Allelic variation in colonization factor CS6 of enterotoxigenic *Escherichia coli* isolated from patients with acute diarrhoea and controls

Subrata Sabui,1 Abhisek Ghosal,1 Sanjucta Dutta,1 Avishek Ghosh,1 T. Ramamurthy,1 James P. Nataro,2 T. Hamabata3 and Nabendu Sekhar Chatterjee1

**Correspondence**
Nabendu Sekhar Chatterjee
chatterjeens@icmr.org.in or nschatterjee@rediffmail.com

1National Institute of Cholera & Enteric Diseases, Kolkata, India
2Center for Vaccine Development, University of Maryland, Baltimore, USA
3International Medical Center of Japan, Tokyo, Japan

Colonization factor antigens (CFAs) are important virulence factors in enterotoxigenic *Escherichia coli* (ETEC). Using a multiplex PCR and RT-PCR, this study tested the presence of common colonization factor-encoding genes and their expression in 50 ETEC strains isolated from stool specimens. The samples were from patients (children) with acute diarrhoea (cases) admitted to the Infectious Disease Hospital (Kolkata, India) and from normal children (controls) under 5 years of age from the community. The results indicated that coli surface antigen 6 (CS6) was the most prevalent CFA (78%) expressed by these ETEC strains. Sequence analysis of both of the CS6 structural genes, i.e. cssA and cssB, in different ETEC isolates revealed the presence of point mutations in a systematic fashion. Based on the analysis of these variations, it was found that CssA had three alleles and CssB had two. Based on the allelic variations, subtyping of CS6 into AIBI, AIIIBII, AIIIBI, AIBII and AIIIBII is proposed. The point mutations in the different alleles were reflected in a partial alteration in the secondary structure of both subunits, as determined by computational analysis. The functional significance of these changes was confirmed with cellular binding studies in Caco-2 cells with representative ETEC isolates. CS6 with AI or AIII allelic subtypes showed a higher binding capacity than AII, whereas BI showed stronger binding than BII. The AII and BII alleles were mostly detected in controls rather than in cases. The antibody specificity of BI and BII also varied due to alteration of the amino acids. Thus, CS6 variants are formed as a result of different allelic combinations of CsaA and CsaB, and these changes at the functional level might be important in the development of an effective ETEC vaccine.

**INTRODUCTION**

Enterotoxigenic *Escherichia coli* (ETEC) is mostly associated with watery diarrhoea in children under 5 years of age and is a major cause of travellers’ diarrhoea in many developing countries (Black, 1993; Gaastra & Svennerholm, 1996; Sack, 1975). ETEC isolates express well-defined heat-labile enterotoxin and/or heat-stable enterotoxin (ST), which are the main virulence factors associated with diarrhoeal infection (Gaastra & Svennerholm, 1996). ETEC colonizes the small intestine with the help of surface antigens, known as colonization factor antigens (CFAs), multiplies in the small intestine and produces toxin(s). Generally, ETEC is classified on the basis of its surface antigen profiles (Wolf et al., 1997), as well as on the production of toxins. Most of the identified CFAs from clinical isolates have been categorized into several groups based on their antigenic variations, structural similarities and amino acid sequences (Sack et al., 1971; Helander et al., 1997; Nataro & Kaper, 1998).

From many epidemiological studies, it has been found that several human ETEC isolates express CFA/I, CFA/II and CFA/IV. CFA/I is a fimbrial antigen (Nataro & Kaper, 1998; Kaper et al., 2004), whereas CFA/II and CFA/IV comprise several types of antigen. CFA/II is composed of coli surface antigen 1 (CS1), CS2 and CS3, whereas CFA/IV is composed of CS4, CS5 and CS6 (Wolf et al., 1989). As well as CFA/I, CFA/II and CFA/IV, other identified CFAs are CFA/III (CS8), CS12, CS14, CS17 and CS21.

**Abbreviations:** CFA, colonization factor antigen; CS, coli surface antigen; ETEC, enterotoxigenic *Escherichia coli*; RT, reverse transcription; ST, heat-stable enterotoxin.

The GenBank/EMBL/DDBJ accession numbers for the complete genomic sequences of cssA and cssB alleles of different representative ETEC isolates are GQ241328–GQ241337.
Conventionally, the CFAs of ETEC isolates are detected using specific mAbs raised against these antigens. This mAb-based assay is difficult to perform during routine surveillance where the availability of the mAbs is limited. Recently, various PCR-based methods have been developed, and their specificities and sensitivities were found to be on a par with those of conventional serological methods (Ghosal et al., 2007; Sjoling et al., 2007). As the initial formulation of single PCRs is labour-intensive, multiplex PCRs have been developed to facilitate the detection of common CFAs during routine surveillance (Rodas et al., 2009; Vidal et al., 2009).

In this study, we used a multiplex PCR assay, which is an extension of our previous work (Ghosal et al., 2007), for quicker and more cost-effective identification of CFAs. Using this assay, we found that CS6 was the most prevalent CFA among the ETEC isolates from diarrhoeal patients. We also identified for the first time different CS6 allelic subtypes, which showed differential cellular binding and antibody specificities.

**METHODS**

**Bacterial strains and growth conditions.** The 50 ETEC isolates used in this study were isolated from patients (children) with acute diarrhoea (cases) from the Infectious Disease Hospital (IDH; Kolkata, India) and from normal children (controls) under 5 years of age from a community-based study, which is part of the ongoing Global Enteric Multicenter Study (University of Maryland, MD, USA), where Kolkata is one of the sites. From each stool specimen, five typical *E. coli* colonies grown on MacConkey agar were screened for ETEC using a multiplex PCR (Nguyen et al., 2005). Detailed clinical information for the 50 ETEC isolates is presented in Table 1. ETEC isolates harbouring *elt* or *elt* and *eltb* genes were grown at 37 °C in Luria–Bertani broth (Difco) and cryo-preserved by adding 15% glycerol before storing at –80 °C. For detection of surface antigens of ETEC isolates, CFA agar (1% Casamino acids, 0.15% yeast extract, 0.005% MgSO₄, 0.005% MnCl₂, 2% agar, pH 7.4) plates were used. One ETEC isolate per patient was used in this study.

‘O’ Serogrouping of ETEC isolates. The ‘O’ serogrouping of ETEC isolates was carried out using an *E. coli* ‘O’ serotyping kit in accordance with the manufacturer’s instructions (Denka Seiken). This kit consists of eight ‘O’ polyvalent and 43 monospecific antisera.

**Caco-2 cell culture.** The Caco-2 cell line was cultured as a monolayer in Eagle’s minimal essential medium (EMEM) at 37 °C with 5% CO₂ using a humidified incubator (HERAcell 150; Thermo Scientific). The medium was supplemented with 10% (v/v) fetal bovine serum (FBS), 2.2 g sodium bicarbonate 1⁻¹, 1% non-essential amino acids (100 ×; MP Biomedical) and 1% (v/v) antibiotics (5000 U penicillin ml⁻¹, 50 µg streptomycin ml⁻¹; Sigma). The Caco-2 cells from passages 20 and 40 were seeded at a concentration of 1.4 x 10⁵ cells cm⁻² in 96-well tissue culture plates (Corning) (Le Blay et al., 2004). The culture medium was replaced every other day and the monolayer was used after the cells reached confluency. A day before the adhesion assay, the culture medium was replaced with antibiotic-free EMEM with 0.5% FBS.

**Isolation of genomic DNA and total RNA and preparation of cDNA from ETEC isolates.** Total genomic DNA was isolated from the overnight ETEC cultures as described previously (Ghosal et al., 2007). To prepare RNA, the ETEC isolates were grown overnight on CFA agar at 37 °C. mRNA was prepared using bacterial cultures with TRIZol (Invitrogen) following the manufacturer’s protocol. Ten micrograms of RNA was treated with RNase-free DNase I (NEB) and stored at –70 °C until further use. Reverse transcription (RT) was performed with 1 µg total RNA using a Reverse Transcription System (Promega) following the manufacturer’s instructions to obtain cDNA from the ETEC isolates.

**Primer design.** The primers used for amplification of each CFA gene used in the multiplex PCR and in the simplex PCRs for the detection of the two structural subunits of CS6 (cssA and cssB) were as reported previously (Ghosal et al., 2007, 2009).

**Multiplex PCR for the detection of CFA genes.** Genomic DNA or the first-strand cDNA was used as a template and amplified by multiplex PCR in an Eppendorf Mastercycler Personal. The multiplex PCR primer sets were designed by calculating similar *Tₘ* values for the primers of the CFA genes. The amplicon sizes of the different CFA genes are shown in Fig. 1. Reaction conditions for different sets of CFA primers were as follows: initial denaturation at 95 °C for 3 min, followed by 35 cycles of denaturation at 95 °C for 1 min, annealing at 54 °C for 45 s and elongation at 72 °C for 1 min 30 s, with a 5 min final extension at 72 °C. mRNA that had not undergone RT and reaction mix without the RT product as template were used as negative controls. Amplified DNA fragments were detected using a previously described method (Ghosal et al., 2007).

**Sequencing of the cssA and cssB subunit genes.** For sequencing, cssA and cssB genes from each CS6-expressing isolate were amplified from the genomic DNA by initial denaturation at 96 °C for 2 min, followed by 25 cycles of denaturation at 96 °C for 10 s, annealing at 50 °C for 5 s and elongation at 60 °C for 4 min, using specific primers for each subunit, as described previously (Ghosal et al., 2009). Sequencing was performed using an ABI PRISM 3700 automated DNA sequencer (Applied Biosystems) and the amplified products were confirmed against GenBank using BLAST (http://www.ncbi.nlm.nih.gov/BLAST). Translation from genomic sequence to amino acid sequence was performed using Gene Runner software. The CLUSTAL W program (http://align.genome.jp/) was used forCssA and CssB sequence alignment.

**Antibody production and antibody response between CS6 subtypes.** Antibodies against the CssA and CssB subunits of CS6 have been described by us previously (Ghosal et al., 2007, 2009). The antibody titre against each subunit and each allelic subtype was analysed using a protocol described previously (Ghosal et al., 2009). Briefly, a microtitre plate was coated with 100 µl (1 ng µl⁻¹) purified CS6 variants. Specific antibodies against CssA (AI type) and CssB (BI type) were added at different dilutions (1:50, 1:100, 1:300, 1:500, 1:1000, 1:5000 and 1:100,000 in PBS) and incubated for 2 h. After washing, the plate was incubated with horseradish peroxidase (HRP)-conjugated secondary antibody for 1 h, and bound antibody was quantified using 3,3′,5,5′-tetramethylbenzidine as a substrate and measuring *A₄₅₀* using a microplate reader (Bio-Rad).

**Bacterial binding assay using Caco-2 cells.** Determination of bacterial binding to Caco-2 cells was performed using an ELISA. Only CS6-expressing ETEC in each representative group was analysed, and was harvested by centrifugation, washed three times with sterile PBS (pH 7.4) and diluted to ~5 x 10⁹ c.f.u. ml⁻¹ in PBS. Biotinylation was carried out as described previously (Sasmal et al., 1995). After three washes with PBS, biotinylated bacterial suspension (100 µl per well) at a fixed concentration (~5 x 10⁶ c.f.u. ml⁻¹) was added to a 96-well
plate coated with a monolayer of Caco-2 cells following a standard
protocol (Le Blay et al., 2004). One hundred microlitres of HRP-
conjugated avidin (1:300 dilution in PBS) was added and incubated
for 1 h. Colour development was determined by measuring the
A450 following a standard procedure.

**Table 1. Clinical information on the ETEC isolates included in this study**

<table>
<thead>
<tr>
<th>ETEC isolate</th>
<th>Date</th>
<th>Case/control</th>
<th>Pathogen(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD173</td>
<td>Jan 2008</td>
<td>Control</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD082</td>
<td>Jan 2008</td>
<td>Case</td>
<td>ETEC, rotavirus</td>
</tr>
<tr>
<td>SD067</td>
<td>Dec 2007</td>
<td>Case</td>
<td>ETEC, <em>Shigella flexneri</em> 6</td>
</tr>
<tr>
<td>SD354</td>
<td>Mar 2008</td>
<td>Control</td>
<td>ETEC, <em>Vibrio cholerae</em> O139</td>
</tr>
<tr>
<td>SD049</td>
<td>Jan 2008</td>
<td>Case</td>
<td>ETEC, rotavirus</td>
</tr>
<tr>
<td>SD213</td>
<td>Feb 2008</td>
<td>Case</td>
<td>ETEC, rotavirus, <em>Giardia lamblia, Campylobacter jejuni</em></td>
</tr>
<tr>
<td>SD284</td>
<td>Apr 2008</td>
<td>Case</td>
<td>ETEC, <em>Giardia lamblia</em></td>
</tr>
<tr>
<td>IDH00424</td>
<td>Apr 2008</td>
<td>Case</td>
<td>ETEC, <em>Vibrio cholerae</em> O1</td>
</tr>
<tr>
<td>IDH00650</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC, <em>Vibrio cholerae</em> O1</td>
</tr>
<tr>
<td>SD439</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC, <em>Giardia lamblia, Cryptosporidium, Campylobacter jejuni</em></td>
</tr>
<tr>
<td>IDH00781</td>
<td>Aug 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD357</td>
<td>Mar 2008</td>
<td>Control</td>
<td>ETEC</td>
</tr>
<tr>
<td>IDH00139</td>
<td>Dec 2007</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>IDH00764</td>
<td>Aug 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD661</td>
<td>Aug 2008</td>
<td>Case</td>
<td>ETEC, <em>Vibrio cholerae</em> O1</td>
</tr>
<tr>
<td>SD519</td>
<td>May 2008</td>
<td>Control</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD140</td>
<td>Dec 2007</td>
<td>Control</td>
<td>ETEC, <em>Vibrio cholerae</em> non-O1, non-O139</td>
</tr>
<tr>
<td>SD018</td>
<td>Dec 2007</td>
<td>Case</td>
<td>ETEC, rotavirus</td>
</tr>
<tr>
<td>SD281</td>
<td>Mar 2008</td>
<td>Case</td>
<td>ETEC, <em>Campylobacter jejuni</em>, norovirus GII</td>
</tr>
<tr>
<td>SD117</td>
<td>Dec 2007</td>
<td>Control</td>
<td>ETEC</td>
</tr>
<tr>
<td>IDH00364</td>
<td>Mar 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>IDH00420</td>
<td>Apr 2008</td>
<td>Case</td>
<td>ETEC, <em>Giardia lamblia</em></td>
</tr>
<tr>
<td>IDH00238</td>
<td>Jan 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD487</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD570</td>
<td>Jul 2008</td>
<td>Control</td>
<td>ETEC, <em>Giardia lamblia, Campylobacter jejuni</em></td>
</tr>
<tr>
<td>IDH00647</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC, <em>Vibrio fluvialis</em></td>
</tr>
<tr>
<td>SD669</td>
<td>Aug 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD453</td>
<td>May 2008</td>
<td>Case</td>
<td>ETEC, norovirus GII</td>
</tr>
<tr>
<td>SD662</td>
<td>Aug 2008</td>
<td>Case</td>
<td>ETEC, rotavirus</td>
</tr>
<tr>
<td>SD332</td>
<td>Feb 2008</td>
<td>Control</td>
<td>ETEC, <em>Giardia lamblia</em></td>
</tr>
<tr>
<td>IDH00427</td>
<td>Apr 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>IDH00838</td>
<td>Sep 2008</td>
<td>Case</td>
<td>ETEC, <em>Giardia lamblia</em></td>
</tr>
<tr>
<td>IDH00696</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD124</td>
<td>Dec 2007</td>
<td>Control</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD260</td>
<td>Feb 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD276</td>
<td>Mar 2008</td>
<td>Case</td>
<td>ETEC, rotavirus</td>
</tr>
<tr>
<td>SD282</td>
<td>Mar 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>IDH00358</td>
<td>Mar 2008</td>
<td>Case</td>
<td>ETEC, <em>Entamoeba histolytica</em></td>
</tr>
<tr>
<td>IDH00428</td>
<td>Apr 2008</td>
<td>Case</td>
<td>ETEC, <em>Cryptosporidium</em></td>
</tr>
<tr>
<td>SD486</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD480</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC, <em>Campylobacter jejuni</em></td>
</tr>
<tr>
<td>SD479</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC, <em>Campylobacter jejuni, Vibrio cholerae</em> non-O, non-O139</td>
</tr>
<tr>
<td>SD535</td>
<td>Jun 2008</td>
<td>Control</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD349</td>
<td>Mar 2008</td>
<td>Control</td>
<td>ETEC</td>
</tr>
<tr>
<td>SD456</td>
<td>May 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>IDH00200</td>
<td>Jan 2008</td>
<td>Case</td>
<td>ETEC, rotavirus, <em>Entamoeba histolytica, Giardia lamblia</em></td>
</tr>
<tr>
<td>IDH00661</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC, <em>Aeromonas</em> sp.</td>
</tr>
<tr>
<td>IDH00757</td>
<td>Apr 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
<tr>
<td>IDH00646</td>
<td>Jul 2008</td>
<td>Case</td>
<td>ETEC, rotavirus, <em>Vibrio fluvialis, Shigella flexneri</em></td>
</tr>
<tr>
<td>IDH00356</td>
<td>Mar 2008</td>
<td>Case</td>
<td>ETEC</td>
</tr>
</tbody>
</table>

**Computational analysis of the secondary structure of CS6 subtypes.** The CS6 subunit secondary structure predictions for each subgroup were carried out using Jpred software (http://www. compbio.dundee.ac.uk/www-jpred/) from the ExPASy proteomics server (http://www.expasy.ch).
RESULTS AND DISCUSSION

Identification of CFA-encoding genes and their expression

We designed a multiplex PCR protocol to detect the 10 CFAs that are the most prevalent adhesins in ETEC isolated from developing countries. This method is a modification of our previously described simplex PCR (Ghosal et al., 2007). We found that all the isolates harboured multiple CFA genes. The CFA types of 50 ETEC isolates are presented in Table 2 and representative agarose gels are shown in Fig. 1. In the multiplex PCR, CFA/I (8%), CFA/II-CS1 (8%), CFA/II-CS2 (4%), CFA/II-CS3 (16%), CFA/III-CS8 (2%), CFA/IV-CS6 (78%), CFA/IV-CS5 (28%), CFA/IV-CS4 (4%), CS14 (24%) and CS17 (22%) were detected among the ETEC isolates. It was found that the CFAs were expressed as a single surface antigen or in combination with more than one antigen. Our study indicated that the most commonly expressed CFA types in RT-PCR in order of frequency were CS6>CS5>CS14>CS17>CS3. Amplification of CFA genes by simplex PCR is time-consuming and labour-intensive. Recent studies have shown that multiplex PCR assays are a rapid and sensitive method for identification of ETEC CFAs (Rodas et al., 2009; Vidal et al., 2009). For this reason, we used a new multiplex PCR assay for the detection of common CFAs so that a greater number of isolates could be accommodated in each assay. Our study indicated that CS6 was the most prevalent CF type in Kolkata. In different geographical areas, the prevalence of CFs in ETEC isolates differs considerably (Grewal et al., 1994), but CS6 has consistently been the most frequently identified (Gaastra & Svennerholm, 1996).

Previous studies with ETEC have shown that CS6 is associated with specific ‘O’ serogroups and enterotoxin types (McConnell et al., 1988; Wolf, 1997; Qadri et al., 2000). In rural Egypt, CS6-expressing ETEC was detected in association with ST and O169, O27, O159 and O78 serogroups among children with diarrhoea (Shaheen et al., 2004, 2009). Among the US military personnel deployed in northern Egypt, CS6-expressing ETEC with ST enterotoxin was the most common (Rockabrand et al., 2006). We found that the O6 and O128 serogroups were the most common in CS6-expressing ETEC found in Kolkata.

Sequence analysis of the prevalent colonization factor CS6

Point mutations in the cssA and cssB subunit genes of CS6 were detected in several ETEC isolates. Protein sequence analysis of these mutations in CS6 indicated that there were many variations in these genes from different isolates (Fig. 2). These variations were first noted in two ETEC isolates by Wolf et al. (1997) but were not characterized in detail. Based on the results obtained with our 50 ETEC isolates, we identified three alleles for cssA and two for cssB.

Comparative protein sequence analysis of the CsaA andCssB subunits of CS6 revealed specific variations in the amino acids as a result of these point mutations (Fig. 2). On the basis of amino acid changes in CsaA, CS6-expressing ETEC isolates were classified into three allelic subtypes, namely AI, AII and AIII. Similarly, the CsbB subunit could be classified into two allelic subtypes, BI and BII. Correlation of amino acid changes between CsaA and CsbB was not found; however, combinations of AIIBI and AIIIBI among ETEC isolates were frequently detected (Table 3). The occurrence of AI (43%) and AIII (48%) was

Fig. 1. Representative multiplex PCR assay results of detection of CFAs from ETEC isolates. (a) Lanes: 1, 100 bp ladder; 2, CS1 (324 bp) and CS3 (264 bp); 3, CS2 (368 bp) and CS3 (264 bp); 4, negative control (no template). (b) Lanes: 1, 100 bp ladder; 2, CS4 (250 bp); 3, CS5 (453 bp) and CS6 (321 bp); 4, negative control (no template). (c) Lanes: 1, 100 bp ladder; 2, CFA/III (437 bp); 3, CFA/I (364 bp) and CS17 (290 bp); 4, CS14 (394 bp).
Table 2. Characteristics of freshly isolated ETEC from patients with diarrhoea in Kolkata

<table>
<thead>
<tr>
<th>ETEC isolate</th>
<th>Enterotoxin gene(s)</th>
<th>Serotype*</th>
<th>Genotypic profile of CFAs</th>
<th>Genes present†</th>
<th>Genes expressed‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD173</td>
<td>elt</td>
<td>O rough</td>
<td>CS4 + CS6</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>SD082</td>
<td>elt</td>
<td>O114</td>
<td>CS5 + CS6</td>
<td>CS5 + CS6</td>
<td></td>
</tr>
<tr>
<td>SD067</td>
<td>elt</td>
<td>O168</td>
<td>CS1 + CS3 + CS6</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>SD354</td>
<td>elt</td>
<td>OUT</td>
<td>CS1 + CS3 + CS4 + CS6</td>
<td>CS3 + CS6</td>
<td></td>
</tr>
<tr>
<td>SD049</td>
<td>elt</td>
<td>OUT</td>
<td>CS5 + CS6</td>
<td>CS5 + CS6</td>
<td></td>
</tr>
<tr>
<td>SD213</td>
<td>elt</td>
<td>OUT</td>
<td>CS6 + CS17</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>SD284</td>
<td>elt</td>
<td>O8</td>
<td>CS6 + CS17 + CFA/I</td>
<td>CS6 + CS17</td>
<td></td>
</tr>
<tr>
<td>IDH00424</td>
<td>elt</td>
<td>OUT</td>
<td>CS5 + CS6 + CS14 + CS17</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>IDH00650</td>
<td>elt</td>
<td>O rough</td>
<td>CS6 + CS14</td>
<td>CS6 + CS14</td>
<td></td>
</tr>
<tr>
<td>SD439</td>
<td>elt</td>
<td>OUT</td>
<td>CS6 + CS14</td>
<td>CS6 + CS14</td>
<td></td>
</tr>
<tr>
<td>IDH00781</td>
<td>elt</td>
<td>O128</td>
<td>CS6 + CS14 + CS17 + CFA/I</td>
<td>CS6 + CS17</td>
<td></td>
</tr>
<tr>
<td>SD357</td>
<td>elt</td>
<td>O114</td>
<td>CS3 + CS5 + CS17</td>
<td>CS17</td>
<td></td>
</tr>
<tr>
<td>IDH00139</td>
<td>elt</td>
<td>O8</td>
<td>CS17 + CFA/I</td>
<td>CFA/I</td>
<td></td>
</tr>
<tr>
<td>IDH00764</td>
<td>elt</td>
<td>OUT</td>
<td>CS4 + CS6 + CS14</td>
<td>CS14</td>
<td></td>
</tr>
<tr>
<td>SD661</td>
<td>elt</td>
<td>O148</td>
<td>CS5 + CS14 + CFA/I</td>
<td>CS5 + CFA/I</td>
<td></td>
</tr>
<tr>
<td>SD519</td>
<td>elt</td>
<td>O114</td>
<td>CS5 + CS6 + CS14</td>
<td>CS5 + CS14</td>
<td></td>
</tr>
<tr>
<td>SD140</td>
<td>est</td>
<td>O128</td>
<td>CS6 + CFA/I</td>
<td>CS6 + CFA/I</td>
<td></td>
</tr>
<tr>
<td>SD018</td>
<td>est</td>
<td>O6</td>
<td>CS2 + CS3 + CS6</td>
<td>CS2 + CS3 + CS6</td>
<td></td>
</tr>
<tr>
<td>SD281</td>
<td>est</td>
<td>O126</td>
<td>CS6 + CS14</td>
<td>CS6 + CS14</td>
<td></td>
</tr>
<tr>
<td>SD117</td>
<td>est</td>
<td>O148</td>
<td>CS6 + CFA/I</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>IDH00364</td>
<td>est</td>
<td>O128</td>
<td>CS5 + CS6 + CS14</td>
<td>CS5 + CS6 + CS14</td>
<td></td>
</tr>
<tr>
<td>IDH00420</td>
<td>est</td>
<td>O25</td>
<td>CS6 + CFA/I</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>IDH00238</td>
<td>est</td>
<td>OUT</td>
<td>CS6 + CFA/I</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>SD487</td>
<td>est</td>
<td>OUT</td>
<td>CS6 + CS14</td>
<td>CS6 + CS14</td>
<td></td>
</tr>
<tr>
<td>SD570</td>
<td>est</td>
<td>O128</td>
<td>CS6 + CS17 + CFA/III</td>
<td>CS6 + CS17</td>
<td></td>
</tr>
<tr>
<td>IDH00647</td>
<td>est</td>
<td>O1</td>
<td>CS5 + CS6 + CS14 + CS17</td>
<td>CS5 + CS6 + CS14 + CS17</td>
<td></td>
</tr>
<tr>
<td>SD669</td>
<td>est</td>
<td>OUT</td>
<td>CS6 + CS14</td>
<td>CS6 + CS14</td>
<td></td>
</tr>
<tr>
<td>SD453</td>
<td>est</td>
<td>O128</td>
<td>CS5 + CS6 + CS14</td>
<td>CS6 + CS14</td>
<td></td>
</tr>
<tr>
<td>SD662</td>
<td>est</td>
<td>O1</td>
<td>CS5 + CS6 + CS14 + CS17</td>
<td>CS5 + CS6 + CS14</td>
<td></td>
</tr>
<tr>
<td>SD332</td>
<td>est</td>
<td>O146</td>
<td>CS1 + CS5</td>
<td>CS1 + CS5</td>
<td></td>
</tr>
<tr>
<td>IDH00427</td>
<td>est</td>
<td>O15</td>
<td>CS5 + CS14 + CS17</td>
<td>CS5 + CS17</td>
<td></td>
</tr>
<tr>
<td>IDH00838</td>
<td>est</td>
<td>OUT</td>
<td>CS17 + CFA/I</td>
<td>CS17</td>
<td></td>
</tr>
<tr>
<td>IDH00696</td>
<td>est</td>
<td>O25</td>
<td>CS5 + CS6 + CS17 + CFA/I</td>
<td>CS6 + CS17</td>
<td></td>
</tr>
<tr>
<td>SD124</td>
<td>elt + est</td>
<td>OUT</td>
<td>CS1 + CS3 + CS6</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>SD260</td>
<td>elt + est</td>
<td>OUT</td>
<td>CS2 + CS3 + CS4 + CS6</td>
<td>CS3 + CS6</td>
<td></td>
</tr>
<tr>
<td>SD276</td>
<td>elt + est</td>
<td>O6</td>
<td>CS1 + CS3 + CS4 + CS6</td>
<td>CS1 + CS3 + CS4 + CS6</td>
<td></td>
</tr>
<tr>
<td>SD282</td>
<td>elt + est</td>
<td>O6</td>
<td>CS1 + CS3 + CS4 + CS6</td>
<td>CS1 + CS3 + CS4 + CS6</td>
<td></td>
</tr>
<tr>
<td>IDH00358</td>
<td>elt + est</td>
<td>O6</td>
<td>CS3 + CS5 + CS6 + CS14</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>IDH00428</td>
<td>elt + est</td>
<td>OUT</td>
<td>CS5 + CS6 + CS14 + CS17</td>
<td>CS5 + CS6 + CS17</td>
<td></td>
</tr>
<tr>
<td>SD486</td>
<td>elt + est</td>
<td>O rough</td>
<td>CS6 + CS14</td>
<td>CS6 + CS14</td>
<td></td>
</tr>
<tr>
<td>SD480</td>
<td>elt + est</td>
<td>OUT</td>
<td>CS6 + CFA/I</td>
<td>CS6§</td>
<td></td>
</tr>
<tr>
<td>SD479</td>
<td>elt + est</td>
<td>O rough</td>
<td>CS5 + CS6</td>
<td>CS5 + CS6</td>
<td></td>
</tr>
<tr>
<td>SD355</td>
<td>elt + est</td>
<td>O1</td>
<td>CS5 + CS6 + CS14</td>
<td>CS5 + CS6</td>
<td></td>
</tr>
<tr>
<td>SD349</td>
<td>elt + est</td>
<td>O6</td>
<td>CS2 + CS3 + CS6 + CFA/III</td>
<td>CS2 + CS3 + CFA/III</td>
<td></td>
</tr>
<tr>
<td>SD456</td>
<td>elt + est</td>
<td>O167</td>
<td>CS4 + CS6 + CS14</td>
<td>CS14</td>
<td></td>
</tr>
<tr>
<td>IDH00200</td>
<td>elt + est</td>
<td>O6</td>
<td>CS1 + CS3 + CS6 + CFA/I</td>
<td>CS1 + CS3 + CS6</td>
<td></td>
</tr>
<tr>
<td>IDH00661</td>
<td>elt + est</td>
<td>O44</td>
<td>CS5 + CS6</td>
<td>CS5 + CS6</td>
<td></td>
</tr>
<tr>
<td>IDH00757</td>
<td>elt + est</td>
<td>O119</td>
<td>CS3 + CS5 + CS6 + CS17</td>
<td>CS3 + CS5 + CS6 + CS17</td>
<td></td>
</tr>
<tr>
<td>IDH00646</td>
<td>elt + est</td>
<td>OUT</td>
<td>CS3 + CS5 + CS6 + CS14 + CS17</td>
<td>CS5 + CS6 + CS17</td>
<td></td>
</tr>
<tr>
<td>IDH00356</td>
<td>elt + est</td>
<td>O148</td>
<td>CS5 + CS6 + CS14 + CFA/I</td>
<td>CS14 + CS17</td>
<td></td>
</tr>
</tbody>
</table>

*OUT, O antigen untypable.
†Genes present were detected by multiplex PCR.
‡Genes expressed were detected by multiplex RT-PCR.
§CS6 gene expressed alone.
Table 3. Characteristics of CS6-expressing ETEC isolates

<table>
<thead>
<tr>
<th>ETEC isolate</th>
<th>Gene expression profile</th>
<th>CS6 subtyping</th>
<th>Phenotypic profile (Western blot)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD439</td>
<td>CS6 + CS14</td>
<td>AIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD284</td>
<td>CS6 + CS17</td>
<td>AIBI</td>
<td>NE</td>
</tr>
<tr>
<td>SD479</td>
<td>CS5 + CS6</td>
<td>AIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD493</td>
<td>CS6 + CS14</td>
<td>AIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD669</td>
<td>CS6 + CS14</td>
<td>AIBI</td>
<td>NE</td>
</tr>
<tr>
<td>SD535</td>
<td>CS5 + CS6</td>
<td>AIBI</td>
<td>NE</td>
</tr>
<tr>
<td>SD662</td>
<td>CS5 + CS6 + CS14</td>
<td>AIBI</td>
<td>+</td>
</tr>
<tr>
<td>IDH00364</td>
<td>CS6</td>
<td>AIBI</td>
<td>+</td>
</tr>
<tr>
<td>IDH00358</td>
<td>CS6</td>
<td>AIBI</td>
<td>+</td>
</tr>
<tr>
<td>IDH00428</td>
<td>CS5 + CS6 + CS14</td>
<td>AIBI</td>
<td>+</td>
</tr>
<tr>
<td>IDH00200</td>
<td>CS1 + CS3 + CS6</td>
<td>AIBI†</td>
<td>+</td>
</tr>
<tr>
<td>IDH00650</td>
<td>CS6 + CS14</td>
<td>AIBI</td>
<td>NE</td>
</tr>
<tr>
<td>IDH00661</td>
<td>CS5 + CS6</td>
<td>AIBI</td>
<td>+</td>
</tr>
<tr>
<td>IDH00781</td>
<td>CS6 + CS17</td>
<td>AIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD124</td>
<td>CS6</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD276</td>
<td>CS1 + CS3 + CS4 + CS6</td>
<td>AIIIBI†</td>
<td>+</td>
</tr>
<tr>
<td>SD082</td>
<td>CS5 + CS6</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD067</td>
<td>CS6</td>
<td>AIIIBI</td>
<td>NE</td>
</tr>
<tr>
<td>SD281</td>
<td>CS6 + CS14</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD213</td>
<td>CS6</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD487</td>
<td>CS6 + CS14</td>
<td>AIIIBI</td>
<td>NE</td>
</tr>
<tr>
<td>SD486</td>
<td>CS6 + CS14</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD480</td>
<td>CS6</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>IDH00424</td>
<td>CS6</td>
<td>AIIIBI</td>
<td>NE</td>
</tr>
<tr>
<td>IDH00238</td>
<td>CS6</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>IDH00696</td>
<td>CS6 + CS17</td>
<td>AIIIBI</td>
<td>NE</td>
</tr>
<tr>
<td>IDH00757</td>
<td>CS3 + CS5 + CS6 + CS17</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>IDH00646</td>
<td>CS5 + CS6 + CS17</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD140</td>
<td>CS6 + CFA/I</td>
<td>AIBII</td>
<td>+</td>
</tr>
<tr>
<td>SD117</td>
<td>CS6</td>
<td>AIBII†</td>
<td>+</td>
</tr>
<tr>
<td>SD049</td>
<td>CS5 + CS6</td>
<td>AIBII</td>
<td>NE</td>
</tr>
<tr>
<td>SD173</td>
<td>CS6</td>
<td>AIIIBI†</td>
<td>+</td>
</tr>
<tr>
<td>SD260</td>
<td>CS3 + CS6</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD354</td>
<td>CS3 + CS6</td>
<td>AIIIBI</td>
<td>+</td>
</tr>
<tr>
<td>SD282</td>
<td>CS1 + CS3 + CS4 + CS6</td>
<td>AIIIBI†</td>
<td>+</td>
</tr>
<tr>
<td>SD570</td>
<td>CS6 + CS17</td>
<td>AIIIBI</td>
<td>NE</td>
</tr>
<tr>
<td>IDH00420</td>
<td>CS6</td>
<td>AIIIBI</td>
<td>NE</td>
</tr>
<tr>
<td>IDH00647</td>
<td>CS5 + CS6 + CS14 + CS17</td>
<td>AIIIBI</td>
<td>NE</td>
</tr>
</tbody>
</table>

*NE, CS6 protein not expressed; +, CS6 protein expressed.
†The GenBank accession numbers for the complete genomic sequences of cssA and cssB alleles of different representative ETEC isolates are GQ241328–GQ241337.
very common, whereas AII (9%) was less so. In isolates having the AI and AIII alleles, variations were found at three positions: A/D39→G, K90→R and K109→R, and A/G39→D, T48→I and A99→T, respectively. In the case of AII, variations were found at nine positions: Q24→R, K32→E, G/D39→A, S46→N, D76→A, D87→E, D117→E, E119→K and P121→S (Fig. 2a). The amino acid at position 39 of the CssA subunit was unique in the sense that each allele had a different amino acid, i.e. G, D and A in the AI, AIII and AII allelic subtypes, respectively.

Fig. 3. Schematic diagram showing the secondary structure of the CssA (a) and CssB (b) allelic subtypes of CS6. The secondary structure was predicted theoretically using the Jpred program from amino acid sequences derived from the DNA sequence. The barrels and arrows denote the α-helices and β-sheets, respectively, with numbers above indicating the amino acid positions. Fn, Fibronectin.

Fig. 4. Determination of antibody specificity among CS6 allelic subtypes. (a) Determination of antibody specificity between CssB allelic subtypes. CssBI and CssBII protein was coated onto 96-well ELISA plates and anti-CssBI antibody at different dilutions was added for determination of antibody specificity. Each bar represents the mean ± SEM for the independent experiments. Shaded and filled bars denote the antibody cross-reactivity of CssBI- and CssBII-expressing ETEC, respectively. (b) Determination of antibody specificity among CssA allelic subtypes. CssAI, CssAII and CssAIII protein was coated onto 96-well ELISA plates and anti-CssAI antibody at different dilutions was added for determination of antibody specificity. Each bar represents the mean ± SEM of the independent experiments. Filled, shaded and open bars denote the antibody specificity of CssAI-, CssAIII- and CssAII-expressing ETEC, respectively.
Our previous study indicated that the C-terminal loop of the CssA subunit of CS6 binds to fibronectin, promoting colonization (Ghosal et al., 2009). CssAI and CssAIII had an identical fibronectin-binding domain. In the case of ETEC isolates with the AII allele, there were three variations (D117→E, E119→K and P121→S) in the fibronectin-binding domain. These alterations did not appear to affect fibronectin binding (data not shown).

Based on the amino acid variations, CssB could be categorized into two allelic subtypes, namely BI and BII. The BI allelic type was prevalent (74% of isolates) among the ETEC isolates in this study. A minor fraction (26%), which belonged to the BII subtype, had variations at six positions: D40→N, A60→V, H80→N, K97→N, S107→P and N138→H (Fig. 2b).

**Secondary structure predictions for the CS6 allelic subtypes**

The secondary structures of the CssA and CssB subunits of different CS6 subtypes were determined from the primary protein sequence using Jpred software (Fig. 3). The region spanning aa 85–96 in the AI and AIII subtypes was predicted to be an α-helix, whereas, in the AII subtypes, the α-helix spanned aa 77–96, increasing the length of the α-helical region (Fig. 3a). The CssBI subtype was mainly rich in β-sheets, and no α-helix was observed. However, in CssBII, an α-helical region appeared from aa 98 to 101 as a result of changes in the amino acid sequence (Fig. 3b). The secondary structure predictions for each subtype of CssA and CssB might have occurred due to amino acid changes in both subunits. As these structural alterations may play...
an important role in their functional activity, we attempted to study the immunogenicity and cellular binding of ETEC isolates expressing different CS6 alleles.

**Immunogenicity of CS6 subtypes**

From computational analysis, it was found that the amino acid variations were mostly clustered into the antigenic regions of the protein, and this variation was reflected in their immunogenicity. The antibody against CssBI showed a twofold decrease in binding to CssBII at a dilution of 1:1000 (Fig. 4a). The reverse was true for the CssBII antibody (data not shown). These results suggested that the amino acid modifications play an important role in protein folding and may be responsible for the differential immunogenicity. Similarly, the peptide antibody (Ghosal et al., 2007) against CssAI interacted equally withCssAIII but fourfold less efficiently with CssAI at a dilution of 1:300 (Fig. 4b). This difference occurred because the amino acid sequence in the peptide antibody-interacting region of CssAI varied from that of CssAI and CssAIII.

**Cellular binding of CS6 allelic variants**

The Caco-2 cell binding assay indicated that the cellular binding of CS6-expressing ETEC with types AIBI and AIIIBI was better than that with the other types (Fig. 5c). Cellular binding was reduced by approximately 50% for isolates expressing CssBII. We assume that the structural alteration in the region from aa 98 to 101 resulted in decreased binding to Caco-2 cells. When the ETEC isolates were pre-incubated with anti-CssAI and anti-CssBI antibodies, the representative isolates showed a significant decrease in binding to Caco-2 cells ($P<0.05$). Interestingly, ETEC isolates pre-incubated with anti-CssBI antibody bound to intestinal cells less efficiently than anti-CssAI antibody-incubated bacteria ($P<0.05$).

It is known that CS6-expressing ETEC isolates can bind fibronectin through Cssa (Ghosal et al., 2009), and binding to polarized cells could be achieved by the CssB subunit as a result of protein–lipid interaction (Jansson et al., 2009). However, the involvement of structural domains of either subunit (if any) needs to be elucidated. Using a panel of ETEC isolates expressing different allelic variants of CS6, we observed differential binding to an intestinal cell line. We found that CS6-expressing ETEC with a combination of AIBI or AIIIBI could bind intestinal cells relatively better than CssBII-expressing isolates. It could be concluded that the fibronectin-binding region was conserved and hence there was no alteration in binding due to the AI or AII alleles. However, the CssB-binding region was not studied in detail. Our data indicated that CssBII had reduced binding efficiency and that the region from aa 98 to 101 in the CssB subunit might be responsible for cellular binding. With the available set of clinical data, we found that the CS6-expressing ETEC isolates with the BII subtype were found more frequently in the control group (71.4%), whereas the BI subtype was more common in the diarrhoeal cases (83.9%). Our study also indicated that CS6-expressing ETEC with the BII subtype showed decreased cellular binding compared with the BI subtype. However, this hypothesis should be further confirmed by the inclusion of an increased number of CS6-expressing ETEC isolated from both cases and controls.

Further studies are in progress to define the functional domains of CssB. We are also confirming our results using other intestinal cell lines, as well as determining the clinical and epidemiological significance of the CS6 subtypes. Nevertheless, the current results support the hypothesis that both subunits play a role in binding to intestinal (Caco-2) cells and that expression of AI or AIII together with the BI allele results in more pronounced binding. This information may help to improve the approach towards producing an effective vaccine against CS6-expressing ETEC.

**ACKNOWLEDGEMENTS**

This study was supported in part by grants from the Program of Founding Research Center for Emerging and Reemerging Infectious Diseases, Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Global Enteric Multicenter Study, University of Maryland, MD, USA. The authors thank Indranil Biswas for technical assistance.

**REFERENCES**


