Prevalence of enteroaggregative *Escherichia coli* and its virulence-related genes in a case–control study among children from north-eastern Brazil

Ila Fernanda Nunes Lima,1 Nadia Boisen,2 Josiane da Silva Quetz,1 Alexandre Havit,1 Eunice Bobo de Carvalho,1 Alberto Melo Soares,1 Noélia Leal Lima,1 Rosa Maria Salani Mota,3 James P. Nataro,2 Richard Littleton Guerrant1,2 and Aldo Ângelo Moreira Lima1,2

Correspondence
Ila F. N. Lima
ilafarm@yahoo.com.br

Received 25 October 2012
Accepted 13 February 2013

1Institute of Biomedicine for Brazilian Semi-Arid and Clinical Research Unit/Center for Global Health, School of Medicine, Federal University of Ceara, Fortaleza, CE, Brazil
2Center for Global Health, Division of Infectious Diseases and International Health, Departments of Medicine and Pediatrics, School of Medicine, University of Virginia, Charlottesville, VA, USA
3Department of Statistics and Applied Mathematics, Science Center, Federal University of Ceara, Fortaleza, CE, Brazil

Enteraggregative *Escherichia coli* (EAEC) is an important agent that causes endemic and epidemic diarrhoeal diseases worldwide. Several EAEC virulence-related genes (VRGs) have been described but their role in the clinical outcome of infection is not completely defined. This study investigated the prevalence of EAEC and potential associations of its VRGs with risk of or protection from diarrhoeal diseases in children from urban communities in north-eastern Brazil. The case–control study included 166 children, who had their stools evaluated for the EAEC diagnostic genes (*aaiC* and *aatA*) using PCR. Positive samples were further analysed by multiplex PCR and identified 18 VRGs. EAEC was found in the same proportion in both groups (41 %). The plasmid-borne gene encoding a hexosyltransferase homologue (*capU*) was the most frequently detected (89.6 %), followed by dispersin protein (*aap*, 58.2 %) and EAEC *HilA* homologue (*eilA*, 57.8 %). The AAF/III fimbrial subunit (*agg3A*) gene was observed at lower frequency (1.5 %). Plasmid-encoded toxin (*pet*) or AAF/II fimbrial subunit (*aafA*) was associated significantly with disease. AAF/IV fimbrial subunit (*agg4A*) or hypothetical plasmid-encoded haemolysin (*orf61*) was detected significantly more in controls than in children with diarrhoea. In addition, one set of genes in combination, *aaiC* and *agg3/4C* but lacking *agg4A* and *orf61* was associated with diarrhoea cases; and another one, *orf61* in the absence of *pet* and *aafA*, was correlated with control children. These data confirm a high prevalence, endemicity and heterogeneity of EAEC strains in the developing urban areas of north-eastern Brazil. Statistical correlation between cases and controls was seen with either isolated or combined sets of genes, suggesting that the pathophysiology of EAEC infection involves a complex and dynamic modulation of several VRGs.

**INTRODUCTION**

Since its first description by Nataro *et al.* (1987) in a prospective study of paediatric diarrhoea in Santiago, Chile, enteraggregative *Escherichia coli* (EAEC) has been increasingly recognized as an agent of diarrhoea. EAEC is probably best known for its role in persistent diarrhoea in children living in developing countries, the context in which its pathogenesis was originally defined (Okeke & Nataro, 2001). However, it has emerged as an important pathogen in outbreaks of acute diarrhoea in developed (Huang *et al.*, 2006; Scavia *et al.*, 2008) and developing countries (Bueris *et al.*, 2007; Kermani *et al.*, 2010; Meng *et al.*, 2011; Pereira *et al.*, 2007), international traveller’s diarrhoea (Paschke *et al.*, 2011) and diarrhoea among patients with human immunodeficiency virus-related infection (Samie *et al.*, 2007).
The pathogenesis of EAEC infection is not fully understood. The major obstacle in identifying the mechanism of pathogenesis for this bacterium is the heterogeneity of strains. While some investigations confirm the clear association of EAEC with diarrhoea in some individuals (Meng et al., 2011; Okeke et al., 2000; Opintan et al., 2010), in many others, it appears to cause subclinical infection or only intestinal colonization (Bueris et al., 2007; Piva et al., 2003). EAEC has also been associated with chronic intestinal inflammation, leading to childhood malnutrition and growth impairment (Steiner et al., 1998).

In recent years several advances have been made in the determination of EAEC pathogenesis, including the development of in vivo models (Harrington et al., 2009; Roche et al., 2010), and the use of previously characterized in vitro models to investigate the effects of EAEC infection on intestinal cells (Strauman et al., 2010) and the repair of cell damage (Carvalho et al., 2012). The complete genome of the prototypical strain EAEC 042 has recently been published (Chaudhuri et al., 2010). Despite this progress, the disease pathophysiology remains obscure, as several candidate virulence-related genes (VRGs) have been described but are not present in all EAEC strains. The prevalence and association of these genes with diarrhoeal diseases are not well established, as they may vary by geographical location (Estrada-Garcia & Navarro-Garcia, 2012; Nataro & Kaper, 1998). A recent large-scale outbreak of diarrhoea and haemolytic uraemic syndrome infecting 4137 individuals and resulting in 54 deaths was caused by a Shiga toxin-producing EAEC O104:H4 with a distinct set of virulence factors (Frank et al., 2011; Muniesa et al., 2012; Scheutz et al., 2011). The importance of combinations of these virulence factors warrants further investigation.

Aggregative adherence fimbriae (AAFs) are the main mucosal adhesins of EAEC, of which at least four variants are known (Bernier et al., 2002; Boisen et al., 2008; Czeczulin et al., 1997; Nataro et al., 1992). The four structural subunits are respectively encoded by aggA (AAF/I), aafA (AAF/II), agg3A (AAF/III) and agg4A (AAF/IV), and they have been shown to be regulated by the transcriptional AraC/XylS activator aggregative adherence regulator (aggR). Pathogenesis and molecular studies suggest the presence of a package of aggR-regulated VRGs (<50 in number) encoded on either chromosomal or virulence plasmids (Dudley et al., 2006; Morin et al., 2010). Under the control of aggR is the anti-aggregation protein gene (aap), formerly known as the EAEC secreted protein U gene (spu), which is present on the pAA-plasmid. The aap gene encodes a secreted low-molecular-mass protein, called dispersin, that promotes dispersal of EAEC on the intestinal mucosa to establish new foci of infection (Sheik et al., 2002). Dispersin secretion is translocated via a system called the EAEC ABC transporter (aat), encoded by aatPABCD (Nishi et al., 2003). Also regulated by aggR is the chromosomal cluster termed the aggR-activated island (aai), aaiA-P, encoding a type VI secretion system (Dudley et al., 2006). Other putative EAEC virulence factors not regulated by aggR are the EAEC heat-stable toxin 1, encoded by the aggregative heat-stable toxin A gene (astA) (Savarino et al., 1993), and a set of toxins termed serine protease autotransporters of Enterobacteriaceae (SPATEs).

SPATEs can be organized phylogenetically into two classes. Members of class I are cytotoxic and include the plasmid-encoded toxin gene (pet) (Navarro-García et al., 1998) and its homologues, secreted autotransporter toxin gene (sat) and Shigella IgA-like protease homologue gene (sigA) (Guyer et al., 2000; Rajakumar et al., 1997). Members of class II SPATEs are non-cytotoxic and include the protein involved in colonization (pic), a mucinase that promotes intestinal colonization (Henderson et al., 1999) and Shigella extracellular protease (sepA), which appears to be involved in tissue invasion (Benjelloun-Touimi et al., 1995). Recently, sepA was associated with clinical illness in a case–control study from Mali (Boisen et al., 2012).

Other EAEC candidate VRGs include capU (cap locus that encodes a protein 50% identical to an rfbU-related lipopolysaccharide biosynthetic gene of E. coli O157:H7) (Czeczulin et al., 1999; Fujiyama et al., 2008), the regulator eilA (EAEC Hia homologue) (Sheik et al., 2006) and hypothetics orf3 (cryptic protein) and orf61 (plasmid-encoded haemolysin) (Boisen et al., 2012).

Molecular approaches, especially DNA-based tests involving PCR, have increasingly been applied to identify diarrhoeagenic E. coli, which was among the first pathogens for which molecular diagnostic techniques were developed (Nataro & Kaper, 1998). In this case–control study, we determined the prevalence of EAEC strains in children with or without diarrhoea, living in a poor urban area from Fortaleza, Ceara, north-eastern Brazil. We used PCR assays to amplify aggR-activated island C (aaiC) and EAEC ABC transporter A (aatA) genes. In addition, samples positive for EAEC were further analysed for the presence of homologous sequences to the putative VRGs associated with the EAEC pathotype.

METHODS

Study site and ethical clearance. The study was conducted in two urban communities in Fortaleza. Gonçalves Dias is a five-block area that houses 1826 people, and Parque Universitario is a community of 11018 inhabitants.

The research protocol was reviewed and approved by the Research Ethics Committee of the Federal University of Ceara, the National Commission on Ethics in Research of Brazil and the Institutional Review Board of the University of Virginia. A consent form was read and signed by all parents or guardians.

Study design. This was a case–control study in children 2–36 months old, as previously described by da Silva Quetz et al. (2010). Briefly, the study included children with (cases) and without (controls) diarrhoea in the last 2 weeks. Diarrhoea was defined as three or more liquid stools in a 24 h period. Stool samples were collected between March and July 2007, during the rainy season, the period which accounts for the higher incidence of diarrhoeal diseases.
in Fortaleza (Façanha & Pinheiro, 2005). A questionnaire with information regarding the socioeconomic and clinical conditions and the occurrence of diarrhoea was completed with the guardians. Stool samples were collected from all children at each household and they were transported within 4 h in ice boxes to the Infectious Diseases Laboratory, Institute of Biomedicine for Brazilian Semi-Arid & Clinical Research Unit/Center for Global Health, Federal University of Ceará. All the samples were aliquoted without dilution and stored at −80 °C until molecular tests were performed.

**DNA extraction.** Bacterial DNA was extracted from all stool samples using the QuickAmp DNA Stool Mini kit (Qiagen) according to the manufacturer’s instructions. The options of incubating the stools with lysis solution at 95 °C and removing the second washing solution in two steps were used for enhancing pathogen DNA extraction. DNA quality and quantity were checked by a spectrophotometric method (BioPhotometer; Eppendorf). Extracted DNA was stored at −20 °C until further use.

**PCR assays.** Molecular diagnosis of EAEC was performed by gene amplification of *aatC* (chromosomal gene) and *aatA* (plasmid gene), which had no apparent homologues within GenBank. Single PCR was performed using AmpliTaq Gold PCR Master Mix (Applied Biosystems), which contains Taq polymerase, dNTPs, MgCl2, and the appropriate buffer. Each PCR tube contained 25 μl reaction mixture composed of 12.5 μl of the master mix, 2.5 μl of each forward and reverse primer solution (in a final concentration of 200 nM), 1–2.5 μl of faecal DNA (this volume was DNA concentration-dependent) and nuclease-free water to complete the final volume. The PCR conditions were one cycle for 5 min at 95 °C, 35 cycles for 20 s at 95 °C, 20 s at 57 °C and 1 min at 72 °C; and a final extension step for 10 min at 72 °C in a thermal cycler (Bio-Rad Laboratories). Bands were visualized and photographed (Chemidoc XRS; Bio-Rad Laboratories) after electrophoresis of an ethidium bromide-stained 1.2 % agarose gel in 1X Tris-acetate-EDTA-buffer. The primers used are described in Table 1. Only the presence of the correctly sized PCR product was interpreted as a positive result. One sample was considered positive for EAEC when it presented one or both researched genes.

All EAEC-positive samples were analysed by additional PCR with the primers and conditions described in Table 1 to amplify fragments of 18 genes encoding putative virulence factors, which were divided into four multiplex reactions. All multiplexes were performed as described by Boisen et al. (2012). PCR products were separated on 2 % pre-cast agarose gels (Life Technologies).

The following EAEC strains were used as positive controls: JM221 (*aggA, sat*) (Mathewson et al., 1987), 042 (*aatA, aggR, aatC, aap, orf3, pic, pet, aatA, aofA, aofC, capU, elIA*) (Nataro et al., 1985), 55909 (*aggA, agg3/4C*) (Bernier et al., 2002), H223-1 (*sigA*) (Czeckulin et al., 1999) and C1010-00 (*aggA, agg3/4C, sat, sepA*) (Olesen et al., 2005). *E. coli* MC1061 (Clermont et al., 2000) and HS (Levine et al., 1978), and water were used as negative controls.

**Data analysis.** Data were typed in duplicate and analysed by two independent investigators using Microsoft Office Access software (Microsoft Corporation). Classification and regression tree (CART software, pro version 6.0; Salford Systems) was used inputting 18 factors of interest as binary (present/absent) independent predictive variables. Case/control status was the binary dependent outcome variable. Chi-squared, Fisher’s exact, Mann–Whitney and odds ratio (OR) tests in addition to logistic regression analysis were used to compare data derived from case and control children depending on data requirements. Statistical analyses were performed using Statistical Package for Social Sciences (SPSS software, version 11.0) and GraphPad Prism (GraphPad software, version 5.01). P-values of ≤0.05 were considered statistically significant.

**RESULTS**

**Characteristics of the study population**

A total of 325 children were screened for enteric pathogens, 43 inhabitants of Gonçalves Dias and 282 of Parque Universitario. Children were from 295 houses, and 268 of those (90.8 %) had one child per house. Only two houses (0.7 %, one in each community) had more than three children. Study households had a median of 5.5 persons who slept in 1.9 rooms and lived in four compartments. Most dwellings were permanent structures built of brick and adobe (98.3 %), supplied with piped water (85.5 %), with an interior flush toilet (83.0 %). About 42.1 % of these families received up to US$ 175.00 per month.

Of the 325 children, 172 (52.9 %) were male and 219 (67.4 %) were more than 12 months of age [median 18 months; only 37 (11.4 %) were between 2 and 6 months of age]. Criteria for inclusion in the case group were met in 83 children (25.5 %). Age-matched controls were randomly selected among the other 242 children of the study. Thus, we analysed data from 83 case children and 83 controls.

**Prevalence of EAEC**

At least one of the two EAEC diagnostic genes analysed in this study was found in 34 (41.0 %) children of the case group. A similar prevalence of EAEC was obtained among children without diarrhoea (34/83, 41.0 %, *P*=1.000, by Pearson’s chi-squared). In total, 68 samples were positive for EAEC and three different patterns of diagnostic genes were observed: (i) only positive for *aatC*, (ii) only positive for *aatA* or (iii) positive for both *aatC* and *aatA*. Of the two EAEC diagnostic genes, *aatC* was the most common. It was detected in 33 (39.8 %) of 83 cases, of which 18 (21.7 %) children presented only this gene and 15 (18.1 %) were positive for both *aatC* and *aatA* genes. Only one (1.2 %) case was positive only for the *aatA* gene. Fifteen (18.1 %) and three (3.6 %) children from the control group had positive PCR for *aatC* or *aatA*, respectively, and 16 (19.3 %) children were positive for the two diagnostic genes. Neither was associated with the case or control groups, as shown in Table 2.

**Individual frequencies of VRGs**

To investigate if other EAEC VRGs could be associated with disease, we performed four multiplex PCR assays targeting 18 genes of 33 cases and 34 controls positive for this bacterium. Individual frequencies of each gene are given in Table 2. Results from amplification reactions of chromosomal and pAA-encoded genes showed that all samples carried at least one of the 18 assayed VRGs. Several genotypes were found in this population. The plasmid-borne gene encoding the hexosyltransferase homologue *capU* was the most frequently detected (89.6 %), followed by *aap* (58.2 %), *elIA* (57.8 %), *aggR* (56.7 %) and hypothetical *orf3* (50.7 %).
Table 1. Description of genes, GenBank accession numbers, primer sequences, size of the obtained products, PCR conditions of the genes used for diagnosis of EAEC and its VRGs

For all PCR conditions, an initial denaturation step (5 min for single PCR, 15 min for multiplex PCR 1, and 2 min for multiplexes 2–4 at 95 °C) and a final extension (10 min at 72 °C) were performed.

<table>
<thead>
<tr>
<th>Target gene (GenBank accession no.)</th>
<th>Type of PCR</th>
<th>Primer sequence (5′→3′)</th>
<th>Amplicon size (bp)</th>
<th>PCR conditions (35 cycles)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnostic genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aaiC – aggR-activated island (FN54766.1)</td>
<td>Single</td>
<td>ATTTGCTCTCAGGCAATTCACACG</td>
<td>215</td>
<td>20 s at 95 °C, 20 s at 57 °C, 1 min at 72 °C</td>
<td>Designed for this study</td>
</tr>
<tr>
<td>aatA – anti-aggregation protein transporter (AY351860)</td>
<td>Single</td>
<td>CTTGGCGAAAGACTGTATCATC</td>
<td>630</td>
<td>20 s at 95 °C, 20 s at 57 °C, 1 min at 72 °C</td>
<td>Schmidt et al. (1995)</td>
</tr>
<tr>
<td><strong>Virulence genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>astA – aggregative heat-stable toxin A, EAST1 (L11241)</td>
<td>Multiplex 1</td>
<td>ATGCCCATACACAGTATATGC</td>
<td>110</td>
<td>60 s at 94 °C, 1.5 min at 58 °C, 1.5 min at 72 °C</td>
<td>Mohamed et al. (2007)</td>
</tr>
<tr>
<td>pet – plasmid-encoded toxin (AF056581)</td>
<td>Multiplex 1</td>
<td>GGCCACAGAATAAGGCTGTTTT</td>
<td>302</td>
<td>60 s at 94 °C, 1.5 min at 58 °C, 1.5 min at 72 °C</td>
<td>Restieri et al. (2007)</td>
</tr>
<tr>
<td>sigA – Shigella IgA-like protease homologue (NC_004337)</td>
<td>Multiplex 1</td>
<td>CCGACTTCTCAGCTTCCGACC</td>
<td>430</td>
<td>60 s at 94 °C, 1.5 min at 58 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2009)</td>
</tr>
<tr>
<td>pic – protein involved in colonization (AF097644)</td>
<td>Multiplex 1</td>
<td>ACTGGATCTTAAAGGCTACGACT</td>
<td>572</td>
<td>60 s at 94 °C, 1.5 min at 58 °C, 1.5 min at 72 °C</td>
<td>Restieri et al. (2007)</td>
</tr>
<tr>
<td>sepA – Shigella extracellular protease (Z48219)</td>
<td>Multiplex 1</td>
<td>GCACAGTCAATATGTCAGGTTTT</td>
<td>794</td>
<td>60 s at 94 °C, 1.5 min at 58 °C, 1.5 min at 72 °C</td>
<td>Restieri et al. (2007)</td>
</tr>
<tr>
<td>sat – secreted autotransporter toxin (AE014075)</td>
<td>Multiplex 1</td>
<td>TCAAGAATGCTCTGATATGATGCTTACGACC</td>
<td>932</td>
<td>60 s at 94 °C, 1.5 min at 58 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2009)</td>
</tr>
<tr>
<td>orf2 – cryptic protein (FN54767.1)</td>
<td>Multiplex 2</td>
<td>CATACACTGCGATTTCTCGCATCAGG</td>
<td>121</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>aap – anti-aggregation protein, dispersin (Z12523)</td>
<td>Multiplex 2</td>
<td>GGTAGGTGGTGGGTATGTAGACG</td>
<td>426</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>aggR – aggregative adherence regulator (Z18751)</td>
<td>Multiplex 2</td>
<td>ATCTTGTGATAGCTGACTGACG</td>
<td>426</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>agg4A – AAF/IV fimbrial subunit (EU637023)</td>
<td>Multiplex 3</td>
<td>TCTACCTGCTGCTGACG</td>
<td>169</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>agg2A – AAF/I fimbrial subunit (Y18149, AY344586)</td>
<td>Multiplex 3</td>
<td>CTACCTGCTGCTGACG</td>
<td>220</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>asfA – AAF/II fimbrial subunit (AF012835)</td>
<td>Multiplex 3</td>
<td>CTCCTGCTGCTGACG</td>
<td>289</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>agg3A – AAF/III fimbrial subunit (AF411067)</td>
<td>Multiplex 3</td>
<td>CACGGTATATGCTGACG</td>
<td>370</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>agg3A/4Ct – usher, AAF/III-IV assembly unit (AF411067, AB255435, EU637023)</td>
<td>Multiplex 3</td>
<td>TCTACCTGCTGCTGACG</td>
<td>409</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
</tbody>
</table>
Table 1. cont.

<table>
<thead>
<tr>
<th>Target gene (GenBank accession no.)</th>
<th>Type of PCR</th>
<th>Primer sequence (5’→3’)</th>
<th>Amplicon size (bp)</th>
<th>PCR conditions (35 cycles)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>aafC – usher, AAF/II assembly unit (AF114828)</td>
<td>Multiplex 3</td>
<td>ACAGGCTTGCGGTAAAAGCGC TTACGGGTACGAGTTTTACGG</td>
<td>491</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>orf61 – plasmid-encoded haemolysin (FN554767.1)</td>
<td>Multiplex 4</td>
<td>AGCTCTGGAAACTGGCCTCTA ACGTCCTGATTTCTGCTT</td>
<td>108</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>eilA – Salmonella HilA homologue (FN554766.1)</td>
<td>Multiplex 4</td>
<td>AGGTCTGGAGCGCGAGTGTT GTAAAACGGTATCCACGACC</td>
<td>248</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
<tr>
<td>capU – hexosyltransferase homologue (AF134403)</td>
<td>Multiplex 4</td>
<td>CAGGCTGTTGCTCAAATGAA GTTCGACATCCTTCCTGCTC</td>
<td>395</td>
<td>50 s at 94 °C, 1.5 min at 57 °C, 1.5 min at 72 °C</td>
<td>Boisen et al. (2012)</td>
</tr>
</tbody>
</table>

*Two forward primers and one reverse primer were used for the amplification of agg3/4C. This primer set was designed to amplify the usher gene from both AAF/III and IV.

Table 2. Prevalence of enteroaggregative E. coli (EAEC) and its virulence-related genes in stool samples from case and control children

The aaiC and aatA genes were used for EAEC diagnosis. Samples were considered positive for EAEC when they presented one or both genes. Comparison of gene frequencies between case and control groups was performed by chi-squared using 2 × 2 contingency tables. Fisher’s exact test was applied when the observed frequency was < 5. The total number of samples (83 cases and 83 controls) was considered for analysis of diagnostic genes. For statistical tests of the virulence genes, only EAEC-positive samples were eligible (33 cases and 34 controls). One positive case did not contain a sufficient amount of DNA sample to perform the tests.

<table>
<thead>
<tr>
<th>EAEC genes</th>
<th>No. of cases (%)</th>
<th>No. of controls (%)</th>
<th>Total no. (%)</th>
<th>OR</th>
<th>Risk estimate (95% CI)</th>
<th>( \chi^2 )</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaiC*</td>
<td>33 (39.8)</td>
<td>31 (37.3)</td>
<td>64 (38.5)</td>
<td>1.11</td>
<td>0.59–2.07</td>
<td>0.10</td>
<td>0.750</td>
</tr>
<tr>
<td>aatA</td>
<td>16 (19.3)</td>
<td>19 (22.9)</td>
<td>35 (21.1)</td>
<td>0.80</td>
<td>0.38–1.70</td>
<td>0.33</td>
<td>0.568</td>
</tr>
<tr>
<td>aggR</td>
<td>17 (51.5)</td>
<td>21 (61.8)</td>
<td>38 (56.7)</td>
<td>0.66</td>
<td>0.25–1.74</td>
<td>0.72</td>
<td>0.397</td>
</tr>
<tr>
<td>aap</td>
<td>18 (54.5)</td>
<td>21 (61.8)</td>
<td>39 (58.2)</td>
<td>0.75</td>
<td>0.28–1.97</td>
<td>0.36</td>
<td>0.549</td>
</tr>
<tr>
<td>orf3</td>
<td>18 (54.5)</td>
<td>16 (47.1)</td>
<td>34 (50.7)</td>
<td>1.35</td>
<td>0.52–3.53</td>
<td>0.38</td>
<td>0.540</td>
</tr>
<tr>
<td>sat</td>
<td>10 (30.3)</td>
<td>16 (47.1)</td>
<td>26 (38.8)</td>
<td>0.49</td>
<td>0.18–1.33</td>
<td>1.98</td>
<td>0.159</td>
</tr>
<tr>
<td>sepA</td>
<td>9 (27.3)</td>
<td>7 (20.6)</td>
<td>16 (23.9)</td>
<td>1.45</td>
<td>0.47–4.48</td>
<td>0.42</td>
<td>0.521</td>
</tr>
<tr>
<td>pic*</td>
<td>7 (21.2)</td>
<td>9 (26.5)</td>
<td>16 (23.9)</td>
<td>0.75</td>
<td>0.25–2.32</td>
<td>0.26</td>
<td>0.614</td>
</tr>
<tr>
<td>sigA*</td>
<td>1 (3.0)</td>
<td>3 (8.8)</td>
<td>4 (6.0)</td>
<td>0.32</td>
<td>0.32–3.28</td>
<td>–</td>
<td>0.613</td>
</tr>
<tr>
<td>pet</td>
<td>6 (18.2)</td>
<td>1 (2.9)</td>
<td>7 (10.5)</td>
<td>7.33</td>
<td>0.83–64.73</td>
<td>–</td>
<td>0.050†</td>
</tr>
<tr>
<td>astA</td>
<td>15 (45.5)</td>
<td>17 (50.0)</td>
<td>32 (47.8)</td>
<td>0.83</td>
<td>0.32–2.18</td>
<td>0.14</td>
<td>0.710</td>
</tr>
<tr>
<td>aafC</td>
<td>4 (12.1)</td>
<td>2 (0.9)</td>
<td>5 (7.5)</td>
<td>4.55</td>
<td>0.48–43.10</td>
<td>–</td>
<td>0.197</td>
</tr>
<tr>
<td>agg3/4C</td>
<td>14 (42.4)</td>
<td>13 (38.2)</td>
<td>27 (40.3)</td>
<td>1.19</td>
<td>0.45–3.16</td>
<td>0.12</td>
<td>0.727</td>
</tr>
<tr>
<td>agg3A</td>
<td>1 (3.0)</td>
<td>0 (0.0)</td>
<td>1 (1.5)</td>
<td>3.19</td>
<td>0.13–81.08</td>
<td>–</td>
<td>0.492</td>
</tr>
<tr>
<td>aafA</td>
<td>4 (12.1)</td>
<td>0 (0.0)</td>
<td>4 (6.0)</td>
<td>10.53</td>
<td>0.54–203.80</td>
<td>–</td>
<td>0.050†</td>
</tr>
<tr>
<td>aggG</td>
<td>3 (9.1)</td>
<td>8 (23.5)</td>
<td>11 (16.4)</td>
<td>0.33</td>
<td>0.78–1.36</td>
<td>–</td>
<td>0.186</td>
</tr>
<tr>
<td>agg4A</td>
<td>2 (6.1)</td>
<td>9 (26.5)</td>
<td>11 (16.4)</td>
<td>0.18</td>
<td>0.03–0.91</td>
<td>–</td>
<td>0.045†</td>
</tr>
<tr>
<td>capU</td>
<td>30 (90.9)</td>
<td>30 (88.2)</td>
<td>60 (89.6)</td>
<td>1.33</td>
<td>0.27–6.48</td>
<td>0.13</td>
<td>0.721</td>
</tr>
<tr>
<td>eilA*</td>
<td>20 (60.6)</td>
<td>17 (50.0)</td>
<td>37 (57.8)</td>
<td>1.54</td>
<td>0.58–4.06</td>
<td>0.76</td>
<td>0.383</td>
</tr>
<tr>
<td>orf61</td>
<td>7 (21.2)</td>
<td>18 (52.9)</td>
<td>25 (37.3)</td>
<td>0.24</td>
<td>0.08–0.70</td>
<td>7.21</td>
<td>0.007†</td>
</tr>
</tbody>
</table>

*EAEC chromosomal genes. All other genes are pAA-plasmid encoded.
†P < 0.05.
Regarding the AAF pilin, AAF/I and AAF/IV, encoded by aggA and agg4A genes, respectively, were the most frequent at 16.4 % each, followed by AAF/II (aafA, 6.0 %) and AAF/III (agg3A, 1.5 %), which was the least frequently detected among all 18 genes studied. The closely related usher for AAF/III and AAF/IV variants (agg3/4C) was detected in 40.3 % of samples. A total of 59.7 % of samples were negative for any described AAF and none was positive for more than one variant.

Among the genes encoding SPATEs, the frequencies of detection were: sat, 38.8 %; pic, 23.9 %; sepA, 23.9 %; pet, 10.5 %; and sigA, 6.0 %.

Considering the frequencies of VRGs between case and control groups, pet and aafA showed significant associations with diarrhoea (P<0.054 and P<0.053, respectively). In addition, two potential protective genes were identified: agg4A and hypothetical orf61. The first was detected in 6.1 % of cases versus 26.5 % of controls [P<0.045, OR=0.179 and 95 % confidence interval (CI)=0.035–0.906] and orf61 was identified in 21.2 % of cases and 52.9 % of controls (P=0.007, OR=0.239, and 95 % CI=0.082–0.699). In the logistic regression analysis, orf61 kept its protective role (P=0.009, OR=0.239 and 95 % CI=0.082–0.699) (Table 2).

**Combinations of VRGs**

To investigate if a specific combination of VRGs could be correlated with disease, we employed CART analysis, which constructs a model in stepwise fashion and the outcome is a combination of factors most strongly associated with cases or controls.

We analysed all genotypic assays of 34 controls and 33 cases. The best CART fit for the dataset is shown in Fig. 1. The analysis suggested two important trait clusters, which were associated with each studied group. Samples harbouring aaiC and agg3/4C but lacking agg4A and orf61 were associated with children with diarrhoea (P<0.002, OR=23.00 and 95 % CI=2.90–219.00). In the logistic regression analysis, orf61 kept its protective role (P<0.009, OR=0.239 and 95 % CI=0.082–0.699) (Table 2).

**Fig. 1. CART model for VRGs in stool samples from case and control children positive for EAEC.** The presence or absence of each gene between case and control groups is given inside each node. Terminal nodes are indicated by the red boxes. The tree is hierarchical in nature. *P≤0.05.
CI=1.267–417.4). In addition, the presence of orf61 in the absence of pet and aafA was associated with the control group (P=0.0006, OR=0.123 and 95 CI=0.0353–0.425).

Characterization of diarrhoeal diseases and its correlation with EAEC

The median duration of diarrhoea in 83 cases was 4.2 days (range 1–18 days). Sick children with EAEC infection did not show differences in the median durations of diarrhoea in comparison with sick children without EAEC infection (4.3 versus 4.2 days, P=0.752, by Mann–Whitney U test, data not shown). Similarly, no differences were seen in the peak number of liquid stools per day (4.0 versus 4.1, P=0.635, by Mann–Whitney U test, data not shown).

Only three of 83 children (3.6%) had diarrhoea lasting longer than 14 days (classified as persistent diarrhoea). Of these, 2/3 (66.7%) were positive for EAEC. Of 80 cases of acute diarrhoea found in this study, 32 (40.0%) were positive for this bacterium. The frequency of persistent diarrhoea did not differ between cases positive and negative for EAEC (P=0.565, by Fisher’s exact test, data not shown).

DISCUSSION

In this case–control study, we examined children 2–36 months of age living in resource-poor urban communities in Fortaleza, individuals who are particularly susceptible to infectious diarrhoea (Keusch et al., 2006). We employed PCR assays to diagnose EAEC directly from stool samples. This molecular technique enables the differentiation of micro-organisms that otherwise appear to be identical by identification of minor differences among the strains. It also identifies VRGs and their products, providing new perspectives for studies involving the epidemiology of diarrhoeal diseases (Keusch et al., 2006; Nataro & Kaper, 1998; Wright & Wynford-Thomas, 1990). The highly heterogeneous characteristics of EAEC are major obstacles for its efficient molecular diagnosis (Nataro & Kaper, 1998). A recently described conserved chromosomal locus, the aaiC gene, was chosen for the development of a specific bacterial primer pair as well as the pAA plasmid-encoded aatA. The aatA gene has been applied in the molecular identification of EAEC since 1990, when Baudry et al. (1990) published the first specific probe (CVD432) for EAEC. Neither gene has been identified among non-EAEC genomes deposited in GenBank (Baudry et al., 1990; Dudley et al., 2006).

The prevalence of EAEC in this population was high (41%). In previous studies performed in industrialized countries, the prevalence of EAEC varied between 2 and 12% (Chan et al., 1994; Cohen et al., 2005; Huppertz et al., 1997; Nataro et al., 2006; Usein et al., 2009). Although diarrhoeal disease is a less important cause of morbidity and mortality in industrialized countries than in developing countries, sporadic diarrhoeal diseases constitute the second most common infectious disease in industrialized areas, with 1–2 episodes per person per year. Besides, they are frequently associated with costly hospital admissions (Guerrant et al., 2001; Steiner et al., 2006). Studies carried out in England, Germany, USA and Romania implicated EAEC as a common bacterial cause of diarrhoea (Chan et al., 1994; Cohen et al., 2005; Huppertz et al., 1997; Nataro et al., 2006; Usein et al., 2009). In developing regions, the percentage of EAEC among diarrhoea cases ranged from 4.5 to 39% (Meng et al., 2011; Ochoa et al., 2011; Okeke et al., 2000; Rüeges et al., 2010). In Brazil, reports found a prevalence of EAEC between 11 and 45% (Bueris et al., 2007; Pereira et al., 2007; Piva et al., 2003; Scaletsky et al., 2002), consistent with the percentage observed in this work.

In the present study, we detected similar EAEC distribution between diarrhoea and non-diarrhoea groups. Similar findings have been made by other researchers (Bueris et al., 2007; Piva et al., 2003). In previous work by our group, studying children from an urban community in Fortaleza, we found a prevalence of EAEC in 38 and 34.8% of children with acute and persistent diarrhoea, respectively (>14 days of duration), and in 34.8% of children representing controls (Lima et al., 2000). We therefore hypothesized that some VRGs or a specific combination of them might be found in EAEC from Fortaleza and possibly be associated with enteric infection symptoms and their risks for dehydration and/or malnutrition, consequences of the EAEC infection previously described (Steiner et al., 1998).

In the last two decades, several bacterial putative VRGs have been suggested to influence the course of EAEC enteric infection (Nataro & Kaper, 1998). This work explored 18 candidate EAEC VRGs in a childhood population living in a poor urban community with a very high prevalence for EAEC colonization and diarrhoeal diseases (Lima et al., 2000). In agreement with previous reports, our samples harboured a diverse range and combination of VRGs. Genes encoding Pet and AAF/II were individually associated with diarrhoea; agg4A (AAF/IV) and the hypothetical orf61 genes were detected significantly more often in children without diarrhoea. Other reports have not found a correlation between the pet gene and the occurrence of diarrhoeal disease (Huang et al., 2007; Pereira et al., 2007; Piva et al., 2003; Regua-Mangia et al., 2009). However, studies in vitro showed that this cytotoxin induces modification in the cytoskeleton followed by cell rounding and detachment of cell monolayers in culture (Betancourt-Sanchez & Navarro-Garcia, 2009). The importance of AAF/II in EAEC pathogenesis was reported by Okeke et al. (2000), which suggested the possibility of using it as a reference marker to identify potentially pathogenic EAEC. Nevertheless, results obtained by Regua-Mangia et al. (2009), analysing EAEC strains isolated from Brazilian children, did not support this suggestion. Surprisingly, the newly described AAF/IV (Boisen et al., 2008) was the most frequently detected AAF, with a similar percentage to that of AAF/I. Furthermore, AAF/IV was associated with control children, data so far not described. The increased number of samples
negative for any known AAF strengthens the wide diversity of EAEC adhesive structures, which include non-fimbrial and uncharacterized fimbrial adhesins. The recently identified orf61 was first investigated by Boisen et al. (2012). Similar to our findings, the authors detected its significant association with control children.

The terms ‘typical’ and ‘atypical’ EAEC have been suggested to classify EAEC strains harbouring or lacking the aggR regulon, respectively (Kaper et al., 2004). In this study the aggR gene was found only in 56.7% of the EAEC-positive samples, and it was not associated with disease. Intriguingly, this regulatory gene controls a number of plasmid genes encoding virulence factors, in addition to pathogenicity islands in the EAEC chromosome involved in several steps of its pathogenesis (Nataro & Kaper, 1998; Okeke & Nataro, 2001). Correlations between aggR and diarrhoea are not consistent. While some authors found significant differences in aggR alone or in combination with cases compared with controls (Huang et al., 2007; Sarantuya et al., 2004), others did not observe any correlation (Boisen et al., 2012; Regua-Mangia et al., 2009). The lack of aggR in the majority of EAEC-positive samples also showed that this gene may not be a good marker to diagnose EAEC samples as suggested by several reports (Cerna et al., 2003; Samie et al., 2007; Sarantuya et al., 2004). Furthermore, we found samples that did not have the aggR gene but were positive for some genes under its control and vice versa. Similar findings were described by Bouzari et al. (2005) and Boisen et al. (2012). These data could be explained by the presence of mutated plasmids in our wild-type strains, and/or by the mosaic nature of EAEC genomes.

CART analysis suggested a trait cluster that indicates virulent EAEC, and another one that is associated with less-pathogenic EAEC. In a recent study by Boisen et al. (2012), characterizing EAEC strains isolated as part of a case–control study of diarrhoea among children in Mali, the investigators identified two gene combinations correlated with diarrhoea: a group harbouring the EAEC heat-stable toxin 1 and the flagellar type H33, and a group carrying several of the typical EAEC VRGs. However, this report did not detect any set of EAEC genes associated with controls.

The variation of EAEC VRGs found in our samples, as well as from different populations of distinct geographical regions of the world, confirmed the genetic heterogeneity of this organism. As most of the VRGs are encoded on the plasmid, this may help to explain the dynamic horizontal acquisition and loss of these VRGs, which favour the variety of genetic profiles found in these EAEC strains (Czeckzulin et al., 1999).

In conclusion, this study has demonstrated a high prevalence of EAEC in children living in a resource-poor urban community of north-eastern Brazil. The overall prevalence of EAEC-positive samples in the diarrhoea group was similar to that of the non-diarrhoea control group. The frequency of 18 different VRGs among these children confirmed the genetic heterogeneity of EAEC. Identification of trait clusters (isolated genes or in combination) correlated with both sick (pet and aafA) and healthy children (aggA and orf61) suggests that the pathophysiology of this enteric infection involves a complex and dynamic modulation of several VRGs. Further efforts are being directed to elucidate if the expressed proteins encoded by the studied VRGs play a role in EAEC infection outcome and pathogenesis.

ACKNOWLEDGEMENTS

This study was supported by the Fogarty International Center (grants 3D43TW006578-04S1, 3D43TW006578-05S1 and 3R24TW007988) and the National Institute of Allergy and Infectious Diseases (NIADD grants 2U01AI026512 and AI-033096), US National Institutes of Health (NIH) and the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq grants 152098/2008-9 and 140242/2009-0) and the Foundation for Support to Scientific and Technological Development from the Ceará government (Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico, FUNCAP grant BCP-0055-00042.01.01/11) of Brazil. We thank field workers Sayonara Alencar, Maria Luzia Melo, Rosânia Silva and Lucia de Fátima Alves; laboratory technicians Maria do Carmo Pinho, Verônica Oliveira, Conceição Nogueira and Leah Barrett; the data management team, José Quirino da Silva Filho, Francisco Sousa Júnior and Charles Melo; and ethics coordinator Fabiana Nascimento. We also thank the participating children and their families.

REFERENCES


