FEBS Letters 472 (2000) 22–26 FEBS 23550

The sequences of enterohemorrhagic *Escherichia coli* and *Yersinia pestis* that are homologous to the enteroaggregative *E. coli* heat-stable enterotoxin gene: cross-species transfer in evolution

Tatsuo Yamamoto*, Ikue Taneike

Department of Bacteriology, School of Medicine, Niigata University, 757 Ichibanchou, Asahimachidori, Niigata, Japan

Received 7 February 2000; received in revised form 21 March 2000

Edited by Takashi Gojobori

Abstract The enteroaggregative *Escherichia coli* heat-stable enterotoxin 1 (EAST1) gene is widely distributed among diarrheagenic *E. coli*. In this study, we examined the sequences of enterohemorrhagic *E. coli* (EHEC) strains by PCR and sequencing. All the EHEC strains possessed the EAST1 gene homologues but with two types of mutations. One of the mutation types was strongly associated with the large outbreak episodes in 1996 in Japan. Sequence comparison showed that the EHEC sequences are a branch of the EAST1 gene sequence family that showed the cross-species transfer in evolution among *E. coli* and *Yersinia pestis*.

Key words: Insertion sequence; Mutation; Cross-species transfer; Enteroaggregative *Escherichia coli* heat-stable enterotoxin 1 gene; Enterohemorrhagic *Escherichia coli*; *Yersinia pestis*

1. Introduction

Enteroaggregative *Escherichia coli* (EAggEC) is associated with pediatric and chronic diarrhea [1,2], traveler's diarrhea [3,4], and chronic diarrhea of AIDS patients [5–8]. Savarino et al. have reported a heat-stable enterotoxin (called EAggEC heat-stable enterotoxin 1, EAST1) as a virulence factor of some EAggEC strains [9]. They have also determined the gene sequence encoding a peptide of 38 amino acids in EAggEC strain 17-2 [10].

The EAST1 gene is an enterotoxin gene distributed among a broad range of diarrhea-associated *E. coli*. It is distributed among EAggEC [9,10], enterotoxigenic *E. coli* (ETEC) [11,12], enteropathogenic *E. coli* (EPEC) [13,14], and diffusely adhering *E. coli* (DAEC) [13–15]. The EAST1 gene family includes one major type of sequence that is distributed widely among the diarrhea-associated *E. coli* strains, i.e. the EAST1 gene sequence of EAggEC strain O42 [14]. In addition, there exists (at least) a variant type of sequence with a non-synonymous nucleotide substitution, i.e. the EAST1 gene sequence of EAggEC strain 17-2 [14]. Although major virulence factors have not been characterized in EAggEC, strain O42 but not strain 17-2 was shown to be pathogenic in volunteer experiments [16].

In Japan, during 1996, large outbreaks and sporadic cases of enterohemorrhagic *E. coli* (EHEC) infections occurred with 17 877 people infected, 1795 patients hospitalized, and 12 deaths (mostly children) [17]. EHEC includes not only serotype O157:H7 (the most prevalent serotype), but other serotypes; O26, O111, O145 have also been implicated in human infections [18]. EHEC was positive for the EAST1 gene sequence in the hybridization assay [13] but negative in the PCR assay [14], giving rise to a discrepancy. In this study, we investigated the EAST1 gene homologues of EHEC strains from outbreaks as well as sporadic cases in Japan, and demonstrated the cross-species transfer of the gene sequence in evolution.

2. Materials and methods

2.1. EHEC strains

Outbreak-derived EHEC strains (19 strains) belonging to serotype O157:H7 included two or three strains each from the following nine large episodes: Urawa City episodes in 1990 (number of patients, 319; sources of infections included well water; EHEC toxin type, Shiga toxin (Stx) 1/Stx2; EHEC pulsed-field gel electrophoresis (PFGE) pattern type, IIa and SK) [17,19]; Obihiro City episode in 1996 (number of patients, 157; source of infections, potato salad; EHEC toxin type, Stx2; EHEC PFGE type, IIIh; EHEC phage type, PT4) [17,20]; Morioka City episode in 1996 (number of patients, 124; sources of infections, pumpkin salad and seafood salad; EHEC toxin type, Stx1/Stx2; EHEC PFGE type, IIj; EHEC phage type, PT14) [17,20]; Gifu City episode in 1996 (number of patients, 379; source of infections, bonito salad; EHEC toxin type, Stx1/Stx2; EHEC PFGE type, Ib; EHEC phage type, PT40) [17,20,21]; Aichi City episode in 1996 (number of patients, 21; source of infections, unknown; EHEC toxin type, Stx1/Stx2; EHEC PFGE type, Ib; EHEC phage type, PT21) [17,20,21]; Sakai City episode in 1996 (number of patients, 9523; source of infections, white radish sprouts; EHEC toxin type, Stx1/Stx2; EHEC PFGE type, IIa; EHEC phage type, PT32; representative strain in this study, S1) [17,20,21]; Okayama City episode in 1996 (number of patients, 468; source of infections, unknown; EHEC toxin type, Stx1/Stx2; EHEC PFGE type, Ic; EHEC phage type, PT40) [17,20,21]; Hiroshima City episode in 1996 (number of patients, 185; source of infections, unknown; EHEC toxin type, Stx1/Stx2; EHEC PFGE type, Ia; EHEC phage type, PT21) [17,20,21]; and Fukuoka City episode in 1996 (number of patients, 48; source of infections, unknown; EHEC toxin type, Stx1/Stx2; EHEC PFGE type, Ia; EHEC phage type, PT21) [17,20,21].

Forty-eight EHEC strains from sporadic cases included 20 strains

*Corresponding author. Fax: (81)-25-227-0762.
E-mail: tatsuoy@med.niigata-u.ac.jp

Abbreviations: EAST1, enteroaggregative *Escherichia coli* heat-stable enterotoxin 1; EAggEC, enteroaggregative *Escherichia coli*; EHEC, enterohemorrhagic *Escherichia coli*; Stx, Shiga toxin; PCR, polymerase chain reaction; PFGE, pulsed-field gel electrophoresis; IS, insertion sequence

PII: S0014-5793(00)01414-9

belonging to serotype O157:H7 (two strains of toxin type Stx1, nine strains of toxin type Stx2 (representative strain in this study, U10), and nine strains of toxin type Stx1/Stx2), one strain belonging to serotype O157:H- (toxin type, Stx1/Stx2), 15 strains belonging to serotype O26:H11 (toxin type, Stx1), 10 strains belonging to serotype O111 (two strains of toxin type Stx1 and eight strains of toxin type Stx1/Stx2), and two strains belonging to serotype O145 (toxin type, Stx1/Stx2). They were isolated during 1996 and 1998.

2.2. PCR primers

PCR primers employed for the detection of the EAST1 gene sequence were: EAST11a (5′-CCATCAACACAGTATATCCGA, corresponding to the 2nd to 8th codons) and EAST11b (5′-GGTCGCGAGTGACGGCTTTGT, corresponding to the 38th (C-terminus) to 32nd codons), generating a 111-base pair (bp) product [10,11]; EAST12a (5′-ACGATATCCTCATCGCCTGTG) and EAST12b (5′-CTGCTGGCCTGCCTCTTCCGT), generating a 203-bp product [11,14]; and EAST13a (5′-AGAACTGCTGGGTATGTGGCT) and EAST13b (5′-GTTGGATAAGCGAAGAACGTG), generating a 393-bp product [11,14]. The primer sequences EAST12a and EAST13a were located 24 and 110 nucleotides upstream, respectively, from the initiation ATG sequence of the EAST1 gene, and the primer sequences EAST12b and EAST13b were located 20 and 124 nucleotides downstream, respectively, from the stop TGA sequence of the EAST1 gene [11,14]. A primer set consisting of SHEAST1a (5′-CTATCAACACGGTGTATCCGG) and SHEAST1b (5′-GGTCGCGGGTGACGGCTTTGT), generating a 111-bp product, was designed on the basis of the determined nucleotide sequence of the outbreak-derived EHEC strains in this study.

2.3. Detection of the EAST1 gene sequence by PCR

For the detection of the EAST1 gene sequence, EHEC strains grown on nutrient agar (Eiken Chemical, Tokyo, Japan) overnight at 37°C were suspended in sterile water at a concentration of ca. 5×10^7/ml and boiled for 5 min. A 2.5-μl portion of the DNA solution was subjected to PCR. *E. coli* DNA prepared essentially as previously described [22] was used for PCR, followed by cloning and sequencing.

Cycling conditions were denaturation for 30 s at 95°C, annealing for 120 s at 55°C, and polymerization for 120 s at 72°C (30 cycles), except for PCR with the primer set consisting of SHEAST1a and SHEAST1b, in which annealing was conducted at 60°C. Amplified products were analyzed by electrophoresis in 2% agarose gel or 5% polyacrylamide gel. ϕX174 RF DNA/*Hae*III fragments (Life Technologies, Gaithersburg, MD, USA) were used as molecular size standards.

2.4. Sequencing

For sequencing, the PCR products were generated from the EHEC strains with the primer set of EAST13a and EAST12b and then cloned into a cloning vector pT7Blue(R) (Novagen Inc., Madison, WI, USA) in accordance with the manufacturer's instructions, and the nucleotide sequence was determined using a *Taq* dyedeoxy terminator cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) and a 373 A DNA sequencer (Applied Biosystems) according to the manual. Sequence analysis of the PCR products was performed using three to four clones of the recombinant plasmids and sequence determination was done on both DNA strands in each case.

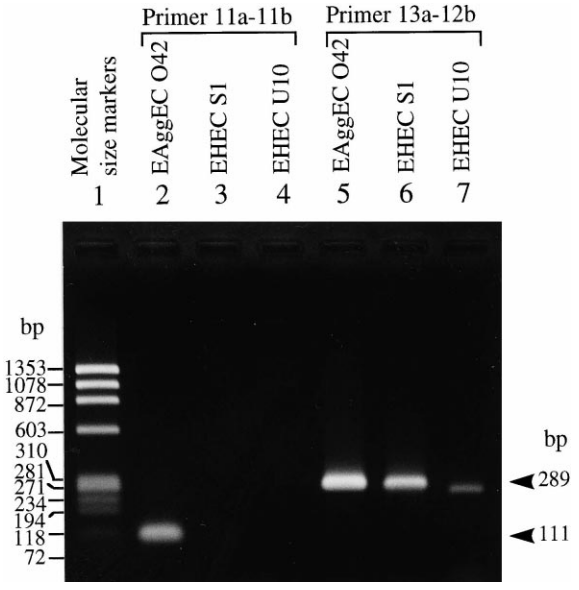

Fig. 1. PCR analysis of the EHEC EAST1 gene sequences. Bacterial DNA amplified with the primer set consisting of EAST11a and EAST11b (see text) is seen in lanes 2–4 and that amplified with the primer set of EAST13a and EAST12b (see text) is depicted in lanes 5–7. PCR products were analyzed by 2.0% agarose gel electrophoresis, and stained with ethidium bromide. Lane l, ϕX174 RF DNA/*Hae*III fragments; lanes 2 and 5, EAggEC strain O42 (carrying the EAST1 gene) used as a control; lanes 3 and 6, outbreak-derived EHEC strain S1; lanes 4 and 7, sporadic case-derived EHEC strain U10.

3. Results

3.1. PCR and EASTl gene sequences of EHEC

Two EHEC strains, S1 (belonging to serotype O157:H7 and isolated from a large outbreak at Sakai City) and U10 (belonging to serotype O157:H7 and isolated from a sporadic case), were examined for PCR and sequencing. They were negative in the PCR when assayed with the primer set of EAST11a and EAST11b (Fig. 1, lanes 3 and 4) or with the primer set of EAST13a and EAST13b (data not shown). However, they produced positive results when assayed with the primer set of EAST12a and EAST12b (data not shown) or with the primer set of EAST13a and EAST12b (Fig. 1, lanes 6 and 7). In this PCR assay, strain S1 gave a band at the position of the expected DNA size (lane 6), while strain U10 gave a faint band of a smaller DNA size (lane 7).

The sequences of the PCR products were determined (Fig.

Table 1
Type of the DNA sequences homologous to the EAST1 gene (SHEAST) of the EHEC strains derived from outbreaks and sporadic cases in Japan

Serotype	Isolation and no. of strains examined	SHEAST type[a]	Frequency (positive/total)
O157:H7	outbreaks ($n=19$[b])	type 1	19/19 (100%)
	sporadic cases ($n=20$)	type 1	11/20 (55%)
	sporadic cases ($n=20$)	type 2	9/20 (45%)
O157:H-	sporadic case ($n=1$)	type 1	1/1
O26:H11	sporadic cases ($n=15$)	type 2	15/15 (100%)
O111	sporadic cases ($n=10$)	type 2	10/10 (100%)
O145	sporadic cases ($n=2$)	type 2	2/2

[a]Bacterial DNA was amplified with the primer set of EAST13a-EAST12b and the PCR products were sequenced (see text). Type 1 and type 2 indicate the sequence identical to the SHEAST of EHEC strains S1 and U10 (shown in Fig. 2), respectively.
[b]Three strains (including strain S1) from outbreaks in Sakai City and two strains each from eight major outbreaks in other cities were examined.

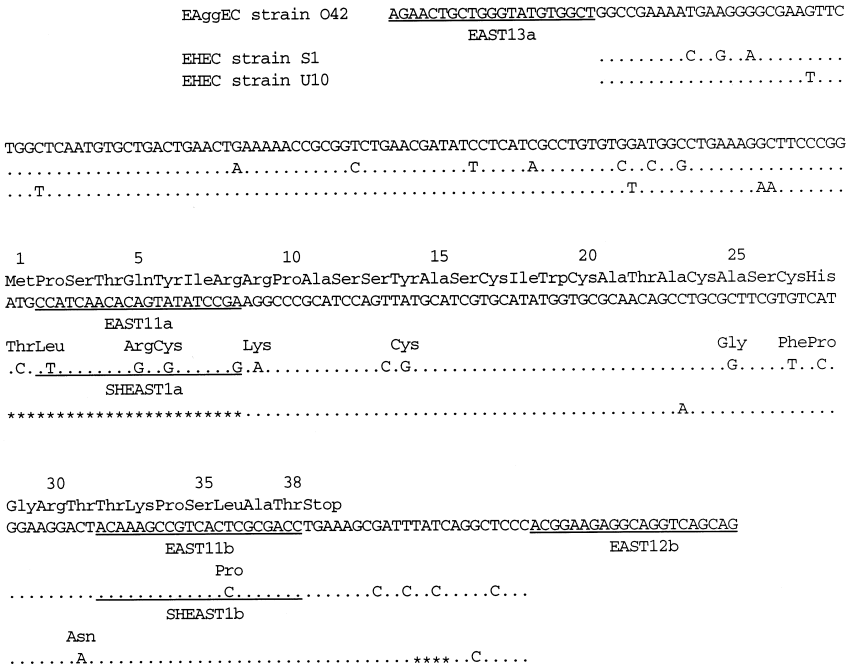

Fig. 2. Nucleotide sequences of the PCR products generated from EHEC strains S1 and U10 with the primer set of EAST13a and EAST12b. The nucleotide sequence of the EAST1 gene of EAggEC strain O42 is from [14] and shown in the first (top) line of the nucleotide sequences in the figure; the deduced amino acid sequence of EAST1 is shown above the nucleotide sequence. The nucleotide sequences determined for EHEC strains S1 and U10 are shown in the second and third (bottom) nucleotide lines, respectively; nucleotides that are identical to those for EAggEC strain O42 are shown as dots and the amino acids predicted from each codon change are shown above the codon change. Asterisks indicate the positions of nucleotide deletion. Underlines indicate the positions of primers EAST13a-EAST12b, EAST11a-EAST11b, and SHEAST1a-SHEAST1b. The primer set of EAST11a-EAST11b detects the EAST1 gene [10,11]. The primer set of SHEAST1a-SHEAST1b detects the unique DNA sequence homologous to the EAST1 gene (SHEAST) of the O157:H7 strains derived from the outbreaks and some sporadic cases (see text).

2). The EHEC strains carried a DNA sequence homologous to the EAST1 gene (SHEAST). The SHEAST of strain S1 showed 89.7% homology to the EAST1 gene sequence. However, of the 12 single base changes found, 10 (83.3%) including that in the initiation ATG sequence were non-synonymous substitutions that cause amino acid change (Fig. 2). Four base changes in the region corresponding to the primer EAST11a resulted in a negative reaction in the PCR assay with the primer set of EAST11a and EAST11b. The SHEAST of strain S1 was designated as type 1 SHEAST.

The SHEAST of strain U10 lacked the first eight codons, including the initiation ATG sequence (Fig. 2); this again resulted in a negative reaction in the PCR assay with the primer set of EAST11a and EAST11b. Of the two single base changes found in the SHEAST region, one (50%) was a non-synonymous substitution. The SHEAST of strain U10 was designated as type 2 SHEAST. The nucleotide sequences upstream of the SHEAST were well conserved in the EHEC strains, especially in strain U10 (Fig. 2).

3.2. Distribution of two types of SHEAST in EHEC strains

The O157:H7 strains derived from different outbreaks in Japan have been distinguished from each other in terms of their PFGE or phage types [20,21]. Two strains each from nine such major outbreaks that occurred in different places in Japan were also examined for SHEAST by PCR (with the primer set of EAST13a and EAST12b) and sequencing. All the determined sequences showed complete agreement with the sequence (type 1 SHEAST) of strain S1 (Table 1).

Next, 47 more EHEC strains isolated from the sporadic cases were examined for SHEAST as above. Eleven of the 20 O157:H7 strains (55%) and one O157:H- strain possessed type 1 SHEAST (Table 1). In contrast, nine of the 20 O157:H7 strains (including strain U10) (45%), all the O26 strains (15 strains), all the O111 strains (10 strains), and two of the O145 strains possessed type 2 SHEAST (Table 1).

3.3. Sequence comparison to the insertion sequence of Yersinia pestis

The EAST1 gene and SHEAST showed a high sequence homology to IS*285* of *Y. pestis* (Fig. 3). The homology in the EAST1 gene-corresponding region (117 bases) was 78.6% between IS*285* and the EAST1 gene, 76.9% between IS*285* and type 1 SHEAST, and 77.4% between IS*285* and type 2 SHEAST (comparison in the sequence except for the deletion).

4. Discussion

The EAST1 gene has been shown to be distributed among several categories of diarrhea-associated *E. coli*; it was found in 41% [13] to 53.3% [14] of strains of EAggEC, 20% [14] to 22% [13] of strains of EPEC, 6.7% of strains of type 1 DAEC [14], 100% of strains of type 2 DAEC [14,15], and 0 to 100% of strains of ETEC, depending upon their adherence factor type [11,12]. In the case of ETEC strain H10407, multiple copies of the EAST1 gene exist on the chromosome as well as on the virulence plasmid, in complete agreement with EAggEC strain O42 in the nucleotide sequence [11,14].

This study demonstrated that all 67 EHEC strains exam-

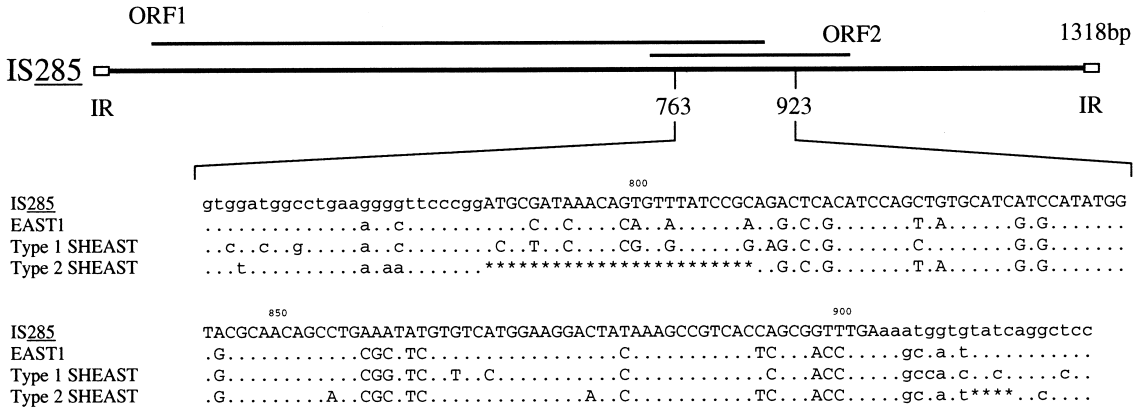

Fig. 3. *Y. pestis* IS*285* sequence that is homologous to the EAST1 gene. The IS*285* sequence including two putative open reading frames (ORF1 and ORF2) shown in the upper part of the figure was taken from [23]. IR, inverted repeats. In the lower part of the figure, the nucleotide sequence of IS*285* is shown in the first (top) line of the nucleotide sequences; the sequence corresponding to the EAST1 gene is shown in capital letters. The nucleotide sequences of the EAST1 gene of EAggEC strain O42, type 1 SHEAST, and type 2 SHEAST are shown in the second, third, and fourth (bottom) nucleotide lines, respectively (they are aligned to the IS*285* sequence in such a way as to generate maximum homology); nucleotides that are identical to those for IS*285* are shown as dots, and asterisks indicate the positions of nucleotide deletion.

ined (belonging to serotypes O157, O26, O111, and O145) possessed the sequences of the EAST1 gene and its upstream region, in confirming previous hybridization data [13]. However, it was found that the EAST1 gene-corresponding region has mutations, which resulted in the negative reaction in the PCR assay with the primer set of EAST11a and EAST11b [14]. Thus, when the EAST1 gene distribution is examined in clinical isolates, the complete gene sequence should be determined.

The determined EHEC sequences (SHEAST) have been classified into two types (type 1 and type 2) and the EHEC strains possessed either type 1 or type 2 SHEAST. The type 1 SHEAST was found primarily in O157:H7 strains isolated from outbreaks that occurred in Japan in 1996, while type 2 SHEAST was linked with sporadic cases. Type 1 SHEAST could serve as a marker for outbreak-associated strains of O157:H7. The sources of the outbreaks of O157:H7 infections in Japan were contaminated water, vegetables, vegetable salad, and seafood salad or sauce. The O157:H7 strains possessing type 1 SHEAST might have a greater potential to cause outbreaks in the environment of Japan, although type 1 SHEAST may not be directly involved in such a virulence potential, and may only be a marker for the presence of other authentic virulence factors. In order to detect EHEC O157:H7 strains with type 1 SHEAST by PCR, the primer set of SHEAST1a and SHEAST1b (generating a 111-bp product) was successfully constructed (Fig. 2).

The sequence homologous to the EAST1 gene was also found in *Y. pestis*. It was located within an IS sequence (IS*285*). Sequence comparison showed that type 1 SHEAST is closer to the EAST1 gene (89.7% homology in sequence) than to the EAST1-like sequence of IS*285* (76.9% homology in sequence). The EAST1 gene, type 1 SHEAST, and the EAST1-like sequence of IS*285* are theoretically changeable to each other by single base substitutions, indicating that they have originated from a common ancestral sequence. The findings thus provide evolutionary evidence of species–species transfer of the EAST1 gene sequences between *E. coli* and *Y. pestis*. The possibility exists that the EAST1 gene sequences are (or have derived from) a part of the genes involved in transposition events.

In the case of SHEAST in EHEC, the mutations, in part, are directed to 'switch off' the translation, and the EAST1 gene-homologous sequence in *Y. pestis* has two stop codons within the sequence, indicating the 'sleeping' gene sequences. Further studies are required to examine whether or not any of the EAST1-like sequences found in EHEC and *Y. pestis* would be predicted to yield a functional polypeptide toxin.

Acknowledgements: This study was supported by a grant-in-aid for scientific research (B) from the Ministry of Education, Science, Sports and Culture, Japan, and a grant from the Organization for Pharmaceutical Safety and Research (OPSR), Japan.

References

[1] Nataro, J.P., Steiner, T. and Guerrant, R.L. (1998) Emerg. Infect. Dis. 4, 251–261.
[2] Huppertz, H.I., Rutkowski, S., Aleksic, S. and Karch, H. (1997) Lancet 349, 1660–1662.
[3] Cohen, M.B., Hawkins, J.A., Weckbach, L.S., Staneck, J.L., Levine, M.M. and Heck, J.E. (1993) J. Clin. Microbiol. 31, 351–353.
[4] Gascón, J., Vargas, M., Quintó, L., Corachán, M., de Anta, M.T.J. and Vila, J. (1998) J. Infect. Dis. 177, 1409–1412.
[5] Joseph, J., Harley, R.A. and Frye, M.D. (1995) N. Engl. J. Med. 332, 273–274.
[6] Mathewson, J.J., Jiang, Z.D., Zumla, A., Chintu, C., Luo, N., Calamari, S.R., Genta, R.M., Steephen, A., Schwartz, P. and DuPont, H.L. (1995) J. Infect. Dis. 171, 1636–1639.
[7] Polotsky, Y., Nataro, J.P., Kotler, D., Barrett, T.J. and Orenstein, J.M. (1997) J. Clin. Microbiol. 35, 1952–1958.
[8] Germani, Y., Minssart, P., Vohito, M., Yassibanda, S., Glaziou, P., Hocquet, D., Berthélémy, P. and Morvan, J. (1998) Am. J. Trop. Med. Hyg. 59, 1008–1014.
[9] Savarino, S.J., Fasano, A., Robertson, D.C. and Levine, M.M. (1991) J. Clin. Invest. 87, 1450–1455.
[10] Savarino, S.J., Fasano, A., Watson, J., Martin, B.M., Levine, M.M., Guandalini, S. and Guerry, P. (1993) Proc. Natl. Acad. Sci. USA 90, 3093–3097.
[11] Yamamoto, T. and Echeverria, P. (1996) Infect. Immun. 64, 1441–1445.
[12] Yamamoto, T. and Nakazawa, M. (1997) J. Clin. Microbiol. 35, 223–227.
[13] Savarino, S.J., McVeigh, A., Watson, J., Cravioto, A., Molina, J., Echeverria, P., Bhan, M.K., Levine, M.M. and Fasano, A. (1996) J. Infect. Dis. 173, 1019–1022.
[14] Yamamoto, T., Wakisaka, N., Sato, F. and Kato, A. (1997) FEMS Microbiol. Lett. 147, 89–95.

[15] Yamamoto, T., Wakisaka, N., Nakae, T., Kamano, T., Serichantalergs, O. and Echeverria, P. (1996) Infect. Immun. 64, 3694–3702.
[16] Nataro, J.P., Deng, Y., Cookson, S., Cravioto, A., Savarino, S.J., Guers, L.D., Levine, M.M. and Tacket, C.O. (1995) J. Infect. Dis. 171, 465–468.
[17] Food Sanitation Division Environmental Health Bureau Ministry of Health and Welfare (1997) Report on enterohemorrhagic *Escherichia coli* O157 infections.
[18] American gastroenterological association (1995) Gastroenterology 108, 1923–1934.
[19] Hamano, S., Nakanishi, Y., Nara, T., Seki, T., Ohtani, T., Oishi, T., Joh, K., Oikawa, T., Muramatsu, Y. and Ogawa, Y. (1993) Acta Paediatr. 82, 454–458.
[20] Watanabe, H., Terajima, J., Izumiya, H., Wada, A., Iyoda, S. and Tamura, K. (1998) Jpn. J. Med. Sci. Biol. 51 (Suppl. 1), S115–S123.
[21] Izumiya, H., Terajima, J., Wada, A., Inagaki, Y., Itoh, K., Tamura, K. and Watanabe, H. (1997) J. Clin. Microbiol. 35, 1675–1680.
[22] Murray, M.G. and Thompson, W.F. (1980) Nucleic Acids Res. 8, 4321–4325.
[23] Filippov, A.A., Oleinikov, P.N., Motin, V.L., Protsenko, O.A. and Smirnov, G.B. (1995) Contrib. Microbiol. Immunol. 13, 306–309.