Genomic Characterization of Enteroaggregative 
Escherichia coli From Children in Mali

Nadia Boisen,1,2 Flemming Scheutz,1 David A. Rasko,3 Julia C. Redman,3 Søren Persson,4 Jakub Simon,5 Karen L. Kotloff,5 Myron M. Levine,5 Samba Sow,6 Boubou Tamboura,6 Aliou Toure,6 Dramane Malle,6 Sandra Panchalingam,5 Karen A. Krogfelt,1 and James P. Nataro2

1Department of Microbiological Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark; 2Department of Pediatrics, University of Virginia School of Medicine, Charlottesville; 3Department of Microbiological Diagnostics, Copenhagen, Denmark; 4Department of Microbiology & Immunology, Institute for Genome Sciences, and 5Departments of Pediatrics and Medicine, Center for Vaccine Development, University of Maryland School of Medicine, Baltimore; and 6Center for Vaccine Development, Bamako, Mali

Background. Enteroaggregative Escherichia coli (EAEC) is a cause of epidemic and sporadic diarrhea, yet its role as an enteric pathogen is not fully understood.

Methods. We characterized 121 EAEC strains isolated in 2008 as part of a case-control study of moderate to severe acute diarrhea among children 0–59 months of age in Bamako, Mali. We applied multiplex polymerase chain reaction and comparative genome hybridization to identify potential virulence factors among the EAEC strains, coupled with classification and regression tree modeling to reveal combinations of factors most strongly associated with illness.

Results. The gene encoding the autotransporter protease SepA, originally described in Shigella species, was most strongly associated with diarrhea among the EAEC strains tested (odds ratio, 5.6 [95% confidence interval, 1.92–16.17]; \( P = .0006 \)). In addition, we identified 3 gene combinations correlated with diarrhea: (1) a clonal group positive for sepA and a putative hemolysin; (2) a group harboring the EAST-1 enterotoxin and the flagellar type H33 but no other previously identified EAEC virulence factor; and (3) a group carrying several of the typical EAEC virulence genes.

Conclusion. Our data suggest that only a subset of EAEC strains are pathogenic in Mali and suggest that sepA may serve as a valuable marker for the most virulent isolates.

It is estimated that diarrhea causes at least 1.5 million deaths annually, mostly in children <5 years of age [1]. Although in aggregate the diarrheagenic Escherichia coli (DEC) pathotypes comprise the most common bacterial pathogens worldwide [2], each DEC pathotype is clinically, epidemiologically, and pathogenetically distinct. For some pathotypes, the key virulence factors are known, at least in part, whereas for other pathotypes, the key virulence genes and how they coordinately function in the setting of enteric disease remain elusive.

The enteroaggregative E. coli (EAEC) pathotype has been implicated in travelers’ diarrhea [3], in endemic diarrhea among children in both industrialized [4] and resource-poor countries [5], and in persistent diarrhea among individuals infected with human immunodeficiency virus. A recent outbreak of Shiga toxin–producing EAEC highlights its pathogenic potential [6]. Despite this, the molecular epidemiology of EAEC infection remains unclear, largely due to imperfect recognition of the true pathogenic factors within the broadly defined pathotype.

Most EAEC strains colonize the intestinal mucosa via the aggregative adherence fimbriae (AAFs), which include at least 4 major antigenic variants [7–10]. AAFs are transcriptionally regulated by an AraC/XylS family activator called AggR [7, 11]. AggR is also required for expression of genes encoding dispersin (the aap gene), the Aat dispersin translocator [12], and the...
## Table 1. Primers Used for the 4 Multiplex Polymerase Chain Reactions (PCRs) and 3 Monoplex PCRs, Description of Target Gene, Product Size in Base Pairs, Annealing Temperature, and Concentration of the Primers

| Multiplex PCR | Gene/Target | Description of Target | Primer Sequence (5' - 3') | PCR Product, bp | Annealing Temperature/Primer Concentration (°C, pmol/μL) | GenBank Accession No. |
|---------------|-------------|-----------------------|---------------------------|----------------|----------------------------------------------------------|----------------------|
| 1             | astA        | EAST-1 heat-stable toxin | ATGCCATCAACACAGTATAT [22] | 110            | 58/20 L11241                                              |                     |
|               |             |                       | GCGAGTGAGGCTTGTGAGT [22]  |                |                                                          |                     |
|               | pet         | Plasmid-encoded toxin  | GGCACAGATAAAGGAGGTTTT [23] | 302            | 58/25 AF06681                                             |                     |
|               |             |                       | CCTCTGGTTGACGACGATAC [23] |                |                                                          |                     |
|               | sigA        | IgA protease-like homolog | CGCATTTCTCATTCCTCGCG [19] | 430            | 58/30 NC_004337                                            |                     |
|               |             |                       | CTATCCACAGTGATGTTCCTTGG [19] |              |                                                          |                     |
|               | pic         | Serine protease precursor | ACTGATTTAAAGGCAGGAT [23] | 572            | 58/25 AF097644                                            |                     |
|               |             |                       | GACTAAATCTGCTGTTCAGCCG [23] |                |                                                          |                     |
|               | sepA        | Shigella extracellular protease | GCGATGGAAATATGAGGCCT [23] | 794            | 58/25 Z48219                                              |                     |
|               |             |                       | TTGTCAGATGGGAGGAGAAGCG [23] |                |                                                          |                     |
|               | sat         | Secreted autotransporter toxin [15] | TCAGAAGTCGCGATACTTGG [19] | 932            | 58/25 AE014075                                            |                     |
| 2             | ORF3        | Cryptic protein[a]    | CCATCTTCACACGCTATTCTTA | 121            | 57/35 ...                                                 |                     |
|               |             |                       | CACATCTTAATACCTCACA     |                |                                                          |                     |
|               | aap         | Dispersin, antiaggregation protein [12] | GGACCCGCTCCCAATGATTA [B] | 250            | 57/25 Z32362                                              |                     |
|               |             |                       | GACACCGTGCCTAATGATTA [B] |                |                                                          |                     |
|               | aaiC        | AaiC, secreted protein | TGGTACTACCTGGATAATGGACTTGT [B] | 313            | 57/25 ...                                                 |                     |
|               |             |                       | GACACTCTCTTGTTGTTGACACGA [B] |                |                                                          |                     |
|               | aggR        | Transcriptional activator | GCAATGATTAACGACGATAC [B] | 426            | 57/25 Z18751                                              |                     |
|               |             |                       | CAGCTTCTGACTAGTATACACG [B] |                |                                                          |                     |
|               | aatA        | Dispersin transporter protein | CAGACTTGTGAAAGACTGTATCAT [B] | 642            | 57/35 ...                                                 |                     |
|               |             |                       | CAGTAAATATGTATATGAAATCCGCTGT [B] |            |                                                          |                     |
| 3             | agg4A       | AAF/IV fimbrial subunit | TGAGTTGCGGCTACGTTGA [B] | 169            | 57/25 Z32362                                              |                     |
|               | aggA        | AAF/I fimbrial subunit | CCTCTGCAAGGGTCTAGCT [B] | 220            | 57/20 Y18149                                              |                     |
|               |             |                       | CACACGCTTACGCTAATAGCA [B] |                |                                                          |                     |
|               |             |                       | ACCTGCTCCCATAACCCAGC [B] |                |                                                          |                     |
|               | aafA        | AAF/I fimbrial subunit | CTACTTTATATCAGGGAGGGCTA [B] | 289            | 57/25 AF012835                                            |                     |
|               |             |                       | GGAGGCCAGGAGGTAACTTCG [B] |                |                                                          |                     |
|               | agg3A       | AAF/III fimbrial subunit | CCAGTTTACAGGGGGTTAAGGAA [B] | 370            | 57/25 AF011067                                            |                     |
|               |             |                       | TTGGCTGGAGAATAACACTGGAAGC [B] |            |                                                          |                     |
|               | agg3/4C     | Usher, AAF/III-IV assembly unit | TTCTCAATTACTGGACACGCAAT [B] | 409            | 57/35 AF011067 AN255435 EU637023                         |                     |
|               |             |                       | TTAAATGTGACGCAATTGCAAT [B] |                |                                                          |                     |
|               |             |                       | TCTGACAAATGATTTACCTATGAT [B] |            |                                                          |                     |
| 4             | ORF61       | Plasmid-encoded hemolysin[a] | AGCTTGGAGAATCTGCTCT | 108            | 57/10 ...                                                 |                     |
|               |             |                       | AAGCTGCTTGATTCTGCTT     |                |                                                          |                     |
|               | eilA        | Salmonella HilA homolog | AGGTCTGGAGGCGGAGTT [B] | 248            | 57/30 ...                                                 |                     |
|               |             |                       | GTAAAAGCTGATCCAGGACCC [B] |                |                                                          |                     |
|               | capU        | Hexosyltransferase homolog | CAGCTGTTGCTCCAATGGAA [B] | 395            | 57/25 AF134403                                            |                     |
|               |             |                       | GTCTGACATCTCTCTGCTCT [B] |                |                                                          |                     |
|               | air         | Enteroaggregative immunoglobulin repeat protein [24] | TTACCTCTGCTGCTCCTAAT | 600            | 57/25 ...                                                 |                     |
|               |             |                       | GGTAAATCGCTGTCTTCTT     |                |                                                          |                     |
chromosomal cluster termed Aai, encoding a type VI secretion system [13]. Factors not under AggR control include the Air adhesin, a regulator termed EilA, the EAEC heat-stable toxin EAST-1 (encoded by the astA gene), and a set of toxins termed the serine protease autotransporters of Enterobacteriaceae (SPATEs).

SPATEs have been organized phylogenetically into 2 classes. Members of class 1 are cytotoxic to epithelial cells [14]; class 1 SPATEs found in EAEC strains include the plasmid-encoded toxin (Pet) and its 2 homologs, Sat [15] and SigA [16]. The class 2, or noncytotoxic, SPATEs include Pic, a mucinase that promotes intestinal colonization [17, 18]. As with cytotoxic SPATEs and Pic, we have recently reported that the class 2 SPATE SepA is found commonly among EAEC strains [19]. SepA is a cryptic protease originally described in Shigella species, and is reported to contribute to intestinal inflammation [20]. Importantly, none of these factors are found in all EAEC isolates, and no single factor has ever been consistently implicated in EAEC virulence.

Here, we characterize 121 EAEC strains isolated as part of a case-control study of acute moderate to severe diarrhea among children aged 0–59 months in Mali. We report that the sepA gene and flagellar type H33 are strongly associated with illness, and we define additional sets of virulence genes and factors that are important in this population.

**MATERIALS AND METHODS**

**Study Design**

The strains utilized were isolated in the course of a prospective multicenter case-control study (Global Enteric Multi-Center Study, GEMS) of moderate to severe diarrhea among children <5 years of age. Full details of the GEMS design will be published elsewhere. In brief, children ≤59 months presenting to health centers for care with a complaint of diarrhea within the previous 7 days were considered eligible. Cases were enrolled upon parental consent if they met criteria for moderate to severe diarrhea comprising signs of moderate to severe dehydration (sunken eyes, decreased skin turgor), dysentery (blood in stool), or if they were deemed to require hospitalization or intravenous rehydration. Diarrhea was defined as the passage of ≥3 or more unformed stools within a 24-hour period. A stool sample was obtained at enrollment and analyzed comprehensively for bacterial, viral, and protozoal agents. An age-matched asymptomatic control from the same neighborhood was enrolled for each case; a stool sample was obtained from the control child and analyzed similarly.

**Specimen Processing and Microbiological Analysis**

A single, fresh, whole stool specimen was collected from cases and controls at enrollment for the recovery of potential enteropathogens. Various specific growth media were used for detecting the bacterial pathogens. Up to 3 colonies with the appearance of E. coli on MacConkey agar were selected from each sample and tested using multiplex polymerase chain reaction (PCR) for enterotoxigenic E. coli (ETEC) (heat-labile [LT] and heat-stable [ST] enterotoxins), enteropathogenic E. coli (EPEC) (eae and bfpA), and EAEC (aaiC and aatA). Any colonies that were positive for either aaiC (chromosomally encoded) or aatA (encoded on the pAA plasmid) were considered EAEC for the purposes of this analysis.

**Serotyping**

Somatic (O) and flagella (H) antigens were identified as described elsewhere [21]; the following designations were included: “O rough,” the boiled culture auto-agglutinated, suggesting
absence of O antigen; “O−,” it could not be determined whether the strain produces an O antigen (precipitation with Cetavlon indicates an acidic polysaccharide that could represent capsular K antigen); and “O+,” the O antigen is present but could not be typed. Serotyping was performed at the International Escherichia and Klebsiella Centre (World Health Organization), Department of Microbiological Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark.

Polymerase Chain Reaction
Primers and conditions for detecting sequences encoding 21 putative virulence genes, which are described in Table 1, were used in 4 multiplex reactions. Multiplex PCR 1 was performed as previously described [19], with the addition of primers targeting astA. Multiplexes 2–4 were performed using PCR mastermix (2X) according to the manufacturer’s instructions (Fermentas International), with the addition of 1 μL 50 mM magnesium chloride per 50 μL reaction. A DNA template was prepared by boiling a suspension of 10 isolated colonies in 200 μL distilled water. PCR reaction cycles were as follows: (1) 2 minutes denaturation at 95°C, (2) 50 seconds denaturation at 94°C, (3) annealing for 1.5 minutes, and (4) extension for 1.5 minutes at 72°C with 35 cycles returning to step 2. The final extension was 10 minutes at 72°C. Products were amplified using an Eppendorf Mastercycler Gradient thermal cycler (Eppendorf North America) and separated in 2% agarose gels.

Individual amplification reactions to detect genes designated rmoA, espY2, and shiA were done in 25 μL reaction volumes using crude bacterial cell lysates; PCR reactions were performed as multiplex 2–4. The final extension was 10 minutes at 72°C.

The phylogenetic groups A, B1, B2, and D were determined using triplex PCR methods employing phylogenetic group-specific primers for 2 genes, chuA and yjaA, and a cryptic DNA fragment, TspE4C2. The grouping was coupled to a dichotomous decision tree according to Clermont et al [26].

The following strains where used as controls for detection of target genes: JM221 (aggA, sat) [27], 042 (aatA, aggR, aaiC, aap, ORF3, pic, pet, astA, aafA, aac, air, capU, eilA) [28], 55989 (agg3A, agg3/4C) [8], H223-1 (sigA) [29], C1010-00 (agg4A, agg3/4C, sat, sepA) [30], MC1061 (negative control), J96 (chuA, yjaA) [26], CFT073 (chuA, yjaA, TspE4.C2) [31], C452-97 (TspE4.C2) [32], and EDL933 (chuA) [33].

Genomic Hybridization
Comparative genome hybridization (CGH) was performed on all the sepA-positive strains as well as sepA-negative strains C801-09 and C46-10 and reference EAEC isolates as previously described [34]. The pan-genome microarrays used in this study were designed by FDA-ECSG Array Probe Set Design and represent the genomes of 32 diverse E. coli and Shigella species, as well as 46 enteric plasmid sequences [35]. Initial data analysis was performed with the Gene Chip Operating System suite of tools provided by Affymetrix. Additional analysis was performed using the Affymetrix power tools software. The MAS5 algorithm was utilized with the perfect match and mismatch calculations and a Tau of 0.150 to detect which probes were present or absent. Features that were present or absent in all samples were removed from further analysis. The resulting features, known as the variable gene set, were analyzed using Multiple Experiment Viewer, version 4.5. The cladogram was constructed using the 12 673 variable features in this dataset, which contained hybridization data from 36 strains. The relationship was determined using hierarchical clustering with Pearson correlation, using the absolute distance and complete linkage run with 500 bootstrap calculations.

Statistics
We utilized classification and regression tree (CART) Pro Version 6.0 (Salford Systems) software inputting 21 or 24 factors of interest as binary (present/absent) independent predictive variables along with a continuous “factor total” that was a sum of all factors including flagellum type H33. Case/control status was the binary dependent outcome variable.

RESULTS
Initial Characterization of EAEC Strains
After 1 year of surveillance, EAEC strains were isolated as the sole DEC pathogen from 60 children with diarrhea and 61 asymptomatic controls. The lack of association of EAEC with diarrhea among the cases persisted even when the presence of other potential pathogens, stratifying for age, was considered or when either aatA or aaiC alone or in combination were considered.

One EAEC isolate was selected from each stool sample that yielded EAEC by multiplex PCR. The 121 EAEC strains belonged to diverse serotypes (Table 2; cases listed in Table 2A and controls listed in Table 2B). Examination of the correlation between serotype and case/control status revealed that only flagellum type H33 was significantly more common among cases than controls (12 cases, 2 controls; odds ratio [OR], 5.9; P = .0138) (Table 3). EAEC cases and control strains were localized to similar positions within a previously published general E. coli phylogenetic tree (Tables 2 and 3).

Frequencies of Virulence-Related Genes
In order to assess the roles of putative virulence factors in EAEC epidemiology, we developed 4 multiplex PCR assays for the characterization of 21 genes previously found in EAEC strains. The results of the PCR assays for all strains are listed in Table 2. Of the 21 genes scored, hypothetical ORF3 was the most frequently detected (86%) followed by eilA (85.1%), capU (81.8%), aap (71.9%), aggR (69.4%), and aatA (68.6%) (Table 3). There was a high degree of concordance of these genes, which has been demonstrated previously for the
Table 2a. Characteristics of 60 Enteroaggregative Escherichia coli Strains Isolated From Children With Diarrhea

| Strain   | Serotype | pAA plasmid | Adhesin | Toxins | Chromosome | Phylogentic group | Genomic hybridization | Age* | Virulence factor score* |
|----------|----------|-------------|---------|--------|------------|-------------------|-----------------------|------|------------------------|
| A. Characteristics of 60 EAEC Strains |          |             |         |        |            |                   |                       |      |                        |
| 28-10    | O86      |             |         | H5     |            |                   |                       | 18   | 11                     |
| 34-10    | O29      |             |         | H2     |            |                   |                       | 19   | 13                     |
| 35-10*   | O104     |             |         | H4     |            |                   |                       | 12   | 12                     |
| 36-10    | O101     |             |         | H5     |            |                   |                       | 12   | 12                     |
| 46-10*   | O99      |             |         | H33    |            |                   |                       | 24   | 8                      |
| 678-09*  | O51      |             |         | H5     |            |                   |                       | 38   | 10                     |
| 682-09*  | O104     |             |         | H4     |            |                   |                       | 19   | 13                     |
| 685-09   | Otu      |             |         | H30    |            |                   |                       | 1    | 10                     |
| 587-09*  | O153     |             |         | H30    |            |                   |                       | 7    | 8                      |
| 993-09*  | O21      |             |         | H40    |            |                   |                       | 40   | 8                      |
| 701-09*  | OX184    |             |         | H32    |            |                   |                       | 12   | 8                      |
| 703-09*  | O153     |             |         | H30    |            |                   |                       | 4    | 13                     |
| 705-09   | O73.13    |             |         | H33    |            |                   |                       | 14   | 10                     |
| 708-09   | O73      |             |         | H33    |            |                   |                       | 23   | 10                     |
| 709-09   | O86      |             |         | H30    |            |                   |                       | 4    | 11                     |
| 711-09   | O25      |             |         | H5     |            |                   |                       | 14   | 9                      |
| 713-09   | O15      |             |         | H+     |            |                   |                       | 45   | 4                      |
| 714-09   | O25      |             |         | H5     |            |                   |                       | 12   | 11                     |
| 716-09*  | O34      |             |         | H12    |            |                   |                       | 5    | 9                      |
| 718-09*  | O78      |             |         | H6     |            |                   |                       | 8    | 10                     |
| 719-09*  | O39      |             |         | H49    |            |                   |                       | 9    | 12                     |
| 720-09   | O44      |             |         | H34    |            |                   |                       | 5    | 9                      |
| 721-09*  | O14      |             |         | H4     |            |                   |                       | 27   | 13                     |
| 724-09   | O7       |             |         | H10    |            |                   |                       | 8    | 14                     |
| 727-09   | O99      |             |         | H+     |            |                   |                       | 34   | 7                      |
| 729-09*  | O25      |             |         | H27    |            |                   |                       | 18   | 8                      |
| 730-09   | O153     |             |         | H30    |            |                   |                       | 11   | 12                     |
| 736-09*  | OX184    |             |         | H32    |            |                   |                       | 13   | 7                      |
| 737-09   | O59      |             |         | H30    |            |                   |                       | 8    | 13                     |
| 738-09   | O7       |             |         | H33    |            |                   |                       | 5    | 8                      |
| 739-09   | O99      |             |         | H33    |            |                   |                       | 9    | 11                     |
| 741-09   | O73.128ab|             |         | H33    |            |                   |                       | 7    | 6                      |
| 744-09*  | O162     |             |         | H5     |            |                   |                       | 10   | 10                     |
| 745-09*  | O21      |             |         | H25    |            |                   |                       | 8    | 7                      |
| 756-09   | O157     |             |         | H5     |            |                   |                       | 25   | 4                      |
| 757-09   | O25      |             |         | H9     |            |                   |                       | 9    | 10                     |
| 762-09   | O153     |             |         | H30    |            |                   |                       | 7    | 13                     |
| 764-09*  | O45.71   |             |         | H14    |            |                   |                       | 23   | 11                     |
| 765-09*  | O39      |             |         | H30    |            |                   |                       | 7    | 8                      |
| 768-09*  | O99      |             |         | H33    |            |                   |                       | 9    | 6                      |
| 769-09*  | O7       |             |         | H+     |            |                   |                       | 8    | 10                     |
| 771-09   | O39      |             |         | H49    |            |                   |                       | 22   | 15                     |
| 772-09   | O7       |             |         | H27    |            |                   |                       | 16   | 10                     |
| 773-09   | O44      |             |         | H33    |            |                   |                       | 3    | 10                     |
| 774-09   | O166     |             |         | H15    |            |                   |                       | 23   | 4                      |
| 776-09   | O7       |             |         | H+     |            |                   |                       | 17   | 10                     |
| 777-09   | O104     |             |         | H4     |            |                   |                       | 13   | 10                     |
| 778-09*  | O68      |             |         | H26    |            |                   |                       | 16   | 10                     |
| 780-09   | O63      |             |         | H5     |            |                   |                       | 22   | 7                      |
| 781-09   | O69      |             |         | H33    |            |                   |                       | 23   | 12                     |
| 783-09*  | O18      |             |         | H9     |            |                   |                       | 36   | 13                     |
| 784-09*  | O7       |             |         | H+     |            |                   |                       | 7    | 12                     |
| 785-09*  | O166     |             |         | H+     |            |                   |                       | 31   | 6                      |
| 788-09   | O146     |             |         | H8     |            |                   |                       | 13   | 9                      |
| 796-09*  | O25      |             |         | H27    |            |                   |                       | 5    | 7                      |
| 798-09   | O7       |             |         | H5     |            |                   |                       | 4    | 10                     |
| 800-09   | Otu      |             |         | H19    |            |                   |                       | 7    | 10                     |
| 801-09*  | O153     |             |         | H30    |            |                   |                       | 7    | 10                     |
| Strain   | Serotype  | pAA plasmid | Chromosome | Phylogenetic group | Genomic hybridization | Age<sup>a</sup> | Virulence factor score<sup>b</sup> |
|----------|-----------|-------------|------------|--------------------|-----------------------|-------------------|---------------------|
|          |           | aaA         |            |                    |                       |                   |                     |
| 24-10    | O12      | H           |            |                    |                       | 3                 | 11                  |
| 30-10    | O39, O165 | H           |            |                    |                       | 24                | 4                   |
| 33-10    | O125ab   | H1          |            |                    |                       | 12                | 9                   |
| 38-10    | O39, O165 | H           |            |                    |                       | 9                 | 14                  |
| 677-09   | O+        | H38         |            |                    |                       | 10                | 5                   |
| 678-09   | O73      | H18         |            |                    |                       | 10                | 12                  |
| 680-09   | O18b,23  | H15         |            |                    |                       | 17                | 12                  |
| 681-09   | O14      | H           |            |                    |                       | 17                | 7                   |
| 683-09   | O68      | H9          |            |                    |                       | 15                | 7                   |
| 684-09   | O21      | H           |            |                    |                       | 16                | 11                  |
| 695-09   | O21      | H           |            |                    |                       | 10                | 8                   |
| 699-09   | O+       | H18         |            |                    |                       | 10                | 9                   |
| 699-09   | O117     | H           |            |                    |                       | 21                | 12                  |
| 701-09   | O73      | H18         |            |                    |                       | 36                | 10                  |
| 704-09   | O21      | H           |            |                    |                       | 5                 | 11                  |
| 705-09   | O14      | H           |            |                    |                       | 35                | 6                   |
| 709-09   | O14      | H           |            |                    |                       | 3                | 10                  |
| 709-09   | O141     | H           |            |                    |                       | 15                | 19                  |
| 709-09   | O191     | H           |            |                    |                       | 62                | 11                  |
| 709-09   | O196     | H15         |            |                    |                       | 15                | 14                  |
| 713-09   | O17      | H7          |            |                    |                       | 15                | 19                  |
| 714-09   | O17      | H7          |            |                    |                       | 15                | 19                  |
| 717-09   | O117     | H           |            |                    |                       | 6                 | 11                  |
| 718-09   | O17      | H3          |            |                    |                       | 4                 | 11                  |
| 722-09   | O20      | H11         |            |                    |                       | 35                | 8                   |
| 725-09   | O99      | H3           |            |                    |                       | 21                | 7                   |
| 726-09   | O73,13   | H           |            |                    |                       | 16                | 9                   |
| 731-09   | O73     | H11         |            |                    |                       | 9                 | 12                  |
| 732-09   | O73     | H11         |            |                    |                       | 14                | 10                  |
| 733-09   | O14      | H           |            |                    |                       | 17                | 7                   |
| 734-09   | O194     | H4          |            |                    |                       | 5                 | 13                  |
| 735-09   | O7      | H7          |            |                    |                       | 16                | 13                  |
| 742-09   | O86      | H18         |            |                    |                       | 9                 | 7                   |
| 743-09   | O73      | H11         |            |                    |                       | 3                 | 11                  |
| 746-09   | O128     | H12         |            |                    |                       | 5                 | 14                  |
| 747-09   | O22      | H           |            |                    |                       | 39                | 14                  |
| 748-09   | O132     | H14         |            |                    |                       | 12                | 9                   |
| 748-09   | O7      | H38         |            |                    |                       | 12                | 9                   |
| 752-09   | O86      | H18         |            |                    |                       | 11                | 10                  |
| 753-09   | O51      | H38         |            |                    |                       | 3                 | 10                  |
| 754-09   | O194     | H4          |            |                    |                       | 17                | 14                  |
| 755-09   | O7      | H7          |            |                    |                       | 14                | 13                  |
| 759-09   | O86      | H           |            |                    |                       | 7                 | 11                  |
| 760-09   | O104     | H4          |            |                    |                       | 10                | 19                  |
| 761-09   | O7      | H           |            |                    |                       | 1                 | 19                  |
| 763-09   | O99      | H33         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 6                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 8                 | 14                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |
| 767-09   | O125ac   | H12         |            |                    |                       | 5                 | 12                  |

Black boxes represent positivity. The designations of all Malian strains are preceded by the letter “c” in the text.
Abbreviation: EAEC, enteragreggative Escherichia coli.
* Age in months.
<sup>b</sup> Virulence factor score is based on the number of genes 1 strain is positive for.
<sup>c</sup> Strains analyzed for other virulence genes by comparative genomic hybridization.
plasmid-encoded *aat*, *aggR*, and *aap* genes [7, 12, 36–38]. Sixty-eight percent of the strains were positive for the usher-encoding gene *agg3/4C* (the ushers for AAF/III and AAF/IV variants are closely related). The most frequent AAF pilin gene was that of AAF/I, encoded by *aggA* (26.4%), followed by those of AAF/II (*aafA*), AAF/III (*agg3A*), and AAF/IV (*agg4A*) at 5% each (Table 3). The *agg4A* gene was found more frequently among cases than controls (5 cases and 1 control), although this difference did not reach statistical significance.

A total of 71 strains (58.7%) were negative for a known AAF variant.

Of the 5 genes encoding SPATEs, the most frequent were *sat* (47.1%) and *pic* (46.3%). The least common SPATEs were *pet* (8.3%) and *sigA* (12.4%). The *sepA* gene was found in 25 strains (20.7%): 20 from cases and 5 from controls, yielding an OR of 5.6 (*P* = .0006) (Table 3). Among all the putative virulence factors scored, *sepA* was the only one significantly associated with moderate to severe diarrheal illness.

### Table 3. Distribution of Enteroaggregative *Escherichia coli* Virulence Factors in Cases and Controls

| Virulence Gene | Cases (n = 60) | Controls (n = 61) | Total (N = 121) | Risk Estimate |
|----------------|---------------|------------------|-----------------|---------------|
|                | No. (%)       | No. (%)          | No. (%)         | Odds Ratio    | [95% CI]      | χ2  | P Value |
| *aatA*         | 37 (61.7)     | 46 (75.4)        | 83 (68.6)       | 0.5           | [0.24–1.15]   | 2.7 | .10     |
| *aggR*         | 38 (63.3)     | 46 (75.4)        | 84 (69.4)       | 0.6           | [0.26–1.23]   | 2.1 | .15     |
| *aaiC*         | 32 (53.3)     | 26 (42.6)        | 58 (47.9)       | 1.5           | [0.75–3.15]   | 1.4 | .24     |
| *aap*          | 39 (65.0)     | 48 (78.7)        | 87 (71.9)       | 0.5           | [0.22–1.19]   | 2.8 | .09     |
| ORF3           | 49 (81.7)     | 55 (90.2)        | 104 (86.0)      | 0.5           | [0.17–1.39]   | 1.8 | .18     |
| sat            | 24 (40.0)     | 33 (54.1)        | 57 (47.1)       | 0.6           | [0.28–1.16]   | 2.4 | .12     |
| *sepA*         | 20 (33.3)     | 5 (8.2)          | 25 (20.7)       | 5.6           | [1.92–16.17]  | 11.7| .0006   |
| *pic*          | 29 (48.3)     | 27 (44.3)        | 56 (46.3)       | 1.2           | [0.58–2.41]   | 0.2 | .66     |
| *sigA*         | 8 (13.3)      | 7 (11.5)         | 15 (12.4)       | 1.2           | [0.40–3.51]   | 0.1 | .76     |
| *pet*          | 4 (6.7)       | 6 (9.8)          | 10 (8.3)        | 0.7           | ...           | ... | .74     |
| *astA*         | 32 (53.3)     | 30 (49.2)        | 62 (51.2)       | 1.2           | [0.58–2.41]   | 0.2 | .65     |
| *aafC*         | 5 (8.3)       | 5 (8.2)          | 10 (8.3)        | 1.0           | ...           | ... | >.999   |
| *agg3/4C*      | 42 (70.0)     | 40 (65.6)        | 82 (67.8)       | 1.2           | [0.57–2.63]   | 0.3 | .60     |
| *agg3A*        | 1 (1.7)       | 5 (8.2)          | 6 (5.0)         | 0.2           | ...           | ... | .21     |
| *aafA*         | 3 (5.0)       | 3 (4.9)          | 6 (5.0)         | 1.0           | ...           | ... | >.999   |
| *aggA*         | 11 (18.3)     | 21 (34.4)        | 32 (26.4)       | 0.4           | [0.18–0.99]   | 4.0 | .04     |
| *agg4A*        | 5 (8.3)       | 1 (1.6)          | 6 (5.0)         | 5.5           | ...           | ... | .11     |
| *air*          | 20 (33.3)     | 29 (47.5)        | 49 (40.5)       | 0.6           | [0.26–1.15]   | 2.5 | .11     |
| *capU*         | 48 (80.0)     | 51 (83.6)        | 99 (81.8)       | 0.8           | [0.31–1.99]   | 0.3 | .61     |
| *e1iA*         | 50 (83.3)     | 53 (86.9)        | 103 (85.1)      | 0.8           | [0.28–2.09]   | 0.3 | .58     |
| ORF61          | 28 (46.7)     | 44 (72.1)        | 72 (59.5)       | 0.3           | [0.16–0.72]   | 8.1 | .004    |
| *espY2*        | 13 (21.6)     | 20 (32.8)        | 33 (27.3)       | 0.6           | [0.25–1.28]   | 1.9 | .17     |
| *rmoA*         | 30 (50.0)     | 23 (37.7)        | 53 (43.8)       | 1.7           | [0.30–3.39]   | 1.9 | .17     |
| *shiA*         | 21 (35)       | 22 (36.1)        | 43 (35.5)       | 0.5           | [0.45–2.01]   | 0.01| .92     |

| EAEC Serogroup | Cases (n = 60) | Controls (n = 61) | Total (N = 121) | Risk Estimate |
|----------------|---------------|------------------|-----------------|---------------|
| O99            | 5 (8.3)       | 2 (3.3)          | 7 (5.8)         | 2.7           | ...           | ... | .27     |
| O153           | 6 (10.0)      | 1 (1.6)          | 7 (5.8)         | 6.7           | ...           | ... | .61     |
| H              | 7 (11.7)      | 17 (27.9)        | 24 (19.9)       | 0.3           | [0.12–0.98]   | 4.9 | .04     |
| H5             | 6 (10.0)      | 2 (3.3)          | 8 (6.6)         | 3.3           | ...           | ... | .16     |
| H9             | 5 (8.3)       | 1 (1.6)          | 6 (5.0)         | 5.5           | ...           | ... | .21     |
| H30            | 9 (15.0)      | 4 (6.6)          | 13 (10.7)       | 2.5           | [0.73–8.66]   | 2.2 | .13     |
| H33            | 10 (16.7)     | 2 (3.3)          | 12 (9.9)        | 5.9           | [1.23–28.19]  | 6.7 | .01     |

| Phylogenetic Group | Cases (n = 60) | Controls (n = 61) | Total (N = 121) | Risk Estimate |
|--------------------|---------------|------------------|-----------------|---------------|
| A                  | 23 (38.3)     | 14 (22.9)        | 37 (32.2)       | 2.1           | [0.95–4.61]   | 3.4 | .07     |
| B1                 | 15 (25)       | 13 (21.3)        | 28 (23.1)       | 1.2           | [0.53–2.78]   | 0.2 | .63     |
| B2                 | 8 (13.3)      | 9 (14.7)         | 17 (14)         | 0.9           | [0.32–2.48]   | 0.05| .82     |
| D                  | 14 (23)       | 25 (40.9)        | 39 (32.2)       | 0.4           | [0.19–96]     | 4.3 | .04     |

*P* < .05 is significant. Fisher exact test was applied when the comparisons between cases and controls were <5 observations.

Abbreviation: EAEC, enteroaggregative *Escherichia coli*. 

Genomic Characterization of EAEC • JID 2012:205 (1 February) • 437
Significance of Combinations of EAEC Genes

In addition to considering each factor individually, we pursued a number of approaches to consider the importance of combinations of potential EAEC virulence factors. When crudely considering the collective number of virulence loci present (generating a virulence factor score, VFS), the average number of virulence genes from cases was 8.75 versus 9.5 from control isolates.

To consider combinations of factors, we employed CART analysis, which builds a model in stepwise fashion to yield the combination of factors most strongly associated with the queried outcome. Each branch of a CART output tree ends in a terminal “node”; each observation falls into exactly 1 terminal node; and each terminal node is uniquely defined by a set of rules, such as having or not having a certain factor.

We considered all genotypic and phenotypic assays performed and interrogated the association with case status. Figure 1 illustrates the best CART fit for the dataset. The analysis demonstrates that the presence of sepA, regardless of the presence or absence of any other scored genotype or phenotype.

Figure 1. Classification and regression tree (CART) classification tree topology reveals combinations of factors most strongly associated with moderate to severe diarrhea. We considered all genotypic and phenotypic assays performed: aatA, aggR, aggC, aap, ORF3, sat, sapA, pic, sigA, pet, astA, aafC, agg3/4C, aafA, agg3A, agg4A, air, capU, euA, ORF61, virulence factor score (VFS), and flagellum type H33. Each branch of the CART tree ends in a terminal “node” (red boxes), and each terminal node is uniquely defined by the presence or absence of a predictive factor such as a gene or VFS. The tree is hierarchical in nature. C701-09, C718-09, C801-09, and C46-10 are also shown on the dendrogram.
among the sepA-positive strains, provides a strong association with diarrhea.

Among the sepA-negative strains, CART analysis suggested 2 additional trait clusters that were associated with moderate to severe diarrheal illness: 1 cluster included those strains harboring the flagellum H33 and the toxin EAST-1, whereas a second cluster lacked H33 but featured a VFS of 9, suggesting a combination of typical EAEC factors in addition to the Sat toxin.

Genomic Analyses
We hypothesized that the strain sets belonging to the nodes most strongly associated with diarrhea would reveal the presence of additional virulence determinants, which themselves might explain the observed clinical correlations. We therefore performed CGH analysis using a previously described microarray containing the full genomes of 32 E. coli and Shigella strains and the genes of an additional 46 E. coli plasmids [35]. For this analysis, we chose all 25 sepA-positive strains, 2 additional strains (C46-10 and C801-09) representing CART (Figure 1) nodes 2 (SepA absent, H33 present, EAST-1 present) and 3 (SepA absent, H33 absent, >8.5VFS, Sat present), and a set of archetype EAEC reference strains. Standard cluster analysis was performed on the microarray data (Figure 2). All isolates belonging to a common serotype clustered together in this analysis. Although cluster analysis did not suggest genomic differences discriminating cases and controls, the analysis did suggest that sepA-positive strains segregated into 2 major clusters (indicated as A and B in Figure 2). We chose for further genomic examination archetypal strains representing sepA-positive clusters A

---

Figure 2. Cladogram of comparative genomic hybridization data of sepA-positive isolates (C34-10, C35-10, C679-09C, C682-09, C693-09, C701-09, C703-09, C716-09, C718-09, C719-09, C729-09, C736-09, C745-09, C764-09, C765-09, C769-09, C771-09, C778-09, C783-09, C796-09, C697-09, C707-09, C710-09, C734-09, and C748-09), sepA-negative isolates (C46-10 and C801-09), and reference isolates (C1096, 042, 55989, JM221, 17-2, 34b, 101-1, and H5). Notably, C46-10 was most closely related to Mexican enterotoaggregative Escherichia coli (EAEC) strain JM221 (isolated from an adult traveler to Guadalajara [27]), and strain C801-09 was most closely related to EAEC strain 042, isolated from a child with diarrhea in Lima, Peru [28]. The phylogenetic comparison was performed using the 12,673 variable features of the 36 hybridizations included. The tree is built using a hierarchical clustering with Pearson correlation using both the absolute distance and complete linkage and viewed in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Isolates represented in black are reference isolates, controls are indicated in blue, and cases in red. The serotypes of the strains are indicated to the right. The gray boxes identify clusters of serotypes within the context of the larger tree, indicating that those serogroups are genomically similar.
Table 4. Comparative Genomic Hybridization of Strains C701-09, C718-09, C801-09, and C46-10 Against a Microarray That Comprises the Full Genomes of 32 Escherichia coli and Shigella Strains and the Genes of Additional 46 E. coli Plasmids

| Putative Virulence Gene\(^a\) | Accession No. | Hybridization by Genome |
|-----------------------------|---------------|-------------------------|
|                             |               | Nonpathogenic\(^b\) | C701-09 | C718-09 | C801-09 | C46-10 |
| Adhesins                    |               |                        |         |         |         |         |
| csgA; cryptic curlin major subunit\(^b\) | SBO_2026    | +                       | +       | +       | +       | –       |
| csgA; major curlin subunit\(^b\)    | LF82_0360    | +                       | –       | –       | –       | +       |
| csgC; putative autoagglutination protein\(^b\) | ECUMN_1217  | +                       | +       | +       | +       | –       |
| ecpD; putative chaperone protein EcpD\(^b\) | SBO_0126    | +                       | –       | +       | +       | –       |
| Fimbrial usher family protein\(^b\)    | SbBS512_E2717 | +                       | +       | +       | –       | +       |
| Flu; antigen 43 (Ag43)\(^b\)       | ECUMN_3400    | +                       | –       | –       | +       | +       |
| Hemagglutinin family\(^c\)       | SbBS512_E4026 | –                       | –       | +       | +       | –       |
| Putative AidA-I adhesin-like protein\(^d\) | ECO26_1353  | +                       | –       | –       | +       | –       |
| Putative chaperone protein EcpD   | ECUMN_0137    | –                       | –       | +       | –       | +       |
| Putative fimbrial biogenesis outer membrane usher protein | ECUMN_0019 | –                       | –       | +       | –       | +       |
| Putative fimbrial protein\(^b\)    | SbBS512_E2376 | +                       | +       | +       | –       | +       |
| Putative fimbrial-like protein\(^b\) | SD_Y_0915    | +                       | –       | –       | +       | +       |
| Putative invasin\(^b\)           | EcsMS35_1_1146 | +                       | +       | +       | +       | +       |
| Putative type 1 fimbrial protein | ECSP_0022     | –                       | –       | –       | +       | –       |
| Putative type I fimbrial usher protein | Sfmd\(^b\) | +                       | –       | –       | +       | –       |
| Putative type II secretion protein (GspI-like)\(^b\) | E2348C_2478  | –                       | –       | –       | +       | –       |
| Putative type III secretion protein EpaRb | ECO26_3933  | +                       | +       | –       | +       | +       |
| Putative type III secretion protein EpaQb | ECO103_3428 | +                       | –       | +       | –       | +       |
| Putative type III secretion protein EprJb | ECO26_3933  | +                       | +       | –       | +       | –       |
| Putative type IV secretion protein (VgrG-like) | ECO26_1696  | +                       | +       | +       | –       | +       |
| Non-LEE-encoded Type III Secreted Effector | ECSP_0073 | –                       | –       | –       | +       | +       |
| Putative type II secretion protein (GspI-like)\(^b\) | ECIAI1_3105  | +                       | +       | –       | +       | –       |
| Putative type III secretion protein EpaRb | ECUMN_3195  | +                       | –       | –       | +       | –       |
| T3SS effector-like protein EspL-homolog\(^b\) | ECO111_4829 | +                       | –       | +       | –       | +       |
| TolC; outer membrane channel protein\(^b\) | SDY_3205    | +                       | +       | –       | –       | +       |
| Type III secretion protein EpaQb | ECO26_3940    | +                       | –       | +       | –       | +       |
| Type III secretion protein EpaRb | ECO103_3428  | +                       | –       | +       | –       | +       |
| Type III secretion protein EprJb | ECO26_3933    | +                       | –       | +       | –       | +       |
| Other                       |               |                        |         |         |         |         |
| Hemolysin expression-modulating protein | EC55989_3351 | –                       | +       | +       | +       | +       |
| Putative hemolysin expression-modulating protein RmoA | ECUMN_0072  | –                       | +       | +       | –       | +       |
| Putative hemolysin co-regulated protein\(^b\) | SSON_0255   | +                       | –       | –       | +       | +       |
| ShiA-like protein            | ECB_03517     | –                       | –       | –       | –       | –       |

\(^a\) EAEC genes are listed in Table 2A and 2B.

\(^b\) One hundred percent identities with HS and/or K12.

\(^c\) Eighty-eight percent identities with HS.

\(^d\) Fifty-five percent to 62% identities with HS and/or K12.
and B, as well as the 2 additional nodes (2 and 3) indicated by CART analysis (Figure 1).

**Genome Analysis of sepA-Positive Strains**

To represent the *sepA*-positive strain cluster A (Figure 2), we chose strain C718-09 for further genome analysis; to represent cluster B, we chose strain C701-09. Results are presented in Table 4. Genome analysis of strain C-718-09 did not reveal the presence of additional genes that were not also carried by nonpathogenic *E. coli* strains. C701-09 hybridized to an open reading frame that was 99% identical to the rmoA gene encoded on plasmid R100 (GenBank accession number Y13856.1); rmoA encodes a predicted 69 amino acid putative hemolysin expression-modulation protein that is 100% identical to protein RmoA found on plasmid R100 from *E. coli* [39]. The protein sequence of R100 RmoA exhibits 52% identity and 75% amino acid similarity with Hha protein from *E. coli* K12 (GenBank accession number NP_414993).

**Genome Analysis of sepA-Negative Strains**

Strains C46-10 and C801-09 were representative of the 2 *sepA*-negative nodes that were associated with diarrhea in Figure 1.

**C46-10 Genome Analysis**

Strain C46-10 best represented the confluence of factors identified in node 2 (Figure 1), characterized as *sepA* absent, H33 present, and EAST-1 present. By CGH, strain C46-10 hybridized to a large number of genes found among DEC pathotypes and

---

**Figure 3.** Classification and regression tree analysis described in Figure 1, adding the genes *shiA*, *espy2*, *rmoA* (hemolysin expression-modulation protein). See Figure 1 legend for details of analysis.
**Shigella** species (Table 4) and encodes a complete *yfc* gene cluster, which has been proposed to encode a novel usher-chaperone fimbrial adhesin [40]. C46-10 harbored elements of a type VI secretion system homologous to VgrG of *Agrobacterium*. However, this component of the newly described type VI systems is also found among nonvirulent isolates and has not yet been assigned any virulence function among DEC or *Shigella* strains.

C46-10 DNA was found to hybridize with the 347 amino acid ShiA-like protein from *E. coli* strain REL606 (GenBank accession number YP_003046696). The latter protein exhibited 97% identity with the ShiA protein initially described in *Shigella flexneri* 5a strain M90T (GenBank accession number AF141323) [41]. ShiA and related proteins identified in uropathogenic *E. coli* and *Shigella* strains have been found to suppress the inflammatory response in animal models [42].

**C801-09 Genome Analysis**

C801-09 is closely related to the virulent archetype EAEC strain 042 and harbors many of the same virulence genes, including a near-complete plasmid-borne AggR regulon. It represents the most common serotype found in our study (O153:H30). Like C701-09, C801-09 harbored homologs of a large number of adhesins, including the *siieA* locus that is associated with colonization of cattle by *E. coli* strain UMN026. C801-09 also harbored EspY2, a non-LEE-encoded type III secreted effector from *E. coli* O157:H7 strain TW14359 (GenElD: 8214639). Five proteins, EspY1-5 from the *E. coli* O157:H7 Sakai strain, possess an N-terminal WEXSF domain, which has been linked to type III secretion and is conserved in several well-characterized *Salmonella* effectors and in putative effectors from *Edwardsiella* and *Sodalis* [43].

**Screening of the EAEC Collection for Presence of espY2, rmoA, and shiA Genes**

Based on the CGH analysis from strains C46-10, C701-09, C718-09, and C801-09, we inferred that *espY2*, *rmoA*, and *shiA* were the factors most plausibly associated with virulence. Using PCR, we found that 35.5% of the EAEC strains from cases and a similar percentage from controls harbored the *shiA* gene and 27.3% of each group harbored *espY2*. The *rmoA* gene was found in 43.8% of the EAEC strains (50% from cases and 37.7% from controls). None of the 3 genes were independently associated with diarrheal illness (Table 3). However, when we repeated the CART analysis including the *espY2*, *rmoA*, and *shiA* genes (Figure 3), *sepA* once again exhibited a strong association, yet strains that were both *sepA*- and *rmoA*-positive were most strongly associated with disease (13 out of 14 strains positive for this combination were present among cases).

**DISCUSSION**

EAEC is a common diarrheal isolate, yet apart from those outbreak-associated, identification of truly pathogenic strains remains difficult. A large number of virulence factors and combinations have been associated with clinical illness in epidemiologic studies, and it is possible that either the principal determinants of pathogenicity vary by site and population or that the true determinants have not yet been identified.

We report the most detailed genomic characterization of EAEC performed to date, targeting a collection of 121 EAEC strains isolated from children in Mali with or without moderate to severe diarrhea. In agreement with previous reports [44–46], our strains belonged to a diverse range and combination of O:H and phylogenetic types. Although no specific O:H combination was associated with diarrhea, strains expressing the H33 flagellar antigen were found significantly more often in cases than in controls. This association may signify the existence of a specific set of virulence genes in strains of this H type.

To profile the virulence genes of our strain set, we developed and applied 4 multiplex PCR assays targeting 21 putative virulence genes. We found our EAEC strains to be astonishingly diverse. The only factor associated individually with diarrhea in these analyses was the *Shigella* SPATE toxin *SepA*. Recognizing that pathogenicity represents the concerted action of multiple virulence factors, which can sort independently throughout the *E. coli* population, we assessed combinations of virulence factors using CART analysis. This analysis reinforced the association of *sepA* with diarrhea, independent of any of the other 20 genes scored. We then performed comprehensive genomic analyses on the *sepA*-positive strains using CGH against a reference set of *E. coli* genomes. These studies identified the hemolysin expression-modulating protein RmoA as commonly present in combination with SepA and served to strengthen the association of SepA with clinical illness (Figure 3). Our data demonstrate the importance of strains encoding a combination of virulence factors (here SepA and RmoA), although additional factors may colocalize with these genes.

Among the *sepA*-negative strains, CART analysis suggested 2 combinations of factors that indicate virulent strains (Figure 1). CGH analysis of strains representative of these combinations (strains C801-09 and C46-10) revealed 2 additional factors: T3SS effector EspY2 and ShiA, the latter being associated with modulation of the inflammatory response. However, screening the complete strain set for the presence of these 2 factors, followed by revised CART analysis, did not suggest that these 2 genes strengthened the association with moderate to severe diarrhea.

The association of the toxin EAST-1 with diarrhea only occurred among strains that lacked the majority of the AggR regulon, suggesting that they may require virulence factors not yet apparent; these may occur predominantly in strains harboring flagellar type H33 (Figure 1). EAST-1-positive strains have previously been implicated in pediatric diarrhea [47], so these strains may warrant continued investigation.

The terms *typical EAEC* and *atypical EAEC* have been suggested to refer to EAEC strains harboring or lacking AggR,
respectively. Some studies have demonstrated an association of typical EAEC with diarrhea [5, 48]. We did not observe any correlation of AggR regulon genes with moderate to severe illness in this study. It is possible that our focus on moderate to severe diarrhea overlooks mild illness due to EAEC and that true determinants of pathogenicity are not recognized. Alternatively, illness may be obscured by epidemiologic factors, such as previous exposure. Also, we note that our EAEC definition included two AggR-related genes, potentially introducing strain selection bias.

This study is notable for the association of the Shigella virulence factor SepA with clinical illness, an association that persisted when the effects of other pathogens were considered (OR, 5.6; \( P = .0006; \) data not shown). SepA was first described by Benjelloun-Touimi et al [20] and is a prominent extracellular protein secreted by \( S. \) flexneri strains. SepA is produced during infection [20] and has been shown to confer increased epithelial cell exfoliation from human intestinal explants infected with \( S. \) flexneri [49]. We also note that SepA is produced by the Shiga toxin–producing outbreak strain from Germany in 2011. The previously unsuspected role of SepA in EAEC warrants further investigation.

Leveraging a large epidemiologic study and powerful genomic techniques, our study sheds additional light on the complex nature of diarrheagenic \( E. \) coli genomes and their association with human disease.

Notes

Acknowledgments. We thank Susanne Jesspern and Pernille Gynoese, Statens Serum Institut, for technical support regarding serotyping and Dr Steen Ethelberg for statistical advice.

This paper is dedicated to the memory of Dr Bernadette Baudry, whose seminal description of the EAEC probe provided the first, and still unfulfilled, promise of the PCR method.

Financial support. This study was supported by the Bill & Melinda Gates Foundation, US Public Health Service grant AI33096 to J. P. N. and Danish Council for Strategic Research grant 2101-07-0023 to K. A. K. N. B.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Black RE, Cousens S, Johnson HL, et al. Child Health Epidemiology Reference Group of WHO and UNICEF. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 2010; 375:1969–87.
2. Wanke CA. To know \( E. \) coli is to know bacterial diarrheal disease. Clin Infect Dis 2001; 32:1710–2.
3. Adachi JA, Jiang ZD, Mathewson JJ, et al. Enterogauggerative \( E. \) coli as a major etiologic agent in traveler’s diarrhea in 3 regions of the world. Clin Infect Dis 2001; 32:1706–9.
4. Tompkins DS, Hudson MJ, Smith HR, et al. A study of infectious intestinal disease in England: microbiological findings in cases and controls [see comments]. Commun Dis Public Health 1999; 2:108–13.
5. Okeke IN, Lamikanra A, Czezelin J, Dubovsky F, Kaper JB, Nataro JP. Heterogeneous virulence of enterogauggerative \( E. \) coli strains isolated from children in Southwest Nigeria. J Infect Dis 2000; 181:252–60.
6. Rasko DA, Webster DR, Sahl JW, et al. Origins of the \( E. \) coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med 2011; 365:709–17.
7. Boisen N, Struve C, Scheutz F, Krogfelt KA, Nataro JP. New adhesin of enterogauggerative \( E. \) coli related to the Afa/Dra/AAF family. Infect Immun 2008; 76:3281–92.
8. Bernier C, Gounon P, Le Bouguenec C. Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enterogauggerative \( E. \) coli as a sensitive probe for detecting the AAF-encoding operon family. Infect Immun 2002; 70:4302–11.
9. Nataro JP, Yi Kang D, Giron JA, Savarino SJ, Kothary MH, Hall R. Aggregative adherence fimbria II expression in enterogauggerative \( E. \) coli requires two unlinked plasmid regions. Infect Immun 1993; 61:1126–31.
10. Nataro JP, Deng Y, Maneval DR, German AL, Martin WC, Levine MM. Aggregative adherence fimbriae I of enterogauggerative \( E. \) coli mediate adherence to HEp-2 cells and hemagglutination of human erythrocytes. Infect Immun 1992; 60:2297–304.
11. Elias WP Jr, Czezelin J, Nataro JP, Trabulsi LR, Nataro JP. Organization of biogenesis genes for aggregative adherence fimbria II defines a virulence gene cluster in enterogauggerative \( E. \) coli. J Bacteriol 1999; 181:1779–85.
12. Sheikh J, Czezelin JR, Harrington S, et al. A novel dispersin protein in enterogauggerative \( E. \) coli. J Clin Invest 2002; 110:1329–37.
13. Dudley EG, Thomson NR, Parkhill J, Morin NP, Nataro JP. Proteomic and microarray characterization of the AggR regulon identifies a phoU pathogenicity island in enterogauggerative \( E. \) coli. Mol Microbiol 2006; 61:1267–82.
14. Henderson IR, Hicks S, Navarro-Garcia F, Elias WP, Philips AD, Nataro JP. Involvement of the enterogauggerative \( E. \) coli plasmid-encoded toxin in causing human intestinal damage. Infect Immun 1999; 67:5338–44.
15. Guyer DM, Henderson IR, Nataro JP, Mobley HL. Identification of sat, an autotransporter toxin produced by uropathogenic \( E. \) coli. Mol Microbiol 2000; 38:53–66.
16. Rajakumar K, Sasakawa C, Adler B. Use of a novel approach, termed island probing, identifies the Shigella flexneri she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. Infect Immun 1997; 65:4606–14.
17. Harrington SM, Sheikh J, Henderson IR, Ruiz-Perez F, Cohen PS, Nataro JP. The Pic protease of enterogauggerative \( E. \) coli promotes intestinal colonization and growth in the presence of mucin. Infect Immun 2009; 77:2465–73.
18. Henderson IR, Czezelin J, Eslava C, Noriega F, Nataro JP. Characterization of pic, a secreted protease of Shigella flexneri and enterogauggerative \( E. \) coli. Infect Immun 1999; 67:5587–96.
19. Boisen N, Ruiz-Perez F, Scheutz F, Krogfelt KA, Nataro JP. Short report: high prevalence of serine protease autotransporter cytotoxins among strains of enterogauggerative \( E. \) coli. Am J Trop Med Hyg 2009; 80:294–301.
20. Benjelloun-Touimi Z, Sansonetti PJ, Parsot C. SepA, the major extracellular protein of \( Shigella flexneri \): autonomous secretion and involvement in tissue invasion. Mol Microbiol 1995; 17:123–35.
21. Orskiv F, Orskov I. \( E. \) coli serotyping and disease in man and animals. Can J Microbiol 1992; 38:699–704.
22. Mohamed JA, Huang DB, Jiang ZD, et al. Association of putative enterogauggerative \( E. \) coli virulence genes and biofilm production in isolates from travelers to developing countries. J Clin Microbiol 2007; 45:121–6.
23. Restieri C, Garriss G, Locas MC, Doozins CM. Autotransporter-encoding sequences are phylogenetically distributed among \( E. \) coli clinical isolates and reference strains. Appl Environ Microbiol 2007; 73:1553–62.
24. Sheikh J, Dudley EG, Sui B, Tamboura B, Suleman A, Nataro JP. Ei1A, a HLA-like regulator in enteropathogenic Escherichia coli. Mol Microbiol 2006; 61:338–50.
25. Lloyd AL, Smith SN, Eaton KA, Molineux H. Uropathogenic Escherichia coli suppresses the host inflammatory response via pathogenicity island genes sisa and sib. Infect Immun 2009; 77:5322–33.
26. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the Escherichia coli phylogenetic group. Appl Environ Microbiol 2000; 66:4555–8.
27. Mathews JW, Oberhelman RA, Dupont HL, Javier de la Cabada F, Garibay EV. Enteroadherent Escherichia coli as a cause of diarrhea among children in Mexico. J Clin Microbiol 1987; 25:1917–9.
28. Nataro JP, Baldini MM, Kaper JB, Black RE, Bravo N, Levine MM. Detection of an adherence factor of enteropathogenic Escherichia coli with a DNA probe. J Infect Dis 1985; 152:560–5.
29. Czezulin J, Whittam TS, Henderson IR, Navarro-Garcia F, Nataro JP. Phylogenetic analysis of enteropathogenic and diffusely adherent Escherichia coli. Infect Immun 1999; 67:2692–9.
30. Olsen B, Neumann J, Bottiger B, et al. Etiology of diarrhea in young children in Denmark: a case-control study. J Clin Microbiol 2005; 43:3636–41.
31. Welch RA, Burland V, Plunkett G 3rd, et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. Proc Natl Acad Sci U S A 2002; 99:17020–4.
32. Johnson JR, Scheutz F, Ullerud P, Kuskowski MA, O'Bryan TT, Kuskowski MA. Phylogenetic and pathotypic comparison of concurrent urine and rectal Escherichia coli isolates from men with febrile urinary tract infection. J Clin Microbiol 2005; 43:3895–900.
33. Perna NT, Plunkett G 3rd, Burland V, et al. Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. Nature 2001; 409:529–33.
34. Sahl JW, Lloyd AL, Redman JC, et al. Genomic characterization of asymptomatic Escherichia coli isolated from the neobladder. Microbiology 2011; 157(Pt 4):1088–102.
35. Fang H, Xu J, Ding D, et al. An FDA bioinformatics tool for microbial genomics research on molecular characterization of bacterial foodborne pathogens using microarrays. BMC Bioinformatics 2010; 11(Suppl 6):S4.
36. Kahali S, Sarkar B, Rajendran K, et al. Virulence characteristics and molecular epidemiology of enteropathogenic Escherichia coli isolates from hospitalized diarrheal patients in Kolkata, India. J Clin Microbiol 2004; 42:4111–20.
37. Jiang ZD, Greenberg D, Nataro JP, Steffen R, DuPont HL. Rate of occurrence and pathogenic effect of enteropathogenic Escherichia coli virulence factors in international travelers. J Clin Microbiol 2002; 40:4185–90.
38. Cerna JF, Nataro JP, Estrada-Garcia T. Multiplex PCR for detection of three plasmid-borne genes of enteropathogenic Escherichia coli strains. J Clin Microbiol 2003; 41:2138–40.
39. Niiro JM, Prenafeta A, Miquel E, Torrades S, Juarez A. Sequence, identification and effect on conjugation of the rmpA gene of plasmid R100-1. FEMS Microbiol Lett 1998; 169:59–66.
40. Korea CG, Badouraly R, Prevost MC, Ghigo JM, Beloin C. Escherichia coli K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. Environ Microbiol 2010; 12:1957–77.
41. Ingersoll MA, Moss JE, Weinrauch Y, Fisher PE, Groisman EA, Zychlinsky A. The ShIA protein encoded by the Shigella flexneri SHI-2 pathogenicity island attenuates inflammation. Cell Microbiol 2003; 5:797–807.
42. Ingersoll MA, Zychlinsky A. ShIA abrogates the innate T-cell response to Shigella flexneri infection. Infect Immun 2006; 74:2317–27.
43. Tobe T, Beaton SA, Taniguchi H, et al. An extensive repertoire of type III secretion effectors in Escherichia coli O157 and the role of lambdoid phages in their dissemination. Proc Natl Acad Sci U S A 2006; 103:14941–6.
44. Vial PA, Robins-Browne R, Lior H, et al. Characterization of enteroadherent-aggregative Escherichia coli, a putative agent of diarrheal disease. J Infect Dis 1988; 158:70–9.
45. Qadri F, Haque A, Faruque SM, Bettelheim KA, Robins-Browne R, Albert MJ. Hemagglutinating properties of enteropathogenic Escherichia coli. J Clin Microbiol 1994; 32:510–4.
46. Yamamoto T, Echeverria P, Yokota T. Drug resistance and adherence to human intestines of enteropathogenic Escherichia coli. J Infect Dis 1992; 165:744–9.
47. Vila J, Gene A, Vargas M, Gascon J, Latorre C, Jimenez de Anta MT. A case-control study of diarrhea in children caused by Escherichia coli producing heat-stable enterotoxin (EAST-1). J Med Microbiol 1998; 47:889–91.
48. Sarantuya J, Nishi J, Wakimoto N, et al. Typical enteropathogenic Escherichia coli is the most prevalent pathotype among E. coli strains causing diarrhea in Mongolian children. J Clin Microbiol 2004; 42:133–9.
49. Corun E, Flamant M, Aubert P, et al. Characterisation of early mucosal and neuronal lesions following Shigella flexneri infection in human colon. PLoS One 2009; 4:e713.
50. Baudry B, Savarino SJ, Vial P, Kaper JB, Levine MM. A sensitive and specific DNA probe to identify enteropathogenic Escherichia coli, a recently discovered diarrheal pathogen. J Infect Dis 1990; 161:1249–51.