylvania, USA

[7]Instituto de Investigación Nutricional, Lima, Peru

[8]Universidad Peruana de Ciencias Aplicadas, Lima, Peru

The aim of this study was to determine the frequency and allele associations of locus of enterocyte effacement encoded *esp* and *tir* genes among 181 enteropathogenic *Escherichia coli* (EPEC) strains (90 diarrhoea-associated and 91 controls) isolated from Peruvian children under 18 months of age. We analysed *espA*, *espB*, *espD* and *tir* alleles by PCR-RFLP. EPEC strains were isolated with higher frequency from healthy controls (91/424, 21.7 %) than from diarrhoeal samples (90/936, 9.6 %) ($P<0.001$); 28.9 % of diarrhoeal and 17.6 % of control samples were typical EPEC (tEPEC). The distribution of *espA* alleles (alpha, beta, beta2 and gamma) and *espD* alleles (alpha, beta, gamma and a new variant, *espD*-N1) between tEPEC and atypical EPEC (aEPEC) was significantly different ($P<0.05$). *espD*-alpha was more common among acute episodes ($P<0.05$). *espB* typing resulted in five alleles (alpha, beta, gamma and two new sub-alleles, *espB*-alpha2 and *espB*-alpha3), while *tir*-beta and *tir*-gamma2 were the most common intimin receptor subtypes. Seventy-two combinations of *espA*, *espB*, *espD* and *tir* alleles were found; the most prevalent combination was *espA*-beta, *espB*-beta, *espD*-beta, *tir*-beta (34/181 strains), which was more frequent among tEPEC strains ($P<0.05$). Our findings indicate that there is a high degree of heterogeneity among EPEC strains isolated from Peruvian children and that aEPEC and tEPEC variants cluster.

Received 2 April 2012
Accepted 4 April 2012

INTRODUCTION

Enteropathogenic *Escherichia coli* (EPEC) is an important group of diarrhoeal pathogens of young children living in developing countries (Chen & Frankel, 2005; Scaletsky *et al.*, 2002). EPEC is characterized as typical or atypical depending on the presence of the EPEC adherence factor (EAF) virulence plasmid encoding bundle-forming pili, which are associated with a pattern of intestinal epithelial cell attachment known as localized adherence (Donnenberg *et al.*, 1992). A key characteristic of all EPEC is a chromosomally located pathogenicity island named the locus of enterocyte effacement (LEE). Genes located on the LEE encode structural components of a type III secretion–translocation apparatus, factors enabling the bacterium to adhere intimately to intestinal epithelial cells (*eae* and *tir*),

The GenBank/EMBL/DDBJ accession numbers for the sequences of the new *espB* (*espB*-alpha3 and *espB*-alpha2), *espD* (*espD*-N1) and *tir* (*tir*-alpha2) alleles are JN571730, JN571731, JN571732 and JN571733, respectively.

Abbreviations: aEPEC, atypical EPEC; EPEC, enteropathogenic *Escherichia coli*; LEE, locus of enterocyte effacement; tEPEC, typical EPEC.

Supplementary tables are available with the online version of this paper.

secreted and effector proteins, chaperones and transcriptional regulators (Ler, GrlR/A) (Deng *et al.*, 2004). The coordinated expression of these genes finally causes attaching-and-effacing lesions, histopathological lesions characterized by intimate bacterial adherence to the host cell plasma membrane, leading to destruction of the enterocyte microvilli and induction of cytoskeletal rearrangements beneath adherent bacteria (Frankel *et al.*, 1998; Nataro & Kaper, 1998; Wong *et al.*, 2011). The effector proteins are Tir, EspG, EspF, EspZ, Map and EspH; and the translocators are EspA, EspD and EspB, required for translocating the effectors into host cells (Dean & Kenny, 2009). Tir, EspB and EspD are transferred into the host cells through the translocation machinery formed by the type III secretion system. Translocated EspB and EspD are integrated into the cytoplasmic membrane of the target cells and form a pore that allows other molecules to enter these cells (Goffaux *et al.*, 2001). Analysis of the LEE region shows that the core LEE is largely conserved, particularly among genes encoding the type III secretion system, whereas genes encoding effector proteins display a higher degree of variability (Müller *et al.*, 2009). In addition to the high levels of sequence polymorphism observed in the *eae* gene (Blanco *et al.*, 2006; Contreras *et al.*, 2010; Lacher *et al.*, 2006), allelic variants in the *tir* (alpha, beta, gamma and gamma2), *espA* (alpha, beta, beta2 and gamma), *espB* (alpha, beta and gamma) and *espD* (alpha, beta and gamma) genes have also been described (Afset *et al.*, 2008; China *et al.*, 1999; Garrido *et al.*, 2006; Goffaux *et al.*, 2001; Nielsen & Andersen, 2003; Yuste *et al.*, 2008). In contrast, the *esc* (*E. coli* secreted components) and *sep* genes are more conserved (Goffaux *et al.*, 2001). The differentiation of *eae*, *tir* and *esp* alleles is an important tool for EPEC typing as well as in epidemiological and clonal studies (Garrido *et al.*, 2006; Yuste *et al.*, 2008). Few studies have evaluated the associations between the *tir*, *espA*, *espB* and *espD* allele variants (Afset *et al.*, 2008; Garrido *et al.*, 2006). In the present study, we classified EPEC strains based on associations of *esp* and *tir* alleles using PCR-RFLP and investigated the relationships between these *esp*–*tir* associations and the characteristics of the diarrhoeal episodes.

METHODS

Strains. The 181 EPEC strains examined in this study were isolated from a cohort epidemiological study of diarrhoea in infants from 2 to 18 months of age (Contreras *et al.*, 2010; Ochoa *et al.*, 2009). In the study, 1360 stools samples were obtained from 1034 children; 936 samples were from children with diarrhoea (case patients) and 424 were from children without diarrhoea (controls). Five lactose-positive colonies were analysed by real-time PCR to detect diarrhoeagenic *E. coli* as described previously (Guion *et al.*, 2008). One hundred and twenty of the 181 strains were analysed previously for detecting variability in *eae*, *bfpA* and *perA* genes and the association with clinical characteristics (Contreras *et al.*, 2010).

Clinical data and serotyping. Clinical information on the diarrhoeal episodes obtained during the cohort study was analysed by a modified Vesikari score (Ruuska & Vesikari, 1990) to quantify the severity of the episode. The score included: duration of diarrhoea in days (0–3 points), maximum number of stools per day during the episode (1–3), number of days with vomiting (0–3), maximum number of emesis per day during the episode (0–3), presence of fever (0–1), dehydration (0–3) and treatment (0–2). The maximum possible score was 18. A 'mild' score was 0–8 points, a 'moderate' score was 9–14 points and a 'severe' score was 15–18 points. In the analysis, moderate and severe scores were grouped together because of the low number (n=3) of patients with a severe score.

The serotyping was performed at the *E. coli* Reference Center at the Pennsylvania State University according to standard methods (Contreras *et al.*, 2010).

Detection of virulence genes. EPEC isolates were examined for *bfpA*, *espA*, *espB*, *espD* and *tir* genes by PCR using the primers and conditions listed in Table 1. PCR for the five genes was performed in a 25 µl reaction mixture containing 2.5 µl 2.5 mM of each dNTP (Promega), 2.5 µl 10× buffer with 15 mM $MgCl_2$ (Kappa Biosystems), 0.5 U *Taq* polymerase (Kappa Biosystems) and 2 µl DNA template. PCR amplification was performed in a thermocycler (iCycler; Bio-Rad) under the conditions listed in Garrido *et al.* (2006), Guion *et al.* (2008) and Lacher *et al.* (2006) (also see Table S1 available in JMM Online). The mixture was held at 72 °C for 7 min after the final cycle before cooling at −20 °C. Amplified products were analysed by using 1.5 % agarose gel electrophoresis and visualized by staining with ethidium bromide. PCR products were analysed by RFLP. The positive PCR control was tEPEC strain E2348/69 and the negative control was *E. coli* C600.

PCR-RFLP. The *in silico* RFLP tool of the EcMLST website (http://www.shigatox.net/ecmlst/cgi-bin/insilicorflp) was used to predict the restriction patterns of all the genes except *bfpA* (Table 1). One restriction enzyme digest was used for all genes analysed except for

Table 1. Expected RFLPs of *esp* PCR amplicons

The digestion patterns are shown (bp).

espA allele	*Alu*I	*Dde*I	*espB* allele	*Bst*UI	*espD* allele	*Bst*UI	*tir* allele	*Dde*I
alpha	237	173, 64	alpha	444, 147, 135	alpha	335, 335	alpha	201, 157, 90, 85
beta	175, 62	143, 64, 34	beta	523, 135, 56	beta	467, 203	beta	530
beta2	103, 72, 62	173, 64	gamma	523, 78, 56, 45	gamma	377, 275	gamma	236, 192, 99
gamma	175, 62	207, 30	*espB*-alpha2	444, 279	*espD*-N1	670	gamma2	376, 157
							tir-alpha2	291, 157, 85
			espB-alpha3	714			*tir*-N2	291, 147, 89

espA (digested with two separate restriction enzymes). Digestion was performed in 20 µl reaction mixtures at a final concentration of 10 U enzyme (New England BioLabs), 1× reaction buffer, 15.0 µl unpurified PCR product and distilled water to complete the final volume. The samples were incubated overnight at the temperature indicated by each enzyme manufacturer. After incubation, 15 µl each digest was visualized on ethidium bromide-stained 3.0 % agarose gels by illumination with UV light.

Sequencing of the *esp* and *tir* genes. The strains that shared an undefined RFLP pattern were sequenced and analysed to establish the genetic relationship with known alleles. PCR products were purified with the QIAquick PCR Purification kit (Qiagen). Sequencing was performed by Macrogen (Korea) using an automatic DNA sequencer (Applied Biosystems 3730XL) and used for phylogenetic analyses. DNA sequences of *esp* genes were edited with BioEdit v4.8.10 (http://www.mbio.ncsu.edu/bioedit/bioedit.html).

Statistical analysis. The allelic frequencies obtained in each group and the allelic distributions in each population were compared using the program GenALEx 6.3 and GenPop on the web tool (http://genepop.curtin.edu.au/). The comparisons between groups were made using the chi-square or Fisher's exact test. *P*-values <0.05 were considered significant.

Nucleotide sequence accession numbers. The GenBank accession numbers for the sequences of the new *espB* (*espB*-alpha3 and *espB*-alpha2), *espD* (*espD*-N1) and *tir* (*tir*-alpha2) alleles are JN571730, JN571731, JN571732 and JN571733, respectively.

RESULTS

Prevalence of *esp*, *tir* and *bfpA* genes

A total of 181 *E. coli* strains were classified as EPEC based on the presence of *eae* and absence of *stx1/stx2* by PCR. EPEC strains were isolated with higher frequency from healthy controls (91/424, 21.5 %) than from diarrhoea samples (90/936, 9.6 %) ($P<0.0001$). Based on the presence of *bfpA*, 42/181 (23.2 %) strains were classified as tEPEC [26/90 (28.9 %) diarrhoea samples and 16/91 (17.6 %) controls, $P=0.08$].

The *espA* gene was present in 88/90 (97.8 %) diarrhoea samples and 89/91 (97.8 %) controls (Table 2). RFLP typing resulted in four previously reported *espA* alleles: *espA*-alpha [48 strains, 5/42 (11.9 %) in tEPEC vs 43/139 (30.9 %) in aEPEC; $P<0.05$)], *espA*-beta (57 strains), *espA*-beta2 (64 strains) and *espA*-gamma (8 strains). The distribution of these alleles between diarrhoea samples and controls was similar (Table 2), although the difference in the distribution of these alleles among tEPEC and aEPEC was marginally significant.

For *espB*, the primers failed to amplify in 11 strains. Of the *espB*-positive strains (170/181, 93.9 %), typing resulted in three previously reported *espB* alleles, *espB*-alpha (53 strains), *espB*-beta (52 strains) and *espB*-gamma (15 strains), and two new sub-alleles designated *espB*-alpha2 (41 strains) and *espB*-alpha3 (9 strains). The distribution of these alleles was similar between diarrhoea samples and controls, and between tEPEC and aEPEC strains (Table 2).

The *espD*-positive strains were subtyped as *espD*-beta (64 strains), *espD*-N1 (40 strains), *espD*-gamma (28 strains) and *espD*-alpha (19 strains). A high number of strains were *espD*-negative (30/181, 16.6 %). There was no difference in the distribution of *espD* alleles among diarrhoea samples versus controls, while the distribution of *espD* alleles was different between tEPEC vs aEPEC, specifically *espD*-N1 was more common in aEPEC ($P<0.05$).

The *tir* subtyping identified *tir*-beta (52/181, 28.7 %) and *tir*-gamma2 [6/42 (14.3 %) in tEPEC vs 45/139 (32.4 %) in aEPEC; $P<0.05$] as the most common intimin receptor subtypes, followed by *tir*-alpha [14/42 (33.3 %) in tEPEC vs 22/139 (15.8 %) in aEPEC; $P<0.05$], *tir*-alpha2 (21/181,11.6 %) and *tir*-gamma (7/181, 3.9 %). Fourteen strains showed negative PCRs for *tir* with the primers used in this study (Table 2).

Combinations among alleles of LEE genes

Seventy-two combinations of variants of the LEE genes *espA*, *espB*, *espD* and *tir* were present among the 181 EPEC strains (GenAlex) (Tables S2 and S3). Two of these combinations were more prevalent: *espA*-beta, *espB*-beta, *espD*-beta, *tir*-beta (34/181, 18.8 %); and *espA*-alpha, *espB*-alpha2, *espD*-N1, *tir*-gamma2 (21/181, 11.6 %). The first LEE gene combination was associated with tEPEC ($P<0.05$). Overall, the distribution of the combinations of all alleles described (*espA*, *espB*, *espD* and *tir*) was different between tEPEC and aEPEC using GenPop (http://genepop.curtin.edu.au/) ($P<0.05$) (Table S2).

Analysis of sequenced LEE genes

The RFLPs were designed based on the allele sequences available in GenBank for all LEE genes (*espA*, *espB*, *espD* and *tir*) (Table S2). However, many strains did not show a previously described RFLP pattern assigned to the known alleles. The two *espB*, one *espD*, and one *tir* nontypable RFLP patterns were designated *espB*-alpha2, *espB*-alpha3, *espD*-N1 and *tir*-alpha2, respectively. The PCR products of the strains with these RFLP patterns were sequenced. Analysis of the sequenced regions of the *espB*-alpha2 and *espD*-N1 amplicons revealed them to be identical to the *espB* and *espD* sequences found in the complete genome of O111 : H^- strain 11128. The *espB*-alpha3 allele was most similar to the *espB* allele of strain 11128 with 18 polymorphic sites within the sequenced region of the amplicon. The *tir*-alpha2 allele was most similar to the allele found in an O119 : H9 strain (GenBank accession no. AJ633129) with 2 nt substitutions. A new allele found in an ONT:H19 strain named *tir*-N2 (GenBank accession number AB288104) was also identified.

Serotypes

One hundred and twenty of 181 EPEC strains were previously serotyped (Contreras *et al.*, 2010). An O serogroup was identified in 105/181 (58 %) of the strains

Table 2. Distribution of *esp* alleles among tEPEC and aEPEC strains and among diarrhoea and control samples

Genes	Allele	Diarrhoea [*n* (%)] (*n*=90)	Control [*n* (%)] (*n*=91)	tEPEC [*n* (%)] (*n*=42)	aEPEC [*n* (%)] (*n*=139)	All EPEC [*n* (%)] (*n*=181)
espA alleles	*espA*-alpha	24 (26.7)	24 (26.4)	5 (11.9)	43 (30.9)*	48 (26.1)
	espA-beta	30 (33.3)	27 (29.7)	15 (35.7)	42 (30.2)	57 (31.5)
	espA-beta2	33 (36.7)	31 (34.1)	22 (52.4)	42 (30.2)	64 (35.4)
	espA-gamma	1 (1.1)	7 (7.7)	0 (0.0)	8 (5.8)	8 (4.4)
	Negative	2 (2.2)	2 (2.2)	0 (0.0)	4 (2.9)	4 (2.2)
espB alleles	*espB*-alpha	26 (26.7)	27 (29.7)	17 (40.5)	36 (25.9)	53 (29.3)
	espB-beta	31 (34.4)	21 (23.1)	14 (33.3)	38 (27.3)	52 (28.7)
	espB-gamma	8 (8.9)	9 (9.9)	2 (4.8)	13 (9.4)	15 (8.3)
	espB-alpha2	16 (17.8)	25 (27.5)	7 (16.7)	34 (24.5)	41 (22.7)
	espB-alpha3	5 (5.6)	4 (4.4)	1 (2.4)	8 (5.8)	9 (4.9)
	Negative	6 (6.7)	5 (5.5)	1 (2.4)	10 (7.2)	11 (6.1)
espD alleles	*espD*-alpha	7 (7.8)	12 (13.2)	7 (16.7)	12 (8.6)	19 (10.5)
	espD-beta	38 (42.2)	26 (28.6)	19 (45.2)	45 (32.4)	64 (35.4)
	espD-gamma	13 (14.4)	15 (16.5)	3 (7.1)	25 (18.0)	28 (15.5)
	espD-N1	18 (20)	22 (24.2)	4 (9.5)	36 (25.9)*	40 (22.1)
	Negative	14 (15.6)	18 (19.8)	9 (21.4)	21 (15.1)	30 (16.6)
tir alleles	*tir*-alpha	18 (20)	18 (19.8)	14 (33.3)*	22 (15.8)	36 (19.9)
	tir-beta	33 (36.7)	19 (20.9)	13 (31.0)	39 (28.1)	52 (28.7)
	tir-gamma	2 (2.2)	5 (5.5)	1 (2.4)	6 (4.3)	7 (3.9)
	tir-gamma2	21 (23.3)	30 (33.0)	6 (14.3)	45 (32.4)*	51 (28.2)
	tir-alpha2	8 (8.9)	13 (14.3)	7 (16.7)	14 (10.1)	21 (11.6)
	Negative	8 (8.9)	6 (6.6)	1 (2.4)	13 (9.4)	14 (7.7)

*$P<0.05$ for the comparison of allele distribution between tEPEC and aEPEC.

investigated. The remaining strains were nontypable with the O antisera used (76 strains). Forty-one strains belonged to classical EPEC serogroups (Trabulsi *et al.*, 2002). The other typable strains belonged to 36 different serogroups. The most common O serogroups were O55 (14 strains), O111 (7 strains) and O119 (6 strains). Thirty different H-types were detected; the most common flagellar types were H27 (20 strains), H8 (16 strains), H7 (14 strains) and H19 (13 strains). Six strains were non-motile and the H-type was not identified in 11 strains. No single serotype was found in more than 3 % of the strains investigated.

Association with clinical characteristics

To determine whether clinical characteristics of the diarrhoea episodes might be related to specific combinations of *esp* and *tir* alleles, single EPEC infections (without other pathogens, e.g. *Shigella*, *Campylobacter*, etc.) were evaluated (60 EPEC strains from diarrhoea). Disease severity, as suggested by the Vesikari score, was similar in the diarrhoea episodes associated with the different allele combinations (data not shown). Among all the diarrhoeal episodes (unique and coinfections), 56 cases had a duration of <7 days, while 34 cases lasted $\geqslant 7$ days. There were no clear differences in the distribution of *espA*, *espB* or *tir* alleles among diarrhoea episodes of longer duration ($\geqslant 7$ days) and shorter episodes (data not shown). For *espD*, the allele distribution was significantly different in relation to the duration of the episode. Specifically, *espD*-alpha was more common in episodes of <7 days compared with episodes $\geqslant 7$ days (12.5 % vs 0 %, respectively; $P<0.05$).

DISCUSSION

The differentiation of *tir*, *espA*, *espB* and *espD* alleles of the LEE pathogenicity island represents a potential tool for EPEC typing in pathogenesis, epidemiological and immunological studies. In this work, we defined a PCR-RFLP for typing of four alleles of *espA*, five alleles of *espB*, four alleles of *espD* and six alleles of *tir*.

aEPEC (139/181, 76.8 %) constituted the majority of *eae*-positive isolates in this study and was common both in patients with diarrhoea (64/90, 71.1 %) and in controls (75/91, 82.4 %), as has been recently reported in developing and developed countries (Ochoa *et al.*, 2008).

In the present study, we identified four variants for *espA*; we found that *espA*-beta2 was the more frequent allele, whereas in other studies *espA*-alpha or *espA*-beta were more frequent variants (Afset *et al.*, 2006; Garrido *et al.*, 2006; Goffaux *et al.*, 2001). As reported by Garrido *et al.* (2006), we found that *espB*-alpha was the most frequent allele; *espB*-beta was the second most frequent allele in this study. In other studies, *espB*-beta was the most frequent allele detected, but the strains were principally isolated from animals (Afset *et al.*, 2008; Goffaux *et al.*, 2001; Yuste *et al.*, 2008). For *espD*, the most frequent allele was *espD*-beta (35.4 %) (Yuste *et al.*, 2008), in contrast to other studies where most strains were not typable for *espD* (Garrido *et al.*, 2006). For the *tir* alleles, *tir*-beta (28.7 %) (Goffaux *et al.*, 2001; Yuste *et al.*, 2008) was the most common *tir* variant, in contrast to a previous study that found *tir*-alpha to be the most frequent (~44 %) (Garrido *et al.*, 2006). These differences in the prevalence of different alleles/sub-alleles may be related to the different geographical origin of the samples. In addition to the high variability observed in these genes, some EPEC strains were PCR-negative for the genes evaluated in this study (Table 2). These PCR-negative results may be due to the absence of the genes in these strains, or mainly because the primers used do not hybridize to the template DNA due to the high sequence variability of these genes in these specific strains.

Although no homologous associations were observed between *eae* and *bfpA*/*perA* alleles in these strains (Contreras *et al.*, 2010), associations of the results of *tir* and *eae* alleles (located in the fifth polycistronic operon termed LEE5) (Deng *et al.*, 2001) showed that 16 % of the EPEC strains were *eae*-beta and *tir*-beta and 23 % were *eae*-theta and *tir*-gamma2. These two associations were found principally in aEPEC (data not shown). This association between *tir* and *eae* alleles was maintained in association with the *esp* genes. Twenty-seven of 34 (79 %) of the strains with associations *espA*-beta, *espB*-beta, *espD*-beta, *tir*-beta were *eae*-beta, and 14 of 21 (67 %) strains with associations *espA*-alpha, *espB*-alpha2, *espD*-N1, *tir*-gamma2 were *eae*-theta. These results were similar to those obtained in isolates from animals (Yuste *et al.*, 2008). The associations of the *tir* and *esp* genes in EPEC strains were associated with variants of the *eae* gene but not with the origin of the strains.

We found a high number of LEE allele combinations compared to reports from other authors (Afset *et al.*, 2008; China *et al.*, 1999; Goffaux *et al.*, 2001; Nielsen & Andersen, 2003; Yuste *et al.*, 2008). China *et al.* (1999) and Goffaux *et al.* (2001) found only four LEE profiles in human and bovine attaching and effacing E. coli strains, Nielsen & Andersen (2003) identified seven LEE gene combinations in verocytotoxin-producing *E. coli* calf strains (20 strains) and Garrido *et al.* (2006) reported 12 combinations of these LEE genes among 25 human and animal Shiga toxin-producing *E. coli* and EPEC strains studied. The combination *espA*-beta, *espB*-beta, *espD*-beta, *tir*-beta was the most frequent (18.8 %); it has been noted in reports of strains isolated from animal and human sources (China *et al.*, 1999; Garrido *et al.*, 2006; Goffaux *et al.*, 2001; Yuste *et al.*, 2008) (Table S3). The most frequent homologous association (*espA*-beta, *espB*-beta, *espD*-beta, *tir*-beta) of the *esp* genes was closely related to *eae*-beta (28/34, 82 %), similar to the results obtained in other studies (Goffaux *et al.*, 2001; Yuste *et al.*, 2008). The second most frequent association observed was *espA*-alpha, *espB*-alpha2, *espD*-N1, *tir*-gamma2 (Table S3) and was associated with *eae*-theta. Most of the *esp* and *tir* associations were heterologous (alleles in combination with more than just one variant of the other LEE genes), consistent with the finding of Afset *et al.* (2008). They suggest that these different combinations within the LEE may be due to

Table 3. Serogroup and variants of *esp* and *tir* genes in EPEC strains isolated from children with diarrhoea and controls

No.	Alleles				Diarrhoea (*n*)* (*n*=90)	Control (*n*)* (*n*=91)
	espA	*espB*	*espD*	*tir*		
1	beta	beta	beta	beta	O5, O15 (2), O26, O98, O111 (2), O119, O123, O128 (3), O153, ONT (6) and X09	O15, O26, O97, O111, O153, O157, ONT (7) and OX09
2	alpha	alpha2	N1	gamma2	O2, O5, O55, O91, OX09 and NT (3)	O34, O49, O51, O69, O82, O91 and ONT (7)
3	beta2	alpha	beta	alpha2	O55, O126 and ONT (2)	O55 (3), O113 and O126
4	beta2	alpha	alpha	alpha	O119 (2), ONT and OR	O119 (2) and ONT (2)
5	beta2	alpha	Negative	alpha	O20 and O55	ONT and O111
6	beta2	alpha	gamma	alpha	O13 and ONT	O153, O55 and OR
7	gamma	gamma	gamma	gamma	O8	O33, O55 (2) and ONT
8	Other				O2, O13, O25, O33, O55 (4), O69 (2), O75, O80, O85, O88 (2), O91, O101, O108, O111, O116, O119 (2), O146, O157, O171, OX10, OX18, OR and ONT (21)	O2, O13, O34, O35, O55, O73, O89, O101, O111, O116, O108, O128, O134, O142 (3), O153, OR, OX9 and ONT (25)

*n, No. of isolates; only expressed if >1.

horizontal exchange of smaller or entire LEE sequences between strains (Afset *et al.*, 2008; Castillo *et al.*, 2005).

Variants of the LEE genes have been associated with the characteristics of colonization and tissue trophism (Torres *et al.*, 2005). Such differences might influence the ability of the bacteria to induce diarrhoea. However, except for a marginally significant association of *espD*-alpha with diarrhoea episodes <7 days, none of the LEE gene variants were significantly associated with diarrhoea.

There were some differences between tEPEC and aEPEC in the distributions of the alleles, principally in *espA*-alpha (related to aEPEC), *espD*-N1 (related to aEPEC), *tir*-alpha (related to tEPEC) and tir-gamma2 (related to aEPEC).

These EPEC strains are very heterogeneous, some belonging to the classical EPEC O serogroups (O26, O55, O111, O119, O125, O126, O127 and O128) and frequently the O and H antigens (Afset *et al.*, 2008; Bando *et al.*, 2009; Dulguer *et al.*, 2003; Trabulsi *et al.*, 2002; Vieira *et al.*, 2001). The present work showed that classical EPEC O serogroups had different combinations of *esp* and *tir* genes (Table 3). Similar results were observed in the other studies. There was no relationship between the distribution of serogroups among the isolates from diarrhoea and controls, nor between aEPEC and tEPEC strains.

The large variability in allele types suggests that this could be used for evaluation of outbreaks for the better characterization of the strains isolated. However, it is clear from this study that allelic typing does not identify organisms that are unusually virulent.

ACKNOWLEDGEMENTS

This work was partially funded by the National Institutes of Health, USA, Public Health Service awards 1K01TW007405 (T.J.O) and R01-HD051716 (T.G.C); Agencia Española de Cooperación Internacional (AECID), Spain, Programa de Cooperación Interuniversitaria e Investigación Científica con Iberoamérica (D/019499/08, D/024648/09 and D/030509/10); and Fondo de Investigaciones Sanitarias, Spain [J.R. has a Miguel Servet fellowship (CP05/0130)].

REFERENCES

Afset, J. E., Bruant, G., Brousseau, R., Harel, J., Anderssen, E., Bevanger, L. & Bergh, K. (2006). Identification of virulence genes linked with diarrhea due to atypical enteropathogenic *Escherichia coli* by DNA microarray analysis and PCR. *J Clin Microbiol* **44**, 3703–3711.

Afset, J. E., Anderssen, E., Bruant, G., Harel, J., Wieler, L. & Bergh, K. (2008). Phylogenetic backgrounds and virulence profiles of atypical enteropathogenic *Escherichia coli* strains from a case-control study using multilocus sequence typing and DNA microarray analysis. *J Clin Microbiol* **46**, 2280–2290.

Bando, S. Y., Andrade, F. B., Guth, B. E., Elias, W. P., Moreira-Filho, C. A. & Pestana de Castro, A. F. (2009). Atypical enteropathogenic *Escherichia coli* genomic background allows the acquisition of non-EPEC virulence factors. *FEMS Microbiol Lett* **299**, 22–30.

Blanco, M., Blanco, J. E., Dahbi, G., Mora, A., Alonso, M. P., Varela, G., Gadea, M. P., Schelotto, F., González, E. A. & Blanco, J. (2006).

Typing of intimin (*eae*) genes from enteropathogenic *Escherichia coli* (EPEC) isolated from children with diarrhoea in Montevideo, Uruguay: identification of two novel intimin variants (μB and ξR/β2B). *J Med Microbiol* **55**, 1165–1174.

Castillo, A., Eguiarte, L. E. & Souza, V. (2005). A genomic population genetics analysis of the pathogenic enterocyte effacement island in *Escherichia coli*: the search for the unit of selection. *Proc Natl Acad Sci U S A* **102**, 1542–1547.

Chen, H. D. & Frankel, G. (2005). Enteropathogenic *Escherichia coli*: unravelling pathogenesis. *FEMS Microbiol Rev* **29**, 83–98.

China, B., Goffaux, F., Pirson, V. & Mainil, J. (1999). Comparison of *eae*, *tir*, *espA* and *espB* genes of bovine and human attaching and effacing *Escherichia coli* by multiplex polymerase chain reaction. *FEMS Microbiol Lett* **178**, 177–182.

Contreras, C. A., Ochoa, T. J., Lacher, D. W., DebRoy, C., Navarro, A., Talledo, M., Donnenberg, M. S., Ecker, L., Gil, A. I. & other authors (2010). Allelic variability of critical virulence genes (*eae*, *bfpA* and *perA*) in typical and atypical enteropathogenic *Escherichia coli* in Peruvian children. *J Med Microbiol* **59**, 25–31.

Dean, P. & Kenny, B. (2009). The effector repertoire of enteropathogenic *E. coli*: ganging up on the host cell. *Curr Opin Microbiol* **12**, 101–109.

Deng, W., Li, Y., Vallance, B. A. & Finlay, B. B. (2001). Locus of enterocyte effacement from *Citrobacter rodentium*: sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens. *Infect Immun* **69**, 6323–6335.

Deng, W., Puente, J. L., Gruenheid, S., Li, Y., Vallance, B. A., Vázquez, A., Barba, J., Ibarra, J. A., O'Donnell, P. & other authors (2004). Dissecting virulence: systematic and functional analyses of a pathogenicity island. *Proc Natl Acad Sci U S A* **101**, 3597–3602.

Donnenberg, M. S., Girón, J. A., Nataro, J. P. & Kaper, J. B. (1992). A plasmid-encoded type IV fimbrial gene of enteropathogenic *Escherichia coli* associated with localized adherence. *Mol Microbiol* **6**, 3427–3437.

Dulguer, M. V., Fabbricotti, S. H., Bando, S. Y., Moreira-Filho, C. A., Fagundes-Neto, U. & Scaletsky, I. C. (2003). Atypical enteropathogenic *Escherichia coli* strains: phenotypic and genetic profiling reveals a strong association between enteroaggregative *E. coli* heat-stable enterotoxin and diarrhea. *J Infect Dis* **188**, 1685–1694.

Frankel, G., Phillips, A. D., Rosenshine, I., Dougan, G., Kaper, J. B. & Knutton, S. (1998). Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. *Mol Microbiol* **30**, 911–921.

Garrido, P., Blanco, M., Moreno-Paz, M., Briones, C., Dahbi, G., Blanco, J., Blanco, J. & Parro, V. (2006). STEC-EPEC oligonucleotide microarray: a new tool for typing genetic variants of the LEE pathogenicity island of human and animal Shiga toxin-producing *Escherichia coli* (STEC) and enteropathogenic *E. coli* (EPEC) strains. *Clin Chem* **52**, 192–201.

Goffaux, F., China, B. & Mainil, J. (2001). Organisation and in vitro expression of *esp* genes of the LEE (locus of enterocyte effacement) of bovine enteropathogenic and enterohemorrhagic *Escherichia coli*. *Vet Microbiol* **83**, 275–286.

Guion, C. E., Ochoa, T. J., Walker, C. M., Barletta, F. & Cleary, T. G. (2008). Detection of diarrheagenic *Escherichia coli* by use of melting-curve analysis and real-time multiplex PCR. *J Clin Microbiol* **46**, 1752–1757.

Lacher, D. W., Steinsland, H. & Whittam, T. S. (2006). Allelic subtyping of the intimin locus (*eae*) of pathogenic *Escherichia coli* by fluorescent RFLP. *FEMS Microbiol Lett* **261**, 80–87.

Müller, D., Benz, I., Liebchen, A., Gallitz, I., Karch, H. & Schmidt, M. A. (2009). Comparative analysis of the locus of enterocyte effacement and its flanking regions. *Infect Immun* **77**, 3501–3513.

Nataro, J. P. & Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* **11**, 142–201.

Nielsen, E. M. & Andersen, M. T. (2003). Detection and characterization of verocytotoxin-producing *Escherichia coli* by automated 5′ nuclease PCR assay. *J Clin Microbiol* **41**, 2884–2893.

Ochoa, T. J., Barletta, F., Contreras, C. & Mercado, E. (2008). New insights into the epidemiology of enteropathogenic *Escherichia coli* infection. *Trans R Soc Trop Med Hyg* **102**, 852–856.

Ochoa, T. J., Ecker, L., Barletta, F., Mispireta, M. L., Gil, A. I., Contreras, C., Molina, M., Amemiya, I., Verastegui, H. & other authors (2009). Age-related susceptibility to infection with diarrheagenic *Escherichia coli* among infants from periurban areas in Lima, Peru. *Clin Infect Dis* **49**, 1694–1702.

Ruuska, T. & Vesikari, T. (1990). Rotavirus disease in Finnish children: use of numerical scores for clinical severity of diarrhoeal episodes. *Scand J Infect Dis* **22**, 259–267.

Scaletsky, I. C., Fabbricotti, S. H., Silva, S. O., Morais, M. B. & Fagundes-Neto, U. (2002). HEp-2-adherent *Escherichia coli* strains associated with acute infantile diarrhea, São Paulo, Brazil. *Emerg Infect Dis* **8**, 855–858.

Torres, A. G., Zhou, X. & Kaper, J. B. (2005). Adherence of diarrheagenic *Escherichia coli* strains to epithelial cells. *Infect Immun* **73**, 18–29.

Trabulsi, L. R., Keller, R. & Tardelli Gomes, T. A. (2002). Typical and atypical enteropathogenic *Escherichia coli*. *Emerg Infect Dis* **8**, 508–513.

Vieira, M. A., Andrade, J. R., Trabulsi, L. R., Rosa, A. C., Dias, A. M., Ramos, S. R., Frankel, G. & Gomes, T. A. (2001). Phenotypic and genotypic characteristics of *Escherichia coli* strains of non-enteropathogenic *E. coli* (EPEC) serogroups that carry EAE and lack the EPEC adherence factor and Shiga toxin DNA probe sequences. *J Infect Dis* **183**, 762–772.

Wong, A. R., Pearson, J. S., Bright, M. D., Munera, D., Robinson, K. S., Lee, S. F., Frankel, G. & Hartland, E. L. (2011). Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. *Mol Microbiol* **80**, 1420–1438.

Yuste, M., Orden, J. A., De La Fuente, R., Ruiz-Santa-Quiteria, J. A., Cid, D., Martínez-Pulgarín, S. & Domínguez-Bernal, G. (2008). Polymerase chain reaction typing of genes of the locus of enterocyte effacement of ruminant attaching and effacing *Escherichia coil*. *Can J Vet Res* **72**, 444–448.