RESEARCH ARTICLE

Conservation and global distribution of non-canonical antigens in Enterotoxigenic *Escherichia coli*

F. Matthew Kuhlmann[1]*, John Martin[1,2], Tracy H. Hazen[3], Tim J. Vickers[1], Madeline Pashos[1], Pablo C. Okhuysen[4], Oscar G. Gómez-Duarte[5], Elizabeth Cebelinski[6], Dave Boxrud[6], Felipe del Canto[7], Roberto Vidal[7,8], Firdausi Qadri[9], Makedonka Mitreva[1,2], David A. Rasko[3], James M. Fleckenstein[1,10]

1 Department of Medicine, Division of Infectious Diseases, Washington University School of Medicine, St. Louis, Missouri, United States of America, **2** McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri, United States of America, **3** Department of Microbiology and Immunology and Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, United States of America, **4** The Department of Infectious Diseases, Infection Control and Employee Health, University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States of America, **5** Department of Pediatrics, Division of Infectious Diseases, University at Buffalo, The State University of New York, Buffalo, New York, United States of America, **6** Minnesota Department of Health, St. Paul, Minnesota, United States of America, **7** Programa de Microbiología y Micología, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Santiago, Chile, **8** Instituto Milenio de Inmunonología e Inmunoterapia, Facultad de Medicina, Universidad de Chile, Santiago, Chile, **9** Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh, **10** Medicine Service, Veterans Affairs Medical Center, St. Louis, Missouri, United States of America

* fkuhlman@wustl.edu

Abstract

Background

Enterotoxigenic *Escherichia coli* (ETEC) cause significant diarrheal morbidity and mortality in children of resource-limited regions, warranting development of effective vaccine strategies. Genetic diversity of the ETEC pathovar has impeded development of broadly protective vaccines centered on the classical canonical antigens, the colonization factors and heat-labile toxin. Two non-canonical ETEC antigens, the EtpA adhesin, and the EatA mucinase are immunogenic in humans and protective in animal models. To foster rational vaccine design that complements existing strategies, we examined the distribution and molecular conservation of these antigens in a diverse population of ETEC isolates.

Methods

Geographically diverse ETEC isolates (n = 1159) were interrogated by PCR, immunoblotting, and/or whole genome sequencing (n = 46) to examine antigen conservation. The most divergent proteins were purified and their core functions assessed *in vitro*.

Results

EatA and EtpA or their coding sequences were present in 57.0% and 51.5% of the ETEC isolates overall, respectively; and were globally dispersed without significant regional

differences in antigen distribution. These antigens also exhibited >93% amino acid sequence identity with even the most divergent proteins retaining the core adhesin and mucinase activity assigned to the prototype molecules.

Conclusions

EtpA and EatA are well-conserved molecules in the ETEC pathovar, suggesting that they serve important roles in virulence and that they could be exploited for rational vaccine design.

Author summary

Enterotoxigenic *Escherichia coli* (ETEC) are a significant cause of childhood diarrhea in low-middle income countries for which there is currently no vaccine. ETEC vaccine approaches to date have focused on antigens known as colonization factors (CFs) and the pathovar-defining toxins. However, both antigenic diversity of CFs and the lack of characterized CF antigens in many strains has hindered deployment of broadly protective ETEC vaccines. Nevertheless, more recent studies have identified additional ETEC virulence molecules that could compliment canonical vaccine strategies. Here, we demonstrate that two proteins, the EtpA adhesin and EatA mucinase, are present in a diverse, global collection of isolates, and are among the most highly conserved ETEC pathovar-specific antigens to be identified, providing additional incentive to explore their utility in the rational design of broadly protective ETEC vaccines.

Introduction

Enterotoxigenic *Escherichia coli* (ETEC) are a genetically heterogeneous group of pathogens defined by the production of plasmid-encoded heat-labile (LT) and/or heat-stable (ST) enterotoxins [1]. Diarrheal illness caused by ETEC ranges from mild to severe cholera-like disease [2]. ETEC is a significant cause of severe diarrhea in young children of low-middle income countries [3, 4], where it leads to tens of thousands of deaths each year. The burden of illness expands substantially with hundreds of millions of less severe cases of diarrhea attributable to ETEC [5]. Childhood infection with ETEC [5–8] is associated with substantial morbidity and post-diarrheal sequelae including environmental enteropathy, malnutrition, growth stunting, and cognitive impairment [9–12]. The strong association of malnutrition and deaths due to other infectious diseases further amplifies the potential impact of an effective vaccine [4, 12, 13].

Given that a large proportion of the world population remains without ready access to clean water or basic sanitation, vaccines targeting ETEC and other common enteric pathogens, remain a high priority in efforts to prevent childhood diarrheal illness [14]. The canonical approach to ETEC vaccine development is based on a classical virulence paradigm in which the bacteria adhere to the small intestine via plasmid-encoded colonization factors (CFs) [15] allowing delivery of their ST and/or LT enterotoxin payloads causing secretory diarrhea. Although the relationship of anti-CF antibodies to protection following ETEC infections in endemic areas has been difficult to establish [16], controlled human infection studies convincingly demonstrate that passive immunization against CFs, or CF tip adhesin molecules can prevent symptomatic ETEC diarrhea [17]. Although most ETEC appear to encode CFs or CF-

like chaperone-usher-pilus operons [18] 20–50% of isolates lack any of the 25 established and immunologically distinct CFs that have been characterized to date [19, 20], potentially confounding development of a broadly protective vaccine focused exclusively on CFs, and prompting the exploration of additional antigens, including toxoids that could collectively afford broad-based protection [21–23].

Human challenge studies demonstrated that the ETEC H10407 strain cured of a virulence plasmid encoding CFA/I was avirulent, engendering early enthusiasm for targeting CFs [24]. Notably, subsequent genetic and molecular pathogenesis studies identified two additional pathovar-specific, non-canonical virulence loci on this plasmid; *eatA*, which encodes a member of the serine protease autotransporters of *Enterobacteriaceae* (SPATE) family [25], and the *etpBAC* two-partner secretion adhesin locus [26]. EatA has been shown to facilitate interaction with epithelial cells by limiting accumulation of the secreted EtpA adhesin [27] and by degrading MUC2, the major mucin secreted by goblet cells in the intestinal mucosa [28]. EtpA is a large (170kD) extracellular glycoprotein that forms a molecular bridge from the bacterial surface to intestinal epithelia where it interacts with abundant N-acetylgalactosamine (GalNAc) residues of mucosal glycoproteins [29]. Recently, EtpA was shown to interact preferentially with GalNAc residues presented as the terminal glycan on the A blood group, thereby accelerating bacterial adhesion and toxin delivery. This interaction may underlie the enhanced disease severity following ETEC infections in blood group A hosts [6, 30].

To date, studies of EtpA and EatA suggest that these non-canonical antigens could complement classical approaches to ETEC vaccine development. Each of the proteins have been shown to protect against ETEC infection in a murine model of intestinal colonization [28, 31] and are immunologically recognized following experimental human challenge [32] or natural infection with ETEC [33].

Because underlying plasticity of *E. coli* genomes can complicate vaccine development [34], antigenic conservation across a broad representation of isolates is of paramount importance in ETEC antigen selection. Although an earlier study, focused primarily on isolates from Bangladesh, suggested that both the *eatA* and *etpBAC* loci were present in a phylogenetically distributed collection of ETEC isolates [33], further studies are required to assess the global distribution of each locus and the functional conservation of the respective proteins within the genetically diverse ETEC pathovar.

In the present study, we examined the molecular and functional conservation of EatA and EtpA in a large, unbiased collection of validated ETEC strains isolated from cases of symptomatic diarrhea in the Global Enteric Multicenter Study (GEMS) [4] complemented with isolates from regions of the Western Hemisphere not represented in GEMS. The interrogation of this global collection of isolates demonstrates that EtpA and EatA are among the most highly conserved ETEC pathovar-specific virulence molecules discovered to date, findings that should foster further exploration of their role in disease and as protective antigens.

Methods

Ethics statement

All bacterial samples analyzed in these studies were de-identified and obtained from previously archived existing collections outlined in S5 Table.

Sample acquisition

To obtain a diverse and unbiased collection, well-validated ETEC isolates were obtained from a variety of sources. These included 883 ETEC isolates from children with moderate-severe diarrhea in the Global Enteric Multicenter Study (GEMS) obtained through the University of

Maryland [4]. Toxin characterization for these isolates was performed at the study site as previously described [35] and verified using PCR with primers described in S1 Table. CF profiles were determined by Roberto Vidal at Universidad de Chile using polymerase chain reaction as described elsewhere [20]. Eight ETEC isolates were obtained from U.S. and Canadian adults with traveler's diarrhea while visiting Mexico [36], and an additional 40 isolates were provided by the Minnesota Department of Health [37]. Those from Bangladesh (icddr,b) included 171 previously collected isolates [33] and 50 additional isolates collected in the course of routine 2% molecular surveillance studies of acute watery diarrhea conducted at icddr,b. Additional clinical isolates from Chile [38], Colombia [39, 40], and elsewhere had also been previously characterized by PCR and/or western blotting (S2 Table).

Antigen detection by PCR and immunoblotting

ETEC were grown overnight in 3 ml Luria Broth at 37˚C from frozen glycerol stock stored at -80˚C. Total nucleic acids were extracted using the Wizard Genomic DNA Purification Kit (Promega catalog #A1120) according to the provided protocol and resuspended in 200 μl water. PCR was performed using primers in S1 Table and described in supporting methods. Supernatants from the overnight cultures used for nucleic acid extraction underwent trichloroacetic acid precipitation and immunoblotting as previously described [33]. Full details of the immunoblotting protocol are available in supporting methods.

Adjudication of discordant results

Samples with discordant PCR/immunoblot data underwent repeat culture from the same stock and DNA was extracted using phenol/chloroform extraction and ethanol precipitation (supporting methods). Immunoblotting was also performed on the repeat culture. A third assay on a separate culture was performed if repeat testing did not confirm our prior results. If the third assay again gave differing or ambivalent results, such as non-specific PCR products, the isolate was not included in our analysis (S1 Fig). EatA positive samples with negative PCR and positive western blots were included as positive in our final analysis, as negative PCR data may result from minor sequence divergence at the primer binding site(s).

Genome sequencing

A total of 46 EtpA positive strains representing diverse geographic origins, CFs, and toxin profiles were selected for Illumina whole genome sequencing (S3 Table). Phenol/chloroform extraction was performed on pelleted cells from 3 ml overnight cultures; *etpA* presence was confirmed by PCR prior to whole genome sequencing. Automated dual indexed libraries were constructed with 600ng of genomic DNA utilizing the KAPA Hyper PCR-free Library Kit (KAPA Biosystems) on the SciClone NGS instrument (Perkin Elmer) targeting 350bp inserts. The concentration of each library was accurately determined through qPCR (Kapa Biosystems) in order to make 250pM library dilutions for the HiSeqX platform. 2x151bp paired end sequence data generated approximately 2.2 Gb per sample. Sequencing runs were completed according the manufacturer's recommendations (Illumina Inc, San Diego, CA). A subset of strains (S3 Table) underwent individual strain or pooled, indexed PacBio sequencing. SMRTbell libraries were prepared starting with 4ug material per each of 12 samples. Samples were mechanically sheared using the Covaris AFA system and then size selected for 10kb fragments using PacBio's BluePippin system. 500ng per sample of fragmented material (as quantified by Qubit fluorometer) were used to build a single, pooled library. This barcoded library was sequenced on a single SMRT cell using PacBio Sequel 2.0 chemistry. Individual libraries were de-multiplexed from the pool after sequencing using PacBio's SMRT

Link. The H10407 and Jurua 18/11 samples followed the same method through shearing, but then 1ug of 10kb fragmented material was used to generate 2 additional libraries that were sequenced on the PacBio RSII using P6-C4 chemistry and assembled using HGAP4 default settings within the SMRT assembler. A hybrid assembly of the 12 indexed samples was performed using the SPAdes assembler. Trimmomatic cleaned, paired end Illumina data and PacBio Sequel subread data for each sample were fed to the SPAdes assembler using default assembly parameters. Job parameters included '- -threads 4' and '- -memory 29', recruiting 4 cpus and 29Gb of memory per hybrid assembly. These hybrid assemblies proved more contiguous than earlier HGAP 4 assemblies (using default HGAP4 arguments) over the *etpBAC* locus.

Sequence alignment and phylogenetic trees

Whole genome sequence was trimmed using Trimmomatic version 0.36 [41], then mapped to the H10407 p948 plasmid as a reference (GenBank: FN649418.1) using HISAT2 version 2.1.0 (https://ccb.jhu.edu/software/hisat2/index.shtml) using default settings. Consensus sequence was obtained for the H10407 *etpA* gene (GenBank: NC_017724) using the integrated genomics viewer (http://software.broadinstitute.org/software/igv/) default settings [42]. Raw reads from previously published isolates were downloaded from the GenBank's SRA and were mapped to the same reference plasmid (Assembly: ASM21047v1) to confirm the veracity of our method (S3 Table, Bioproject PRJNA526881). Whole genome consensus sequence was uploaded to European Bioinformatics Institute and phylogenetic trees were generated using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and visualized using FigTree version 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Recombinant protein purification

The most divergent strains were selected based on their observed distance in assembled phylogenetic trees described above. jf4927 and jf4894 (S4 Table) expression strains were grown overnight in 100 ml Luria Broth with 100 μg/ml ampicillin and 15 μg/ml chloramphenicol then diluted to 4 L until reaching an OD 600 of 0.6. Expression was induced with 0.0002% (w/v) arabinose for 6 hours at 37˚C. After centrifugation, culture supernatants were concentrated using 100kD Molecular Weight Cutoff tangential-flow filter (Millipore). The secreted his-tagged protein was purified using immobilized metal affinity chromatography as described [30, 43]. Strain jf5003 expressing EatA from strain 700241 was utilized for recombinant protein expression and purification as previously described with the omission of ammonium sulfate precipitation [28].

Functional assays

EtpA binding to blood group A-expressing HT-29 cells and the CRISPR α1-3-*N*-acetylgalacto-saminyltransferase deletion line (IE6), generating an isogenic blood group O line [44], was assessed using immunofluorescence (supporting methods) and fluorescence signal was quantified using Volocity software (v6.3).

The rate of recombinant EatA (30 μg) cleavage of N-Succinyl-Ala-Ala-Pro-Leu p-nitroanilide, a common substrate for serine proteases, was performed as previously described without the inclusion of $ZnCl_2$ [25] while MUC2 cleavage was performed as described [28] (supporting methods).

Statistics

A master database of metadata was maintained in Excel 2010 and imported into SPSS (v24) for analysis including determinations of means and 95% confidence intervals (S2 Table). Chi-square analysis was used to determine differences in *etpA*/EtpA or *eatA*/EatA distribution between geography, study origin, and each colonization factor or toxin. Under or over-representation for plasmid associated genes was determined using binomial distributions in Excel (2016) and adjusted for multiple comparisons with a value of <0.05 determining significance (https://www.sdmproject.com/utilities/?show=FDR).

Results

Global conservation of EatA and EtpA

Rational vaccine design necessitates a thorough understanding of antigenic conservation throughout a diverse population. Given the extraordinary genetic plasticity of *E. coli* [45], the diversity of canonical ETEC target antigens [19, 20], and the worldwide distribution of ETEC, we sought an unbiased and global approach to comprehensively determine the molecular distribution of these candidate antigens. Utilizing isolates collected from the GEMS study [4] which included 7 sites in Africa and Asia, complemented with additional isolates obtained from patients in Bangladesh, Chile [38], Colombia [39, 40], the Minnesota Department of Health [37], and Mexico [36], we employed genetic and immunologic detection methods to determine the distribution of both antigens (S2 Table). There was strong concordance between the identification of either *etpA* or *eatA* genes by PCR with production of the corresponding protein determined by immunoblotting (r = 0.82 for *eatA*/EatA and 0.92 for *etpA*/EtpA). Overall, these antigens were identified in at least half of all isolates with the *eatA* gene and/or EatA protein present in 57.0% (95% CI 54.2–59.9%) and *etpA* gene and/or EtpA protein detected in 51.5% (48.6–54.4%). Combined, 73.3% (70.7–75.8%) of all isolates encode either EatA or EtpA (Table 1).

Geographical variation in ETEC antigen conservation may occur for multiple reasons including founder effects, clonality due to localized epidemics, or sampling within a limited timeframe. However, we found no appreciable difference in the occurrence of either antigen according to geographical distribution (Table 1, *eatA*/EatA *p* = 0.80, *etpA*/EtpA *p* = 0.81, either antigen *p* = 0.42, Chi-squared testing). Similarly, variation between studies could also arise for technical reasons such as differences in isolation procedures, culture or storage techniques, and temporal trends during sample collection. Nevertheless, we found no differences in the distribution of these antigens based on the collection (S5 Table, Chi-square testing for *eatA*/EatA *p* = 0.706, *etpA*/EtpA *p* = 0.214, or either *p* = 0.220).

Table 1. Distribution of coding regions and/or protein expression for EatA and EtpA.

Region	Total Isolates	*eatA* or EatA			*etpA* or EtpA			*eatA*/EatA or *etpA*/EtpA		
		N	%	95%CI	N	%	95%CI	N	%	95%CI
Asia	524	301	57.4	53.2–61.7	278	53.1	48.8–57.3	391	74.6	70.9–78.4
Africa	516	297	57.6	53.3–61.8	258	50.0	45.7–54.3	378	73.3	69.4–77.1
North, Central, and South America	105	55	52.4	42.7–62.1	54	51.4	41.7–61.2	70	66.7	57.5–75.8
Europe or Unknown	14	8	57.1	27.5–86.8	7	50.0	20.0–80.0	10	71.4	44.4–98.5
Total	1159	661	57	54.2–59.9	597	51.5	48.6–54.4	849	73.3	70.7–75.8

*N = Number of positive isolates for specified condition

https://doi.org/10.1371/journal.pntd.0007825.t001

To account for potential plasmid loss that could impact determination of antigen distribution, we verified the presence of genes encoding ETEC-defining toxins (S1 Fig). Eighty-four of the original 1243 isolates (6.8%) were determined not to be ETEC or to have lost their toxin-encoding plasmids, consistent with prior estimates of the frequency of plasmid loss due to culture passage alone [16, 46]. The majority of excluded strains were from the GEMS dataset (77 of 84, 91.7%).

EatA and EtpA conservation relative to colonization factors and toxins

EatA or EtpA could augment current vaccine strategies by expanding antigenic valency and by targeting different virulence mechanisms [47, 48]. To further assess the utility of EatA and EtpA as candidate antigens, we determined their conservation relative to each of the major ETEC colonization factors. In general, we found that either EtpA or EatA were conserved among isolates which expressed major colonization factors (CFA/I, CS1, CS2, CS3, and CS7). However, similar to earlier surveys conducted in Bangladesh [33], EtpA was underrepresented in strains that express CS4, CS5, and CS6, Fig 1). Encouragingly, while 23.9% of strains lacked an identifiable CF, *eatA*/EatA or *etpA*/EtpA were found in nearly half of these isolates (115 of 254, 45.3%), suggesting that these antigens could complement canonical approaches by expanding antigenic coverage. We also examined the distribution of EatA or EtpA in combination with individual CFs having a baseline occurrence of >10% (Fig 2). As predicted, given the underrepresentation of EtpA with Cs6 expressing isolates, the greatest increases are observed when combining CS6 with EtpA. Similar to the canonical CFs [20], we also found that the *eatA* and *etpBAC* loci were more commonly found in ST and ST/LT strains than those producing only LT.

EtpA sequence and functional conservation

The large and globally diverse strain collection affords the opportunity to identify sequence variation that may alter either the function or immunogenic epitopes of EatA and EtpA. We

CF or Toxin		Positive Isolates	(%)	Total Isolates	EatA			EtpA			Either EatA or EtpA			
					N	(%)	95%CI	N	(%)	95%CI	N	(%)	95%CI	
colonization factors	CFA/I	166	(15.6)	1062	150	(90.4)	85.8-94.9	154	(92.8)	88.8-96.8	156	(94.0)	90.3-97.6	100%
	Cs1	53	(5.0)	1062	44	(83.0)	72.6-93.5	46	(86.8)	77.4-96.2	48	(90.6)	82.4-98.7	
	Cs2	59	(5.6)	1062	50	(84.8)	75.3-94.2	55	(93.2)	86.6-99.8	57	(96.6)	91.9-100.0	
	Cs3	118	(11.1)	1062	88	(74.6)	66.6-82.6	110	(93.2)	88.6-97.8	115	(97.5)	94.6-100.0	
	Cs4	16	(1.5)	1062	12	(75.0)	51.2-98.8	2	(12.5)	0-30.7	12	(75.0)	51.2-98.8	
	Cs5	54	(5.1)	1062	47	(87.0)	77.8-96.3	6	(11.1)	2.5-19.8	47	(87.0)	77.8-96.3	
	Cs6	295	(27.8)	1062	208	(70.5)	65.3-75.7	43	(14.6)	10.5-18.6	227	(77.0)	72.1-81.8	
	Cs7	53	(5.0)	1062	50	(94.3)	87.9-100.0	48	(90.6)	82.4-98.7	50	(94.3)	87.9-100.0	
	Cs8	2	(0.9)	215	2	(100.0)	ND	0	(0.0)	ND	2	(100.0)	ND	
	Cs12	25	(2.8)	881	7	(28.0)	9.1-46.9	7	(28.0)	9.1-46.9	8	(32.0)	12.4-51.7	
	Cs14	147	(13.8)	1062	63	(42.9)	34.8-51.0	109	(74.2)	67.0-81.3	126	(85.7)	80.0-91.4	
	Cs17	38	(3.6)	1062	19	(50.0)	33.3-66.7	22	(57.9)	41.5-74.3	22	(57.9)	41.5-74.3	
	Cs18	2	(5.9)	34	1	(50.0)	0-100.0	1	(50.0)	0-100.0	1	(50.0)	0-100.0	
	Cs19	1	(2.9)	34	0	(0.0)	ND	0	(0.0)	ND	0	(0.0)	ND	
	Cs21	278	(26.2)	1062	215	(77.3)	72.4-82.3	212	(76.3)	71.2-81.3	243	(87.4)	83.5-91.3	
	No CF	254	(23.9)	1062	55	(21.7)	16.6-26.8	89	(35.0)	29.1-41.0	115	(45.3)	39.1-51.4	
toxins	LT and ST	432	(37.3)	1159	295	(68.3)	63.9-72.7	205	(47.5)	42.7-52.2	352	(81.5)	77.8-85.2	
	LT Only	326	(28.1)	1159	121	(37.1)	31.8-42.4	118	(36.2)	31.0-41.4	157	(48.2)	42.7-53.6	
	ST Only	401	(34.6)	1159	245	(61.1)	56.3-65.9	274	(68.3)	63.8-72.9	340	(84.8)	81.3-88.3	0%

Fig 1. EatA or EtpA co-existence within ETEC strains expressing unique colonization factors and toxin types. Number for column headings represents the total number positive for the specified condition. Isolates expressing both CFs and the designated non-canonical antigen are depicted.

https://doi.org/10.1371/journal.pntd.0007825.g001

Fig 2. Percentage of ETEC strains expressing unique antigens alone or in combination with each other. The percentage of ETEC strains expressing either antigen is presented. Only CFs found in ≥ 10% of the collection are included. Antigen combinations were determined for isolates expressing either antigen.

https://doi.org/10.1371/journal.pntd.0007825.g002

performed whole genome sequencing (Illumina) of 46 EtpA-expressing strains of which 29 co-expressed EatA and supplemented our collection with sequence available in GenBank (S3 Table). The selected strains represent diverse geographies, toxin profiles, and CFs in an attempt to maximize potential sequence heterogeneity.

Notably, the overall EtpA sequence identity was high (93.6–100%, N = 56) and comparable to sequence diversity of *eltAB* genes, encoding the heat-labile toxin of ETEC [20, 49]. Likewise, when the minor sequence variations in *etpA* sequences were mapped onto the prototype sequence of the molecule from H10407 (Accession NC_017724) and used to generate different phylogenic clusters, there was no clear association of any particular sequence cluster with geographic location (Fig 3). Using whole genome phylogrouping [50], each of the sequenced genomes mapped to phylogroups A and B1 (S2 Fig), where we observed unexpected yet significant overrepresentation of *E. coli* phylogroups within EtpA clusters, suggesting certain phylogroups are permissive to or co-evolved with distinct ETEC plasmids.

Short-read sequencing of *etpA* can be confounded by multiple, large repetitive sequences located at the 3' region of the gene spanning 3.34 kb. Therefore, to confirm the sequence of the *etpA* repeat region, we performed additional long read sequencing (PacBio) of *etpA* from ETEC strains H10407, Jurua 18/11, and 100169. The resulting *de novo* assemblies were then compared to sequence assemblies generated using short-read (Illumina) technology. H10407 and Jurua 18/11 were individually sequenced by PacBio and assembled using HGAP4 default settings with mean read length of 10.8 kb for H10407 and 12.0 kb for Jurua 18/11. Indexed PacBio sequencing of 100169 yielded sub-reads of 3.8 kb as described in materials and methods. Very few differences were observed between the sequences generated by PacBio vs Illumina methodologies with (96.7% of bases identical for 1001696, 98.7 for Jurua 18/11, and 100% for H10407). Collectively, these data suggest that EtpA exhibits very little sequence variation over its geographic distribution.

EtpA from H10407 binds preferentially to N-acetylgalactosamine expressed in the context of A blood group glycans on the intestinal mucosa, likely accounting for the increased disease severity observed in individuals with this blood group [6, 30]. To assess functional conservation of the protein, we cloned the most divergent *etpA* genes, expressed and purified the corresponding recombinant EtpA proteins, and examined their interaction with blood group A glycans. Similar to the H10407 prototype molecule, EtpA from these divergent strains exhibited preferential binding to blood group A carbohydrates expressed on the surface of enterocytes (Fig 4), suggesting functional conservation across a broad representation of ETEC.
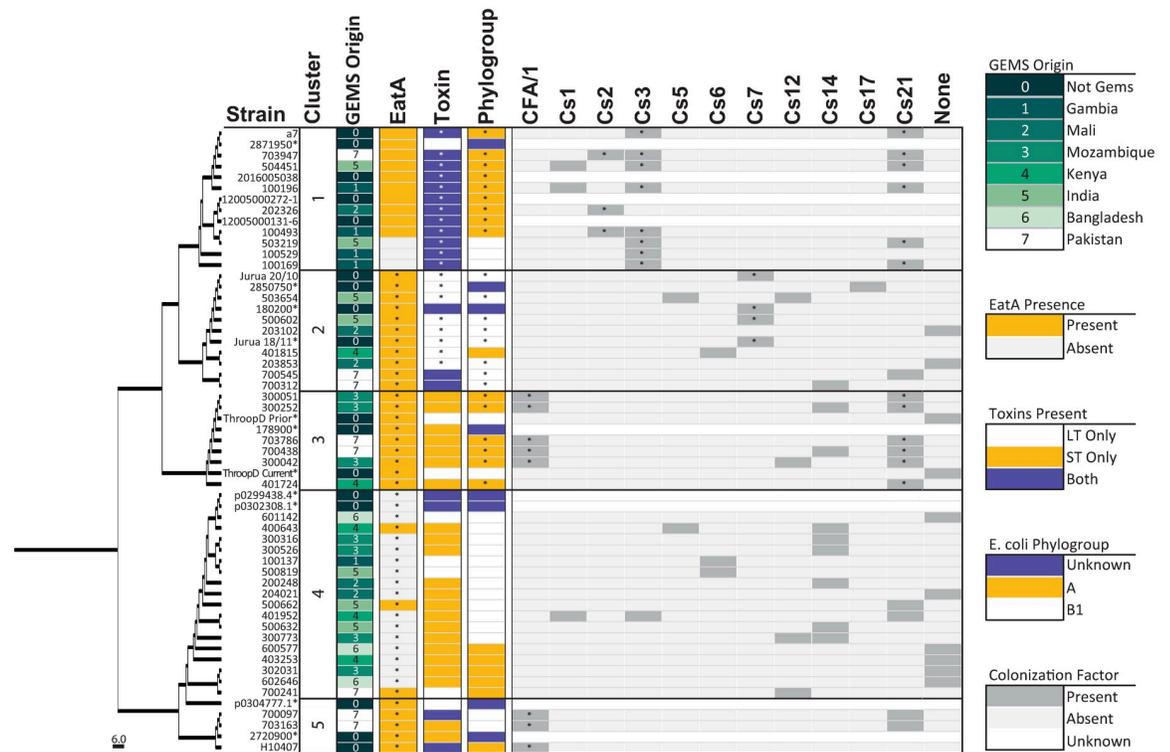
Fig 3. Phylogenetic distribution of EtpA sequences and associated metadata. Five clusters of EtpA were identified using Clustal Omega and FigTree software. Isolates marked with an asterisk denote sequences downloaded from the NCBI as external references (S3 Table). Within boxes, asterisks represent significant under (*e.g.*, EatA absence in cluster 4) or over representation of the specified variable between clusters as calculated using binomial distributions for each variable. For CFs, "None" designates no CF was identified by PCR. Unknown means we have no data to support the presence or absence of a CF or were not included in the phylogroup analysis. Significant *p*-values were adjusted for multiple comparisons (* = $p < 0.05$).

https://doi.org/10.1371/journal.pntd.0007825.g003

EatA sequence and functional conservation

We also assessed potential sequence and functional differences in EatA expressing isolates. The percent amino acid identities for EatA ranged from 94.3 to 100% (N = 38) relative to the H10407 reference sequence. The EatA sequences cluster in a similar fashion to EtpA, a finding we expect as both genes are encoded on the same plasmid in H10407 (Fig 5). Importantly, evaluation of the predicted EatA protein sequence demonstrated complete conservation of the catalytic residues comprised of H134, D162, and S267 (S3 Fig). Similarly, comparison of five available *eatA* sequences from *etpA*-negative genomes in Genbank (S3 Table) [51] demonstrated >96.6% sequence identity and retention of the catalytic triad.

The most divergent *eatA* sequence was cloned and recombinant passenger domain of EatA (rEatAp) was then purified to assess proteolytic activity. We first determined the proteolytic activity of EatA using the synthetic peptide Suc-Ala-Ala-Pro-Leu (AAPL) coupled to p-nitroanilide which when enzymatically cleaved produces the readily detectable yellow indicator, 4-nitroaniline [25]. We found that similar to EatA from H10407, rEatAp from ETEC strain 700241 cleaved AAPL, (Fig 6A). Likewise, rEatAp derived from H10407 and 700241 degraded MUC2 (Fig 6B), providing further evidence that the minor degrees of sequence divergence observed within the predicted EatA peptide sequence did not affect functional activity of the enzyme.
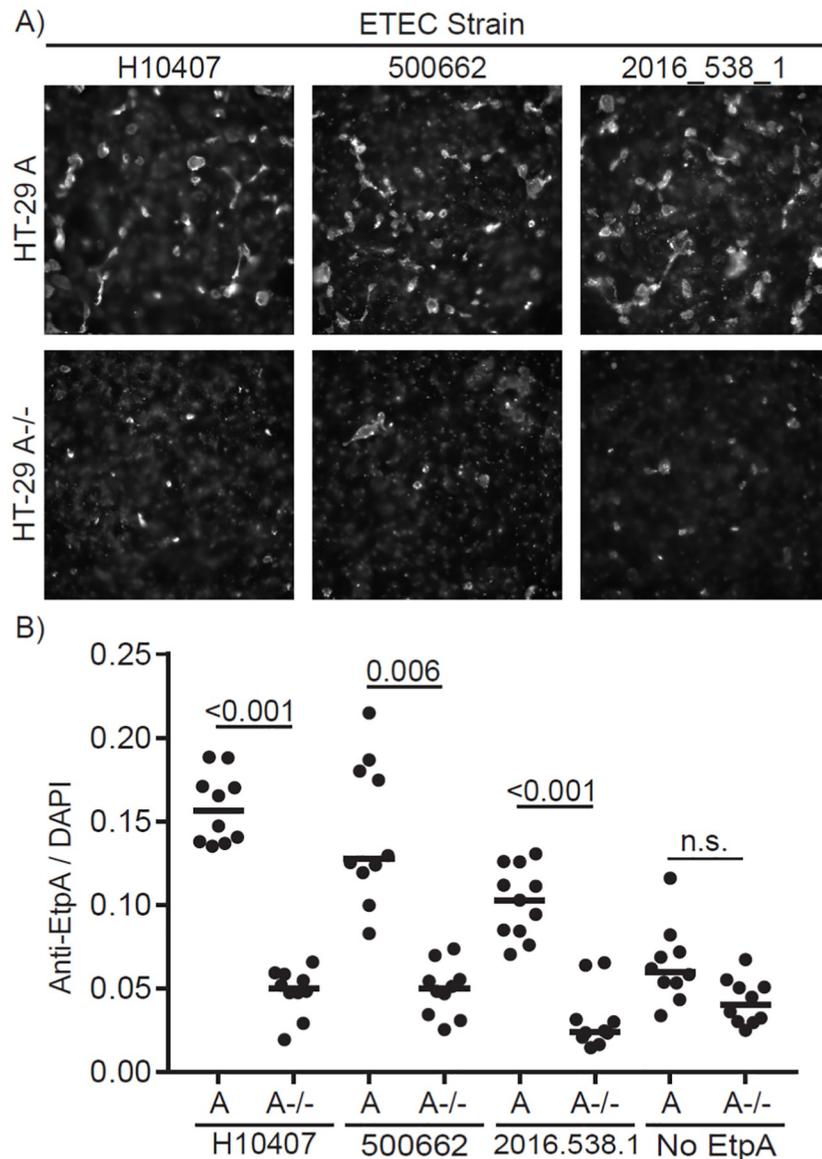
Fig 4. EtpA function is conserved. (A) Immunofluorescence microscopy of EtpA bound to HT-29 WT cells expressing Blood Group A sugars or HT29 A-/- CRISPR deletion mutant generating functional blood group O cells, EtpA detected with anti-EtpA antibodies. (B) Quantitation of mean fluorescent values normalized to DAPI (nuclei) signal using Volocity software. Statistical differences determined by Kruskal-Wallis testing followed by Dunn's test for multiple comparisons with $p < 0.05$ considered significant.

Discussion

The recognized diversity of canonical target molecules and the potential complexity inherent in formulating broadly protective ETEC vaccines has driven the pursuit of novel strategies [52, 53]. Emerging pathogenesis studies suggest that additional surface-expressed antigens could complement ETEC canonical CF-centered approaches. Genetic loci encoding two secreted virulence proteins not currently targeted in ETEC vaccines, EatA and EtpA, were initially discovered in ETEC H10407 on the same plasmid that encodes CFA/I [54]. Studies thus far have demonstrated that these antigens contribute to ETEC virulence, are highly immunogenic during natural and experimental infections, and afford protection against infection in an animal
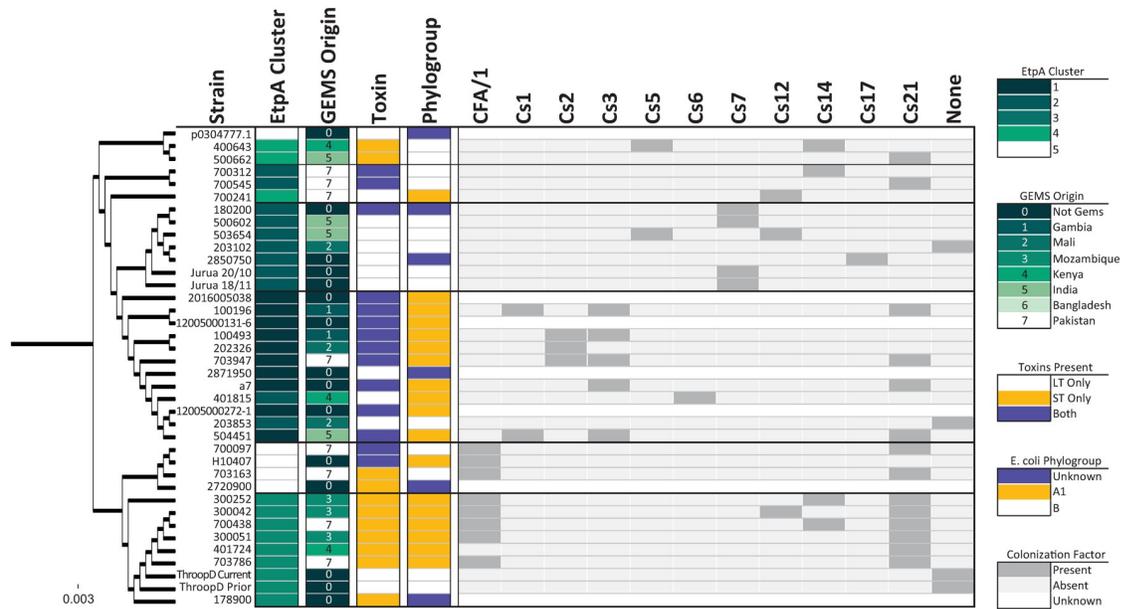
Fig 5. Phylogenetic distribution of EtpA sequences and associated metadata. Phylogenetic differences between EatA isolates and associated metadata using the same methods as in Fig 1. No statistics were determined for this data set due to small numbers and inherent bias in the data.
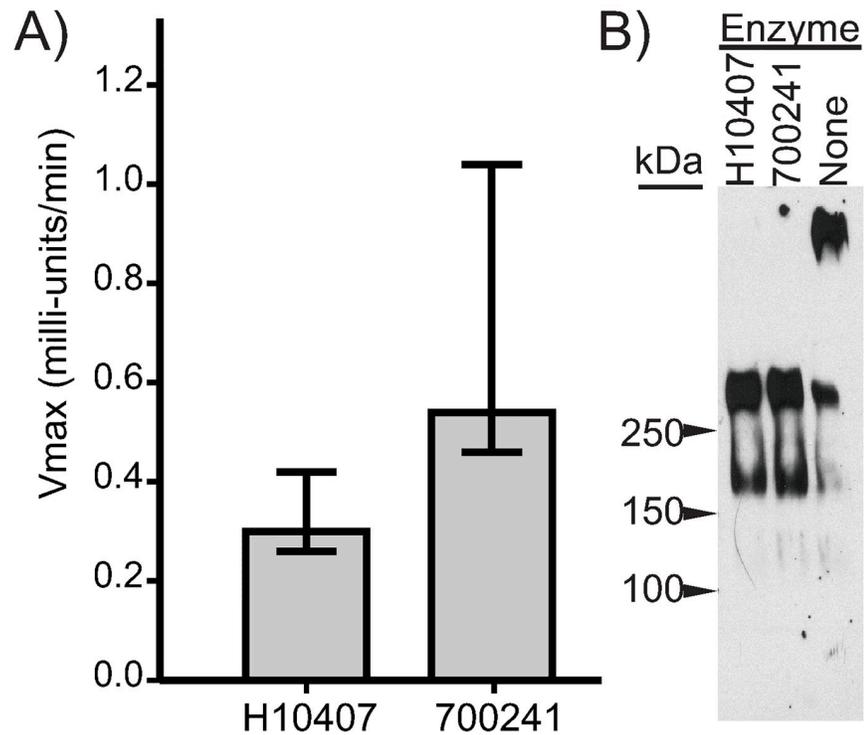
Fig 6. Mucinase activity of EatA is retained. (A) Purified recombinant EatA degrades the synthetic protein Suc-Ala-Ala-Pro-Leu-p-anilimide in a colorimetric assay measuring the rate of release of p-anilimide. 2 independent experiments with 3 replicates total, medians with 95% CI are shown, (p = 0.1, Mann-Whitney U). (B) Degradation of MUC2 by recombinant EatA from H10407 and strain 700241. Purified EatA and MUC2 from LS174T cells (~540 kDa) were incubated together with mucin degradation (~300 kDa) assessed by western blotting, 2 independent experiments performed.

model [28, 31–33]. These promising initial results led to current efforts to ensure that these antigens are sufficiently conserved across isolates to warrant further investigation of their utility as vaccine antigens. As molecular conservation remains an essential benchmark in evaluating candidate antigens, the present studies were performed to provide more accurate estimates of the global distribution and molecular sequence conservation of EtpA and EatA. To avoid limitations of earlier studies derived from one geographic location or biased collections where the intent is to maximize diversity [33, 55–57], we interrogated well-validated ETEC isolates of known provenance from symptomatic illness collected over a broad geographic distribution. Our findings affirm the conclusions from the prior studies and expand the knowledge base to more diverse geographies.

The studies reported here suggest that both EtpA and EatA are among the most highly conserved ETEC pathovar-specific antigens described to date. Moreover, by sequencing a geographically and phenotypically diverse subset of ETEC isolates we found that both antigens exhibit substantial molecular sequence conservation and retain core virulence functions across a range of isolates. Ongoing work seeks to identify conserved and protective epitopes for rational vaccine design. Collectively, these features could simplify rational vaccine design.

No ETEC pathovar-specific protein, including either the LT or ST toxins that define ETEC, is universally conserved in every strain. Because ST-producing strains appear to predominate among ETEC associated with symptomatic illness, recent studies have focused on the prevalence of canonical vaccine target antigens among ST+ETEC. Similar to canonical CF targets, we also find that *eatA* and *etpA* are more frequently associated with ST-producing strains relative to LT-only isolates. Both loci were originally identified on the p948 plasmid in ETEC H10407 which encodes *eatA*, *etpA*, and *sta2* (STh), while the p666 plasmid encodes *sta1* (STp) and genes for LT (*eltA* and *eltB*). Although it is possible that the *eatA* and *etpBAC* loci are similarly linked on plasmids in other strains, the resolution of present sequencing data is not sufficient to support the contention that these loci are inextricably linked to those encoding STh.

It is important to examine the findings of the present study relative to other potential target molecules. Data from recent immunoproteome studies have demonstrated that ETEC infection induces mucosal antibody responses to EatA, EtpA, as well as canonical ETEC vaccine antigens [32]. Notably, these studies also detected immune responses to highly conserved, chromosomally encoded proteins, such as flagellin [58], making them potentially attractive candidate antigens that could protect against multiple *E. coli* pathovars in addition to ETEC. Whether vaccination with these core antigens, which are shared with the small population of commensal *E. coli*, would adversely impact the human intestinal microbiota is currently unknown [59]. Likewise, it is not known whether the immune response to these conserved proteins that follows natural infections with ETEC could contribute to non-diarrheal sequelae linked to these infections. At present, selectively targeting a compilation of pathovar-specific antigens could offer a rational approach to a vaccine that affords broad-based coverage.

The ideal strategy for incorporation of these antigens into the development of a highly efficacious and broadly protective vaccine will need to be determined. However, EtpA and EatA could be adopted in a multivalent approach that targets complementary virulence characteristics and extends antigenic valency. Similarity in virulence strategies between ETEC and *Bordetella pertussis* [60] may offer an important template for rational vaccine development. Both pathogens produce toxins which induce cyclic nucleotides (pertussis toxin vs LT/ST), utilize two-partner secretion hemagglutinin exoprotein adhesin molecules (filamentous hemagglutinin [FHA] vs EtpA), fimbrial adhesins (pertussis fimbriae vs CFs), and surface expressed or secreted autotransporter molecules important for virulence (pertactin vs EatA). Present acellular pertussis vaccines that incorporate toxoids, fimbriae, pertactin, and FHA [61] could therefore provide a valuable paradigm for the rational design of ETEC vaccines that combine novel

antigens highlighted here with emerging LT and ST toxoids, and CFs selected to achieve broad coverage against a diverse population of ETEC.

The broad representation of EatA and EtpA in a diverse and geographically distributed population of ETEC isolated from symptomatic cases of diarrheal illness would appear to support recent *in vitro*, and animal studies, as well as controlled human infection model data suggesting that these antigens play an important role in the molecular pathogenesis of disease. These findings combined with the established immunogenicity of these proteins should encourage further examination of their role as protective antigens.

Supporting information

S1 Supporting methods.
(PDF)

S1 Table. Primers used in this publication.
(PDF)

S2 Table. Database of isolates analyzed in this study.
(XLSX)

S3 Table. List of sequenced isolates and their GenBank IDs.
(XLSX)

S4 Table. Bacterial strains and plasmids in this publication.
(PDF)

S5 Table. Antigen distribution by study origin.
(PDF)

S1 Fig. Flow diagram depicting exclusion of samples. Isolates not meeting PCR criteria for inclusion are defined as having non-specific amplification of toxins despite a minimum of 3 assays. Isolates were determined not to be ETEC if they lacked detectable toxin in the original report and on repeat testing at Washington University. All of these prior isolates were reported to express ETEC specific CFs without toxins. The remaining 51 strains lacked detectable toxin at Washington University, suggesting loss of plasmid. UoB is University of Buffalo (Colombian isolates), MNDOH is Minnesota Department of Health.
(PDF)

S2 Fig. Phylogenomic analysis of EtpA-expressing ETEC strains sequenced in this study compared with a collection of previously sequenced ETEC reference genomes and diverse *E. coli* and *Shigella* genomes. A maximum likelihood phylogeny was generated from 217,248 SNPs relative to the genome of *E. coli* IAI39 as a reference. The newly sequenced EtpA-expressing ETEC genomes are indicated in red. *E. coli* phylogroups (A, B1, B2, D, E, and F) are indicated on the interior of the phylogeny, while ETEC phylogenomic lineages (L1 through L21) that contain an EtpA-expressing isolate are indicated on the exterior of the phylogeny. The presence of *eatA* and *etpA* is indicated by blue and orange squares (see inset legend). Bootstrap values ≥80 are indicated by a gray circle over each respective node.
(PDF)

S3 Fig. EatA sequence alignment. Sequence alignment for EatA expressing isolates (S3 Table) included in this study. Arrows highlight the conserved catalytic triad H124, D164 (page 2), and S267 (page 3). The EatA passenger domain is annotated below the sequence on pages 1 and 10.
(PDF)

Acknowledgments

The authors would like to thank Sharon Tennat for helping to provide the GEMS isolates and all thee GEMS investigators for their tireless efforts. We also appreciate Bruce Rosa's assistance in the statistical analysis of the genomic data.

Author Contributions

Conceptualization: F. Matthew Kuhlmann, Makedonka Mitreva, James M. Fleckenstein.

Data curation: F. Matthew Kuhlmann, John Martin, Pablo C. Okhuysen, Oscar G. Gómez-Duarte, Elizabeth Cebelinski, Dave Boxrud, Felipe del Canto, Roberto Vidal, Firdausi Qadri.

Formal analysis: F. Matthew Kuhlmann, John Martin.

Funding acquisition: F. Matthew Kuhlmann, Makedonka Mitreva, James M. Fleckenstein.

Investigation: F. Matthew Kuhlmann, Tim J. Vickers, Madeline Pashos.

Methodology: F. Matthew Kuhlmann, David A. Rasko, James M. Fleckenstein.

Project administration: F. Matthew Kuhlmann, James M. Fleckenstein.

Resources: Pablo C. Okhuysen, Oscar G. Gómez-Duarte, Elizabeth Cebelinski, Dave Boxrud, Felipe del Canto, Roberto Vidal, Firdausi Qadri.

Software: John Martin, Tracy H. Hazen, David A. Rasko.

Supervision: Makedonka Mitreva, David A. Rasko, James M. Fleckenstein.

Validation: F. Matthew Kuhlmann.

Visualization: F. Matthew Kuhlmann, Tracy H. Hazen, David A. Rasko, James M. Fleckenstein.

Writing – original draft: F. Matthew Kuhlmann, James M. Fleckenstein.

Writing – review & editing: F. Matthew Kuhlmann, Tim J. Vickers, Felipe del Canto, Roberto Vidal, Makedonka Mitreva, David A. Rasko, James M. Fleckenstein.

References

1. Fleckenstein JM, Kuhlmann FM. Enterotoxigenic Escherichia coli Infections. Curr Infect Dis Rep. 2019; 21(3):9. https://doi.org/10.1007/s11908-019-0665-x PMID: 30830466.

2. Levine MM. Escherichia coli that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. J Infect Dis. 1987; 155(3):377–89. https://doi.org/10.1093/infdis/155.3.377 PMID: 3543152.

3. Liu J, Platts-Mills JA, Juma J, Kabir F, Nkeze J, Okoi C, et al. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. Lancet. 2016; 388(10051):1291–301. https://doi.org/10.1016/S0140-6736(16)31529-X PMID: 27673470.

4. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet. 2013; 382(9888):209–22. https://doi.org/10.1016/S0140-6736(13)60844-2 PMID: 23680352.

5. Kotloff KL, Nasrin D, Blackwelder WC, Wu Y, Farag T, Panchalingham S, et al. The incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal episodes among infants and children residing in low-income and middle-income countries: a 12-month case-control study as a follow-on to the Global Enteric Multicenter Study (GEMS). Lancet Glob Health. 2019; 7(5):e568–e84. https://doi.org/10.1016/S2214-109X(19)30076-2 PMID: 31000128.

Non-canonical ETEC antigen conservation

6. Qadri F, Saha A, Ahmed T, Al Tarique A, Begum YA, Svennerholm AM. Disease burden due to enterotoxigenic Escherichia coli in the first 2 years of life in an urban community in Bangladesh. Infect Immun. 2007; 75(8):3961–8. https://doi.org/10.1128/IAI.00459-07 PMID: 17548483.

7. Black RE, Brown KH, Becker S. Effects of diarrhea associated with specific enteropathogens on the growth of children in rural Bangladesh. Pediatrics. 1984; 73(6):799–805. Epub 1984/06/01. PMID: 6374599.

8. Platts-Mills JA, Taniuchi M, Uddin MJ, Sobuz SU, Mahfuz M, Gaffar SA, et al. Association between enteropathogens and malnutrition in children aged 6–23 mo in Bangladesh: a case-control study. Am J Clin Nutr. 2017; 105(5):1132–8. Epub 2017/04/07. https://doi.org/10.3945/ajcn.116.138800 PMID: 28381477.

9. Collaborators GBDCoD. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017; 390(10100):1151–210. https://doi.org/10.1016/S0140-6736(17)32152-9 PMID: 28919116.

10. DALYs GBD, Collaborators H. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017; 390 (10100):1260–344. https://doi.org/10.1016/S0140-6736(17)32130-X PMID: 28919118.

11. Denno DM, Tarr PI, Nataro JP. Environmental Enteric Dysfunction: A Case Definition for Intervention Trials. Am J Trop Med Hyg. 2017; 97(6):1643–6. https://doi.org/10.4269/ajtmh.17-0183 PMID: 29016294.

12. Anderson JDt, Bagamian KH, Muhib F, Amaya MP, Laytner LA, Wierzba T, et al. Burden of enterotoxigenic Escherichia coli and shigella non-fatal diarrheal infections in 79 low-income and lower middle-income countries: a modelling analysis. Lancet Glob Health. 2019; 7(3):e321–e30. https://doi.org/10.1016/S2214-109X(18)30483-2 PMID: 30784633.

13. Schaible UE, Kaufmann SH. Malnutrition and infection: complex mechanisms and global impacts. PLoS Med. 2007; 4(5):e115. https://doi.org/10.1371/journal.pmed.0040115 PMID: 17472433.

14. Hosangadi D, Smith PG, Kaslow DC, Giersing BK, Who E, Shigella Vaccine Consultation Expert G. WHO consultation on ETEC and Shigella burden of disease, Geneva, 6-7th April 2017: Meeting report. Vaccine. 2018. https://doi.org/10.1016/j.vaccine.2017.10.011 PMID: 29352598.

15. Gaastra W, Svennerholm AM. Colonization factors of human enterotoxigenic Escherichia coli (ETEC). Trends Microbiol. 1996; 4(11):444–52. https://doi.org/10.1016/0966-842x(96)10068-8 PMID: 8950814.

16. Steinsland H, Valentiner-Branth P, Gjessing HK, Aaby P, Molbak K, Sommerfelt H. Protection from natural infections with enterotoxigenic Escherichia coli: longitudinal study. Lancet. 2003; 362(9380):286–91. https://doi.org/10.1016/S0140-6736(03)13971-2 PMID: 12892959.

17. Savarino SJ, McKenzie R, Tribble DR, Porter CK, O'Dowd A, Cantrell JA, et al. Prophylactic Efficacy of Hyperimmune Bovine Colostral Antiadhesin Antibodies Against Enterotoxigenic Escherichia coli Diarrhea: A Randomized, Double-Blind, Placebo-Controlled, Phase 1 Trial. J Infect Dis. 2017; 216(1):7–13. Epub 2017/05/26. https://doi.org/10.1093/infdis/jix144 PMID: 28541500.

18. Sahl JW, Sistrunk JR, Baby NI, Begum Y, Luo Q, Sheikh A, et al. Insights into enterotoxigenic Escherichia coli diversity in Bangladesh utilizing genomic epidemiology. Sci Rep. 2017; 7(1):3402. Epub 2017/06/15. https://doi.org/10.1038/s41598-017-03631-x PMID: 28611468.

19. von Mentzer A, Tobias J, Wiklund G, Nordqvist S, Aslett M, Dougan G, et al. Identification and characterization of the novel colonization factor CS30 based on whole genome sequencing in enterotoxigenic Escherichia coli (ETEC). Sci Rep. 2017; 7(1):12514. https://doi.org/10.1038/s41598-017-12743-3 PMID: 28970563.

20. Vidal RM, Muhsen K, Tennant SM, Svennerholm A-M, Sow SO, Sur D, et al. Colonization factors among enterotoxigenic Escherichia coli isolates from children with moderate-to-severe diarrhea and from matched controls in the Global Enteric Multicenter Study (GEMS). PLOS Neglected Tropical Diseases. 2019; 13(1):e0007037. https://doi.org/10.1371/journal.pntd.0007037 PMID: 30608930

21. Bourgeois AL, Wierzba TF, Walker RI. Status of vaccine research and development for enterotoxigenic Escherichia coli. Vaccine. 2016; 34(26):2880–6. https://doi.org/10.1016/j.vaccine.2016.02.076 PMID: 26988259.

22. Qadri F, Ahmed F, Ahmed T, Svennerholm AM. Homologous and cross-reactive immune responses to enterotoxigenic Escherichia coli colonization factors in Bangladeshi children. Infect Immun. 2006; 74 (8):4512–8. https://doi.org/10.1128/IAI.00474-06 PMID: 16861637.

23. Taxt AM, Diaz Y, Aasland R, Clements JD, Nataro JP, Sommerfelt H, et al. Towards Rational Design of a Toxoid Vaccine against the Heat-Stable Toxin of Escherichia coli. Infect Immun. 2016; 84(4):1239–49. https://doi.org/10.1128/IAI.01225-15 PMID: 26883587.

24. Satterwhite TK, Evans DG, DuPont HL, Evans DJ Jr. Role of Escherichia coli colonisation factor antigen in acute diarrhoea. Lancet. 1978; 2(8082):181–4. https://doi.org/10.1016/s0140-6736(78)91921-9 PMID: 78384.

25. Patel SK, Dotson J, Allen KP, Fleckenstein JM. Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic Escherichia coli. Infect Immun. 2004; 72(3):1786–94. https://doi.org/10.1128/IAI.72.3.1786-1794.2004 PMID: 14977988.

26. Fleckenstein JM, Roy K, Fischer JF, Burkitt M. Identification of a two-partner secretion locus of enterotoxigenic Escherichia coli. Infect Immun. 2006; 74(4):2245–58. https://doi.org/10.1128/IAI.74.4.2245-2258.2006 PMID: 16552055.

27. Roy K, Kansal R, Bartels SR, Hamilton DJ, Shaaban S, Fleckenstein JM. Adhesin degradation accelerates delivery of heat-labile toxin by enterotoxigenic Escherichia coli. J Biol Chem. 2011; 286(34):29771–9. https://doi.org/10.1074/jbc.M111.251546 PMID: 21757737.

28. Kumar P, Luo Q, Vickers TJ, Sheikh A, Lewis WG, Fleckenstein JM. EatA, an immunogenic protective antigen of enterotoxigenic Escherichia coli, degrades intestinal mucin. Infect Immun. 2014; 82(2):500–8. https://doi.org/10.1128/IAI.01078-13 PMID: 24478066.

29. Kumar P, Kuhlmann FM, Bhullar K, Yang H, Vallance BA, Xia L, et al. Dynamic Interactions of a Conserved Enterotoxigenic Escherichia coli Adhesin with Intestinal Mucins Govern Epithelium Engagement and Toxin Delivery. Infection and immunity. 2016; 84(12):3608–17. Epub 2016/11/01. https://doi.org/10.1128/IAI.00692-16 PMID: 27736776.

30. Kumar P, Kuhlmann FM, Chakroborty S, Bourgeois AL, Foulke-Abel J, Tumala B, et al. Enterotoxigenic Escherichia coli blood group A interactions intensify diarrheal severity. J Clin Invest. 2018; 125 (5):3298–311. https://doi.org/10.1172/JCI97659 PMID: 29771685.

31. Luo Q, Vickers TJ, Fleckenstein JM. Immunogenicity and Protective Efficacy against Enterotoxigenic Escherichia coli Colonization following Intradermal, Sublingual, or Oral Vaccination with EtpA Adhesin. Clin Vaccine Immunol. 2016; 23(7):628–37. https://doi.org/10.1128/CVI.00248-16 PMID: 27226279.

32. Chakraborty S, Randall A, Vickers TJ, Molina D, Harro CD, DeNearing B, et al. Human Experimental Challenge With Enterotoxigenic Escherichia coli Elicits Immune Responses to Canonical and Novel Antigens Relevant to Vaccine Development. J Infect Dis. 2018; 218(9):1436–46. https://doi.org/10.1093/infdis/jiy312 PMID: 29800314.

33. Luo Q, Qadri F, Kansal R, Rasko DA, Sheikh A, Fleckenstein JM. Conservation and immunogenicity of novel antigens in diverse isolates of enterotoxigenic Escherichia coli. PLoS Negl Trop Dis. 2015; 9(1): e0003446. https://doi.org/10.1371/journal.pntd.0003446 PMID: 25629897.

34. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J Bacteriol. 2008; 190(20):6881–93. Epub 2008/08/05. https://doi.org/10.1128/JB.00619-08 PMID: 18676672.

35. Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, Farag TH, van Eijk A, et al. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. Clin Infect Dis. 2012; 55 Suppl 4:S232–45. https://doi.org/10.1093/cid/cis753 PMID: 23169936.

36. Mohamed JA, DuPont HL, Flores J, Palur H, Nair P, Jiang ZD, et al. Single nucleotide polymorphisms in the promoter of the gene encoding the lipopolysaccharide receptor CD14 are associated with bacterial diarrhea in US and Canadian travelers to Mexico. Clin Infect Dis. 2011; 52(11):1332–41. https://doi.org/10.1093/cid/cir228 PMID: 21596674.

37. Medus C, Besser JM, Juni BA, Koziol B, Lappi V, Smith KE, et al. Long-Term Sentinel Surveillance for Enterotoxigenic Escherichia coli and Non-O157 Shiga Toxin-Producing E. coli in Minnesota. Open Forum Infect Dis. 2016; 3(1):ofw003. https://doi.org/10.1093/ofid/ofw003 PMID: 26913288.

38. Montero D, Vidal M, Pardo M, Torres A, Kruger E, Farfan M, et al. Characterization of enterotoxigenic Escherichia coli strains isolated from the massive multi-pathogen gastroenteritis outbreak in the Antofagasta region following the Chilean earthquake, 2010. Infect Genet Evol. 2017; 52:26–9. https://doi.org/10.1016/j.meegid.2017.04.021 PMID: 28442437.

39. Gomez-Duarte OG, Romero-Herazo YC, Paez-Canro CZ, Eslava-Schmalbach JH, Arzuza O. Enterotoxigenic Escherichia coli associated with childhood diarrhoea in Colombia, South America. J Infect Dev Ctries. 2013; 7(5):372–81. https://doi.org/10.3855/jidc.2667 PMID: 23669426.

40. Guerra JA, Romero-Herazo YC, Arzuza O, Gomez-Duarte OG. Phenotypic and genotypic characterization of enterotoxigenic Escherichia coli clinical isolates from northern Colombia, South America. Biomed Res Int. 2014; 2014:236260. https://doi.org/10.1155/2014/236260 PMID: 24877071.

41. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404.

42. Cavener DR. Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucleic Acids Res. 1987; 15(4):1353–61. https://doi.org/10.1093/nar/15.4.1353 PMID: 3822832.

43. Fleckenstein JM, Roy K. Purification of recombinant high molecular weight two-partner secretion proteins from Escherichia coli. Nat Protoc. 2009; 4(7):1083–92. https://doi.org/10.1038/nprot.2009.87 PMID: 19707189.

44. Kuhlmann FM, Santhanam S, Kumar P, Luo Q, Ciorba MA, Fleckenstein JM. Blood Group O-Dependent Cellular Responses to Cholera Toxin: Parallel Clinical and Epidemiological Links to Severe Cholera. Am J Trop Med Hyg. 2016; 95(2):440–3. https://doi.org/10.4269/ajtmh.16-0161 PMID: 27162272.

45. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of Escherichia coli: Comparative genomic analysis of E-coli commensal and pathogenic isolates. Journal of Bacteriology. 2008; 190(20):6881–93. https://doi.org/10.1128/JB.00619-08 PMID: 18676672

46. Levine MM, Ristaino P, Sack RB, Kaper JB, Orskov F, Orskov I. Colonization factor antigens I and II and type 1 somatic pili in enterotoxigenic Escherichia coli: relation to enterotoxin type. Infect Immun. 1983; 39(2):889–97. PMID: 6131869.

47. Smith J, Lipsitch M, Almond JW. Vaccine production, distribution, access, and uptake. Lancet. 2011; 378(9789):428–38. https://doi.org/10.1016/S0140-6736(11)60478-9 PMID: 21664680.

48. Melvin JA, Scheller EV, Miller JF, Cotter PA. Bordetella pertussis pathogenesis: current and future challenges. Nature Reviews Microbiology. 2014; 12:274. https://doi.org/10.1038/nrmicro3235 PMID: 24608338

49. Joffre E, von Mentzer A, Abd El Ghany M, Oezguen N, Savidge T, Dougan G, et al. Allele variants of enterotoxigenic Escherichia coli heat-labile toxin are globally transmitted and associated with colonization factors. J Bacteriol. 2015; 197(2):392–403. https://doi.org/10.1128/JB.02050-14 PMID: 25404692.

50. Sahl JW, Beckstrom-Sternberg SM, Babic-Sternberg J, Gillece JD, Hepp CM, Auerbach RK, et al. The In Silico Genotyper (ISG): an open-source pipeline to rapidly identify and annotate nucleotide variants for comparative genomics applications. bioRxiv. 2015:015578. https://doi.org/10.1101/015578

51. Hazen TH, Nagaraj S, Sen S, Permala-Booth J, Del Canto F, Vidal R, et al. Genome and Functional Characterization of Colonization Factor Antigen I- and CS6-Encoding Heat-Stable Enterotoxin-Only Enterotoxigenic Escherichia coli Reveals Lineage and Geographic Variation. mSystems. 2019; 4(1). https://doi.org/10.1128/mSystems.00329-18 PMID: 30944874.

52. Fleckenstein JM. Providing Structure to Enterotoxigenic Escherichia coli Vaccine Development. J Infect Dis. 2017; 216(1):1–3. https://doi.org/10.1093/infdis/jix146 PMID: 28541516.

53. Poole ST, McVeigh AL, Anantha RP, Lee LH, Akay YM, Pontzer EA, et al. Donor strand complementation governs intersubunit interaction of fimbriae of the alternate chaperone pathway. Mol Microbiol. 2007; 63(5):1372–84. https://doi.org/10.1111/j.1365-2958.2007.05612.x PMID: 17302815.

54. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, Cunningham AF, et al. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407. J Bacteriol. 2010; 192(21):5822–31. https://doi.org/10.1128/JB.00710-10 PMID: 20802035.

55. Bhakat D, Debnath A, Naik R, Chowdhury G, Deb AK, Mukhopadhyay AK, et al. Identification of common virulence factors present in enterotoxigenic Escherichia coli isolated from diarrheal patients in Kolkata, India. J Appl Microbiol. 2018; 126(1):255–65. https://doi.org/10.1111/jam.14090 PMID: 30153365.

56. Del Canto F, Valenzuela P, Cantero L, Bronstein J, Blanco JE, Blanco J, et al. Distribution of classical and nonclassical virulence genes in enterotoxigenic Escherichia coli isolates from Chilean children and tRNA gene screening for putative insertion sites for genomic islands. J Clin Microbiol. 2011; 49(9):3198–203. https://doi.org/10.1128/JCM.02473-10 PMID: 21775541.

57. von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, et al. Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution. Nat Genet. 2014; 46(12):1321–6. https://doi.org/10.1038/ng.3145 PMID: 25383970.

58. Roy K, Hamilton D, Ostmann MM, Fleckenstein JM. Vaccination with EtpA glycoprotein or flagellin protects against colonization with enterotoxigenic Escherichia coli in a murine model. Vaccine. 2009; 27 (34):4601–8. Epub 2009/06/16. https://doi.org/10.1016/j.vaccine.2009.05.076 PMID: 19523914.

59. Naili I, Vinot J, Baudner BC, Bernalier-Donadille A, Pizza M, Desvaux M, et al. Mixed mucosal-parenteral immunizations with the broadly conserved pathogenic Escherichia coli antigen SslE induce a robust mucosal and systemic immunity without affecting the murine intestinal microbiota. Vaccine. 2019; 37 (2):314–24. Epub 2018/12/07. https://doi.org/10.1016/j.vaccine.2018.10.008 PMID: 30503655.

60. Melvin JA, Scheller EV, Miller JF, Cotter PA. Bordetella pertussis pathogenesis: current and future challenges. Nat Rev Microbiol. 2014; 12(4):274–88. Epub 2014/03/13. https://doi.org/10.1038/nrmicro3235 PMID: 24608338.

61. Gustafsson L, Hallander HO, Olin P, Reizenstein E, Storsaeter J. A controlled trial of a two-component acellular, a five-component acellular, and a whole-cell pertussis vaccine. N Engl J Med. 1996; 334 (6):349–55. https://doi.org/10.1056/NEJM199602083340602 PMID: 8538705.