Microbiology (1999), **145**, 1005–1014 Printed in Great Britain

The large plasmids of Shiga-toxin-producing *Escherichia coli* (STEC) are highly variable genetic elements

Werner Brunder, Herbert Schmidt, Matthias Frosch and Helge Karch

Author for correspondence: Werner Brunder. Tel: +49 931 201 3981. Fax: +49 931 201 3445. e-mail: wbrunder@hygiene.uni-wuerzburg.de

Institut für Hygiene und Mikrobiologie der Universität Würzburg, Josef-Schneider-Str. 2, D-97080 Würzburg, Germany

Shiga-toxin-producing *Escherichia coli* (STEC) of different serotypes are known to harbour large plasmids. The aim of this study was to investigate, using the example of the plasmid-encoded serine protease EspP, whether these plasmids are a uniform genetic element present in STEC. Examination of 201 diarrhoeagenic *E. coli* strains using a newly developed *espP*-specific PCR showed that *espP* is specific for STEC and present in 57% of STEC belonging to 16 different serotypes. The *espP* genes of the 16 STEC serotypes varied to a certain extent, as shown by nucleotide sequence and restriction enzyme analyses, but the DNA regions adjacent to the *espP* gene were completely different. When two further STEC-plasmid markers, the catalase-peroxidase gene *katP* and the enterohaemorrhagic *E. coli*-haemolysin gene EHEC-*hlyA* were included, many combinations of the three markers were found, depending in part on the serotype. In addition, strains possessing none of the three markers still harboured large plasmids. In the most prevalent STEC serogroup, O157, it was observed that the plasmid of sorbitol-fermenting STEC O157:H$^-$ lacks the *espP* and *katP* genes although both genes are present in the plasmid of the non-sorbitol-fermenting STEC O157:H7. The EHEC-*hlyA* gene, however, is present in both. In conclusion, this study shows that the large plasmids of STEC are not uniform genetic elements but heterogeneous in both their gene composition and arrangement.

Keywords: Shiga-toxin-producing *Escherichia coli*, plasmids, variability, serine protease, *espP*

INTRODUCTION

Shiga-toxin-producing *Escherichia coli* (STEC) are a well-known cause of severe human diseases such as haemorrhagic colitis and haemolytic–uraemic syndrome (HUS). STEC harbour plasmids of various sizes, the most common being 75–100 kbp (Bopp *et al.*, 1987; Ostroff *et al.*, 1989; Fratamico *et al.*, 1993; Paros *et al.*, 1993).

Plasmids are also found in other pathogroups of diarrhoeagenic *E. coli*. They often carry genes responsible for the expression of their group-specific pathogenicity, e.g. the heat-labile and heat-stable toxins of enterotoxinogenic *E. coli* (ETEC), the invasion factors of enteroinvasive *E. coli* (EIEC) or the aggregative adherence fimbriae mediating the typical phenotype of enteroaggregative *E. coli* (EAEC) (for a review see Nataro & Kaper, 1998). Plasmids or plasmid-encoded virulence factors were also used to define and differentiate the *E. coli* pathogroups. For example, possession of the so-called EAF (EPEC adherence factor) plasmid is used to characterize enteropathogenic *E. coli* (EPEC) (Nataro & Kaper, 1998). However, identification or classification of such plasmids has often been based only on their size or on the presence of a small DNA fragment in the plasmid using hybridization or PCR. Examples are a PCR for the detection of EAEC plasmids (Schmidt *et al.*, 1995a), probes specific for the EPEC EAF plasmid (Nataro *et al.*, 1985; Girón *et al.*, 1993) or the CVD419 probe for the large plasmid of STEC (Levine *et al.*,

Abbreviations: EAEC, enteroaggregative *E. coli*; EHEC, enterohaemorrhagic *E. coli*; EIEC, enteroinvasive *E. coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxinogenic *E. coli*; HUS, haemolytic–uraemic syndrome; IS, insertion sequence; NSF, non-sorbitol-fermenting; SF, sorbitol-fermenting; STEC, Shiga-toxin-producing *E. coli*.

The EMBL accession numbers for the sequences reported in this paper are AJ010390 and AJ007716.

1987). Apart from these examples, there is little information on the variability of the plasmids, or on their mobility or interchangeability between *E. coli* strains of different patho- or serotypes. As a rule, only the plasmid of one or two prototype strains from a particular group has been studied and characterized to any extent.

Recently, the nucleotide sequence of the large plasmid of an STEC O157:H7 strain (RIMD 0509952), derived from an outbreak in Japan in 1996, was completely determined (Makino *et al.*, 1998). Our group investigated, at both the molecular and functional level, plasmid pO157 of STEC O157:H7 reference strain EDL933, isolated during one of the first reported outbreaks of *E. coli* O157:H7 in 1982 (O'Brien *et al.*, 1983; Riley *et al.*, 1983). We showed that this prototype plasmid carried the genetic information for the EHEC (enterohaemorrhagic *E. coli*) haemolysin (Schmidt *et al.*, 1995b, 1996) which is responsible for the so-called enterohaemolytic phenotype. In addition, we found genes encoding a periplasmic catalase-peroxidase (*katP*, Brunder *et al.*, 1996) and a complex type II secretion system (*etp*; Schmidt *et al.*, 1997). Recently, we described an extracellular serine protease, termed EspP, encoded by pO157 (Brunder *et al.*, 1997). We showed that EspP is able to cleave pepsin and coagulation factor V and have suggested therefore that EspP might be an accessory virulence factor exacerbating haemorrhagic colitis. Djafari *et al.* (1997) reported a similar protein (PssA) encoded by a large plasmid of a bovine STEC O26:H$^-$ strain and showed that it is cytotoxic for Vero cells. Comparison of the two sequences revealed eight base differences within the ORF. One of these led to an amino acid substitution at amino acid position 1113 where PssA had a valine residue instead of the phenylalanine residue in EspP.

In the present study, we investigated the prevalence of the *espP* gene in a large number of clinical STEC isolates of a wide variety of serotypes, and in other diarrhoeagenic *E. coli*. Furthermore, we analysed the diversity of the *espP* gene within and between STEC serotypes and its linkage with other markers of pO157. The results of this study may contribute to our understanding of the structure and diversity of STEC and their plasmids.

METHODS

Bacterial isolates. A total of 156 STEC isolates belonging to 30 serotypes were investigated (see Table 1). The strains were isolated between January 1995 and December 1996 from individual human stool samples collected in Germany. A few well-characterized STEC reference strains and isolates from bovine faeces, meat and raw milk were included in the study. Serotyping of *E. coli* was performed according to Bockemühl *et al.* (1992). D-Sorbitol fermentation was examined on D-sorbitol-MacConkey agar (Oxoid).

Of the 83 STEC isolates belonging to serogroup O157, 48 possessed the H7 flagellar antigen whereas 35 were H$^-$. Sixty-nine of the O157 strains harboured the stx_2 gene and 14 both the stx_1 and stx_2 genes. Seventy-five per cent of the O157 strains were from patients suffering from HUS, the others from patients with diarrhoea or asymptomatic carriers. We also investigated 20 STEC strains belonging to serogroup O26 (11 isolates carried H11, nine isolates were H$^-$), 12 of which harboured the stx_2 gene and eight the stx_1 gene. Half of the strains in this serogroup were from patients suffering from HUS. Of eight strains of serogroup O111, five possessed the stx_1 gene, two the stx_2 gene and one both genes. Eleven isolates belonging to serotype O103:H2 were also tested, 10 of which harboured the stx_1 gene alone and one both the stx_1 and stx_2 genes. One to three strains of STEC serogroups rarely isolated from humans were included in the study.

Strains of other diarrhoeagenic *E. coli* pathogroups (EPEC, ETEC, EIEC and EAEC) were obtained from our strain collection. *E. coli* K-12 laboratory strains DH5α and HB101 (Gibco-BRL) were used as hosts for cloned DNA. STEC O157:H7 prototype strain EDL933 (O'Brien *et al.*, 1983; Riley *et al.*, 1983) and its plasmid-cured derivative EDL933-cu (Tzipori *et al.*, 1987) were used as control strains for PCR and hybridization experiments.

PCR. PCR for detection of *espP* sequences was performed using the GeneAmp 9600 PCR System (Perkin Elmer). Primers esp-A (5′-AAACAGCAGGCACTTGAACG-3′) and esp-B (5′-GGAGTCGTCAGTCAGTAGAT-3′) were designed to amplify a 1830 bp fragment of the *espP* gene. Amplification was carried out in a total volume of 50 µl containing 5 µl bacterial cell suspension (about 10^3 bacterial cells suspended in 0·85%, w/v, sodium chloride), 30 pmol each primer, 200 µM each dNTP, 5 µl 10 × GeneAmp PCR buffer II, 3 µl 25 mM $MgCl_2$ and 2 units Ampli*Taq* DNA polymerase (Perkin Elmer). After an initial denaturation step of 5 min at 94 °C the samples were subjected to 30 cycles, each consisting of 30 s at 94 °C (denaturing), 60 s at 56 °C (annealing) and 150 s at 72 °C (extension). The reaction was completed with a final extension step of 5 min at 72 °C. The PCR products were analysed by agarose gel electrophoresis and ethidium bromide staining. Each reaction was conducted at least twice in independent reactions.

PCR for detection of the *katP* gene and the EHEC-*hlyA* gene was performed as described previously (Schmidt *et al.*, 1995b; Brunder *et al.*, 1996). PCRs for the investigation of DNA regions adjacent to the *espP* gene were performed using primers web-5 (5′-CAGAAGTATATGCCTCGGATGA-3′), web-6 (5′-AAATATGCCCTTGTTCTGTGC-3′), web-7 (5′-GGCCGAACGGGTGTGGATGT-3′), web-8 (5′-TCGATTTCGCTGCCGTGACTGT-3′) and web-9 (5′-TGCTTAGCCAGACGGACCTTA-3′). See Fig. 3 for the binding sites of these primers. Reaction conditions for primer pairs web-5/web-6 and web-9/web-6 were identical to that used for the *espP* PCR; for primer pair web-7/web-8, the annealing temperature was 60 °C instead of 56 °C.

General recombinant DNA techniques. Plasmid DNA was purified with Nucleobond AX cartridges (Macherey-Nagel) according to the instructions of the supplier. Purification of DNA from agarose gels was performed using a Prep-A-Gene kit (Bio-Rad). Restriction enzymes and T4 DNA ligase were purchased from Gíbco-BRL and New England Biolabs. Plasmid pK18 (Pridmore, 1987) was used as the cloning vector. Restriction enzyme analysis, ligation and transformation were conducted according to standard procedures (Sambrook *et al.*, 1989). Southern-hybridization experiments were performed using a digoxigenin labelling and detection kit (Boehringer Mannheim) according to the manufacturer's instructions.

Nucleotide sequencing. Nucleotide sequencing was performed with an automatic sequencer (model 373A, Applied Biosystems) as described previously (Schmidt *et al.*, 1995b). Both strands of the DNA were sequenced stepwise using

customized oligonucleotide primers (ARK Scientific). Each base was determined on average three times.

Sequence analysis was conducted with the HUSAR program package (Heidelberg Unix Sequence Analysis Resources, German Cancer Research Centre, Heidelberg, Germany) as well as with the DNASIS program (Hitachi Software).

Colony-blot hybridization. This was performed using a digoxigenin labelling and detection kit (Boehringer Mannheim). Probes were prepared by random labelling of purified *espP*, *katP* or EHEC-*hlyA* PCR products, which were obtained using DNA of STEC O157:H7 strain EDL933 as the template.

Bacteria were picked onto Luria agar plates and incubated for 16 h at 37 °C. The plates were cooled at 4 °C for 30 min. Sheets of nylon membrane (Zeta-Probe GT; Bio-Rad) were placed onto the plates for 1 min. Cells were subsequently lysed by alkali as follows. The membrane was laid onto a Whatman 3MM paper soaked with 0·5 M NaOH and 1·5 M NaCl and incubated for 20 min at room temperature. Then the filter was dried on Whatman paper at room temperature for 5 min. The procedure above was repeated twice, then the filter was placed onto a Whatman paper soaked with a solution of 0·5 M NaOH and 0·2% Triton X-100 and incubated for 15 min. After that, the filter was neutralized using 1 M Tris/HCl, pH 7·5, 1·5 M NaCl. Finally, the membrane was washed for 5 min in 2 × SSC (300 mM NaCl, 30 mM sodium citrate), air-dried and baked for 2 h at 80 °C. Hybridization was carried out under stringent conditions according to the manufacturer's instructions. Each strain was tested at least twice.

RESULTS

Prevalence of *espP* in STEC of various serotypes

To investigate the prevalence of *espP* sequences in a large number of STEC isolates, we developed a PCR strategy for detection of the *espP* gene. PCR primers esp-A and esp-B were designed to amplify a 1830 bp fragment of the *espP* gene (Fig. 1). To ensure a high specificity in the PCR, the primer-binding sites were selected from regions where there was a low level of similarity between the *espP* gene and the homologous genes *espC* from EPEC (Stein *et al.*, 1996), *sepA* from *Shigella flexneri* (Benjelloun-Touimi *et al.*, 1995) and *tsh* of avian pathogenic *E. coli* (Provence & Curtiss, 1994).

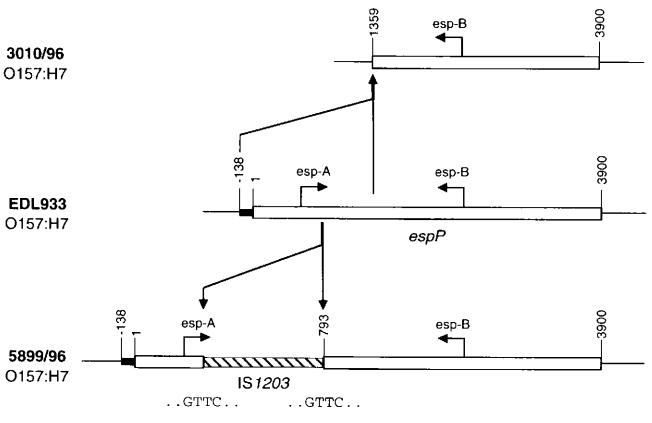

Fig. 1. Variation of the *espP* gene (white box) in STEC strains of serotype O157:H7. Strain 3010/96 exhibited a partial deletion of the gene. The *espP* gene of strain 5899/96 is interrupted by insertion of IS*1203* (hatched box). Binding sites of the PCR primers esp-A and esp-B and the IS target-site duplication in strain 5899/96 are shown. Sequence positions in base pairs are given with reference to the start codon of *espP*.

Table 1. Prevalence of *espP* in diarrhoeagenic *E. coli* determined by *espP*-PCR

Pathogroup/serotype	Total	*espP* positive (%)
STEC O157:H7/H⁻ (NSF)	63	61 (96%)*
STEC O157:H⁻ (SF)	20	0 (0%)
STEC O26:H11/H⁻	20	12 (60%)
STEC O111:H2/H⁻	8	1 (12%)
STEC O103:H2	11	0 (0%)
STEC of other serotypes†	34	15 (44%)
EPEC	15	0 (0%)
ETEC	5	0 (0%)
EIEC	5	0 (0%)
EAEC	20	0 (0%)

* The two strains negative by *espP* PCR (3010/96 and 5899/96) hybridized with the *espP* probe. Strain 5899/96 gave a PCR product of 3·1 kbp instead of the expected product of 1·8 kbp (see text for further comments).

† Other serotypes included O8:H14/H19/H⁻, O22:H8, O25:H14/H⁻, O52:H19, O55:H⁻, O69:H⁻, O77:H⁻, O98:H8, O104:H16, O113:H4/H⁻, O128:H⁻, O129:H⁻, O145:H⁻, O156:H27, ONT:H14/H⁻ and Orough:H11/H⁻.

We first tested the procedure on *E. coli* K-12 laboratory strains harbouring the cloned *espP* gene (DH5α/pB9 and DH5α/pB9-5; Brunder *et al.*, 1997) and negative control strains DH5α, HB101 and DH5α/pK18 (vector control). The STEC O157:H7 reference strain EDL933 and its plasmid-cured derivative EDL933-cu were also included in this evaluation. As expected, strains DH5α/pB9, DH5α/pB9-5 and EDL933, which harbour the *espP* gene, gave a positive PCR reaction, whereas the control strains were all negative in the *espP* PCR.

Next, we examined 156 STEC strains and 45 strains of other diarrhoeagenic *E. coli* (EPEC, EAEC, ETEC and EIEC) using *espP* PCR (see Table 1 and Methods). Of the 83 O157 strains, 73% possessed the *espP* gene. It was of interest that when these strains were subdivided according to their capability to ferment sorbitol there was a close correlation with the occurrence of the *espP* gene: 96% of the classical non-sorbitol-fermenting (NSF) STEC O157:H7/H⁻ strains possessed the *espP* gene, whereas the 20 strains of sorbitol-fermenting (SF) STEC O157:H⁻ tested were all negative in the *espP* PCR. The occurrence of the *espP* gene was less common

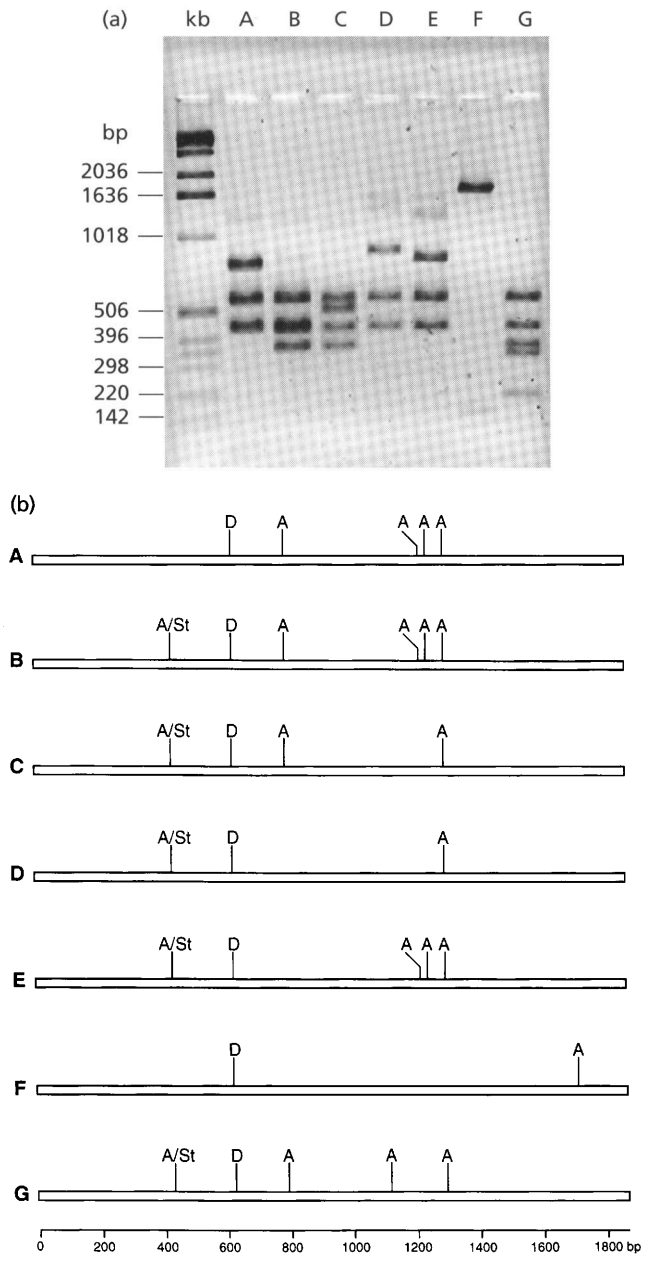

Fig. 2. RFLP within *espP* genes of STEC strains of various serotypes. (a) PCR products obtained with primers esp-A and esp-B were restricted by *AluI* and separated on a 2% (w/v) agarose gel. kb, 1 kb DNA ladder; A–G, restriction types. Restriction type E is distinguished from the similar type A by the presence of a *SstI* restriction site. (b) *AluI* (A), *DraI* (D) and *SstI* (St) restriction sites within the *espP* PCR products. For restriction types A and B, the sites are defined by the nucleotide sequence of the STEC O157 *espP* gene (type A) and the STEC O26 *pssA* gene (type B). Putative sites of the other restriction types were deduced from the known sequences and from double digests.

in STEC of non-O157 serotypes than in NSF STEC O157:H7/H$^-$ (Table 1). None of the 45 EPEC, EAEC, ETEC and EIEC strains was positive by *espP* PCR, indicating that *espP* is not present in diarrhoeagenic *E. coli* other than STEC (Table 1).

Table 2. *AluI/SstI* RFLP of the *espP* PCR product

Serotype	No. of strains of restriction type:						
	A	B	C	D	E	F	G
O157:H7/H$^-$ (NSF)	9	0	0	0	0	0	0
O26:H11/H$^-$	0	11	0	0	0	0	0
O145:H$^-$	0	2	0	0	0	0	0
O111:H$^-$	0	1	0	0	0	0	0
O22:H8	0	0	2	0	0	0	0
O69:H$^-$	0	0	1	0	0	0	0
Orough:H11	0	0	1	0	0	0	0
O156:H27	0	0	0	1	0	0	0
O55:H$^-$	0	0	0	0	2	0	0
O52:H19	0	0	0	0	1	0	0
O8:H19	0	0	0	0	1	0	0
ONT:H14/H$^-$	0	0	0	0	0	3	0
O98:H8	0	0	0	0	0	0	1

To exclude the possibility that the PCR failed to detect the genes due to minor sequence variations in the primer-binding sites, we investigated all strains by colony-blot hybridization using an *espP*-specific probe. With the exception of STEC O157:H7 strains 3010/96 and 5899/96, which were positive by hybridization but negative in the *espP* PCR, hybridization of all the strains mentioned above gave the same results as those obtained by PCR.

Deletions and insertions in *espP* genes of NSF STEC O157:H7

Strains 3010/96 and 5899/96 were examined in detail in order to explain the conflicting results obtained using *espP* PCR and colony-blot hybridization. Strain 3010/96 was negative in the *espP* PCR but exhibited hybridization with the *espP* probe. Strain 5899/96, however, was positive using hybridization, but the product of the *espP* PCR was approximately 3·1 kbp in length instead of 1·8 kbp. To investigate the nature of the variation in these unusual strains, we hybridized their plasmid DNA with the *espP*-specific probe and cloned the hybridizing *Bam*HI restriction fragments of 7 kbp (3010/96) and 9·9 kbp (5899/96), into pK18.

Restriction analysis and nucleotide sequencing of the recombinant plasmids revealed that strain 3010/96 had a deletion within the *espP* gene covering position −138 to 1359 with reference to the ATG start codon of *espP*. Since the deleted DNA fragment contained the primer-binding site for primer esp-A, this deletion is the cause of the negative PCR reaction (see Fig. 1).

Strain 5899/96, however, was shown to carry a copy of insertion sequence IS*1203* (Paton & Paton, 1994), introduced into the *espP* gene at position 793. This insertion has apparently been caused by a regular transposition event, since we detected direct repeats of the sequence GTTC at the ends of the insertion,

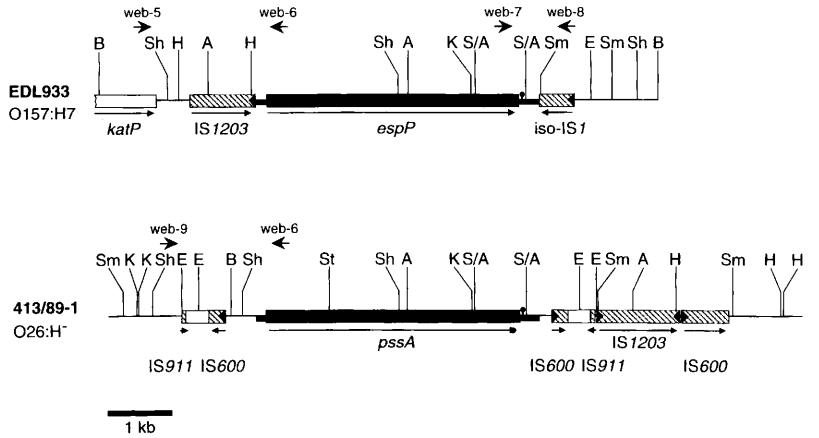

Fig. 3. Restriction maps of the *espP* region in pO157 of O157:H7 strain EDL933 (Brunder *et al.*, 1997) and the *pssA* region in the plasmid of O26:H⁻ strain 413/89-1 (Djafari *et al.*, 1997). The highly homologous *espP* and *pssA* genes, including their promoter and terminator regions, are shown in black. Adjacent sequences with homology to ISs are shown as hatched boxes, together with their inverted terminal repeats (black triangles) and putative transcription direction (small-headed arrows). The black 'lollipops' indicate the hairpin-loop structures of the rho-independent terminators of the *espP*/*pssA* genes. The *katP* gene present in strain EDL933 is depicted as a white box. Large-headed arrows indicate binding sites of the oligonucleotide primers used to investigate the sequences adjacent to *espP*/*pssA*. Restriction enzyme sites are shown: A, *Acc*I; B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; S, *Sal*I; Sm, *Sma*I; Sh, *Sph*I; St, *Sst*I.

Table 3. Linkage of the plasmid genes *espP* (e), *katP* (k) and EHEC-*hlyA* (h) in STEC strains of various serotypes

The number of strains exhibiting the respective marker combination is shown in relation to the serotype.

Genotype	Serotype					
	NSF O157:H7/H⁻	SF O157:H⁻	O26:H11/H⁻	O111:H2/H⁻	O103:H2	Other
$e^+ k^+ h^+$	**63**	0	**13**	**1**	0	**1**
$e^+ k^+ h^-$	0	0	0	0	0	0
$e^+ k^- h^+$	0	0	0	0	0	**11**
$e^+ k^- h^-$	0	0	0	0	0	**3**
$e^- k^+ h^+$	0	0	0	**1**	**6**	0
$e^- k^+ h^-$	0	0	0	0	0	**5**
$e^- k^- h^+$	0	**20**	**7**	**5**	**5**	**6**
$e^- k^- h^-$	0	0	0	**1**	0	**8**
Total	**63**	**20**	**20**	**8**	**11**	**34**

representing the target-site duplication generated by transposition (Fig. 1).

RFLP of *espP* genes

To investigate the diversity of the *espP* locus across a broad spectrum of STEC serotypes, we analysed *espP* PCR products obtained from 36 strains using restriction enzyme digestion with *Acc*I, *Dra*I, *Eco*RV, *Alu*I and *Sst*I. The subset of strains was representative for all those serotypes in which we had detected the *espP* gene.

Restriction with *Dra*I revealed the same pattern for all strains tested. In contrast, restriction with *Sst*I revealed two different patterns, one of which corresponded to that predicted by the nucleotide sequence of *espP* derived from strain EDL933 (EMBL accession number X97542; no *Sst*I site within the PCR product), the other to that of the STEC O26 *pssA* sequence (Djafari *et al.*, 1997; see Fig. 2b). We could identify seven distinct restriction patterns using *Alu*I (restriction types A–G, see Fig. 2). Restriction type E can be discriminated from the similar type A by the presence of one *Sst*I restriction site. All NSF O157:H7/H⁻ isolates showed restriction type A (Table 2). This pattern was not seen in any of the other serotypes analysed. A second restriction pattern, type B, was present in all STEC isolates of serogroup O26 tested and also in two strains of serotype O145:H⁻ and one O111:H⁻ strain. Like restriction type B, types C and E also included strains of different serotypes but no serotype contained more than one restriction type of the *espP* PCR product (Table 2). Restriction analysis using *Acc*I and *Eco*RV revealed the same pattern for all strains tested, with the exception of strains of serotype ONT:H14/H⁻ (data not shown).

Table 4. Southern hybridization of plasmid DNA with probes specific for *espP*, *katP* and EHEC-*hlyA*

ND, not determined; Ø, no hybridizing plasmid fragments detected.

Serogroup	Genotype	Strain	*espP* (kbp)*	*katP* (kbp)†	*hlyA* (kbp)*
NSF O157	$e^+ k^+ h^+$	3958/95	8·6	9·7	12
SF O157	$e^- k^- h^+$	3072/96	ND	ND	15
O26	$e^+ k^+ h^+$	7662/96	13	4·3	18
	$e^- k^- h^+$	6061/96	ND	ND	12
O111	$e^+ k^+ h^+$	4993/96	9·8	27	18
	$e^- k^+ h^+$	94127588	ND	8·2	12
	$e^- k^- h^+$	6037/96	ND	ND	13
O103	$e^- k^+ h^+$	5714/96	ND	5·5	12
	$e^- k^- h^+$	1858/96	ND	ND	20
O145	$e^+ k^+ h^+$	5157/96I	9	4·4	12
	$e^+ k^- h^+$	3985/96	9	ND	20
ONT	$e^+ k^- h^+$	0071/96	9·8	ND	28
	$e^- k^+ h^-$	1867/96	ND	30	ND
	$e^- k^- h^+$	6481/96	ND	ND	24
Orough	$e^+ k^- h^+$	7137/95	Ø	ND	Ø
	$e^+ k^- h^-$	6705/95	6·2	ND	ND
O55	$e^+ k^- h^-$	3379/96	25	ND	ND
	$e^- k^- h^+$	ED44	ND	ND	9
O25	$e^- k^+ h^-$	4817/96	ND	6	ND
	$e^- k^- h^+$	6561/95	ND	ND	Ø
O156	$e^+ k^- h^+$	6781/95	3·9	ND	20
O22	$e^+ k^- h^+$	4094	25	ND	10
O52	$e^+ k^- h^+$	ED99	9	ND	14
O69	$e^+ k^- h^+$	4014/95	24	ND	9·5
O8	$e^+ k^- h^+$	3651/96	Ø	ND	24
O98	$e^+ k^- h^-$	89/94	30	ND	ND
O129	$e^- k^+ h^-$	3378/96	ND	6	ND
O104	$e^- k^- h^+$	4823/96	ND	ND	9
O113	$e^- k^- h^+$	4395/96	ND	ND	9
O128	$e^- k^- h^+$	3560/96	ND	ND	23

* DNA restricted with *Bam*HI.

† DNA restricted with *Sma*I.

Variability at the *espP* boundaries

In addition to the *espP* gene itself, we analysed the DNA regions adjacent to the gene. Comparing the nucleotide sequences upstream and downstream of *espP* and *pssA*, we found that the region of high homology extended from 138 bp upstream of the ORF to 254 bp downstream of the stop codon. This region included the putative promoter sequences and also the rho-independent terminator described (Brunder *et al.*, 1997; Djafari *et al.*, 1997; see Fig. 3). Beyond this nearly identical DNA section, the two sequences are completely different. Whereas the *espP* gene is flanked by incomplete insertion sequences (ISs) homologous to IS*1203* and iso-IS*1* (Brunder *et al.*, 1997), the *pssA* gene is flanked by two identical elements with inverted orientation and composed of remnants of IS*600* and IS*911* (Djafari *et al.*, 1997). A complete IS*1203*-like sequence and a further IS*600* remnant occur further downstream on *pssA* (Fig. 3). We conclude therefore, that in the large plasmid of the STEC O26:H^- strain and the STEC O157:H7 strain EDL933 the almost identical genes *pssA* and *espP*, together with their regulatory sequences, are inserted at completely different sites.

To investigate the *espP* boundaries in a greater number of strains, several PCR primers were constructed (see Fig. 3). Using primer pairs web-5/web-6 and web-7/web-8, the expected PCR products were obtained for 10/10 O157:H7/H^- strains tested, whereas none of the non-O157 strains gave a PCR product. It is concluded, therefore, that the large plasmids of NSF STEC O157:H7/H^- are uniform in the DNA regions adjacent to *espP*. Another primer pair (web-9/web-6) was constructed based on the sequence upstream of the *pssA* gene (see Fig. 3). Using this primer pair, PCR products of

the expected length were obtained with 8/11 of the O26 strains tested and also with 2/2 O145:H$^-$ and 1/1 O111:H$^-$ strains. The other *espP*-positive serotypes were all negative by PCR using the three primer pairs, indicating that the *espP* gene of these strains is located within a DNA region different from that described for STEC O157 and STEC O26. Further studies are necessary in order to elucidate the nature of the *espP* boundaries in these serotypes.

Linkage of *espP* with other pO157 genes

In a further step we investigated the relationship between the presence of the *espP* gene in STEC and the presence of *katP* and the EHEC-haemolysin gene EHEC-*hlyA*, which are known to be also located on plasmid pO157 of STEC O157:H7 strain EDL933 (Schmidt *et al.*, 1995b; Brunder *et al.*, 1996). The genes were also detected by PCR and colony-blot hybridization. Table 3 shows the frequency of all possible combinations of *espP* (e), *katP* (k) and EHEC-*hlyA* (h) in STEC strains of specific serotypes. Whereas all NSF STEC O157:H7/H$^-$ harboured all three genes, the SF O157:H$^-$ strains tested were all negative for *espP* and *katP* but positive for EHEC-*hlyA* (by PCR). This is a notable difference between the two groups in addition to the difference in the ability to ferment sorbitol. In STEC strains of serogroup O26, the combinations e$^+$ k$^+$ h$^+$ and e$^-$ k$^-$ h$^+$ occurred, whereas in strains of serotype O103:H2, the predominant combinations were e$^-$ k$^+$ h$^+$ and e$^-$ k$^-$ h$^+$ (Table 3). In STEC of other serotypes, the three genes were found in almost all possible combinations.

To prove that all markers are located on plasmids, we performed Southern-blot hybridization of plasmid DNA using probes specific for *espP*, *katP* and EHEC-*hlyA* . A representative strain for each marker combination and serogroup was investigated. As shown in Table 4, the genes were detectable on plasmids in all strains with the exception of strain 3651/96 (O8:H19), 7137/95 (Orough:H11) and 6561/95 (O25:H14). In strain 3651/96, EHEC-*hlyA* was detected on a 24 kbp plasmid fragment whereas the *espP* gene, which is also present, was not detectable on plasmid DNA. It seems likely that in the three cases, the genes (*espP* and/or EHEC-*hlyA*) are located on the bacterial chromosome. All STEC strains harboured large plasmids, including the nine strains found to be negative for the markers tested. Taken together, the findings clearly demonstrate that STEC large plasmids are a heterogeneous group of genetic elements.

DISCUSSION

Studies on putative virulence factors of STEC encoded by their plasmids have been hitherto confined to a few prototype strains and plasmids. The prototypes have been of serotype O157:H7 because of the prominence of this serotype in outbreaks of STEC infections (Karch *et al.*, 1987, 1998; Tzipori *et al.*, 1987; Toth *et al.*, 1990; Fratamico *et al.*, 1993). In view of the increasing prevalence of non-O157 serotypes in human STEC diseases (Johnson *et al.*, 1996; Caprioli *et al.*, 1997), however, there is now a need for comprehensive data on the virulence properties of these STEC strains. Moreover, virulence genes, including those encoded by plasmids, are used as targets in diagnostic methods and strain subtyping (Nataro & Kaper, 1998). The present study showed that plasmids of STEC, other than the classical serotype O157:H7, possess high variability.

Initially, we analysed the prevalence of *espP* in a large variety of STEC serotypes and other diarrhoeagenic *E. coli* and observed a high specificity for STEC among all diarrhoeagenic *E. coli* pathogroups. However, the gene was not detected in about 40% of STEC strains tested. The presence or absence of *espP* was, in part, correlated with the serotype, but there were also serotypes (e.g. O26:H11/H$^-$) in which only some strains possessed the gene. Some differences in STEC plasmids are to be expected when considering this single marker. Therefore, we included other plasmid markers (*katP* and EHEC-*hlyA*) and analysed the correlation between the presence of these two additional markers and that of *espP*. It became apparent that the gene composition is highly variable in STEC plasmids. In contrast to the uniformity present in the group of NSF STEC O157:H7/H$^-$, other serotypes showed nearly all possible combinations of these genes. The fact that plasmids were also detected in the nine STEC strains lacking the three genes indicates that the absence of these genes is not necessarily an indication of a plasmid-less STEC strain.

The SF STEC O157:H$^-$ are of particular interest. Strains of this group were isolated in Germany and the Czech Republic from patients suffering from HUS and diarrhoea (Karch *et al.*, 1990; Gunzer *et al.*, 1992; Bielaszewska *et al.*, 1998). In contrast to STEC O157:H7, they ferment sorbitol and exhibit β-glucuronidase activity (Gunzer *et al.*, 1992) and, as shown in the present study, the plasmids of this clone are quite distinct from those of NSF STEC O157:H7/H$^-$. Whereas the haemolysin gene EHEC-*hlyA* is present in both, the plasmids of SF O157:H$^-$ strains lack the *espP* and *katP* genes. Since NSF *E. coli* O157:H7 and SF O157:H$^-$ are both thought to be derived from a common EPEC-like O55:H7 ancestor (Feng *et al.*, 1998), one can hypothesize that they have acquired different plasmids after their evolutionary split into two lineages. However, it is also possible that one common plasmid, harboured by the ancestor and carrying the EHEC-*hlyA* , *espP* and *katP* genes, has lost the *espP* and *katP* markers in the SF O157:H$^-$ clone. The third possibility is that the ancestral plasmid has acquired the two markers only in the NSF O157:H7 lineage, possibly enhancing the survival and spread of this clone.

One important question is how the acquisition or loss of virulence genes like *espP* can occur during the evolution of new pathogens. Although there is little known about the source of the *espP* gene, there are several homologous genes present in *E. coli* pathogroups and other enteric pathogens, for example the *espC* gene of EPEC (Stein *et al.*, 1996), the *tsh* gene of avian pathogenic *E. coli*

(Provence & Curtiss, 1994), the *pet* gene of EAEC (Eslava *et al.*, 1998) and the *sepA* gene of *Shigella flexneri* (Benjelloun-Touimi *et al.*, 1995). Interestingly, some of these (*sepA* and *pet*) are also located on large plasmids, whereas *espC* is on the chromosome. The localization of genes on plasmids renders them capable of horizontal transfer, should these plasmids be self-conjugable or mobilizable. The plasmid of STEC O157:H7 is reported to be transfer-deficient (Hales *et al.*, 1992; Makino *et al.*, 1998) but it is probable that it was transferable during some earlier phase of evolution. However, the EspP homologues of other *E. coli* pathogroups share only about 60% sequence homology (Brunder *et al.*, 1997; Eslava *et al.*, 1998), indicating that there is no exchange of these genes between *E. coli* pathogroups at present. Differences in habitats of the groups or incompatibility of plasmids occurring in the various groups may be the cause of this transfer barrier.

In addition to the observation that different combinations of the STEC plasmid genes exist, an important finding of our studies was that there were variations in the DNA flanking the *espP* gene in plasmids of a number of serotypes. By comparing the nucleotide sequences we showed that the nearly identical genes *espP* and *pssA* occur at very different sites on plasmids of STEC O157:H7 and O26:H$^-$, respectively. Furthermore, PCR-based investigations suggested that *espP* genes are located at even more diverse sites in other STEC serotypes. The question arises as to how such diversity is created. One possible clue could be the presence of ISs, or remnants of them, near the *espP* and *pssA* genes. The fact that the genes in the two serotypes are flanked by ISs of different types might indicate that the genes have arisen from different sources. ISs could also be responsible for the transfer of *espP* genes to diverse DNA regions in a transposon-like manner. Partial deletion of the IS could then lead to the stabilization of the gene within the plasmid.

Furthermore, DNA regions with homology to IS in the vicinity of the gene might act as a substrate for homologous recombination events. The abundance of such IS in the bacterial genome makes it likely that an exchange of genes between different plasmids, or plasmids and the chromosome, could occur by such a mechanism. It is of interest that in three strains we could not detect the *espP* and/or the EHEC-*hlyA* gene on plasmid DNA using Southern-blot hybridization although these strains were positive by PCR and colony-blot hybridization. Although these unusual strains also possess plasmids it seems likely that the genes in these strains are located on the chromosome.

ISs also contributed to the formation of the *espP*-variant strains in the otherwise uniform STEC O157:H7 group. Insertion of IS*1203* (Paton & Paton, 1994) into the *espP* gene of strain 5899/96 inactivated the gene. Similarly, insertion of an IS followed by homologous recombination with a similar sequence in the vicinity could be the cause of the partial deletion of *espP* seen in strain 3010/96 (Craig, 1996). The detection of the two *espP* mutant strains described in our study provides an example for a clonal turnover which may have taken place in the recent evolution of *E. coli* O157:H7.

The *espP* gene itself also shows variability between strains of different serotypes. As we have shown using *Alu*I and *Sst*I restriction of the PCR products, there are at least seven restriction types among *espP* genes of STEC from the 16 *espP*-possessing serotypes investigated. These sequence variations, some of which are restricted to one serogroup only, may trace the evolution of the gene within distinct phylogenetic STEC groups. Interestingly, strains of serogroups O26, O111 and O145 are similar, not only in regard to their *espP* RFLP but also in regard to their *espP* boundaries. On the other hand, there are other strains within the O26 serogroup showing identical *espP*-RFLP patterns but different sequences adjacent to the *espP* gene. However, none of the serotypes tested possessed *espP* genes of more than one restriction type. Strains with *espP* restriction type A, B or D also possessed the chromosomal *eae* gene, whereas those of restriction types C, E, F and G were *eae* negative (data not shown).

Based on possession of the *eae* gene and according to the results of DNA hybridization studies, Boerlin *et al.* (1998) classified STEC plasmids into two categories, type I and type II. They hypothesized that type I plasmids were present in STEC before divergence of the serotypes comprising the *eae*-positive phylogenetic lineage. Our studies however, show that these type I plasmids are not as uniform as supposed by Boerlin *et al.*, who investigated the EHEC-*hlyA* gene and the sequences upstream of the EHEC-haemolysin operon. Based on our investigations on *espP* and *katP* genes, it is evident that high diversity also exists in the type I plasmids of *eae*-positive STEC strains. This not only applies to the combinations of the investigated plasmid markers, where marked variations were found in *eae*-positive strains of serogroups O157, O26 and O103, but also to the DNA regions adjacent to the *espP* gene.

In summary, our study showed that the large plasmids of STEC are not as uniform as previously supposed. Further efforts are necessary to elucidate the mechanisms of plasmid variability and to determine how variations in the virulence-gene composition influence the pathogenicity of individual STEC strains.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB 479). We thank Stefanie Amersbach for excellent technical assistance and Jürgen Scheef for critical reading of the manuscript.

REFERENCES

Benjelloun-Touimi, Z., Sansonetti, P. J. & Parsot, C. (1995). SepA, the major extracellular protein of *Shigella flexneri*: autonomous secretion and involvement in tissue invasion. *Mol Microbiol* **17**, 123–135.

Bielaszewska, M., Schmidt, H., Karmali, M. A., Khakhria, R., Janda, J., Bláhová, K. & Karch, H. (1998). Isolation and charac-

terization of sorbitol-fermenting Shiga toxin (Verocytotoxin)-producing *Escherichia coli* O157:H$^-$ strains in the Czech republic. *J Clin Microbiol* **36**, 2135–2137.

Bockemühl, J., Aleksic, S. & Karch, H. (1992). Serological and biochemical properties of Shiga-like toxin (verocytotoxin)-producing strains of *Escherichia coli*, other than O-group 157, from patients in Germany. *Zentbl Bakteriol* **276**, 189–195.

Boerlin, P., Chen, S., Colbourne, J. K., Johnson, R., de Grandis, S. & Gyles, C. (1998). Evolution of enterohemorrhagic *Escherichia coli* hemolysin plasmids and the locus for enterocyte effacement in Shiga toxin-producing *E. coli*. *Infect Immun* **66**, 2553–2561.

Bopp, C. A., Greene, K. D., Downes, F. P., Sowers, E. G., Wells, J. G. & Wachsmuth, I. K. (1987). Unusual verotoxin-producing *Escherichia coli* associated with hemorrhagic colitis. *J Clin Microbiol* **25**, 1486–1489.

Brunder, W., Schmidt, H. & Karch, H. (1996). KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* **142**, 3305–3315.

Brunder, W., Schmidt, H. & Karch, H. (1997). EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Mol Microbiol* **24**, 767–778.

Caprioli, A., Tozzi, A. E., Rizzoni, G. & Karch, H. (1997). Non-O157 Shiga toxin-producing *Escherichia coli* infections in Europe. *Emerg Infect Dis* **3**, 578–579.

Craig, N. L. (1996). Transposition. In *Escherichia coli and Salmonella typhimurium*, pp. 2339–2362. Edited by F. C. Neidhardt and others. Washington, DC: American Society for Microbiology.

Djafari, S., Ebel, F., Deibel, C., Krämer, S., Hudel, M. & Chakraborty, T. (1997). Characterization of an exported protease from Shiga toxin-producing *Escherichia coli*. *Mol Microbiol* **25**, 771–784.

Eslava, C., Navarro-García, F., Czeczulin, J. R., Henderson, I. R., Cravioto, A. & Nataro, J. P. (1998). Pet, an autotransporter enterotoxin from enteroaggregative *Escherichia coli*. *Infect Immun* **66**, 3155–3163.

Feng, P., Lampel, K. A., Karch, H. & Whittam, T. S. (1998). Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. *J Infect Dis* **177**, 1750–1753.

Fratamico, P. M., Bhaduri, S. & Buchanan, R. L. (1993). Studies on *Escherichia coli* serotype O157:H7 strains containing a 60-MDa plasmid and on 60-MDa plasmid-cured derivatives. *J Med Microbiol* **39**, 371–381.

Girón, J. A., Donnenberg, M. S., Martin, W. C., Jarvis, K. G. & Kaper, J. B. (1993). Distribution of the bundle-forming pilus structural gene (*bfpA*) among enteropathogenic *Escherichia coli*. *J Infect Dis* **168**, 1037–1041.

Gunzer, F., Böhm, H., Rüssmann, H., Bitzan, M., Aleksic, S. & Karch, H. (1992). Molecular detection of sorbitol-fermenting *Escherichia coli* O157 in patients with hemolytic-uremic syndrome. *J Clin Microbiol* **30**, 1807–1810.

Hales, B. A., Hart, C. A., Batt, R. M. & Saunders, J. R. (1992). The large plasmids found in enterohemorrhagic and enteropathogenic *Escherichia coli* constitute a related series of transfer-defective Inc F-IIA replicons. *Plasmid* **28**, 183–193.

Johnson, R. P., Clarke, R. C., Wilson, J. B. & 10 other authors (1996). Growing concerns and recent outbreaks involving non-O157:H7 serotypes of verotoxigenic *Escherichia coli*. *J Food Prot* **59**, 1112–1122.

Karch, H., Heesemann, J., Laufs, R., O'Brien, A. D., Tacket, C. O. & Levine, M. M. (1987). A plasmid of enterohemorrhagic *Escherichia coli* O157:H7 is required for expression of a new fimbrial antigen and for adhesion to epithelial cells. *Infect Immun* **55**, 455–461.

Karch, H., Schmidt, H. & Brunder, W. (1998). Plasmid-encoded determinants of *Escherichia coli* O157:H7. In *Escherichia coli* O157:H7 *and Other Shiga Toxin-Producing E. coli Strains*, pp. 183–195. Edited by J. B. Kaper & A. D. O'Brien. Washington, DC: American Society for Microbiology.

Karch, H., Wiß, R., Gloning, H., Emmrich, P., Aleksic, S. & Bockemühl, J. (1990). Hämolytisch-urämisches Syndrom bei Kleinkindern durch Verotoxin-produzierende *Escherichia coli*. *Dtsch Med Wochenschr* **115**, 489–495.

Levine, M. M., Xu, J. G., Kaper, J. B., Lior, H., Prado, V., Tall, B., Nataro, J., Karch, H. & Wachsmuth, K. (1987). A DNA probe to identify enterohemorrhagic *Escherichia coli* of O157:H7 and other serotypes that cause hemorrhagic colitis and hemolytic-uremic syndrome. *J Infect Dis* **156**, 175–182.

Makino, K., Ishii, K., Yasunaga, T. & 14 other authors (1998). Complete nucleotide sequences of 93-kb and 3·3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. *DNA Res* **5**, 1–9.

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

Nataro, J. P., Baldini, M. M., Kaper, J. B., Black, R. E., Bravo, N. & Levine, M. M. (1985). Detection of an adherence factor of enteropathogenic *Escherichia coli* with a DNA probe. *J Infect Dis* **152**, 560–565.

O'Brien, A. D., Lively, T. A., Chen, M. E., Rothman, S. W. & Formal, S. B. (1983). *Escherichia coli* O157:H7 strains associated with haemorrhagic colitis in the United States produce a *Shigella dysenteriae* 1 (SHIGA) like cytotoxin. *Lancet* **1**, 702.

Ostroff, S. M., Tarr, P. I., Neill, M. A., Lewis, J. H., Hargrett-Bean, N. & Kobayashi, J. M. (1989). Toxin genotypes and plasmid profiles as determinants of systemic sequelae in *Escherichia coli* O157:H7 infections. *J Infect Dis* **160**, 994–998.

Paros, M., Tarr, P. I., Kim, H., Besser, T. E. & Hancock, D. D. (1993). A comparison of human and bovine *Escherichia coli* O157:H7 isolates by toxin genotype, plasmid profile, and bacteriophage lambda-restriction fragment length polymorphism profile. *J Infect Dis* **168**, 1300–1303.

Paton, A. W. & Paton, J. C. (1994). Characterization of IS*1203*, an insertion sequence in *Escherichia coli* O111:H$^-$. *Gene* **150**, 67–70.

Pridmore, R. D. (1987). New and versatile cloning vectors with kanamycin-resistance marker. *Gene* **56**, 309–312.

Provence, D. L. & Curtiss, R., III. (1994). Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. *Infect Immun* **62**, 1369–1380.

Riley, L. W., Remis, R. S., Helgerson, S. D. & 9 other authors. (1983). Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med* **308**, 681–685.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schmidt, H., Knop, C., Franke, S., Aleksic, S., Heesemann, J. & Karch, H. (1995a). Development of PCR for screening of enteroaggregative *Escherichia coli*. *J Clin Microbiol* **33**, 701–705.

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

Schmidt, H., Kernbach, C. & Karch, H. (1996). Analysis of the EHEC *hly* operon and its location in the physical map of the large

plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* **142**, 907–914.

Schmidt, H., Henkel, B. & Karch, H. (1997). A gene cluster closely related to type II secretion pathway operons of Gram-negative bacteria is located on the large plasmid of enterohemorrhagic *Escherichia coli* O157 strains. *FEMS Microbiol Lett* **148**, 265–272.

Stein, M., Kenny, B., Stein, M. A. & Finlay, B. B. (1996). Characterization of EspC, a 110-kilodalton protein secreted by enteropathogenic *Escherichia coli* which is homologous to members of the immunoglobulin A protease-like family of secreted proteins. *J Bacteriol* **178**, 6546–6554.

Toth, I., Cohen, M. L., Rumschlag, H. S., Riley, L. W., White, E. H., Carr, J. H., Bond, W. W. & Wachsmuth, I. K. (1990). Influence of the 60-megadalton plasmid on adherence of *Escherichia coli* O157:H7 and genetic derivatives. *Infect Immun* **58**, 1223–1231.

Tzipori, S., Karch, H., Wachsmuth, K. I., Robins-Browne, R. M., O'Brien, A. D., Lior, H., Cohen, M. L., Smithers, J. & Levine, M. M. (1987). Role of a 60-megadalton plasmid and Shiga-like toxins in the pathogenesis of infection caused by enterohemorrhagic *Escherichia coli* O157:H7 in gnotobiotic piglets. *Infect Immun* **55**, 3117–3125.

Received 4 September 1998; revised 10 December 1998; accepted 20 January 1999.