Microbiology (2001), **147**, 3149–3158 Printed in Great Britain

The elements of the locus of enterocyte effacement in human and wild mammal isolates of *Escherichia coli*: evolution by assemblage or disruption?

Luisa Sandner,[1] Luis E. Eguiarte,[1] Armando Navarro,[2] Alejandro Cravioto[2] and Valeria Souza[1]

Author for correspondence: Valeria Souza. Tel: +525 622 9006. Fax: +525 622 8995. e-mail: souza@servidor.unam.mx

Departamento de Ecología Evolutiva, Instituto de Ecología[1] and Deptamento de Salud Pública, Facultad de Medicina[2], Universidad Nacional Autónoma de México (UNAM), Apartado Postal 70-275, México D. F. 04510, Mexico

***Escherichia coli* is an excellent model for studying the evolution of pathogenicity since within one species various genes can be found in pathogenic islands and plasmids causing a wide spectrum of virulence. A collection of 122 strains from different human and wild mammal hosts were analysed by PCR and Southern hybridization for the presence of a subset of the genes included in the LEE (locus of enterocyte effacement). In the PCR analysis, two markers (*cesT/eae* and *espB* genes) were found together in more strains (25·4%) than either were found alone. The *cesT/eae* gene was less frequently found alone (8·2%) than was the *espB* gene (15·6%). Four regions of the LEE were analysed in a subsample of 25 strains using Southern hybridization. The four regions were all present (44%), all absent (12%) or present in different combinations (44%) in a given strain. The flanking regions of the LEE showed the highest rate of hybridization (in 72% of the strains). The results indicate that the LEE is a dynamic genetic entity, both the complete gene cluster and the individual genes. The genes that comprise this locus seem to be horizontally acquired (or lost) in an independent way and may control other functions in non-pathogenic *E. coli* lineages. In this way, horizontal transfer may allow the gradual stepwise construction of gene cassettes facilitating coordinate regulation and expression of novel functions.**

Keywords: enterobacteriacae, horizontal transfer, pathogenic islands, wild isolates

INTRODUCTION

Escherichia coli is a common member of the commensal microbial community of the large intestine of mammals and birds (Souza *et al*., 1999). Nevertheless, some *E. coli* strains are capable of causing an impressive variety of diseases in humans and animals, including diarrhoea, haemorrhagic colitis, dysentery, haemolytic uraemic syndrome, septicaemia, bladder and kidney infections, pneumonia, neonatal meningitis, bacteraemia in children and adults with AIDS, and pyelonephritis (Nataro & Kaper, 1998). These characteristics make *E. coli* an excellent model to study the evolution of pathogenicity in bacteria, since within one species different degrees of virulence and a wide spectrum of pathogenicity can be found encoded by various genes and groups of genes.

McDaniel *et al*. (1995) first reported a large region of DNA that encodes all the determinants that elicit the A-E (attaching-effacing) lesion. This region, called LEE (locus of enterocyte effacement), is present in some of the pathogenic groups of *E. coli*, [enteropathogenic (EPEC) and enterohaemorrhagic (EHEC) strains] but it is not present in the 'normal' *E. coli* flora (i.e. K-12) or in the enterotoxigenic (ETEC), enteroinvasive (EIEC), enteroaggregative (EAEC) and in the urinary tract (UTI) *E. coli* strains (Nataro & Kaper, 1998). The LEE region is a pathogenic island (PAI) of around 35624 bp, with a G+C content of 38·36 mol% (Elliot *et al*., 1998). It is

Abbreviations: A-E, attaching-effacing; EAEC, enteroaggregative *E. coli*; EHEC, enterohaemorrhagic *E. coli*; EIEC, enteroinvasive *E. coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxigenic *E. coli*; LEE, locus of enterocyte effacement; PAI, pathogenic island; UPGMA, unweighted pair group method with arithmetic means; UTI, urinary tract *E. coli*.

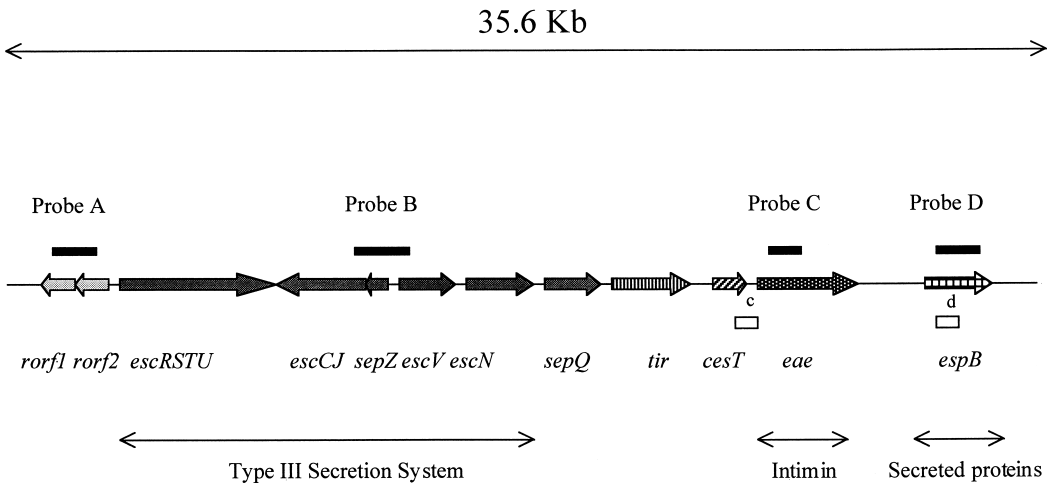

Fig. 1. Schematic organization of the LEE locus of *E. coli* E2348/69 showing the target genes of this study (GenBank accession no. AFO22236). Gene products are shown as horizontal arrows. Hybridization probes sites are shown as black bars [A, a segment of 2870 bp (genes *rorf1* and *rorf2*), B, a segment of 2948 bp (genes *escJ, rorf8, sepZ, orf12* and most of *escV*), C, a segment of 1050 bp (gene *eae*) and D, a segment of 2300 bp (gene *espB*)]. PCR amplification regions are shown as white rectangles (*cesT/eae* PCR product c comprises a 333 bp segment and *espB* PCR product d an internal 387 bp segment).

generally inserted into the chromosome at ~ 82 min, where the gene encoding the tRNA for selenocysteine (*selC*) is located (McDaniel *et al*., 1995). Nevertheless, there is evidence that in some EPEC and EHEC strains it is inserted elsewhere in the chromosome (Wieler *et al*., 1997), such as within the gene encoding the tRNA for phenylalanine at 94 min (*pheU*; Reid *et al*., 2000). Among the 41 predicted ORFs in the LEE there are some genes that could be considered important for interaction with the host, such as the *esc*, *sep*, *tir*, *cesT*, *eae* and *esp* genes (Elliot *et al*., 1998). The *esc* and *sep* gene clusters encode the proteins responsible for the export of virulence factors that belong to a type III secretion system (Jarvis *et al*., 1995). The *tir* gene encodes a translocated intimin receptor (Kenny *et al*., 1997). The *cesT* gene may encode a chaperone protein and is highly conserved in all the A-E positive strains, probably because it contains the transcriptional start of *eae* (Elliot *et al*., 1998). The *eae* gene encodes the 94–97 kDa intimin protein, that is responsible for the intimate attachment of the bacterium to the epithelium (Cravioto *et al*., 1979; Jerse *et al*., 1990). The *esp* genes are necessary to activate epithelial signal-transduction pathways (Kenny & Finlay, 1995). Mutations in the *eae* gene in EPEC and EHEC types produce strains deficient in the production of the A-E lesion and thus strains with attenuated virulence (Jerse & Kaper, 1991; Tzipori *et al*., 1995). Mutations in the *espB* gene eliminate the formation of the A-E lesion (Kenny & Finlay, 1997).

Donnenberg & Kaper (1992) have described the induction of the A-E lesion in a three-stage model of EPEC pathogenesis. This model involves initial attachment of the bacteria to the host epithelial cell, signal transduction and phosphorylation of host proteins, and finally, an intimate attachment of the bacteria to the apical surfaces of the epithelial cells.

There are several possible scenarios regarding the origin and evolution of the LEE. On one hand, the observation that the organization and functions of the LEE are conserved among studied *E. coli* strains (Finlay & Abe, 1998; Kaper, 1998; Taylor *et al*., 1998; Beltrametti *et al*., 1999; Hartland *et al*., 1999) and that the locus has a different G + C proportion to the rest of the chromosome (Elliot *et al*., 1998) suggests that this group of genes was transferred as a cassette from some other bacterial species and was inserted as a whole in the *E. coli* chromosome. On the other hand, there is increasing evidence of horizontal transfer of some genes of the LEE, recombination within this locus (Finlay & Falkow, 1997; Wieler *et al*., 1997; Ismaili *et al*., 1998; McGraw *et al*., 1999) as well as recombination with other pathogenic genes (Lederberg, 1998). These later observations suggest that this locus was assembled by multiple independent recombination events (Lawrence, 1997; Reid *et al*., 2000).

The aim of this study was to explore the evolutionary dynamics of some of the genes involved in the A-E lesion in *E. coli*. To do so, we first confirmed the presence of the *cesT/eae* and *espB* regions, by means of PCR amplifications, in a sample of 122 bacterial strains associated with wild mammals and humans. To understand how the LEE was assembled in evolutionary terms, we studied the upstream flanking region, the *esc* region, the *eae* region and the *espB* region of the locus using Southern hybridization. We also explored the

Table 1. Number and percentage (in parentheses) of *E. coli* strains, classified by serotype, that amplified for two PCR markers

Numbers in bold are totals.

Serotype	No.	*cesT/eae* (c)	*cesT/eae* only	*espB* (d)	*espB* only	Both markers	Neither marker
EAEC	**5**	**5** (100)	**2** (40·0)	**3** (60·0)	**0**	**3** (60·0)	**0**
Animal EAEC	4	4 (100)	2 (50·0)	2 (50·0)	0	2 (50·0)	0
Human EAEC	1	1 (100)	0	1 (100)	0	1 (100)	0
EHEC	7	**6** (85·7)	**2** (28·5)	**4** (57·1)	**0**	**4** (57·1)	**1** (14·3)
Animal EHEC	2	2 (100)	1 (50·0)	1 (50·0)	0	1 (50·0)	0
Human EHEC	5	4 (80·0)	1 (20·0)	3 (60·0)	0	3 (60·0)	1 (20·0)
EIEC	**4**	**0**	**0**	**0**	**0**	**0**	**4** (100)
Animal EIEC	0	0	0	0	0	0	0
Human EIEC	4	0	0	0	0	0	4 (100)
EPEC	**19**	**18** (94·7)	**1** (5·3)	**17** (89·5)	**0**	**17** (89·5)	**1** (5·3)
Animal EPEC	6	6 (100)	0	6 (100)	0	6 (100)	0
Human EPEC	13	12 (92·3)	1 (7·7)	11 (84·6)	0	11 (84·6)	1 (7·7)
ETEC	**8**	**0**	**0**	**0**	**0**	**0**	**8** (100)
Animal ETEC	3	0	0	0	0	0	3 (100)
Human ETEC	5	0	0	0	0	0	5 (100)
UTI	**4**	**1** (25·0)	**1** (25·0)	**0**	**0**	**0**	**3** (75·0)
Animal UTI	0	0	0	0	0	0	0
Human UTI	4	1 (25·0)	1 (25·0)	0	0	0	3 (75·0)
Non-pathogenic	**75**	**11** (14·7)	**4** (5·3)	**26** (34·7)	**19** (25·3)	**7** (9·3)	**45** (60·0)
Animal NP	71	10 (14·1)	3 (4·2)	25 (35·2)	18 (25·4)	7 (9·9)	43 (60·6)
Humans NP	4	1 (25·0)	1 (25·0)	1 (25·0)	1 (25·0)	0	2 (50·0)
Total	**122**	41 (33·6)	10 (8·2)	50 (40·9)	19 (15·6)	31 (25·4)	62 (50·8)

Table 2. Number and percentage (in parentheses) of *E. coli* strains, classified by host order, that produced an amplification product for the *cesT/eae* and *espB* PCR primers

Order	No.	*cesT/eae* (c)	*cesT/eae* only	*espB* (d)	*espB* only	Both markers	Neither marker
Artiodactyla	6	0	0	3 (50·0)	3 (50·0)	0	3 (50·0)
Carnivora	20	12 (60·0)	3 (15·0)	11 (55·0)	2 (10·0)	9 (45·0)	6 (30·0)
Cetacea	3	0	0	0	0	0	3 (100)
Chiroptera	7	3 (42·9)	1 (14·3)	3 (42·9)	1 (14·3)	2 (28·6)	3 (42·9)
Lagomorpha	5	0	0	0	0	0	5 (100)
Marsupialia	3	1 (33·3)	1 (33·3)	0	0	0	2 (66·7)
Perissodactyla	6	0	0	1 (16·7)	1 (16·7)	0	5 (83·3)
Primates	40	22 (55·0)	5 (12·5)	19 (47·5)	2 (5·0)	17 (42·5)	16 (40·0)
Rodentia	28	3 (10·7)	0	13 (46·4)	10 (35·7)	3 (10·7)	15 (53·6)
Sirenida	2	0	0	0	0	0	2 (100)
Xenarthra	2	0	0	0	0	0	2 (100)
Total	122	41 (33·6)	10 (8·2)	50 (40·9)	19 (15·6)	31 (25·4)	62 (50·8)

distribution of the LEE genes on a UPGMA-derived (Souza *et al*., 1999) MLEE-based dendrogram. Associations between the PCR LEE markers and the bacterial serotypes, the host diet and the host order were explored. Finally, the pattern of distribution of the Southern markers within single strains was analysed in order to find associations of the LEE genes inside the locus. We report here that the LEE is a dynamic entity, both at the level of the LEE region and within single genes of the LEE. We hypothesize that horizontal

Table 3. Number and percentage (in parentheses) of *E. coli* strains, classified by host diet, that produced an amplification product for the *cesT/eae* and *espB* PCR primers

Diet	No.	*cesT/eae* (c)	*cesT/eae* only	*espB* (d)	*espB* only	Both markers	Neither marker
Carnivorous	10	7 (70·0)	3 (30·0)	4 (40·0)	0	4 (40·0)	3 (30·0)
Granivorous	12	0	0	6 (50·0)	6 (50·0)	0	6 (50·0)
Hematofagous	2	2 (100)	1 (50·0)	1 (50·0)	0	1 (50·0)	0
Herbivorous	27	3 (11·1)	1 (3·7)	7 (25·9)	5 (18·5)	2 (7·4)	19 (70·4)
Insectivorous	2	0	0	1 (50·0)	1 (50·0)	0	1 (50·0)
Nectarivorous	3	1 (33·3)	0	1 (33·3)	0	1 (33·3)	2 (66·7)
Omnivorous	63	28 (44·4)	5 (7·9)	30 (47·6)	7 (11·1)	23 (36·5)	28 (44·4)
Piscivorous	3	0	0	0	0	0	3 (100)
Total	122	41 (33·6)	10 (8·2)	50 (40·9)	19 (15·6)	31 (25·4)	62 (50·8)

Table 4. Chi squared tests of associations between the presence of PCR amplified pathogenic markers and the bacterial serotype and genealogy, the host diet and Mammalian order

	χ^2	d. f.	*P*
Pathogenic markers – serotype	91·998	18	< 0·001
Pathogenic markers – host order	53·153	30	0·006
Pathogenic markers – host diet	44·292	21	0·002
Pathogenic markers – *E. coli* genealogy	18·574	6	0·005

transfer may have played a role in the assembly or disruption of this PAI. Moreover, we analyse the associations between bacterial serotype, bacterial genealogy, host order and diet, and the presence of the LEE-encoded pathogenic genes. These results confirm, among other things, the use of the serotype as an adequate marker to identify pathogenic strains in *E. coli*.

METHODS

Bacterial strains. Bacterial strains from non-diarrhoeic wild mammals of 11 different orders are part of the collection described by Souza *et al*. (1999). Bacterial strains associated with humans were provided by the School of Medicine of the National Autonomous University of Mexico and were collected from the faeces of healthy babies and patients with diarrhoea. These strains were processed and identified as described by Souza *et al*. (1999). A total of 122 strains were analysed in the present study: 86 associated with wild mammals from Mexico and 36 associated with humans [including a prototypic EPEC strain kindly provided by J. B. Kaper (Center for Vaccine Development, University of Maryland, MD, USA): E2348/69 (strain number 5063 in this study)]. All strains were collected in Mexico with exception of strain 5063.

The strains were serotyped for antigens O and H following the protocol described by Orskov & Orskov (1984) and their pathogenic type was determined (see supplementary data available at http://mic.sgmjournals.org).

PCR characterization of the *cesT/eaeA* and *espB* genes. Two regions of the LEE were analysed by PCR: *cesT/eae* and *espB* genes. Chromosomal DNA was extracted by heating the strains at 94 °C in distilled water for 10 min. PCR was used to amplify two segments of the LEE: a 333 bp segment of the *cesT/eae* gene (region c in Fig. 1) [from bp 24543 to 24897 of the LEE, GenBank accession AFO22236 (Elliot *et al*., 1998)], that includes approximately half of *cesT*, a spacer region of 59 bp and the initial 48 bp of the *eae* gene, and a 387 bp internal segment of the *espB* gene (region d in Fig. 1) [from bp 32408 to 32816 of the LEE, GenBank accession AFO22236 (Elliot *et al*., 1998)] were amplified (Fig. 1). Primers were designed using the Oligo 5.0 sequence analysis program (available from http://dapsas1.weizmann.ac.il/faq/pcr_local.html) and GenBank sequences [GenBank accession U32312 and Z21555 (Zhao *et al*., 1995)]. Thirty amplification cycles (90 °C, 1 min; 56 °C, 1 min; 72 °C, 1 min) were carried out using 1·5 U *Taq* DNA polymerase (Perkin Elmer). Twenty five picomoles of each of the forward (5′-GTTTGCAGAGAATGGTGGCCC-3′) and reverse (5′-TAGCTTATGCTTGTGCCGGGT-3′) *cesT/eae* primers and for the forward (5′-GCCGCTCTGATTGGTGGTGCT-3′) and reverse (5′-TGGCGTTGAACCGGAAATCCT-3′) *espB* primers were used. The PCR products were separated by gel electrophoresis, stained with ethidium bromide and visualized using UV light. The prototypic EPEC strain E2348/69 was used as a positive control.

Southern DNA hybridization. From a subsample of 25 strains selected from the 122 strain collection, total chromosomal DNA was isolated using a phenol:chloroform:isoamyl alcohol extraction method and then digested with 3 U *Eco*RI. The fragments were separated on 1% agarose gels. The capillary transfer method (Southern, 1975) of single stranded DNA onto a nylon membrane (Boehringer Mannheim) was performed. Four probes, kindly supplied by J. B. Kaper

Table 5. PCR and Southern hybridization results for 25 strains of *E. coli* associated with wild mammals and humans

Strain	Host species	PCR*		Southern hybridization†				Pathogenic type
		c *cesT/eae*	d *espB*	A *rorf1 rorf2* region	B *esc sep* region	C *eae* region	D *espB* region	
36	*Leptonycteris nivalis*	1	1	1	1	1	1	EPEC
68	*Sigmodon mascotensis*	1	1	1	0	0	1	NP
90	*Canis latrans*	1	1	0	1	0	0	EAEC
95	*Peromyscus megalops*	1	1	1	1	1	1	EPEC
97	*Peromyscus boylii*	0	1	0	0	0	0	NP
270	*Panthera onca*	1	1	1	1	1	1	EPEC
271	*Panthera onca*	1	1	1	1	1	1	EPEC
272	*Urocyon cinereoargenteus*	1	1	1	1	0	0	EAEC
286	*Dipodomys merriami*	0	1	1	0	0	1	NP
288	*Dipodomys merriami*	0	1	1	0	0	1	NP
808	*Equus caballus*	0	0	1	0	0	0	NP
815	*Bassariscus astutus*	1	1	1	1	1	1	NP
816	*Bassariscus astutus*	1	1	1	1	1	1	NP
820	*Perognathus penicillatus*	0	1	1	0	0	1	NP
830	*Canis latrans*	1	1	1	1	1	1	EHEC
1639	*Alouatta palliata*	1	1	1	1	1	1	EPEC
2055	*Tayassu tajacu*	0	1	0	0	0	0	NP
2395	*Dasypus* sp.	0	0	1	0	0	1	ETEC
3517	*Homo sapiens*	1	1	0	0	0	1	EPEC
3645	*Homo sapiens*	0	0	0	0	0	0	UTI
3653	*Homo sapiens*	1	1	1	1	0	1	EPEC
3681	*Homo sapiens*	0	0	1	0	0	1	EHEC
3689	*Homo sapiens*	1	1	1	1	1	1	EHEC
3693	*Homo sapiens*	1	1	1	1	1	1	EPEC
5063	*Homo sapiens*	1	1	1	1	1	1	EPEC
Total	25 strains	16	21	20	14	11	19	

*PCR region c (*cesT/eae*) and d (*espB*) are 333 and 387 bp, respectively.

†Southern blot probes A, B, C and D are 2870, 2948, 1050 and 2300 bp, respectively.

(Center for Vaccine Development, University of Maryland, MD, USA), were used to test the presence of four different regions of the LEE locus: an upstream flanking region of the locus, genes *rorf1* and *rorf2* (probe A); the *esc* region, genes *escJ*, *rorf8*, *sepZ*, *orf12* and most of *escV* (probe B); the *eae* region (probe C); and the *espB* region (probe D) (Kaper, 1998) (Fig. 1). Hybridization was performed using the DIG High Prime DNA Labelling and Detection Starter Kit II (Boehringer Mannheim) and hybridization products were detected by luminescence. Stringent washing conditions were employed (2× SSC, 0·1% SDS, 42 °C, 15 min and 0·1× SSC, 0·1% SDS, 42 °C, 15 min).

Construction of the dendrogram. To visualize the pattern of distribution of the pathogenic LEE markers (*cesT/eae* and *espB*) in the *E. coli* strains, a dendrogram based on 11 loci obtained by MLEE (multilocus enzyme electrophoresis) was used (Souza *et al*., 1999). Data for the 36 human strains not described in a previous analysis were added. The UPGMA (unweighted pair group method with arithmetic means) dendrogram was constructed using the PAUP program (D. L. Swofford, 2001, Version 4; Sinauer Associates, Sunderland, MA, USA), using the proportion of mismatches as a distance (Ochman & Selander, 1984; Whittam, 1990). We performed 1000 bootstrap replications on the UPGMA analysis. Bootstrap values above 50% are shown on the tree.

Statistical analysis. To determine whether a statistical association exists between bearing the pathogenic traits (LEE genes *cesT/eae* and *espB*), the serotype and genealogy of the bacterial strains, and the host order and diet, chi squared tests were performed. The observed and expected frequencies were compared, in contingency tables, for the presence of each PCR marker alone, the presence of both together and the absence of both markers, versus the host group, the bacterial serotype, the host diet and the strain genealogy. These analyses were carried out under the assumption of independence (Hamburg, 1979) of the pathogenic markers with the source of the strains and with the genealogy of the strains.

RESULTS

Bacterial serotypes

The pathogenic group (EPEC, EHEC, ETEC, EIEC, EAEC or UTI) of each strain was determined based on its serotype (Nataro & Kaper, 1998). Of the 122 strains, 38·5% presented pathogenic serotypes, 26·2% from

human strains and 12·3% from wild isolates (see supplementary data at http://mic.sgmjournals.org). The pathogenic serotypes were more common in isolates from carnivores (40%) than from any other animal host. ETEC serotypes were common in isolates from Xenarthra and Rodentia. Humans were found to contain all pathogenic serotypes. This is not surprising, as the human-derived strains were chosen to include most of the pathogenic groups from Mexico. The most common pathogenic serotype was EPEC in both humans and animal strains. EIEC and UTI serotypes were present only in human strains (see supplementary data).

PCR detection of the *cesT/eae* and *espB* genes

The PCR amplification results were analysed according to the bacterial serotype (Table 1), the host order (Table 2) and the host diet (Table 3). The PCR analysis indicates that the *cesT/eae* (c) and *espB* (d) genes are not always together in a given strain (see supplementary data). Almost half of the strains (40·9%) amplified at least one of the markers. In general, both markers were found together (25·4%) more often than separated. *cesT/eae* (8·2% alone) was less common than *espB* (15·6% alone, Tables 1, 2 and 3). Many of the EPEC strains (89·5%) showed the presence of both pathogenic markers (LEE genes *cesT/eae* and *espB*). Sixty percent of the EAEC strains showed both pathogenic markers as did 57·1% of EHEC strains. Only 9·3% of the non-pathogenic strains showed both pathogenic markers. ETEC and EIEC strains showed none of the pathogenic markers (Table 1). We found an association between the pathogenic markers and the serotype of the bacteria. ($\chi^2 =$ 91·998, 18 d.f., $P = < 0{\cdot}0001$; Table 4), which suggests that serotype is a good predictor of the presence of the LEE region in *E. coli*.

The PCR product for *espB* was present at high frequencies in the *E. coli* isolated from certain orders of mammals where the PCR for *cesT/eae* was not found (i.e. Perissodactyla and Artiodactyla). Both markers of the LEE were present in *E. coli* isolated from Carnivora, Rodentia, Chiroptera and Primates. The markers occurred in a higher proportion of *E. coli* isolated from primates and carnivores (Table 2). There is significant association between the presence of pathogenic genes and the order of the host ($\chi^2 = 53{\cdot}153$, 30 d.f., $P =$ 0·006; Table 4).

There was also an association between both markers of the LEE and the host diet (Table 3), as the markers were present in a higher proportion in omnivorous and carnivorous mammals than in other diet groups ($\chi^2 =$ 44·292, 21 d.f., $P = 0{\cdot}002$; Table 4). Strains from granivorous and insectivorous diets had only the *espB* marker.

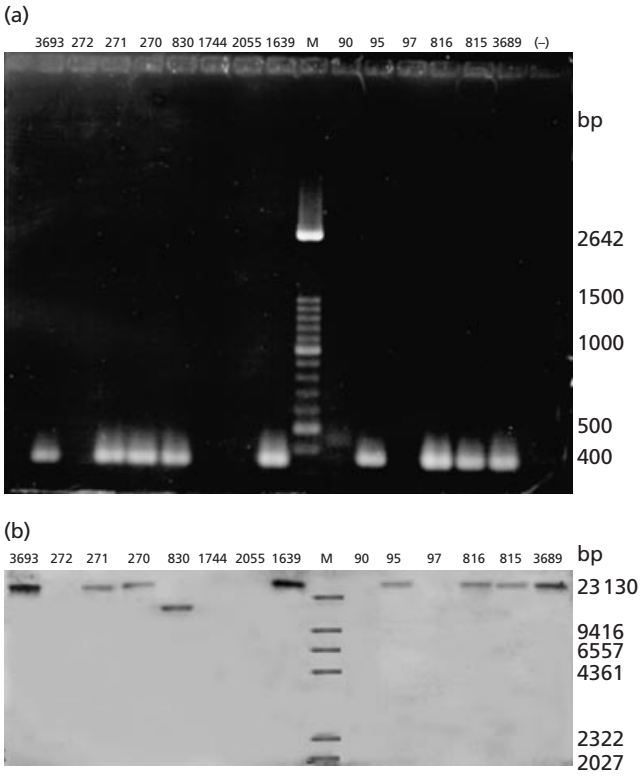

Fig. 2. (a) PCR products of 14 *E. coli* isolates from wild mammals and humans amplified with the primers for the c region of the LEE. The molecular size marker (M) is a ladder of 100 bp (Boehringer Mannheim). (b) Autoradiograph of Southern blot hybridization using probe C (gene *eae*) for the same strains tested in (a). The molecular size marker (M; lambda DNA/*Hin*dIII; Gibco-BRL) was added to the photograph using the pattern observed in the agarose gel.

Southern DNA hybridization

To test the hypothesis of the independent distribution of the components of the LEE in our sample, a more detailed analysis of a randomly chosen subsample of 25 strains using DNA Southern blotting and hybridization was performed. Four regions of the LEE were analysed: an upstream flanking region of the locus, genes *rorf1* and *rorf2* (probe A); the *esc* region, genes *escJ*, *rorf8*, *sepZ*, *orf12* and most of the *escV* gene (probe B); the *eae* region (probe C); and the *espB* region (probe D) (Fig. 1).

Fig. 3. UPGMA dendrogram showing the relationships between *E. coli* strains associated with wild mammals and humans as deduced from the genetic distance of 11 allozymes. The presence of pathogenic markers resulting from PCR amplification are shown as letters: c for the PCR *cesT/eae* marker and d for the PCR *espB* marker. Capital letters A, B, C and D correspond to the presence of Southern hybridization markers: genes *rorf1* and *rorf2*, *esc–sep* region, *eae* region and *espB* region of the LEE, respectively. The pathogenic type of each strain is also shown. The main genetic groups (I, II and III) are shown in Roman numerals. Bootstrap values above 50% are shown.

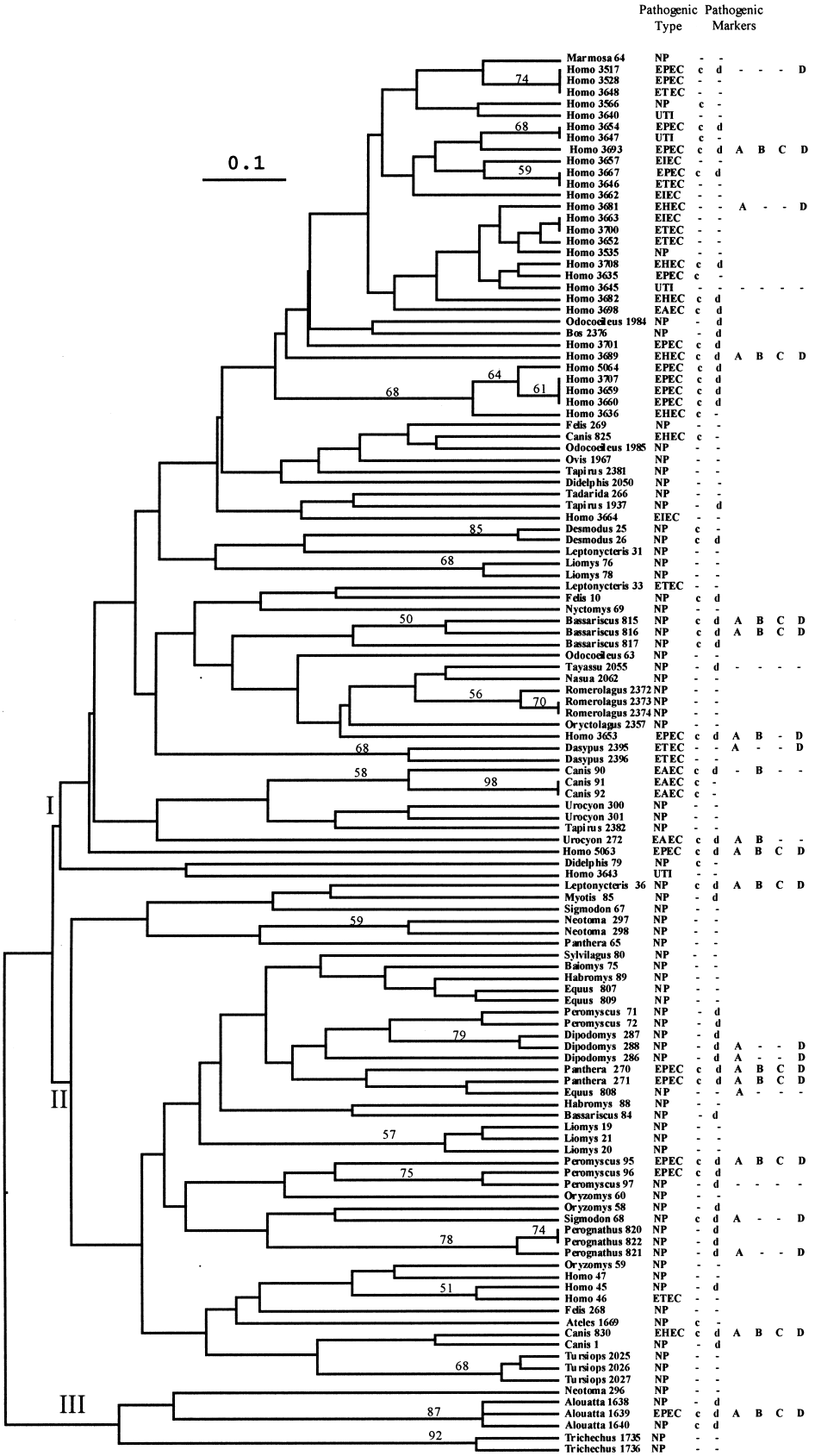

Fig. 3. For legend see opposite.

Southern blotting revealed that in a given strain, each of the component sections of the LEE can be present independently of the others. The results indicate eight different patterns of hybridization, Twelve percent of the strains showed no hybridization at all, 4% of the strains hybridized only with A probe, 4% only with probe B and 4% only with probe D. Four percent hybridized with probes A and B, 24% hybridized only with probes A and D, 44% with all probes and 4% with A, B and D (Table 5). Probe C was never detected alone. Probes A and D had the highest proportions of hybridization (72% of the strains), while probes B and C had lower proportions (56% and 44% of the strains, respectively). In general, the results of the PCR and Southern blotting experiments are congruent (Fig. 2 and Table 5) ($r = 0{\cdot}57$, $P < 0{\cdot}005$). However PCR markers for *cesT/eae* (c) and *espB* (d) are not always congruent with probes C and D. A negative result for PCR and a positive result for the Southern blot could be the consequence of variability at the PCR priming sites. The reverse situation could be explained by the presence of very divergent alleles of the genes (sequence data of *espB* confirms this hypothesis; V. Souza, unpublished data). Nevertheless, the PCR and Southern blot data suggest that the LEE region is not always a complete cassette, and that the screened LEE genes may be found in different combinations. (Fig. 3, Tables 1, 2, 3 and 5).

UPGMA dendrogram

The presence and absence of both PCR markers *cesT/eae* (c) and *espB* (d) and Southern blot (probes A to D) results were mapped in the dendrogram generated by MLEE data (Fig. 3). Three groups were defined: I, II and III. Group I includes most of the human-associated strains. Group II is characterized by strains from rodents, and group III, the most divergent group, includes manatee and monkey strains. The *espB* gene alone was more common in group II, while *cesT/eae* alone was more common in group I. Both genes together were observed scattered throughout the tree. A chi squared test indicated a significant association between the *E. coli* groups and the PCR LEE markers ($\chi^2 = 18{\cdot}574$, 6 d.f., $P = 0{\cdot}005$; Table 4).

DISCUSSION

These results indicate that, contrary to the view proposed elsewhere (Elliot *et al.* 1998; Kaper, 1998; Nataro & Kaper, 1998), the LEE region is a dynamic entity, both at the gene-cluster level (as it is acquired by different independent lineages of *E. coli*), and at the gene level (as different gene patterns within the locus are observed). These observations are interesting for two reasons: they may suggest that the individual genes of the LEE region have other functions when they are not clustered and the data illustrate the process of LEE evolution and assembly. The different genes that comprise the LEE may have evolved independently and served different functions in non-pathogenic bacterial strains. These genes may have been brought together by horizontal transfer in a gradual manner, until their subsequent integration into a PAI (Lawrence, 1997).

There is evidence (Nataro & Kaper, 1998) that EPEC and EHEC serotypes are good markers of this particular genetic background. However, even in human pathogens such as EPEC strain 3517 and EHEC strain 3681, the LEE region is not complete since these strains lack the genes for the secretion system (Table 5, Fig. 1) and are unable to produce the A-E lesion in Hep-2 cells (A. Navarro, unpublished data). In fact, when an adherence test in Hep-2 cells was performed, none of the strains from wild mammals were able to elicit the A-E lesion, while most of the clinical isolates from humans tested positive for the lesion (A. Navarro, unpublished data). The fact that some strains have only some components of the LEE, reinforces the idea that the locus is dynamic. All these data suggest that the entire PAI is not present due to a single genetic event; rather, it is possible that fractions of it were sequentially adapted to one or more intermediate lineages (de la Cruz & Davies, 2000). A combination of several genes may be particularly effective in enabling a given strain to become a successful pathogen. Such a pathogenic strain could then spread in an epidemic way (Maynard-Smith, 1992). The LEE cluster of genes can move together to different lineages of *E. coli*. Data on the *eae* gene suggests that the LEE was assembled independently at least twice in the different EPEC-EHEC groups (McGraw *et al.*, 1999; Reid *et al.*, 2000). The *eae* sequence analysis on these strains (V. Souza, unpublished data), along with current data, suggests a model for the evolution of pathogenicity where horizontal transfer is important both in the initial construction of the original gene complex (i.e. the PAI) and in the latter dispersal of the assembled complex. A possible mechanism of horizontal transfer of the LEE (and/or its components) is by transduction, since the PAI usually inserts into a phage insertion site, the *selC* gene (McDaniel *et al.*, 1995). This patchwork model has been suggested by different authors to account for the origin of new pathogenic clones that can cause novel diseases (de la Cruz & Davies, 2000; Lederberg, 1998), and in particular, in the assembly of the O-antigen complex studied by Reeves (1992, 1993).

It is clear that there are other possible models to explain the evolution of the LEE region. One alternative is that the LEE was originally acquired as a complete entity and that the strains that have only part of it have lost portions of the locus by deletion or chromosomal rearrangement. The G+C content of the LEE is 38·36 mol% (Elliot *et al.*, 1998), which is different from the mean *E. coli* G+C content of 50·8 mol% (Blattner *et al.*, 1997). This suggests that *E. coli* may have recently acquired the LEE. The fact that 72% of the subsample of 25 strains has the flanking regions of the LEE (probes A and D) suggests the possibility of disruption of the inserted cassette.

An adaptive role of the components of the LEE is suggested by the significant association between the pathogenic genes, estimated by PCR, and the host order and diet. This may be due to the differences in the

intestinal epithelia of the hosts, suggesting that the A-E lesion cannot form in the intestines of herbivores. Each gene may have important functions associated with the host biology or diet and this function may not necessarily be correlated with pathogenicity.

The UPGMA dendrogram shows that most of the *E. coli* strains associated with humans, pathogenic or not, are located in group I. Nevertheless, even closely related strains in that group may have different combinations of the LEE elements. In general, no clear pattern was seen in the distribution of pathogenic traits along the dendrogram. However, the interpretation of this analysis should be regarded as preliminary, as in general the bootstrap values are weak. Low bootstrap values are expected given the high number of strains (122) and the low number of characters (11 loci) used (Li, 1997). Also, these bootstrap values could be low as a consequence of genetic recombination among *E. coli* strains, as has been suggested by several authors (Peek *et al*. 2001; V. Souza, unpublished data). Moreover, our data suggest that this genetic complex may be dispersed into different genotypic backgrounds by lateral transfer.

This model reinforces the role of recombination in shaping the genetic structure and evolution of *E. coli*, as suggested by several authors (Dykhuizen & Green, 1991; Guttman, 1997). Our study, and that of Reid *et al*. (2000), indicates that different lineages of *E. coli* have acquired the same pathogenic genes several times, including complete PAIs and plasmid-encoded genes. It is suggested that these PAIs may be then horizontally transferred to different *E. coli* lineages, which result in the generation of novel pathogenic strains. In this case the pathogenic capabilities are the sum of different factors. However, the PAI may be unstable due to high rates of internal recombination. This may cause the high diversity observed in the *eae* genes (McGraw *et al*., 1999) and *tir* genes (Paton *et al*., 1998) but may also cause the disruption of the island, turning a pathogenic strain into a commensal *E. coli* strain.

ACKNOWLEDGEMENTS

We wish to acknowledge Dr J. B. Kaper for his prototypic strain and the LEE probes he kindly shared with us. Dr José Luis Puente, Oscar Rodríguez, Dr Guilermo Dávila, Laura Espinoza, Amanda Castillo and Aldo Valera for technical assistance. We thank Claudia Silva and Martha Rocha for their comments and moral support, and Jordan Golubov and Arturo Silva for the statistical analysis. Several colleagues in Mexico helped in assembling the collection of strains used in this paper. The list is too long to be included here, but we greatly thank them. This research was supported by DGAPA, UNAM IN-218698 and CONACYT project 27557-M to V. Souza, and was written during a sabbatical of V. Souza and L. Eguiarte at the University of California at Irvine, with support of CONACYT and DGAPA, UNAM.

REFERENCES

Beltrametti, F., Kresse, A. U. & Guzmán, C. A. (1999). Transcriptional regulation of the *esp* genes of enterohaemorrhagic *Escherichia coli*. *J Bacteriol* **181**, 3409–3418.

Blattner, F. R., Plunkett, I. G., Bloch, C. A. & 14 other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462.

Cravioto, A., Gross, R. J., Scotland, S. M. & Rowe, B. (1979). An adhesive factor found in strains of *Escherichia coli* belonging to the traditional infantile enteropathogenic serotypes. *Curr Microbiol* **3**, 95–99.

de la Cruz, F. & Davies, J. (2000). Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol* **8**, 128–133.

Donnenberg, M. S. & Kaper, J. B. (1992). Enteropathogenic *Escherichia coli*. *Infect Immun* **60**, 3953–3961.

Dykhuizen, D. E. & Green, L. (1991). Recombination in *Escherichia coli* and the definition of biological species. *J Bacteriol* **173**, 7257–7268.

Elliot, S. J., Wainwright, L. A., McDaniel, T. K., Jarvis, K. G., Deng, Y. K., Lai, L. C., MacNamara, B. P., Donnenberg, M. S. & Kaper, J. B. (1998). The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol* **28**, 1–4.

Finlay, B. B. & Abe, A. (1998). Enteropathogenic *Escherichia coli* interactions with host cells. *Jpn J Med Sci Biol* **51**, 91–100.

Finlay, B. B. & Falkow, S. (1997). Common themes in microbial pathogenicity revisited. *Mol Biol Microbiol Rev* **61**, 136–169.

Guttman, D. S. (1997). Recombination and clonality in natural populations of *Escherichia coli*. *Trends Ecol Evol* **12**, 16–22.

Hamburg, M. (1979). *Basic Statistics: a Modern Approach*, 2nd edn. New York: Harcourt Brace Jovanovich.

Hartland, E. L., Batchelor, M., Delahay, R. M., Hale, C., Matthews, S., Dougan, G., Knutton, S., Connerton, I. & Frankel, G. (1999). Binding of intimin from enteropathogenic *Escherichia coli* to Tir and to host cells. *Mol Microbiol* **32**, 151–158.

Ismaili, A., McWhirter, E., Handelsman, M. Y. C., Brunton, J. L. & Sherman, P. M. (1998). Divergent signal transduction responses to infection with attaching and effacing *Escherichia coli*. *Infect Immun* **66**, 1688–1696.

Jarvis, K. G., Girón, J. A., Jerse, A. E., Mc Daniel, T. K., Donnenberg, M. S. & Kaper, J. B. (1995). Enteropathogenic *Escherichia coli* contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation. *Proc Natl Acad Sci U S A* **92**, 7996–8000.

Jerse, A. E. & Kaper, J. B. (1991). The *eae* gene of enteropathogenic *Escherichia coli* encodes a 94-kilodalton membrane protein, the expression of which is influenced by the EAF plasmid. *Infect Immun* **59**, 4302–4309.

Jerse, A. E., Yu, J., Tall, B. D. & Kaper, J. B. (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 U S A* **87**, 7839–7843.

Kaper, J. B. (1998). The locus of enterocyte effacement pathogenicity island of shiga toxin-producing *Escherichia coli* O157:H7 and other attaching and effacing *E. coli*. *Jpn J Med Sci Biol* **51**, 101–107.

Kenny, B. & Finlay, B. B. (1995). Protein secretion by enteropathogenic *Escherichia coli* is essential for transducing signals to epithelial cells. *Proc Natl Acad Sci U S A* **92**, 7991–7995.

Kenny, B. & Finlay, B. B. (1997). Intimin dependent binding of enteropathogenic *Escherichia coli* to host cells triggers novel signaling events, including tyrosine phosphorylation of phospholipase C-1. *Infect Immun* **65**, 2528–2536.

Kenny, B., DeVinney, R., Stein, M., Reinscheid, D. J., Frey, E. A. &

Finlay, B. B. (1997). Enteropathogenic *Escherichia coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* **91**, 511–520.

Lawrence, J. G. (1997). Selfish operons and speciation by gene transfer. *Trends Microbiol* **5**, 355–359.

Lederberg, J. (1998). Emerging infection: an evolutionary perspective. *Emerg Infect Dis* **4**, 366–371.

Li, W.-H. (1997). Molecular phylogenetics: methods. In *Molecular Evolution*, pp. 144–146. Sunderland, MA: Sinauer.

McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S. & Kaper, J. B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci U S A* **92**, 1664–1668.

McGraw, E. A., Li, J., Selander, R. K. & Whittam, T. S. (1999). Molecular evolution and mosaic structure of α, β, and γ intimins of pathogenic *Escherichia coli*. *Mol Biol Evol* **16**, 12–22.

Maynard-Smith, J. (1992). Analysing the mosaic structure of genes. *J Mol Evol* **34**, 126–129.

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

Ochman, H. & Selander, R. K. (1984). Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* **157**, 690–693.

Orskov, F. & Orskov, I. (1984). Serotyping of *Escherichia coli*. In *Methods in Microbiology*, vol. 14. Edited by T. Bergan. London: Academic Press.

Paton, A. W., Manning, P. A., Woodrow, M. C. & Paton, J. C. (1998). Translocated intimin receptors (Tir) of Shiga-toxigenic *Escherichia coli* isolates belonging to serogroup O26, O111 and O157 react with sera from patients with hemolitic-uremic syndrome and exhibit marked sequence heterogeneity. *Infect Immun* **66**, 5580–5586.

Peek, A. S., Souza, V., Eguiarte, L. E. & Gaut, B. S. (2001). Selection and recombination at the type 1-fimbrial major subunit (fimA) in natural isolates of *Escherichia coli*. *J Mol Evol* **52**, 193–204.

Reeves, P. R. (1992). Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiol Lett* **100**, 509–516.

Reeves, P. R. (1993). Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. *Trends Genet* **9**, 17–22.

Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K. & Whittam, T. S. (2000). Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406**, 64–67.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503.

Souza, V., Rocha, M., Valera, A. & Eguiarte, L. E. (1999). Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. *Appl Environ Microbiol* **65**, 3373–3385.

Taylor, K. A., O'Connell, C. B., Luther, P. W. & Donnenberg, M. S. (1998). The *espB* protein of enteropathogenic *Escherichia coli* is targeted to the cytoplasm of infected HeLa cells. *Infect Immun* **66**, 5501–5507.

Tzipori, S., Gunzer, F., Donnenberg, M. S., de Montigny, L., Kaper, J. B. & Donohue-Rolfe, A. (1995). The role of *eaeA* gene in diarrhea and neurological complications in a gnotobiotic piglet model of enterohaemorrhagic *Escherichia coli* infection. *Infect Immun* **63**, 3621–3627.

Whittam, T. S. (1990). ETDIV and ETCLUS programs. Pennsylvania, PA: Pennsylvania State University.

Wieler, L. H., McDaniel, T. K., Whittam, T. S. & Kaper, J. B. (1997). Insertion site of the locus of enterocyte effacement in enteropathogenic and enterohaemorrhagic *Escherichia coli* differs in relation to the clonal genealogy of the strains. *FEMS Microbiol Lett* **156**, 49–53.

Zhao, S., Mitchell, S. E., Meng, J., Doyle, M. P. & Kresovich, S. (1995). Cloning and nucleotide sequence of a gene upstream of the *eaeA* gene of enterohaemorrhagic *Escherichia coli* O157:H7. *FEMS Microbiol Lett* **133**, 35–39.

Received 29 November 2000; revised 2 July 2001; accepted 20 July 2001.