Colonization Factors of Enterotoxigenic \textit{Escherichia coli} Isolated from Children in North India

Halvor Sommerfelt, Hans Steinsland,* Harleen M. S. Grewal, Gloria Inés Viboud,* Nita Bhandari, Wim Gaastra, Ann-Mari Svennerholm, and Maharaj K. Bhan

Colonization factor antigens (CFAs) mediate attachment of enterotoxigenic \textit{Escherichia coli} (ETEC) to the intestinal mucosa and induce protective immunity against ETEC diarrhea. ETEC strains \((n = 111)\) isolated from North Indian children from 1985 to 1989 were examined for CFAs and putative colonization factors (PCFs). CFA/IV was the most common factor \((26\%)\), followed by \textit{Escherichia coli} surface antigen 17 \((CS17)\) \((19\%)\), CFA/II \((14\%)\), PCFO166 \((7\%)\), and CFA/II \((5\%)\), while 24\% of the isolates were negative for CFAs and PCFs. Among the strains producing heat-stable and heat-labile toxin \((ST^\prime LT^\prime)\) strains, the STal gene was strongly associated with the absence of known CFAs and PCFs, making the STal-LT' isolates an interesting target for the identification of previously undescribed factors. Repetitive sequence–based polymerase chain reaction revealed that the CS17 strains, although clonally related, represented endemically circulating strains with a diversity greater than that of the CFA/I strains, which showed a substantial clonal clustering.

Enterotoxigenic \textit{Escherichia coli} (ETEC) is a major cause of diarrhea in children of developing countries. ETEC colonization factor antigens (CFAs) and putative colonization factors (PCFs), jointly referred to as colonization factors (CFs), mediate bacterial attachment to the intestinal mucosa \([1-3]\), while enterotoxins induce a net secretion of electrolytes and water into the gut lumen \([4-6]\). The toxins encountered in human ETEC diarrhea are the heat-labile enterotoxin \((LT)\) LTh and the heat-stable enterotoxins \((ST)\) STaI (or STh) and STal (or STp). In contrast to LT, STaI and STal are poor immunogens but may become immunogenic when coupled to appropriate carrier molecules \([7, 8]\). The toxoid of the inactivated ETEC vaccine \([9]\), CTB, elicits a substantial anti-LT immunity but leaves ST untargeted by the induced immune response. Even so, CTB confers a considerable, albeit short-lived protection against diarrhea associated with ST'TLT' ETEC \([10]\).

CFA/I is a single fimbrial structure \([11]\); CFA/II is composed of three distinct coli surface antigens (CSs) in various permutations. Thus, strains bearing CFA/II express CS3 alone or in combination with CS1 or CS2 fimbriae \([12]\). Likewise, CFA/IV strains express CS6 alone or together with CS4 or CS5 fimbriae \([2]\). During the last few years, several new CFs have been discovered, including PCFO159:H4, PCFO166, CS7, CS17, CFA/III, PCFO20, and very recently CS19 \([13-18]\).

Although O antigens may play a role in the protection against infection with ETEC of homologous O groups \([4]\), CFAs induce protective antibacterial immunity against strains with the same CFAs but with different O antigens \([1, 2, 5, 19, 20]\). Since ETEC carrying a particular CF are scattered over a variety of serogroups \([21-23]\) and immunity to a large extent is CF-specific \([1, 2, 4, 5, 19, 24]\), the most important of the somatic antigens to be included in a vaccine are the CFs that occur with high frequencies on ETEC in the target populations \([4]\). Detailed information on the distribution of ETEC carrying the various CFs in different geographic regions is accordingly a prerequisite for the design of effective vaccines \([1, 4, 5]\). This information is incomplete, especially with regard to the recently described PCFs. Reliable information on the distribution of CFs on ETEC isolated from children in India is lacking.

The proportion of ETEC expressing the different adhesins varies considerably from study to study. In an examination of ETEC isolates from Thailand and Bangladesh, strains producing CFA/I constituted a large proportion \((\sim 25\%)\) \([21]\). A high
percentage of the Bangladeshi strains produced CFA/IV (≈25%), whereas this CFA was uncommon among the Thai isolates [21]. Most studies indicate that CFA/IV as well as CFA/I are common CFs of ETEC [21–23, 25], although a Peruvian strain collection without CFA/I-producers has been described [22]. CFA/II* ETEC seem to be present in most strain collections, overall in a somewhat lower frequency (5%–15%) than other CFAs [21, 22, 26], although CFA/II was the most common CFA (27%) in a recent cohort study of children in Chile [27].

Similarly, the relative frequency of ETEC without identifiable CFs varies from study to study [22, 25, 28]. Such isolates, in the present report referred to as 0-ETEC, may either produce hitherto undescribed CFs or may have lost their ability to express known CFs [11, 21, 26, 29–33].

Using nucleic acid probes and monoclonal and polyclonal antibody (MAB)-based inhibition ELISAs and slide agglutination tests, we examined ETEC isolated from North Indian children during 1985–1989 for markers of CFA/I, CFA/II, CFA/III, CFA/IV, CS7, CS17, CS19, PCFOI66, PCFOI59:H4, and PCFO20 genes. The study is an attempt to delineate strategies for defining important ETEC CFs and to suggest antigens for inclusion in a future ETEC vaccine.

Materials and Methods

Subjects and bacterial isolates. The children, 1 month to 3 years of age, from whom ETEC were isolated participated in 6 different studies of diarrheal disease during 1985–1989 (table 1). Children participating in studies 1, 2, 4, and 5 were referred to the oral rehydration unit at the pediatric service of the All India Institute of Medical Sciences (AIIMS), which is a referral hospital for the city of New Delhi, which has a total population of ≈8 million. Study 3 was done in the village of Anaagpur-Palla, which is a community health center in Dakshinpuri, a periurban slum of ≈50,000 inhabitants, located 8 km from AIIMS.

For the 80 ETEC-positive subjects participating in studies 1–3, epidemiologic and microbiologic data were available. Each of the studies involved detection of enteric pathogens for ≥12 consecutive months. There was no prominent clustering in time or place of ETEC-associated diarrhea, as ETEC was isolated throughout the year, although somewhat more frequently during the hot season (April through September).

The most frequently isolated pathogens in studies 1–6 were rotavirus, ETEC, Salmonella and Shigella species, enteropathogenic E. coli, and Campylobacter jejuni [34–40]. The proportion of diarrheal episodes in which ETEC was isolated varied from 24% in study 1 to 12% in study 5. Of the 80 ETEC-positive stool samples obtained during studies 1–3 for which detailed microbiologic data were available [34–36], 21 were positive for rotavirus, while 2 samples contained salmonellae and 2 others (1 of which also contained rotavirus) yielded shigellosa. The remaining 56 specimens yielded ETEC as the only enteropathogen.

An ETEC isolate or strain was defined as an ETEC clone recovered from a child’s stool sample at a given time point. Two clones of ETEC with different toxin gene profiles obtained from the same specimen were considered distinct and were termed double isolates.

In total, 111 strains were examined, of which 98 were single isolates from 98 children and 4 were double isolates from 2 children, all with diarrhea. The remaining 9 isolates were obtained from 9 healthy children, 2 of whom also yielded single isolates during diarrheal illness at another time point. Thus, the 111 ETEC strains were isolated from 109 specimens obtained from 107 children, of whom 100 suffered from diarrhea at the time of ETEC isolation.

It may be argued that the specimens obtained from the 9 healthy controls and those containing pathogens that, in young children, show a higher pathogenicity and virulence than ETEC, such as Shigella species and rotavirus [6], should be excluded from a study addressing potential vaccine antigens. However, in endemic situations, passively transferred as well as acquired immunity that

---

Table 1. Sources and characteristics of enterotoxigenic E. coli (ETEC) isolates.

| Study no. | Study period (mo/year) | Type of study | Subject category | No. of subjects studied | Age (mo) of ETEC infection | No. of specimens/
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 [34]</td>
<td>4/86–9/87</td>
<td>Cross-sectional/hospital-based</td>
<td>Hospitalized urban boys with AD</td>
<td>92</td>
<td>3–24</td>
<td>10</td>
</tr>
<tr>
<td>3 [36]</td>
<td>4/85–4/86</td>
<td>Prospective/community-based</td>
<td>Rural children with AD or PD and healthy controls</td>
<td>452</td>
<td>1–31</td>
<td>11.5</td>
</tr>
<tr>
<td>5 [38, 39]</td>
<td>2/85–7/86</td>
<td>Cross-sectional/hospital-based</td>
<td>Urban boys with PD</td>
<td>124</td>
<td>4–23</td>
<td>9</td>
</tr>
<tr>
<td>1–6</td>
<td>11/84–8/89</td>
<td></td>
<td></td>
<td>1220</td>
<td>1–36</td>
<td>10</td>
</tr>
</tbody>
</table>

NOTE. AD, acute diarrhea; PD, persistent diarrhea; ref, reference; mo, month.

* 35 single isolates and 1 double isolate.

† 14 single isolates and 1 double isolate.
prevent disease but not colonization play a central role in the relatively low pathogenicity of many established enteropathogens, including *Vibrio cholerae* and ETEC [4–6, 24, 41]. Thus, ETEC isolated from healthy controls and from children with other pathogens are likely to represent the isolates circulating in the population under study and were therefore included in the analyses.

In the description of the distribution of toxin and CF gene markers in ETEC, the ETEC strain was used as the observational unit. Thus, all 107 single isolates and both clones of each of the 2 double isolates were included. In contrast, in the analyses of the relationships between toxin and CF gene markers on one hand and epidemiologic data on the other, the specimen was used as the observational unit. Accordingly, all 107 single isolates but only the first clone of each of the 2 double isolates were included.

Detection of toxin and CF genes and phenotypic identification of CFs in ETEC. The ETEC isolates were identified and classified with regard to their toxin gene profiles using previously described colony hybridization assays [42]. The strains were stored in stab agars at room temperature for up to 2 years before they were frozen at −70°C suspended in Greave’s solution (50 g of bovine serum albumin, 50 g of sodium glutamate, 100 mL of glycerol/L, pH 6.9) at ~10^8 bacteria/mL.

The isolates were categorized using nonradioactive hybridization assays for CFA/I, CS1, CS2, CS4, CS6, PCFO166, CS17, and CS19 genes [18, 43, 44]. In parallel, the products of these structural genes were identified phenotypically using highly specific MAb and polyclonal antibody—based inhibition ELISAs, slide agglutination, and immunoblotting techniques [18, 25, 28, 44, 45]. Because hybridization assays for CS3, CS5, CS7, PCFO159:H4, and PCFO20 genes were lacking, these factors were identified only phenotypically [17, 25, 28]. A positive CF hybridization or immunodetection test was considered to indicate the presence of the corresponding structural CF gene.

In ETEC, spontaneous loss of enterotoxigenicity is often accompanied by the loss of CF structural or regulatory genes during storage and repeated subcultures [21, 26, 29, 30, 33]. Accordingly, the clones that were tested for CF gene markers were simultaneously reexamined for the presence of toxin genes. If the examined clones showed evidence of loss of toxin gene, additional clones with the original toxin profiles were revied and reexamined for markers of CF genes.

Genotyping. To examine the clonal relatedness among ETEC strains with the two most prevalent single CFs, analysis using repetitive sequence–based polymerase chain reaction (rep-PCR) with two sets of primers, REP1R/REP2-I and ERIC1R/ERIC2 [46], was undertaken. ETEC strains with the CFs in question but isolated in another country were included for comparison. Rep-PCR was also used to examine O-ETEC strains with toxin profiles associated with the absence of CF genes for their relatedness to strains expressing CFs belonging to the CFA/I-CS1-CS2-CS4-PCFO166-CS17-CS19 family of fimbrial proteins [18].

A simplification of the whole cell bacterial lysate rep-PCR was used after we verified that the method yielded results identical to those of rep-PCR of chromosomal DNA isolated and purified by standard methods [46]. Thus, overnight culture was bypassed, and the bacteria stored at −70°C were used as the source of DNA. Bacterial suspension in Greave’s solution (1 µL) was added to a mixture of 18.7 µL of distilled water, 5 µL dimethyl sulfoxide, and 10 µL of 5X reaction buffer (83 mM ammonium sulfate, 335 mM TRIS-HCl, 33.5 mM magnesium chloride, 33.5 µM EDTA, 150 mM β-mercaptoethanol, and 850 µg/mL bovine serum albumin, pH 8.8) and kept at 98°C for 15 min in an automated DNA thermal cycler (Perkin-Elmer Cetus, Norwalk, CT). Thereafter, the tubes were chilled on ice and centrifuged at 15,000 g for 1 min.

While the tubes were kept on ice, 15.3 µL of a second solution containing the primers (100 pmol of each), the dNTPs (62.5 pmol of each: dATP, dCTP, dGTP, and dTTP; Pharmacia Biotech, Piscataway, NY), and Taq polymerase (2 U; Perkin-Elmer Cetus) was added. PCR amplifications were done with an initial denaturation (4 min at 95°C), followed by 31 cycles of denaturation (1 min at 95°C), annealing (1 min at 40°C and 52°C for the REP1R/REP2-I and the ERIC1R/ERIC2 primer pairs, respectively), and extension (8 min at 65°C; the final extension step lasted 16 min). Negative control reactions, in which 1 µL of water was added instead of the bacterial suspension, were included for each rep-PCR experiment. To fractionate the PCR amplification products, electrophoresis was done in 1.5% agarose gels, visualizing 10–24 ethidium bromide–stained DNA fragments.

The degree of genetic identity between 2 strains was defined as the number of identically sized bands divided by the total number of bands amplified from both strains, taking the REP1R/REP2-I and the ERIC1R/ERIC2 PCR products into account. Accordingly, in an instance in which 2 strains yielded an identical 15-bp REP1R/REP2-I PCR pattern, while ERIC1R/ERIC2 PCR showed that 2 of 15 bands from 1 strain were absent in the amplification products of the other strain but the remaining 13 bands were identical, the genetic identity was calculated to be [(15 + 15) + (13 + 13)]/(15 + 15) + (13 + 13) = 0.97 or 97%. Bacterial isolates were said to belong to a single clone (100% identity) only if REP1R/REP2-I and ERIC1R/ERIC2 PCR yielded uniform patterns. Strains were considered to belong to the same clonal cluster when the identity was >75%, while ≤75% identity was interpreted to indicate lack of clonal relatedness, analogous to the principles described for distinguishing bacteria on the basis of gel migration patterns of restriction enzyme–digested chromosomal DNA [47].

Serotyping. Serotyping was done by F. Scheutz at the International Escherichia coli Laboratory, Statens Seruminstitut, Copenhagen, on a subset of the CS17* strains to look for a possible relationship between rep-PCR patterns and serotypes.

Statistical analyses. Yates’s corrected χ² or Fisher’s 2-tailed exact test, as appropriate, was used to compare frequencies of CF gene markers in different groups of ETEC isolates. McNemar’s test for correlated proportions was used to compare the sensitivities of genotypic and phenotypic CF identification methods [48].

Results

The phenotypic and genotypic assays showed a complete concordance in the identification of markers for CFA/I, CS1, CS2, CS4, CS17, and CS19 genes (table 2). On the other hand, only 5 of 8 strains that were positive in the PCFO166 hybridization assay were positive in the PCFO166 ELISA (P = .25), while only 6 of 12 CS6 gene probe–positive strains were positive in the CS6 ELISA (P = .03).
down of all markers for enterotoxin and CF genes is shown in table 2. Overall, the most common factor was CFA/IV (26%, 29/111), followed by CS17 (19%, 21/111), CFA/I (14%, 16/111), PCFO166 (7%, 8/111), and CFA/II (5%, 5/111), making CS17 the most common single CF. Of the 111 isolates, 27 (24%) were O-ETEC. Of these, 52% (14/27) were ST-LT+ and 40% (11/27) were STaI-LT+; the remaining 2 isolates were STaII-LT- and STaII-LT-.

The proportion of O-ETEC was significantly higher among ST-LT+ (36%, 11/30) and ST-LT+ (31%, 14/45) than among ST-LT- isolates (6%, 2/34; \( P = .004 \) and \( .01 \), respectively). Among the ST-LT+ strains, the STaI-LT+ profile was strongly associated with the absence of markers for known CFs (\( P < .001 \)). The 11 STaI-LT+ O-ETEC strains were isolated during studies 2–6, while the 18 STaII-LT+ isolates, all with markers for known CF genes, were isolated during studies 1–5 (table 2).

The distributions of CF gene markers in ETEC isolates obtained during the studies that contributed 15 ETEC+ specimens are compared in table 3. Isolates with CFA/I markers and O-ETEC were strongly represented, CFA/I being the most frequent CF in ETEC strains obtained during study 1 and, along with CFA/I, in study 2. CS17, on the other hand, was the most common CF in ETEC isolates of studies 3 and 5, while only a single CS17 strain was isolated from children participating in study 1 or 2.

Of the 13 STaII-LT+ CFA/I-producers, 12 strains, isolated during April 1986 to May 1989 from children participating in studies 1–3, yielded the same rep-PCR pattern and were accordingly considered to represent a single clone. The remaining strain (201 C) isolated from a child in study 1 showed a 98% identity to this dominating pattern. The 13 strains were all clonally distinct from the STaII-LT+ CFA/I strain 325542-1, isolated in 1980 in Bangladesh [26]. The 3 STaI-LT+ CFA/I strains yielded a single rep-PCR pattern, which was distinct from that of the 13 STaII-LT+ CFA/I isolates. Again, an STaII-LT+ CFA/I strain (252582-1), isolated in 1980 in Bangladesh [26], was clonally unrelated to the 3 Indian isolates.
Table 3. Distribution of colonization factor (CF) gene markers of enterotoxigenic E. coli from studies 1, 2, 3, and 5.

<table>
<thead>
<tr>
<th>CF</th>
<th>No. (%) of isolates from study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 22)</td>
</tr>
<tr>
<td>CFAlIV</td>
<td>9 (40.9)</td>
</tr>
<tr>
<td>CS7</td>
<td>1 (4.6)*</td>
</tr>
<tr>
<td>CFAlI</td>
<td>5 (22.7)</td>
</tr>
<tr>
<td>PCFO166</td>
<td>2 (9.0)</td>
</tr>
<tr>
<td>CSAII</td>
<td>1 (4.6)</td>
</tr>
<tr>
<td>CSA7</td>
<td>1 (4.6)</td>
</tr>
<tr>
<td>CSA9</td>
<td>0 (0)</td>
</tr>
<tr>
<td>None</td>
<td>3 (13.6)</td>
</tr>
</tbody>
</table>

NOTE: 1 isolate was included per specimen.

Yates’s corrected $\chi^2$ or Fisher’s 2-tailed exact test, as appropriate, was used to compare frequencies of CF gene markers between studies. $P$ = .05 except for following: * vs. §, $P = .032$; ** vs. §, $P = .017$.

Rep-PCR showed that the 21 CS17 strains were clonally unrelated to the CFA/I-producers and that 14 strains, isolated between April 1985 and April 1986 during studies 3–5, represented a single clone. The remaining 7 strains, isolated during studies 1, 3, 4, and 5, represented 6 separate clones, except for 1 within the same clonal cluster, and had a genetic identity to the dominating CS17 clone ranging from 90% (strains 435C and 316C from studies 1 and 5, respectively) to 97%. Strain F595C, obtained during study 3, showed only a 58% identity to the main pattern and was therefore considered to be clonally distinct (figure 2). The CS17 strains of the present study were all clonally unrelated to the reference CS17 strain E20738A, which was isolated in Zaire in 1980 [22].

The O-ETEC strains with toxin gene profiles associated with the absence of CF gene markers (i.e., ST+LT+ and ST-LT+) were scrutinized for rep-PCR patterns matching those of CS17 strain F595C and the CS17 strains, respectively. None of the 11 ST+LT+ strains yielded rep-PCR patterns resembling those of strain F595C; on the other hand, 2 of the 14 ST-LT+ strains yielded a rep-PCR pattern with a 97% identity to the main CS17 pattern (figure 3).

Of the CS17-producers, 4 of 14 strains with the main CS17 rep-PCR pattern and the 7 strains with other patterns were serotyped. Of these 11 strains, all but 2 were of serotype O8:H9; strains 316C and F222A belonged to serotypes O107:H27 and O146:H19, respectively.

Discussion

In accordance with the findings of most other investigations [22, 23, 25, 28], a considerable proportion (59%) of the isolates examined in the present study expressed ST, either alone (32%) or in combination with LT (27%). In children living in ETEC-endemic areas, ST+LT- and ST+LT+ ETEC seem to have a higher pathogenicity than ST-LT+ ETEC [27]. Accordingly, future ETEC vaccines may benefit from incorporating an ST toxoid [1, 4, 8].

We have previously shown that identifying the genetic potential to produce a certain CF rather than the expressed CF protein may improve sensitivity [30, 32, 44]. In the present study, colony hybridization showed a higher sensitivity than inhibition ELISA [28, 30, 43, 44] for the identification of ETEC isolates with PCFO166 and CS6 structural genes; the difference in sensitivity reached a level of statistical significance only for CS6. The detection of any additional silent CF genes must await the development of probes for CS3, CS5, CS7, PCFO159:H14, and PCFO20.

Since this is the first extensive report describing CS17 as the most common single CF of ETEC, overall second only to CFA/IV, epidemiologic and genetic analyses of the relatedness between the CS17 strains were undertaken to evaluate whether the high prevalence could be explained by a focal outbreak. The CS17 strains were collected during 4 different studies from children resident in New Delhi and in a village in the neighboring state of Haryana [36]. The CS17 strains obtained from the rural children in Haryana were collected over a 1-year period (April 1985 to April 1986). Furthermore, we have isolated CS17 ETEC from North Indian children with diarrhea after 1989 (data not shown). CS17 strains were, in other words,
Figure 3. ERIC1R/ERIC2 (lanes 2–4) and REP1R-1/REP2-1 (lanes 5–7) repetitive sequence–based polymerase chain reaction (rep-PCR) patterns of enterotoxigenic E. coli (ETEC) strain F46A (lanes 2, 5), which represents main coli surface antigen 17 (CS17) rep-PCR pattern. Two O-ETEC strains, F5A (lanes 3, 6) and F14C (lanes 4, 7), yielded rep-PCR pattern that was 97% identical to that of main CS17 pattern, suggesting that these ST-LT+ strains represent previous CS17-producers that had lost their structural CS17 genes. Lane 1: Haelll-digested ϕX174 DNA molecular size marker; fragment sizes in bp are shown at left. DNA fragments were separated on 1.5% agarose gel and stained with ethidium bromide.

present in the child population living in urban New Delhi and rural Haryana over a period of >5 years.

Moreover, the CS17 strains from the Haryana study yielded three different rep-PCR patterns and altogether 7 different rep-PCR patterns were observed among the 21 CS17-producers. Accordingly, we conclude that, although there was a considerable clonal relatedness between the CS17 strains, they do not stem from a single outbreak but represent endemically circulating strains infecting children in the study populations.

The considerable clonal clustering among the CS17 strains was matched by an even greater homogeneity among the 16 CFA/I ETEC strains, yielding only three different rep-PCR patterns. CFA/I ETEC are an established cause of endemic childhood diarrhea in developing countries.

A single clone constituted 67% (14/21) of the CS17 ETEC strains. Similarly, 92% (12/13) of the STall-LT+ CFA/I isolates yielded a uniform rep-PCR pattern despite being isolated over an extended period and from children in different sites. This suggests that, even in a typical endemic situation, a few strains have a considerable ecologic advantage over others. Although both CFA/I and CS17 are plasmid-encoded [11, 18] and the CFA/I operon is flanked by insertion sequences [49], the rather remarkable clonal clustering indicates that the exchange of CFA/I- and CS17-encoding sequences between strains is limited. This may be a reflection of the non–self-transmissible nature of the large plasmids harboring the CF genes [50] or a recipient factor that may restrict horizontal plasmid [51] or transposon-mediated gene transfer (or both).

The 4 isolates representing the dominant CS17 clone belonged to the most common serotype of CS17 ETEC, O8:H9 [22]. Of interest, 1 of the 2 strains with a 90% identity to the main rep-PCR pattern and the genetically distinct strain F222A were of different serotypes, while the remaining 5 CS17 strains that did not represent the main clone were of serotype O8:H9. In sum, isolates belonging to different serotypes yielded distinct rep-PCR patterns, while diverse rep-PCR patterns could be obtained from strains with the same serotype. Although a considerably larger number of isolates needs to be examined, this indicates that rep-PCR may have a larger discriminatory power than serotyping. Incidentally, to our knowledge, this is the first report of an O107:H27 strain expressing CS17.

The strong association between the STaI gene and the lack of markers for known CF genes among the STall+LT+ strains could have been explained by outbreaks caused by a limited number of STall+LT+ and STaI+LT+ clones. Keeping in mind the strong association between the presence of enterotoxin and CF genes on one hand and clonal clustering on the other for CS17 and CFA/I ETEC, the 18 STall+LT+ isolates, with genes encoding four different CFs, represent at least four distinct clones. As can be seen from table 2, the 11 STaI+LT+ O-ETEC strains were obtained during 5 studies, differing in time, place, or both (table 1).

Analysis using rep-PCR has recently shown that the strains represent 7 different clones (Valvatne H, Sommerfelt H, unpublished data). Not counting CS19 as an established CF, since it was discovered during the present study, and strain F595C yielding a distinct rep-PCR pattern, none of the 8 STaI+LT+ clones had markers for established CF genes. Using this conservative approach, there is still a highly significant difference in the proportion of STaI+LT+ and STaI+LT+ ETEC clones that lacked markers for established CF genes (0/8 and 4/4, respectively, $P = .002$). We conclude that the search for undescribed or unexpressed (or both) CFs could benefit from focusing on STaI+LT+ isolates. As other epidemiologic studies of ETEC CFs have not distinguished between the STaI and STaII genes [22, 25, 28], it is not possible to infer that this approach is appropriate in other settings.

The strategy of searching for previously undescribed PCFs among ETEC with a specific toxin profile is not unique. Thus, McConnell et al. [52] examined ST-LT+ ETEC for PCFs, thereby identifying CS17. Our efforts to unearth "new" PCFs on STaI+LT+ O-ETEC recently resulted in the identification and characterization of CS19 on the O8:H9 ETEC strain F595C [18], the first PCF to be described on an ETEC isolate reported to be STaI+LT+. Among previously described 0-ETEC, 1 strain...
matches both the toxin profile and serotype of strain F595C [23], while 9 ST-LT + strains match the serotype but not the toxin profile [22], thereby suggesting that CS19 may be more widespread than the present study indicates. Moreover, the recent discovery of CS20, yet another PCF from the pool of STaI*LT+ 0-ETEC strains (Valvatne H, Sommerfelt H, unpublished data) [53], supports our suggestion of an association between the STaI gene and new PCFs in ST+LT+ ETEC.

In previous studies of human ETEC, in which a differentiation between STaI and STaII genes is reported, STaI+ ETEC are found less frequently than STaII+ ETEC [54–56]. Two of these studies report a lower isolation frequency of STaI+LT+ strains than the one presented here. Moreover, the percentage is also high in other studies of ETEC CFs, ranging from 33% [25, 28] to 84% [22]. Two of our ST-LT + strains yielded a typical CS17 rep-PCR pattern, suggesting that they may have been derived from CS17-producers, analogous to what was recently found for clusters of 0-ETEC strains harboring genes encoding PCFO159:H4 or CS20 (Valvatne H, Sommerfelt H, unpublished data). Many of the CF genes are linked to the ST gene and are subject to spontaneous loss during storage and subculturing in the laboratory [21, 29, 33, 44]. Most probably, a certain proportion of the ST-LT + isolates represent ST-LT + strains that have spontaneously lost their CF genes along with their ST genes. In fact, a 0-ETEC derivative from 1 of the ST-LT + CFA/I strains was found to be ST-LT + (data not shown), suggesting that the structural CFA/I gene was lost along with the ST gene [29]. Compared with the STaI*LT+ strains, the ST-LT + 0-ETEC strains may accordingly constitute a somewhat less interesting pool for the search of previously undescribed CFs.

The observed differences in CF distributions between studies 1, 2, 3, and 5 demonstrate that the choice of CFs to be included in an ETEC vaccine for a region should preferably not be based on analyses of more or less arbitrarily derived ETEC strain banks [21, 22]. Even prospectively collected samples from limited study populations, such as those of study 3 and of other cohort studies [27], may not be representative of the populations at large. Thus, a regional surveillance of several years' duration, ensuring adequate sampling procedures of the target population, may be required for defining candidate vaccine antigens.

The oral inactivated ETEC vaccine currently undergoing phase I and II trials [9] contains important CFAs of this and of several earlier studies [21–23, 25, 27, 28], namely, CFA/I, CFA/II, and CFA/IV. On the basis of the reports that the PCFs have been represented to a limited and variable extent in available ETEC strain collections [22, 25, 28] and of a lack of evidence of PCF-mediated colonization and induction of protective immunity in the intestine, the PCFs were not included in the candidate ETEC vaccines [9]. A recent animal study showed that several of the PCFs, including PCFO159:H4, CS7, CFA/III, and CS17, are powerful immunogens and strongly suggest their role as true colonization factors [3]. Not counting the 0 ST-LT + 0-ETEC strains with a typical CS17 rep-PCR band, 19% (21/111) of the isolates in the current study were CS17-positive; CS17 was also the most common adhesin antigen among a batch of ETEC isolates from Central Africa [22].

In a recent cohort study of Chilean children, Levine et al. [27] found that the LT-only genotype was more frequently encountered than the ST-LT or the ST-only genotype. If further examination of these Chilean ST-LT + strains reveals a considerable proportion of CS17-producers, as was shown by Viboud et al. [28], among Argentinean ST-LT + isolates, and if future studies show that CS17 elicits protective immunity, its inclusion in future vaccine preparations should be considered.

Acknowledgments

We thank Gunnar Kvåle and Rune Nilsen (Centre for International Health), Dag Helland (Laboratory for Biotechnology), and Roald Matre and Tor Hofstad (Department of Microbiology and Immunology, Gade Institute), University of Bergen, for providing excellent laboratory and computer facilities; Rune Male, Rein Aasland, Bjarne Bjorvatn, and Karl-Henning Kalland for constructive advice; Mari Fj Jonzen for skillful technical assistance; Moiyra M. McConnell for providing strain E20738A; and James Versalovic, Håvard Valvatne, and Thearith Koeuth for valuable advice with respect to the rep-PCR.

References


