Journal of Medical Microbiology (2011), **60**, 639–646 **DOI** 10.1099/jmm.0.026666-0

Phylogenetic relationships of Shiga toxin-producing *Escherichia coli* isolated from Peruvian children

C. A. Contreras,[1] T. J. Ochoa,[1,2] J. Ruiz,[3,4] D. W. Lacher,[5] F. P. Rivera,[1] Y. Saenz,[6] E. Chea-Woo,[7] N. Zavaleta,[8] A. I. Gil,[8] C. F. Lanata,[8,9] L. Huicho,[7,10,11] R. C. Maves,[12] C. Torres,[6] C. DebRoy[13] 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]US Food and Drug Administration, Laurel, Maryland, USA

[6]Área de Microbiología Molecular, Centro de Investigación Biomédica La Rioja, Logroño, Spain

[7]Universidad Peruana Cayetano Heredia, Lima, Peru

[8]Instituto de Investigación Nutricional, Lima, Peru

[9]Escuela de Medicina, Universidad Peruana de Ciencias Aplicadas, Lima, Peru

[10]Universidad Nacional Mayor de San Marcos, Lima, Peru

[11]Instituto Nacional de Salud del Niño, Lima, Peru

[12]US Naval Medical Research Center Detachment, Department of Bacteriology, Lima, Peru

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

The aim of this study was to determine the prevalence, virulence factors (*stx*, *eae*, *ehxA* and *astA*) and phylogenetic relationships [PFGE and multilocus sequence typing (MLST)] of Shiga toxin-producing *Escherichia coli* (STEC) strains isolated from four previous cohort studies in 2212 Peruvian children aged <36 months. STEC prevalence was 0.4 % (14/3219) in diarrhoeal and 0.6 % (15/2695) in control samples. None of the infected children developed haemolytic uraemic syndrome (HUS) or other complications of STEC. *stx1* was present in 83 % of strains, *stx2* in 17 %, *eae* in 72 %, *ehxA* in 59 % and *astA* in 14 %. The most common serotype was O26 : H11 (14 %) and the most common seropathotype was B (45 %). The strains belonged mainly to phylogenetic group B1 (52 %). The distinct combinations of alleles across the seven MLST loci were used to define 13 sequence types among 19 STEC strains. PFGE typing of 20 STEC strains resulted in 19 pulsed-field patterns. Comparison of the patterns revealed 11 clusters (I–XI), each usually including strains belonging to different serotypes; one exception was cluster VI, which gathered exclusively seven strains of seropathotype B, clonal group enterohaemorrhagic *E. coli* (EHEC) 2 and phylogenetic group B1. In summary, STEC prevalence was low in Peruvian children with diarrhoea in the community setting. The strains were phylogenetically diverse and associated with mild infections. However, additional studies are needed in children with bloody diarrhoea and HUS.

Received 30 September 2010
Accepted 28 January 2011

Abbreviations: d_N, number of non-synonymous substitutions per non-synonymous site; d_S, number of synonymous substitutions per synonymous site; EHEC, enterohaemorrhagic *E. coli*; HUS, haemolytic uraemic syndrome; MLST, multilocus sequence typing; ST, sequence type; STEC, Shiga toxin-producing *E. coli*.

A table showing additional strains used for MLST sequence comparisons is available with the online version of this paper.

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) has emerged as a group of foodborne pathogens that can cause severe human disease, such as haemolytic uraemic syndrome (HUS) (Banatvala *et al.*, 2001; Nataro & Kaper, 1998). Enterohaemorrhagic *E. coli* (EHEC), a subclass of STEC, is

also capable of causing haemorrhagic colitis. STEC produces two phage-encoded cytotoxins called Shiga toxins (encoded by *stx1* and *stx2*). In addition to toxin production, STEC frequently possesses other virulence factors such as intimin (*eae*) (Boerlin *et al.*, 1999), a haemolysin (EHEC-HlyA; *ehxA*) (Paton & Paton, 1998; Schmidt *et al.*, 1995) and the enteroaggregative *E. coli* heat-stable enterotoxin EAST1 (*astA*) (Girardeau *et al.*, 2005; Vaz *et al.*, 2004).

Although human STEC strains belong to a large number of serotypes, most outbreaks and sporadic cases of haemorrhagic colitis and HUS are caused by serotype O157 : H7. As non-O157 STEC strains are more prevalent in animals and as contaminants in foods, humans are probably exposed more often to these strains. STEC serogroups have been classified into five seropathotypes (A–E) according to incidence and association with HUS and outbreaks (Karmali *et al.*, 2003). STEC can be classified into four phylogenetic groups (B1, A, D and B2) (Clermont *et al.*, 2000; Escobar-Páramo *et al.*, 2004; Girardeau *et al.*, 2005). Based on multilocus sequence typing (MLST), Whittam and co-workers studied the clonal relationships of STEC strains (STEC Reference Center, http://www.shigatox.net/stec/index.html). Two EHEC clonal groups and 11 STEC groups have been identified.

In our experience, HUS is common in Peru. We had a retrospective case series of patients with HUS admitted during the past 10 years at one paediatric hospital in Lima; however, STEC was not looked for adequately (only routine stool cultures were performed) during that time period in Peruvian HUS patients (unpublished data). There is little information on the prevalence, virulence factors and phylogenetic distribution of STEC strains in Peru. The aims of this study were to: (i) determine the prevalence of STEC in diarrhoea and control samples from Peruvian children; (ii) determine the distribution of critical virulence factors (*stx1*, *stx2*, *eae*, *ehxA* and *astA*); and (iii) determine the phylogenetic distribution (by MLST and PFGE) of the isolated STEC strains.

METHODS

Bacterial strains. We determined the prevalence of STEC in 3219 samples from children with diarrhoea and 2695 samples from healthy controls without diarrhoea from four prospective cohort studies conducted previously in 2212 Peruvian children aged <36 months. All studies were in the community setting: three in peri-urban communities of Lima [Villa el Salvador (N. Zavaleta, Instituto de Investigación Nutricional), Chorrillos (Ochoa *et al.*, 2009) and Independencia (E. Chea-Woo, Universidad Peruana Cayetano Heredia)]; and one in the Andean region of the country [Huaraz (C.F Lanata, Instituto de Investigación Nutricional)] (Table 1). STEC strains were identified by the presence of *stx1*, *stx2* and *eae* using a previously validated multiplex real-time PCR system (Guion *et al.*, 2008). For all studies, five lactose-positive colonies isolated from MacConkey plates were used for the PCR assay. The strain STEC W147 ($stx1^+$ $sxt2^+$ eae^+ $ehxA^+$) provided by Dr C. Torres (Universidad de Rioja, Spain) was used as a positive control.

Detection of virulence factors. One *stx1*- and/or *stx2*-positive colony per patient was tested to identify the presence of virulence genes. The sequences of the primers and amplicon sizes are described in Table 2. PCR for the other virulence genes (*ehxA* and *astA*) was performed in a 25 µl reaction mixture containing 2.5 µl each dNTP (2.5 mM; Bioline), 1.5 µl 50 mM $MgCl_2$, 0.5 µl each primer (10 mM; Isogen Life Science), 2.5 µl 10× NH_4 buffer (Bioline), 1.5 U Biotaq DNA Polymerase (Bioline) and 5 µl DNA template. For all amplification reactions, the mixture was heated to 94 °C for 10 min prior to thermocycling (iCycler; Bio-Rad). The mixture was held at 72 °C for 7 min after the final cycle before cooling at −20 °C. Amplified products were analysed by 1.5 % agarose gel electrophoresis and visualized by staining with ethidium bromide.

Serotyping. Serotyping was performed at the *E. coli* Reference Center (Pennsylvania State University, PA, USA) for O (Orskov *et al.*, 1977) and H (Machado *et al.*, 2000) antigen typing. STEC strains were assigned to one of the five seropathotypes (A–E), as described previously (Karmali *et al.*, 2003).

EHEC haemolysin production. EHEC haemolysin production was detected using blood agar base (Difco) supplemented with 10 mM $CaCl_2$ and 5 % defibrinated sheep blood. Plates were incubated at 37 °C and examined after 24 and 48 h for zones of haemolysis around colonies (Vieira *et al.*, 2001).

Clermont's phylogenetic group determination. STEC strains were assigned to Clermont's phylogenetic groups according to the presence or absence of the genes *chuA*, *yjaA* and *tspE4C2* (Clermont *et al.*, 2000).

MLST. MLST was performed on seven conserved housekeeping genes (*aspC*, *clpX*, *fadD*, *icdA*, *lysP*, *mdh* and *uidA*) as described elsewhere (http://www.shigatox.net/mlst). PCR products were purified using a Wizard SV Gel and PCR Clean-Up System (Promega). Sequencing was performed by Macrogen using an automatic DNA 3730xl sequencer (Applied Biosystems) and concatenated for phylogenetic analyses.

Table 1. STEC prevalence in four cohort studies in Peruvian children

Study	No. patients	Age range (months)	Year of study	Prevalence in diarrhoea group (*n*/*N*) (%)	Prevalence in control group (*n*/*N*) (%)
Huaraz, Ancash	485	0–36	1987	4/733 (0.5)	2/467 (0.4)
Villa el Salvador, Lima	313	6–18	2004	4/556 (0.7)	ND
Chorrillos, Lima	1034	2–24	2007–2008	4/1129 (0.4)	7/767 (0.9)
Independencia, Lima	380	12–24	2008–2009	2/801 (0.2)	6/1461 (0.4)
All four studies	2212	0–36	–	14/3219 (0.4)	15/2695 (0.6)

ND, Not determined: control samples were not collected in this study.

Table 2. Sequence of primers used in this study

Gene	Orientation*	Sequence (5′–3′)	Fragment size (bp)	Reference
sxt1	F	CTGGATTTAATGTCGCATAGTG	150	Guion *et al.* (2008)
	R	AGAACGCCCACTGAGATCATC		
stx2	F	GGCACTGTCTGAAACTGCTCC	255	Guion *et al.* (2008)
	R	TCGCCAGTTATCTGACATTCTG		
eae	F	ATGCTTAGTGCTGGTTTAGG	248	Guion *et al.* (2008)
	R	GCCTTCATCATTTCGCTTTC		
ehxA	F	GGTGCAGCAGAAAAAGTTGTAG	1551	Schmidt *et al.* (1995)
	R	TCTCGCCTGATAGTGTTTGGTA		
astA	F	CCATCAACACAG TAT ATCCGA	111	Yamamoto & Echeverria (1996)
	R	GGTCGCGAGTGACGGCTTTGT		
chuA	F	GACGAACCA ACGGTCAGGAT	279	Clermont *et al.* (2000)
	R	TGCCGCCAGTACCAAAGACA		
yjaA	F	TGAAGTGTCAGGAGACGCTG	211	Clermont *et al.* (2000)
	R	ATGGAGAATGCGTTCCTCAAC		
tspE	F	GAGTAATGTCGGGGCATTCA	152	Clermont *et al.* (2000)
	R	CGCGCCAACAAAGTATTACG		

*F, Forward primer; R, reverse primer.

DNA sequence analyses. The sequences were reviewed and edited by visual inspection using Chromas Lite v.2.01 software (Technelysium Pty). After editing, the sequences were exported to BioEdit v.7.0.9 (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html) and aligned with the CLUSTAL W module. Differences of a single nucleotide allowed us to classify the sequences as different alleles. The different alleles of each housekeeping gene were numbered, and allelic profiles or sequence types (STs) were determined based on the seven studied loci. ST designations were assigned in accordance with the numbering system used by the STEC Center at Michigan State University (MI, USA; http://www.shigatox.net/ecmlst/cgi-bin/index). Strains belonging to the same ST were considered to be the same clone; one member of each ST was used in the phylogenetic analyses.

Phylogenetic analyses. The MLST sequences of the strains were combined with those from 33 published *E. coli* and *Shigella* species genomes for comparison (see Supplementary Table S1, available in JMM Online). Sequences were aligned by CLUSTAL W using the MEGALIGN module of the Lasergene software (DNASTAR). Neighbour-joining trees were constructed using the Kimura two-parameter model of nucleotide substitution with MEGA4 software (Tamura *et al.*, 2007), and the inferred phylogenies were each tested with 500 bootstrap replications. Phylogenetic network analysis was conducted with the SplitsTree 4 program (Huson & Bryant, 2006) using the neighbour-net algorithm (Bryant & Moulton, 2004) and untransformed distances (*p*-distances). The Φ_w recombination test (Bruen *et al.*, 2006) as implemented by SplitsTree 4 was used to distinguish recurrent mutation from recombination in generating genotypic diversity. The numbers of synonymous substitutions per synonymous site (d_S) and non-synonymous substitutions per non-synonymous site (d_N) were estimated by the modified Nei–Gojobori method using MEGA4. Allelic sequences were fitted to a nucleotide substitution model using the Datamonkey website (http://www.datamonkey.org/), and the single likelihood ancestor counting method was used to fit a codon model to detect selection on individual codons (Pond & Frost, 2005).

PFGE. Preparation of genomic DNA and PFGE were performed as described previously (Gautom, 1997). Samples were digested with 40 U *Xba*I (Promega), and DNA fragments were resolved in 1 % agarose gels using a CHEF-DR-II system (Bio-Rad Laboratories). Lambda concatemers (New England Biolabs) with a molecular size range of 50–1000 kb were used as DNA size markers. Evaluation of PFGE profiles for similarity was performed using InfoQuest FP v.5 software (Bio-Rad). A UPGMA tree was constructed using Dice similarity indices, complete linkage and optimization: 1 %, position tolerance 1.3 % (Beutin *et al.*, 2005).

RESULTS AND DISCUSSION

Prevalence

We analysed 5914 samples in total. The prevalence of STEC was 0.4 % (14/3219) in diarrhoeal samples and 0.6 % (15/2695) in healthy controls (Table 1). The prevalence of STEC was significantly lower compared with other pathogens. The mean prevalence of the other isolated pathogens (using the same PCR methodology) was: enteroaggregative *E. coli*, 9.9 %; enteropathogenic *E. coli*, 8.5 %; enterotoxigenic *E. coli*, 6.9 %; and diffusely adherent *E. coli*, 4.8 % (T. J. Ochoa, A. Llanos, J. Lee and F. Lopez, unpublished data). To our knowledge, this is the first study of the prevalence of STEC in Peruvian children. This is important because STEC is not routinely looked for in clinical laboratories, even when the child presents with bloody diarrhoea or HUS. The small number of isolated STEC strains was one of the main limitations of this study. The age of the STEC-infected children was 4–36 months (mean 15 months). Among the STEC-positive diarrhoea samples, one was bloody (VES 230-5, isolation date 2 June 2004). Of the 29 STEC strains, 20 were available for further analysis (by MLST and PFGE).

Serotypes and seropathotypes

The typable STEC belonged to 18 serotypes. The most common serogroups were O26 (four strains, 14 %), O111

(three strains, 10 %) and O145 (three strains, 10 %), with similar distribution among the diarrhoea and control samples (Table 3). Infections with some non-O157 STEC types, such as O26 : H11 or H$^-$, O91 : H21 or H$^-$, O103 : H2, O111 : H$^-$, O113 : H21, O117 : H7, O118 : H16, O121 : H19, O128 : H2 or H$^-$, O145 : H28 or H$^-$ and O146 : H21, have been associated with severe illness in humans (Bettelheim, 2007; Coombes *et al.*, 2008) and with a number of outbreaks (Hiruta *et al.*, 2001; McMaster *et al.*, 2001; Werber *et al.*, 2002). The strains included examples of four of the five reported seropathotypes (Karmali *et al.*, 2003). The most common seropathotype was B (13/29, 45 %), which comprised all O26 : H11 (four strains), O111 (H10, H8 and H7; three strains), O103 : H2 (two strains), O145 (H11 and H$^+$; three strains) and O174 : H19 (one strain), and was associated with disease (Girardeau *et al.*, 2005). There were five non-typable strains and four with only H-types, which did not belong to any of the known seropathotype groups.

Distribution of virulence genes

Analysis of the frequency of virulence factors and clonal distribution of STEC is pivotal to improve our understanding of epidemiological characteristics of pathogens that pose a risk to public health. Epidemiological studies, together with *in vivo* and *in vitro* experiments, have revealed that *stx2* (and its variants) is the most important virulence factor associated with severe human disease. STEC producing *stx2* is more commonly associated with serious disease than isolates producing *stx1* or *stx1* plus *stx2* (Boerlin *et al.*, 1999; Louise & Obrig, 1995; Paton & Paton, 1998). In the current study, the majority of strains were *stx1*-producing strains (24/29, 83 %); only 5/29 (17 %) strains carried *stx2*. This fact presumably explains the mild illness found in these infections. There were too few *stx2*-positive isolates to assess its relationship to pathogenesis. Severe diarrhoea (especially haemorrhagic colitis) and HUS

Table 3. Distribution of serotypes, virulence genes and phylogenetic groups of the STEC strains

ND, Not determined; +, positive for the gene; −, negative for the gene; NT, non-typable.

Number	Code-date of isolation	Type of sample	Serotype*	Seropathotype†	*stx1*	*stx2*	*eae*	*astA*	*ehxA*	Haemolysis	Phylogenetic group‡
1	D0157-22OCT07	Control	O25 : H4	E	+	−	+	−	+	−	B1
2	D0135-08AUG07	Control	O26 : H11	B	+	−	+	−	+	+	B1
3	D3394-24SEP07	Control	O26 : H11	B	+	−	+	−	+	+	B1
4	D3089-10APR07	Control	O132 : H29	D	+	−	+	−	+	+	D
5	LF1004-21APR08	Control	O111 : H8	B	+	−	+	−	+	+	B1
6	LF1237-01JUL09	Control	O145 : H11	B	+	−	+	−	+	+	B1
7	D5018-16JAN08	Control	O145 : H$^+$	B	+	−	+	−	+	+	D
8	D5022-22OCT07	Control	O26 : H11	B	+	−	+	−	−	−	B1
9	LF1238-22SEP09	Control	O5 : HNT	C	+	−	+	−	+	−	A
10	LF1171-23FEB09	Control	O111 : H10	B	+	−	+	−	+	−	B1
11	LF1045-25AUG08	Control	O103 : H2	B	+	−	−	−	+	−	B1
12	LF1062-21APR08	Control	ONT : H$^+$	ND	+	−	−	−	−	−	B1
13	S-056-18DEC86	Control	ONT : HNT	ND	+	−	−	−	−	−	D
14	F-002-1-27JAN87	Control	ONT : HNT	ND	−	+	+	+	−	−	A
15	D3251-09MAY08	Control	O117w : H$^+$	D	−	+	−	−	−	−	A
16	D0020-14NOV06	Diarrhoea	O174 : H19	B	+	−	+	+	+	+	B1
17	D3451-31MAY07	Diarrhoea	O88 : H10	E	+	−	+	+	−	−	B1
18	VES 230-5-02JUN04	Diarrhoea	O145 : H$^+$	B	+	−	+	−	+	−	D
19	VES 273-2-03NOV04	Diarrhoea	O26 : H11	B	+	−	+	−	+	+	B1
20	D3157-04APR07	Diarrhoea	ONT : H$^+$	ND	+	−	+	−	+	+	B2
21	D7042-04MAY07	Diarrhoea	O103 : H2	B	+	−	+	−	+	+	B1
22	LF1263-02SEP09	Diarrhoea	O5 : HNT	C	+	−	+	−	+	−	A
23	VES 256-3-28OCT04	Diarrhoea	O111w : H7	B	+	−	+	−	+	+	D
24	H-083-3-06MAY87	Diarrhoea	ONT : HNT	ND	+	−	+	−	−	−	D
25	LF1226-07OCT09	Diarrhoea	O117 : H7	D	+	−	−	−	−	−	B1
26	VES 256-2-28OCT04	Diarrhoea	ONT : H33	ND	+	−	−	+	−	−	A
27	H-083-2-06MAY87	Diarrhoea	ONT : HNT	ND	−	+	+	−	−	−	D
28	H-119-1-16FEB87	Diarrhoea	ONT : HNT	ND	−	+	−	−	−	−	D
29	H-053-04JUN87	Diarrhoea	ONT : H36	ND	−	+	−	−	−	+	B1

*H$^+$, Positive reaction: the group is novel and does not match with known reference standards; w, weak reaction.
†Proposed by Karmali *et al.* (2003) (A–E).
‡Proposed by Clermont *et al.* (2000) (A, B1, B2 and D).

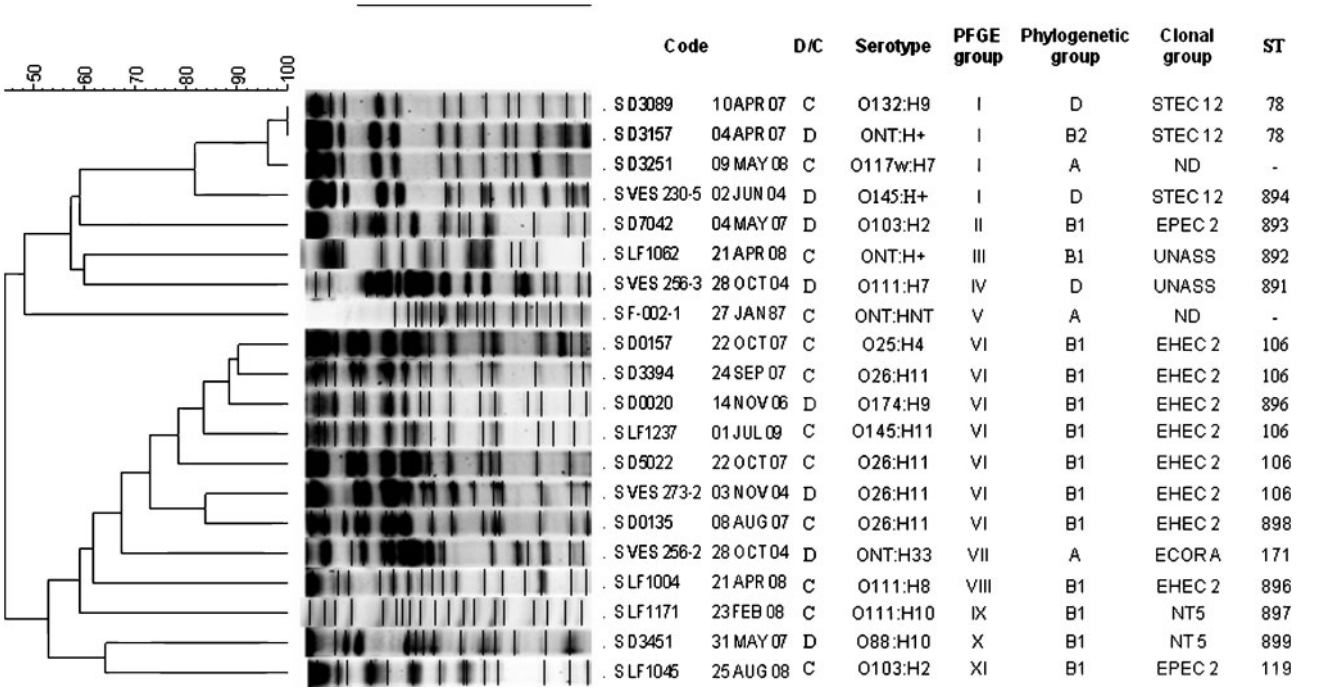

Fig. 1. PFGE profiles and clusters of 20 Peruvian STEC strains. The corresponding MLST ST, clonal group, phylogenetic group and serotype are listed for each strain based on the corresponding pulsed-field pattern (PFGE groups I–XI). D, Diarrhoea sample; C, control sample; S, sample; UNASS, unassigned; ND, not done. Strain D5018-15JAN08 (not shown) was ST895, clonal group STEC 12.

are closely associated with STEC types carrying the *eae* gene for intimin (Boerlin *et al.*, 1999), although a large number of locus of enterocyte effacement-negative STEC have also caused human disease (Bettelheim, 2007). In this study, the *eae* gene was present in 72 % (21/29) of the STEC strains: 19 strains were *stx1*$^+$ *eae*$^+$, five were *stx1*$^+$ *eae*$^-$, three were *stx2*$^+$ *eae*$^-$ and two were *stx2*$^+$ *eae*$^+$. The distribution of frequency of *eae*$^+$ STEC among diarrhoea and control samples was similar.

The *ehxA* gene (enterohaemolysin) was detected in 59 % of the STEC strains (17/29) with a similar distribution among the diarrhoea and control samples. Haemolysis production was present in 11/17 strains positive for the e*hxA* gene; only one strain that was *ehxA*$^-$ was haemolytic (Table 3). The *astA* gene (toxin EAST1) was uncommon (4/29, 14 %), being found in 1/15 control samples (7 %) and 3/14 diarrhoeal samples (21 %) (Table 3). We did not find any significant association between the presence of specific virulence genes and a specific seropathotype as reported by others (Girardeau *et al.*, 2005).

Table 4. Sequence variation in seven MLST loci from 13 STs

Locus	No. sites	No. variable sites	No. alleles	d_S×100 (mean ± SEM)	d_N×100 (mean ± SEM)	Φ_w*
aspC	513	14	5	4.59 ± 1.24	0.00 ± 0.00	0.13
clpX	567	27	7	6.00 ± 1.04	0.00 ± 0.00	0.23
fadD	483	20	7	6.64 ± 1.46	0.00 ± 0.00	0.02
icdA	567	21	7	5.02 ± 1.11	0.00 ± 0.00	0.007
lysP	477	10	5	3.92 ± 1.26	0.00 ± 0.00	1.0
mdh	549	20	8	3.57 ± 0.92	0.25 ± 0.14	0.19
uidA	576	24	6	5.50 ± 1.23	0.41 ± 0.18	0.43
Mean	533.1	19.4	6.4	5.03 ± 1.18	0.09 ± 0.05	–
Total†	3732	136	13	3.86 ± 0.39	0.07 ± 0.03	1.2×10^{-9}

*P value from the Φ_w test for recombination.
†Concatenated sequence data for all 13 STs.

Clermont's phylogenetic group distribution

Recent phylogenetic studies have indicated that STEC/EHEC strains fall principally into phylogenetic groups A, B1 and D (Escobar-Páramo *et al.*, 2004; Girardeau *et al.*, 2005; Ziebell *et al.*, 2008). The strains in this study belonged to phylogenetic group B1 (52 %, 15/29), D (28 %, 8/29), A (17 %, 5/29) and B2 (3 %, 1/29). The most frequent phylogenetic group was B1, consistent with earlier reports (Girardeau *et al.*, 2005). Among the 14 diarrhoeal strains, six belonged to group B1 and five to group D (Fig. 1, Table 3).

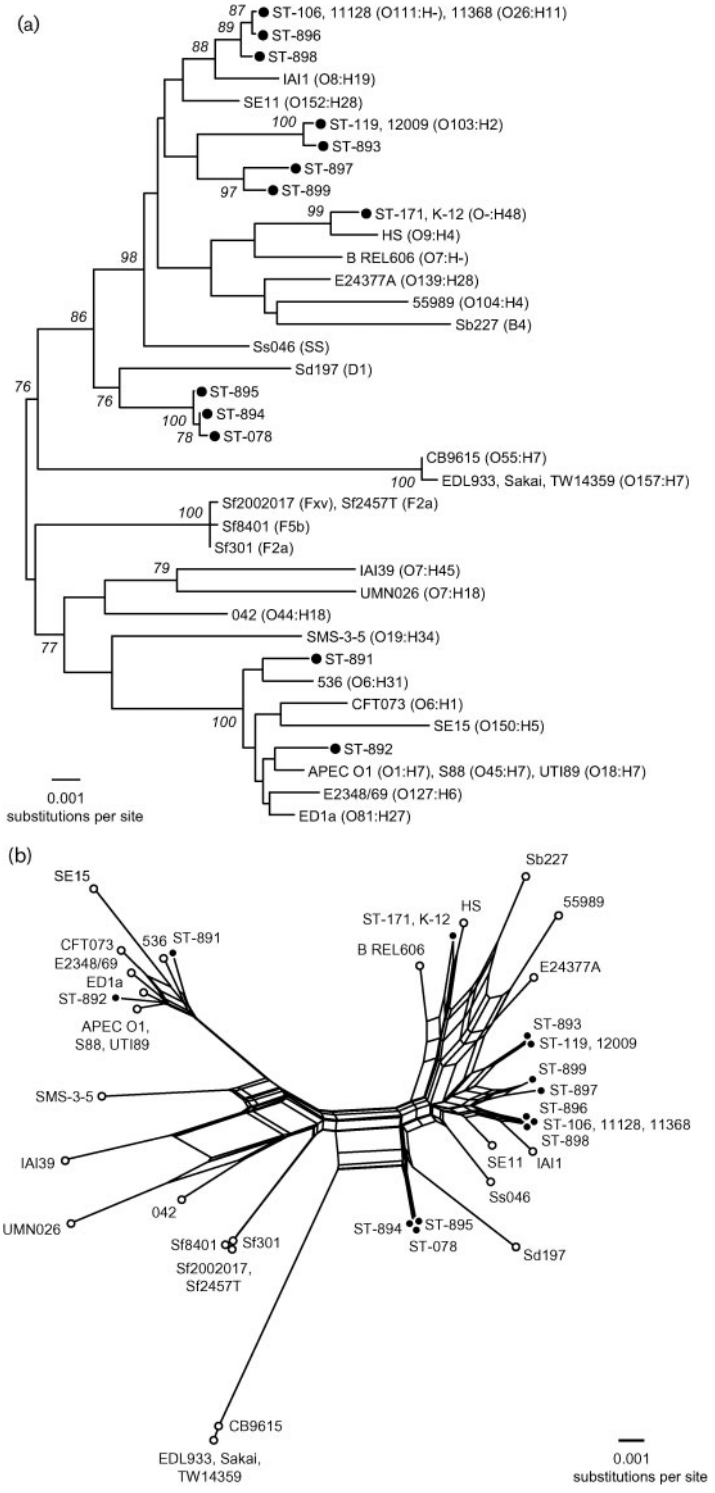

Fig. 2. Phylogenetic relationships among 13 STEC STs. (a) Unrooted phylogenetic tree constructed by a neighbour-joining algorithm based on the Kimura two-parameter model of nucleotide substitution. Bootstrap values greater than 75 % based on 500 replications are given at internal nodes. The serotypes for the published *E. coli* and *Shigella* species genome strains are given in parentheses. (b) Phylogenetic (splits) network based on a neighbour-net algorithm using a *p*-distance matrix. The 13 STEC STs are indicated by filled circles in (a) and (b).

MLST analysis

MLST loci were sequenced in 19 STEC strains. For phylogenetic analyses, the sequenced internal fragments of the seven housekeeping genes were concatenated to yield 3732 nt. MLST analysis resolved a mean of 19.4 variable nucleotide sites per locus, which defined a number of alleles, ranging from five to eight (Table 4). The d_S value ranged from 3.57 % for *mdh* to 6.64 % for *fadD*, with a mean of 5.03 synonymous substitutions per 100 synonymous sites (Table 4). The d_N value per 100 non-synonymous sites was generally an order of magnitude lower than that of d_S, ranging from 0.00 for *aspC*, *clpX*, *fadD*, *icdA* and *lysP* to 0.41 for *uidA*. Tests for natural selection operating on the allelic variation at each MLST locus based on the single likelihood ancestor counting method found no individual sites to be under significant negative or positive selection, indicating that the MLST loci are evolving neutrally.

The distinct combinations of alleles across the MLST loci were used to define 13 multilocus genotypes or STs among the 19 strains. The 13 STs differed on average at 1.2 and 0.2 % of the nucleotide and amino acid sites, respectively. ST106 was the most common multilocus genotype (5/19, 26 % of strains) (Fig. 1). In the phylogenetic tree based on the genetic relationships of the STEC (Fig. 2a), we observed that our strains were closely related to others of the same serotype from other studies. In addition, STEC strains ST106, ST896 and ST898 (EHEC 2) were related in the network, similar to the results observed in the tree based on the PFGE results (Fig. 1). The same results were observed in clonal group STEC 12.

The correlation observed between Whittam's clonal groups, Clermont's phylogenetic groups and some of the serotypes in this study is of interest. EHEC 2 contains serotypes O26 : H11, O111 : H8 (O111 strains are often non-motile or of other H types) and O145 : H11, which are classified as seropathotype B and phylogenetic group B1, as observed by others (Karmali *et al.*, 2003; Ziebell *et al.*, 2008).

The splits network (Fig. 2b) revealed several parallel paths indicative of the presence of phylogenetic incompatibilities in the divergence of clones. Such incompatibilities could arise from recurrent mutation or recombination in MLST loci. To detect recombination, the Φ_w test, which discriminates between recurrent mutation and recombination (Bruen *et al.*, 2006), was used. When applied to the concatenated sequences of the 13 STs, the Φ_w test found significant evidence of recombination (Table 4). Evidence for recombination was also detected among the alleles of *fadD* and *icdA* (Table 4).

PFGE

PFGE typing of 20 STEC strains resulted in 19 pulsed-field patterns.

Comparison of the patterns revealed 11 clusters (I–XI) with a general similarity of 70 % in the UPGMA tree. Each cluster included strains belonging to different serotypes (Fig. 1), with the exception of cluster VI, which exclusively contained seven STEC of clonal group EHEC 2, phylogenetic group B1 and seropathotype B. In addition, the strains of pulsed-field pattern 1 (cluster I) showed the same pattern, belonging to the same clonal group of STEC 2. Most of the strains in this study were from children in separate geographical areas taken on different dates, suggesting that these pathogenic clones may be widespread in Peru.

MLST and PFGE were performed to establish the clonal relationships between representative STEC strains in this study. Both techniques identified strains that shared similar clonal origins (PFGE group VI and EHEC 2; Fig. 1). PFGE was more discriminative than MLST, as each ST was represented by more than one pulsed-field pattern. Differences between MLST and PFGE may be the result of the type of analysis. While PFGE detects multiple differences in the genome, MLST analyses only small fragments of conserved metabolic genes. Therefore, events such as the recent acquisition of virulence factors cannot be detected by MLST; genome sequencing was not carried out in this study.

In summary, STEC prevalence was low in children with diarrhoea in the community setting in Peru. Strains were phylogenetically diverse and associated with mild infections. There was a good correlation between the seropathotypes, clonal groups, PFGE groups and Clermont's phylogenetic groups. However, additional studies are needed in Peruvian children with bloody diarrhoea and HUS to determine the virulence genes and phylogenetic characteristics of more virulent strains.

ACKNOWLEDGEMENTS

The authors wish to thank Maruja Bernal, Rina Meza and David Cepeda for their help in laboratory analysis. This work was partially funded by: Agencia Española de Cooperación Internacional (AECID), Spain, Programa de Cooperación Interuniversitaria e Investigación Científica con Iberoamérica (D/019499/08 and D/024648/09); Institutional Research Funds from Universidad Peruana Cayetano Heredia, Perú (T. J. O.) and Instituto Nacional de Salud del Niño, Peru (L. H.); Military Infectious Disease Research Program work unit # 60000.000.0.B0017 (R. C. M.); Instituto de Investigación Nutricional, Peru (N. Z. and C. F. L.); Programa Miguel Servet (CP05/00130) (J. R.); and the National Institutes of Health, USA, Public Health Service awards 1K01TW007405 (T. J. O.) and R01-HD051716 (T. G. C. and E. C.-W.).

An author of this work (R. C. M.) is an employee of the US Government. This work was prepared as part of his official duties. There is no conflict of interest for any of the authors.

REFERENCES

Banatvala, N., Griffin, P. M., Greene, K. D., Barrett, T. J., Bibb, W. F., Green, J. H., Wells, J. G. & Hemolytic Uremic Syndrome Study Collaborators (2001). The United States National Prospective

Hemolytic Uremic Syndrome Study: microbiologic, serologic, clinical, and epidemiologic findings. *J Infect Dis* **183**, 1063–1070.

Bettelheim, K. A. (2007). The non-O157 Shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. *Crit Rev Microbiol* **33**, 67–87.

Beutin, L., Kaulfuss, S., Herold, S., Oswald, E. & Schmidt, H. (2005). Genetic analysis of enteropathogenic and enterohemorrhagic *Escherichia coli* serogroup O103 strains by molecular typing of virulence and housekeeping genes and pulsed-field gel electrophoresis. *J Clin Microbiol* **43**, 1552–1563.

Boerlin, P., McEwen, S. A., Boerlin-Petzold, F., Wilson, J. B., Johnson, R. P. & Gyles, C. L. (1999). Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. *J Clin Microbiol* **37**, 497–503.

Bruen, T. C., Philippe, H. & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681.

Bryant, D. & Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**, 255–265.

Clermont, O., Bonacorsi, S. & Bingen, E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* **66**, 4555–4558.

Coombes, B. K., Wickham, M. E., Mascarenhas, M., Gruenheid, S., Finlay, B. B. & Karmali, M. A. (2008). Molecular analysis as an aid to assess the public health risk of non-O157 Shiga toxin-producing *Escherichia coli* strains. *Appl Environ Microbiol* **74**, 2153–2160.

Escobar-Páramo, P., Clermont, O., Blanc-Potard, A.-B., Bui, H., Le Bouguénec, C. & Denamur, E. (2004). A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Mol Biol Evol* **21**, 1085–1094.

Gautom, R. K. (1997). Rapid pulsed-field gel electrophoresis protocol for typing of *Escherichia coli* O157 : H7 and other Gram-negative organisms in 1 day. *J Clin Microbiol* **35**, 2977–2980.

Girardeau, J. P., Dalmasso, A., Bertin, Y., Ducrot, C., Bord, S., Livrelli, V., Vernozy-Rozand, C. & Martin, C. (2005). Association of virulence genotype with phylogenetic background in comparison to different seropathotypes of Shiga toxin-producing *Escherichia coli* isolates. *J Clin Microbiol* **43**, 6098–6107.

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.

Hiruta, N., Murase, T. & Okamura, N. (2001). An outbreak of diarrhoea due to multiple antimicrobial-resistant Shiga toxin-producing *Escherichia coli* O26 : H11 in a nursery. *Epidemiol Infect* **127**, 221–227.

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254–267.

Karmali, M. A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., Isaac-Renton, J., Clark, C., Rahn, K. & Kaper, J. B. (2003). Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. *J Clin Microbiol* **41**, 4930–4940.

Louise, C. B. & Obrig, T. G. (1995). Specific interaction of *Escherichia coli* O157 : H7-derived Shiga-like toxin II with human renal endothelial cells. *J Infect Dis* **172**, 1397–1401.

Machado, J., Grimont, F. & Grimont, P. A. (2000). Identification of *Escherichia coli* flagellar types by restriction of the amplified *fliC* gene. *Res Microbiol* **151**, 535–546.

McMaster, C., Roch, E. A., Willshaw, G. A., Doherty, A., Kinnear, W. & Cheasty, T. (2001). Verocytotoxin-producing *Escherichia coli* serotype O26 : H11 outbreak in an Irish crèche. *Eur J Clin Microbiol Infect Dis* **20**, 430–432.

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

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.

Orskov, I., Orskov, F., Jann, B. & Jann, K. (1977). Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol Rev* **41**, 667–710.

Paton, J. C. & Paton, A. W. (1998). Pathogenesis and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Clin Microbiol Rev* **11**, 450–479.

Pond, S. L. K. & Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531–2533.

Schmidt, H., Beutin, L. & Karch, H. (1995). Molecular analysis of the plasmid-encoded hemolysin of *Escherichia coli* O157 : H7 strain EDL 933. *Infect Immun* **63**, 1055–1061.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.

Vaz, T. M. I., Irino, K., Kato, M. A. M. F., Dias, A. M. G., Gomes, T. A. T., Medeiros, M. I. C., Rocha, M. M. M. & Guth, B. E. C. (2004). Virulence properties and characteristics of Shiga toxin-producing *Escherichia coli* in São Paulo, Brazil, from 1976 through 1999. *J Clin Microbiol* **42**, 903–905.

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.

Werber, D., Fruth, A., Liesegang, A., Littmann, M., Buchholz, U., Prager, R., Karch, H., Breuer, T., Tschäpe, H. & Ammon, A. (2002). A multistate outbreak of Shiga toxin-producing *Escherichia coli* O26 : H11 infections in Germany, detected by molecular subtyping surveillance. *J Infect Dis* **186**, 419–422.

Yamamoto, T. & Echeverria, P. (1996). Detection of the enteroaggregative *Escherichia coli* heat-stable enterotoxin 1 gene sequences in enterotoxigenic *E. coli* strains pathogenic for humans. *Infect Immun* **64**, 1441–1445.

Ziebell, K., Konczy, P., Yong, I., Frost, S., Mascarenhas, M., Kropinski, A. M., Whittam, T. S., Read, S. C. & Karmali, M. A. (2008). Applicability of phylogenetic methods for characterizing the public health significance of verocytotoxin-producing *Escherichia coli* strains. *Appl Environ Microbiol* **74**, 1671–1675.