



Published in final edited form as:

*Methods Mol Biol.* 2016 ; 1403: 363–383. doi:10.1007/978-1-4939-3387-7\_19.

## Overcoming Enterotoxigenic *Escherichia coli* Pathogen Diversity: Translational Molecular Approaches to Inform Vaccine Design

James M. Fleckenstein and David A. Rasko

### Abstract

Enterotoxigenic *Escherichia coli* (ETEC) are a genetically diverse *E. coli* pathovar that share in the ability to produce heat-labile toxin and/or heat-stable toxins. While these pathogens contribute substantially to the burden of diarrheal illness in developing countries, at present, there is no suitable broadly protective vaccine to prevent these common infections. Most vaccine development attempts to date have followed a classical approach involving a relatively small group of antigens. The extraordinary underlying genetic plasticity of *E. coli* has confounded the antigen valency requirements based on this approach. The recent discovery of additional virulence proteins within this group of pathogens, as well as the availability of whole-genome sequences from hundreds of ETEC strains to facilitate identification of conserved molecules, now permits a reconsideration of the classical approaches, and the exploration of novel antigenic targets to complement existing strategies overcoming antigenic diversity that has impeded progress toward a broadly protective vaccine. Progress to date in antigen discovery and methods currently available to explore novel immunogens are outlined here.

### Keywords

*Escherichia coli*; Enterotoxigenic; Subunit vaccines; Bacterial genome; Antigenic diversity

## 1 Introduction

### 1.1 Global Importance of Enterotoxigenic *Escherichia coli*

Enterotoxigenic *E. coli* (ETEC) are among the leading causes of diarrheal illness worldwide. These organisms are particularly prevalent in developing countries where basic sanitation and clean water are often limited. Here, these pathogens preferentially affect young children, many of whom continue to succumb to rapid dehydration resulting from severe diarrheal illness [1]. ETEC infections occur following ingestion of contaminated food or water, and these infections have emerged in recent years in the form of large-scale food-borne outbreaks in the industrialized countries including the USA, presumably due to importation of imported food.

### 1.2 Toxins Define the Enterotoxigenic *E. coli* Pathovar

The enterotoxigenic *E. coli* are a diverse collection of pathogens that are defined by the production of at least one of the three diarrheagenic toxins known as heat-labile toxin (LT), or the heat-stable toxins (STh, and STp) [2]. Heat-labile toxin shares approximately 80 %

molecular identity with cholera toxin and both toxins activate production of the second messenger cAMP in target intestinal epithelial cells. Both heat-stable toxins, like the native human intestinal peptide guanylin, bind to the extracellular portion of guanylate cyclase C to stimulate the production of cGMP. Both cAMP [3] and cGMP [4–6] stimulate protein kinases that phosphorylate the cystic fibrosis transmembrane regulatory channel (CFTR), thereby enhancing export of chloride ions into the intestinal lumen. Concurrent inhibition of sodium hydrogen ion exchange results in a net loss of NaCl and water into the intestinal lumen with ensuing watery diarrhea.

All three toxin genes are usually encoded on extrachromosomal plasmids and are frequently flanked by insertion sequences (IS), implicating mobile genetic elements in evolution of these genetically diverse pathogens. Indeed, it has been suggested that the diversity of the ETEC pathotype of diarrheagenic *E. coli* is driven largely by widespread dissemination of toxin gene encoding plasmids among a diverse collection of *E. coli* host strains [7].

### 1.3 The Challenge Posed by Pathogen Diversity

In essence, it appears that a diverse population of *E. coli* can potentially serve as effective hosts for production of these plasmid-encoded toxins. This is perhaps best exemplified by the diversity of serotypes that are represented among the ETEC pathovar [8]. While some specific O and H serotypes are more common, more than 75 O-antigen serogroups and more than 50 different H serogroups are represented in ETEC. A single study of 100 ETEC strains in Egypt identified 59 different O:H combinations [9]. Large-scale whole-genome sequencing of ETEC and other pathogenic *E. coli*, as well as some commensal strains, has provided some additional insight into the nature of this diversity [10, 11]. When considered in general, the *E. coli* pangenome, or collection of all genes present in those genomes sequenced to date, is quite large [12]. Remarkably, as each new genome sequence is analyzed, an estimated 300 unique genes will be added to this “open” pangenome [10, 12]. The appreciable underlying genetic plasticity of *E. coli* coupled with horizontal transfer of essential virulence genes on mobile elements may suggest that the ETEC pathovar will constantly evolve as toxin genes and other essential features are transferred into new host backgrounds.

### 1.4 Limits on ETEC Diversity Imposed by Key Virulence Requirements

Despite the underlying plasticity of *E. coli* genomes, there are two constraints imposed on ETEC genome content. First, all *E. coli* have at their core a collection of approximately 2200 genes that are mostly involved in essential metabolic functions of these organisms [10, 12]. This core subset of genes, common to all *E. coli*, are at least theoretically devoid of viable vaccine targets since they are largely shared with commensal strains that are present in and a component of the natural human gastrointestinal microbiota.

The requirement for other virulence traits in addition to the toxins themselves imposes another potential constraint on pathogen evolution that is more relevant to antigen discovery, and vaccine target selection. While the genes encoding the known enterotoxins define the ETEC pathovar, genes that encode additional features required for effective delivery of toxin payloads to cognate receptors on the epithelial surface can essentially serve as diversity

checkpoint. ETEC must survive ingestion, navigate to lumen of the small intestine, and ultimately interact directly with intestinal enterocytes [13] bringing the pathogen in close proximity to epithelial cell surface receptors for the delivery of LT and ST. The relatively small number of pathovar-specific features in addition to the known toxins suggests that there are only a limited number of ways in which this can be accomplished [10]. Within this context, it should be possible then to discover relatively conserved virulence molecules that are either exclusive to ETEC or shared with other diarrheagenic pathovars, but not commensal strains.

### 1.5 Immunologic and Structural Diversity of Major Vaccine Targets

The best-studied antigens of ETEC to date are the colonization factors (CFs), plasmid-encoded structures that have been a major focus of vaccine development. Perhaps the best published data in support of CFs as vaccine targets come from passive [14, 15] immunization studies demonstrating anti-CF antibodies administered orally to afford significant protection against experimental challenge with homologous strains of ETEC. A number of active vaccination studies with different vaccine ETEC constructs have yielded significant increases in anti-CF antibodies; however overall protection afforded by these vaccines has varied [16–18]. Nevertheless, emerging data suggest that protection can be improved by targeting individual CF fimbrial subunits particularly tip adhesin structures [19, 20].

A variety of different fimbrial, fibrillar, long pilus [21] and small linear fiber [22] colonization factor structures have been described thus far. A formidable challenge to vaccine development based exclusively on colonization factors has been the significant antigenic heterogeneity exhibited by the CFs. Although more than 25 different antigenically distinct CFs have been described to date [23, 24], it has been estimated that approximately 40–50 % of ETEC strains do not make one of these established CFs [25, 26].

### 1.6 Identification of Novel Vaccine Antigens in ETEC

There is no single method for identification of putative vaccine antigens in ETEC, and the technology for antigen identification is evolving rapidly. The approach outlined here encompasses a variety of complementary methods that are currently being used to identify and validate candidate vaccine targets for ETEC. As shown in Fig. 1, these include classical genetic approaches, genomics, and whole-genome sequencing, transcriptomics, proteomics, and immunoproteomics using protein microarray technology. Each of these methodologies has specific merits as well as disadvantages to consider in developing antigen discovery platforms.

### 1.7 Classical Genetic Approaches

Transposon mutagenesis has been widely used in bacterial pathogenesis studies to identify virulence genes involved in pathogen-host interactions. Tn *phoA*, originally introduced by Manoil and Beckwith [27] 30 years ago, has been extensively used for identification of surface antigens in Gram-negative pathogens, including the EatA autotransporter which is secreted by ETEC [28]. Briefly, Tn *phoA* incorporates a truncated version of the gene for alkaline phosphatase (lacking a signal peptide-encoding region) within the Tn5 transposable

element. Transposon “hops” which result in gene fusions of *phoA* with those encoding surface antigens can be detected by colony screening on agar containing the antibiotic in the transposon and an alkaline phosphatase indicator. Original versions of Tn *phoA* included the transposase within the transposable element. This occasionally created problems in attempting to clone and sequence the insertions due to subsequent rearrangements or continued transposition. To avoid these problems and to establish a system where insertions could be easily cloned and sequenced, pTnphoA.ts was developed by placing the transposase gene on a temperature-sensitive suicide plasmid outside of the inverted repeats of the Tn5 transposable element [29] (Fig. 2). This temperature-sensitive “plasposon” has been used to identify a number of candidate antigens in ETEC including EtpA [29]. Because the temperature-sensitive origin of replication is contained within the transposed element, the region of insertion can easily be identified by restriction endonuclease digestion and re-ligation of the target DNA into a recombinant temperature-sensitive plasmid for isolation and subsequent sequencing as outlined in the protocol below.

## 2 Materials

### 2.1 TnphoA.ts Mutagenesis Materials

1. DH10BT1 (pTnphoA.ts) available from Fleckenstein laboratory.
2. Target ETEC strain.
3. Electroporation apparatus, cuvettes.
4. Ampicillin.
5. Luria agar base: Tryptone (1 %), yeast extract (0.5 %), agar (1.5 %).
6. Luria agar plates containing ampicillin at final concentration of 100 µg/ml.
7. 5-Bromo-4-chloro-3-indolyl-phosphate (*p*-toluidine salt) (XP) prepared as a 20 mg/ml stock solution in N,N-dimethylformamide.
8. Gene fusion indicator plates: XP final concentration 40 µg/ml, ampicillin 100 µg/ml (protect from light).
9. Sterile wooden toothpicks.
10. Restriction endonuclease(s), buffer.
11. Genomic DNA preparation kits (e.g., MasterPure DNA Purification Kit, Epicentre).

### 2.2 In Vitro Analysis Materials

1. Gastrointestinal cell line(s) which support bacterial adhesion and/or toxin delivery assays (*see* Table 3) seeded into 96-well tissue culture plates.

2. Luria broth (LB).
3. Sterile culture tubes (Falcon 2059 or equivalent).
4. Sterile Hanks' Balanced Salt Solution (HBSS) (e.g., Life Technologies or equivalent containing Ca<sup>2+</sup>, Mg<sup>2+</sup>).
5. Triton-X-100 [0.1 %] sterile solution in PBS.
6. Luria agar plates.
7. Cyclic nucleotide assay kits.
  - a. cAMP (e.g., Arbor Assays K019).
  - b. cGMP (e.g., Arbor Assays K020).

### 2.3 In Vivo Intestinal Colonization and Vaccine Testing Materials

1. Strain with antibiotic resistance marker in a permissive location in the genome to facilitate counter selection. (jf876 which contains a kanamycin resistance cassette in the *lacZTA* locus that does not contribute to colonization) [13].
2. Mice (adult 5–8-week-old females, e.g., CD-1, Charles River), *n* = 20–30 (at a minimum ten mice will be needed for adjuvant-only controls and ten for adjuvant + vaccine antigen group comparisons).
3. Purified antigen (amount will depend on the route of vaccination).
4. Adjuvant appropriate for route of immunization.
5. Autoclaved food and bedding material.
6. Luria agar culture plates containing antibiotic (e.g., kanamycin (25 µg/ml)).
7. Water containing streptomycin 5 g/L.
8. Famotidine (sterile) for I.P. injection of mice.
9. Saponin 5 % sterile solution in PBS.

## 3 Methods

### 3.1 pTnphoA.ts Plasposon Transformation Steps

1. Grow DH10BT1 (pTnphoA.ts) overnight at 30 °C in Luria broth containing ampicillin, 100 µg/ml.
2. Isolate plasmid DNA.
3. Transform target strain by electroporation.

4. Select transformants overnight at 30 °C, on plates containing ampicillin 100 µg/ml.
5. Pool multiple transformants into single tube (e.g., Falcon 2059) containing 2 ml of Luria broth with ampicillin 100 µg/ml.
6. Grow overnight at 30 °C, and dilute 1:2 in glycerol freezing media.
7. Transformant mix can be preserved at –80 °C for future use.

### 3.2 Generation of Tn *pho A*.ts Mutants

1. Grow transformant mixture or glycerol stock from above at 30 °C overnight in Luria broth containing ampicillin 100 µg/ml.
2. The following morning dilute 1:100 in fresh media (ampicillin 100 µg/ml) and grow with shaking (250 rpm) for 90 min at 30 °C.
3. Plate 100 µl of dilutions (~1:100) onto fresh gene fusion indicator plates and grow at 37 °C overnight.
4. Isolate individual blue colonies from indicator plates with sterile toothpicks and streak-purify onto fresh indicator agar.
5. Incubate at 37 °C overnight.
6. Select isolated, blue colonies for growth overnight in 2 ml of Luria broth containing ampicillin 100 µg/ml.
7. Use 1 ml of overnight culture for isolation of genomic DNA (follow the manufacturer's protocol).
8. Preserve the remaining 1 ml of overnight culture as a frozen glycerol stock.

### 3.3 Tn *pho A*.ts Mutant Screening

Because the Tn5-based transposition occurs largely at random resulting in the production of (at least) thousands of independent mutants, mutagenesis ideally can be coupled with a relatively high-throughput in vitro phenotypic assay. The results of these phenotypic screening assays can then be used to direct cloning and identification of insertion sites.

### 3.4 Identification of Transposon Insertion Sites

Because the transposition element contains a temperature-sensitive origin of replication in addition to the beta lactamase gene between the inverted repeats, recovery of DNA regions flanking the insertion is relatively straightforward.

1. Isolate total genomic DNA (plasmid and chromosomal DNA). (Most commercial genomic DNA preparation kits, e.g., Wizard® Genomic DNA Purification, Promega work well with ETEC strains.)
2. Digest an aliquot of genomic DNA with a restriction endonuclease that does not cut between the inverted repeats of the transposon (for instance, *MluI*).
3. Ligate the DNA with T4 DNA ligase.
4. Transform ligation mixture into commercially available ampicillin-sensitive *E. coli cloning strain* (e.g., DH10BT1, DH5α, Top10), selecting on Luria agar containing ampicillin, 100 µg/ml, at 30 °C.
5. Grow isolated ampicillin-resistant colonies in Luria broth overnight at 30 °C.
6. Isolate plasmid DNA using commercially available plasmid preparation kit that yields DNA of sufficient quality for sequencing.
7. Set up separate sequencing reactions with the primers TnphoA.179 (5′-CC ATCCCATCGCCAATCA-3′) and TnphoA.ts1 (5′-CGAAATTAATACGACTCA-3′).
8. Resulting DNA sequence information can then be used in BLASTN or BLASTX program searches of NCBI databases (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to identify potential homologues.

### 3.5 De Novo Identification of Vaccine Antigens from Whole-Genome Sequences

The cost and time required for sequencing and assembling entire bacterial genomes have declined dramatically since the genome of *Haemophilus influenza* was first assembled now over 20 years ago [30]. This has permitted both the de novo identification of candidate antigens and a recent assessment of the conservation of known antigens and putative vaccine targets in multiple ETEC strains from different geographic regions isolated over time [31].

One approach to the identification of candidate vaccine antigens is the “*in silico*” interrogation of genome sequence data using algorithms or programs that select molecules which have at least a theoretical likelihood of being exposed on the surface of the organism or secreted. These exposed or surface-expressed molecules are at least in principle amenable to neutralization by vaccination. Complex multifactorial investigations involving multiple genomes require training and experience in bioinformatics; however individual genomes can be interrogated using fairly simple Web-based interfaces.

In addition to identification of surface molecules, the approach to characterization of potential novel ETEC vaccine antigens involves an assessment of the degree to which these

antigens are unique to ETEC genomes and to which they are shared with commensal strains. By definition, all ETEC strains make at least one of the known toxins (LT, STh, and/or STp). However, there is no other universally shared antigen that is common to all ETEC that is not also represented in the rest of *E. coli* including the nonpathogenic commensal strains that make up a small portion of the microbiome of most humans [32]. The challenge therefore lies in defining appropriate vaccine targets among the population of antigens that are uniquely pathovar/ETEC associated.

1. General approaches to defining potential vaccine antigens from genome data.

Traditional identification of candidate antigens has relied on empirical microbial pathogenesis studies or genetic techniques outlined above to define surface features that could be exploited in vaccines. However, with the advent of high-throughput whole-genome sequencing, it is now possible to use “reverse vaccinology” [33, 34] to identify candidate antigens by *in silico* interrogation of data from multiple ETEC genomes. To some extent, draft or complete genomes can be interrogated by those without specific training in informatics to identify putative vaccine antigens with publicly available Web-based platforms (Table 1). However, application of these and other algorithms on a broader scale, that potentially involves hundreds of genomes, will certainly require more extensive bioinformatic capabilities. Both approaches follow the same general scheme outlined in Fig. 3.

2. Identification of pathovar-specific features.

The scheme for *in silico* identification of potential candidates essentially involves two main tasks. The first is to identify features of ETEC genomes that are relatively pathovar specific, but which are not shared with commensal *E. coli* strains. A potential pitfall of this analysis is that there are at present a very limited number of true commensal isolates from healthy humans for which DNA sequence data are available (Table 2). Nevertheless, the many *E. coli* genes that provide for essential metabolic functions and core structural elements of these organisms can be digitally subtracted from pathogen genomes. Ideally, candidates would be shared broadly among a diverse population of ETEC. To this end, genome data from several hundred ETEC strains are presently available [11, 31, 35] permitting pan-ETEC genome comparative analyses to identify features that are relatively conserved in this pathovar.



3. Identification of putative vaccine antigens from genome-subtracted data.  
  
Following the identification of conserved, pathovar-specific features, the next major task is to identify those antigens that are potentially surface expressed, and/or which share features in common with known vaccine antigens using the algorithms outlined in Table 1 and Fig. 3. These approaches are complementary. The inclusion of molecules with motifs or domains conserved in other vaccine antigens permits the investigator to capture putative antigens where surface expression may not be obvious, or alternatively to prioritize antigens that share features with effective vaccine targets.

### 3.6 Preclinical Antigen Validation In Vivo

**3.6.1 General Approach to In Vivo Studies**—One of the problems facing investigators hoping to develop effective enteric vaccines is the lack of a small animal model that faithfully recapitulates the nature of the illness in humans. Mice do not develop diarrhea with any of the common enteric pathogens that infect humans, including enterotoxigenic *E. coli*, even at high doses that typically cause serious illness in volunteers. Nevertheless, mice do become colonized with ETEC following oral (gavage) challenge with inocula as small as  $10^3$ – $10^4$  colony-forming units, permitting a straightforward assessment of the impact of vaccination with candidate antigens on intestinal colonization, a critical step in pathogenesis [36]. Studies to date in this model have revealed that colonization of the small intestine is really a very complex phenotype involving a variety of different virulence factors in addition to the fimbrial structures that have been the traditional targets for ETEC vaccines [37–40], thereby affording additional approaches to vaccine development to overcome the limitations of CF-based vaccines. The basic protocol for mouse vaccination studies follows:

1. (~Day 7) Acquire and acclimate mice (at least 1 week prior to experimentation).
2. (Day 0) Vaccinate mice ( $n = 10$ ) with adjuvant and antigen (dose will depend on adjuvant, and route of administration). Vaccinate an equal number with adjuvant-only as control group.
3. (Day 14) Administer first booster vaccination.
4. (Day 28) Administer second booster vaccination.
5. (Day 40) Add streptomycin (5 g/L) to drinking water.
6. (Day 41) Remove streptomycin, and return to regular drinking water. Grow challenge strain overnight.
7. (Day 42) Challenge with  $\sim 10^4$ – $10^5$  colony-forming units of ETEC by gavage. Plate dilutions of inoculum onto selective media.

8. (Day 43) Sacrifice mice, harvest segments of small intestine in saponin, and plate undiluted,  $10^{-1}$ ,  $10^{-2}$ , dilutions onto selective media.
9. Determine cfu/mouse in vaccinated and control groups.

### 3.7 Preclinical Antigen Studies In Vitro

**3.7.1 Bacterial Adhesion Assays**—Effective interaction with intestinal epithelial cells is a key event in ETEC pathogenesis. Bacterial adhesion assays in which bacteria are added to intestinal epithelial cells cultured in vitro have become a mainstay of ETEC molecular pathogenesis investigations [29, 37, 41–44]. Despite the simplicity of these assays, in which bacteria remaining attached to epithelial cells are quantitated after a finite period of incubation, they have been instrumental in characterizing the role of a number of essential virulence factors, including several different adhesins. Likewise, they have been used in a number of preclinical vaccinology studies, where antibodies raised against specific surface antigens are tested for their ability to mitigate bacterial-host interactions [39, 45, 46]. The basic ETEC adhesion assay follows:

1. Plate target epithelial cell line (e.g., Caco-2) into 96-well tissue culture-treated plates.
2. Incubate at 37 °C, 5 % CO<sub>2</sub> to establish confluent monolayers.
3. Inoculate 2 ml of LB media in 15 ml round-bottom tube with frozen glycerol stock of the ETEC testing strain (e.g., H10407).
4. Incubate overnight at 37 °C, 200 rpm.
5. Dilute 1:100 into 2 ml of fresh LB; grow for ~90' to mid-logarithmic phase growth.
6. Immediately prior to addition of bacteria, add antibody against target antigen (s).
7. Inoculate tissue culture wells with 1–2 µl of bacteria per well.
8. Return plate to tissue culture incubator for 1 h.
9. During incubation, plate dilutions of inoculum onto Luria agar.
10. After 1 h, remove plate from tissue culture incubator and wash 4–5× with HBSS, 100 µl/well.
11. Lyse epithelial cells with 0.1 % Triton-x-100 for 5 min.
12. Plate dilutions of lysate in PBS onto Luria agar.
13. The following day count inoculum and output colonies.

14. Express results as % cell-associated bacteria (recovered cfu/input cfu  $\times$  100).

**3.7.2 Toxin Delivery Assays**—ETEC delivery of heat-labile and/or heat-stable enterotoxins, which, respectively, activate cAMP and cGMP production in target intestinal epithelial cells, is the *sine qua* nonvirulence feature that defines this pathovar. Therefore, much can be learned in detailed investigation of the molecular events that culminate in bacterial activation of these cyclic nucleotides. In vitro studies using intestinal epithelial cell lines (Table 3) can be used to investigate the efficiency with which mutant strains lacking candidate virulence features deliver LT and/or ST enterotoxin payloads. Consequently, cAMP and cGMP assays also provide a convenient surrogate marker to gauge the effectiveness of antibodies to individual candidate antigens or to a combination of targets [46] in abrogating toxin delivery. A basic protocol for assessing delivery of the respective toxins follows:

1. Plate target epithelial cell line (e.g., Caco-2) into 96-well tissue culture-treated plates.
2. Incubate at 37 °C, 5 % CO<sub>2</sub>, to establish confluent monolayers.
3. Inoculate 2 ml of LB media in 15 ml round-bottom tube with frozen glycerol stock of the ETEC testing strain (e.g., H10407).
4. Incubate overnight at 37 °C, 200 rpm.
5. Dilute 1:100 into 2 ml of fresh LB; grow for ~90' to mid-logarithmic phase growth.
6. Immediately prior to addition of bacteria, add antibody against target antigen (s).
7. Inoculate tissue culture wells with 1–2  $\mu$ l of bacteria per well.
8. Return plate to tissue culture incubator for ~3 h.
9. Wash three times with pre-warmed tissue culture media.
10. Return to incubator for an additional 2 h.
11. Wash gently with HBSS.
12. Process cells for cyclic nucleotide quantitation following instructions provided in assay.

## Acknowledgments

The work described was supported by Grant Number 2R01AI089894 from the National Institute of Allergy and Infectious Diseases (NIAID), the PATH Enteric Vaccine Initiative (EVI), the Bill and Melinda Gates Foundation (OPP1099494), and the Department of Veterans Affairs. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

## Notes

1. TnphoA.ts mutagenesis notes
  - a. While resident plasmids of ETEC will co-purify with chromosomal DNA in commercial genomic DNA isolation kits, this is not true of commercial plasmid preparation kits. ETEC plasmids are often quite large and do not purify easily with most commercial plasmid kits.
  - b. Re-ligation will generally favor intramolecular ligation; therefore the resulting plasmid should have a single unique restriction site joining the flanking regions. This can be confirmed by restriction endonuclease digestion.
2. Sample strategy: Identification of EtpA as a vaccine antigen.
 

The investigation of EtpA as a candidate vaccine antigen to date has involved many of the strategies outlined above. Therefore, in the following section, we use EtpA to illustrate the application of the different bioinformatic algorithms to vaccine candidate selection.

  - a. Identification of conserved, pathovar-specific antigens
    -

C  
o  
m  
p  
a  
r  
i  
s  
o  
n  
o

f  
s  
e  
q  
u  
e  
n  
c  
e  
o  
f  
t  
h  
e  
l  
a  
r  
g  
e  
p  
l  
a  
s  
m  
i  
d  
o  
f  
E  
T  
E  
C  
H  
1  
0  
4  
0  
7  
(  
p  
9  
4  
8  
)  
(

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

h  
t  
t  
p  
:  
/  
/  
w  
w  
w  
.  
n  
c  
b  
i  
.  
n  
l  
m  
.  
n  
i  
h  
.  
g  
o  
v  
/  
n  
u  
c  
c  
o  
r  
e  
/  
F  
N  
6  
4  
9  
4  
1  
8  
.

1  
)  
w  
i  
t  
h  
t  
h  
e  
s  
e  
q  
u  
e  
n  
c  
e  
d  
g  
e  
n  
o  
m  
e  
s  
o  
f  
H  
S  
,  
S  
E  
1  
1  
,  
S  
E  
1  
5  
,  
a  
n  
d  
*E*  
.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

c  
o  
l  
l  
i  
K  
-  
1  
2  
(  
M  
G  
1  
6  
5  
5  
)  
(  
e  
.  
g  
.  
,  
u  
s  
i  
n  
g  
t  
h  
e  
s  
i  
n  
g  
l  
e  
t  
o  
n  
m  
e  
t  
h  
o  
d



i  
n  
E  
D  
G  
A  
R  
)  
y  
i  
e  
l  
d  
s  
a  
l  
i  
s  
t  
o  
f  
o  
n  
l  
y  
3  
3  
c  
a  
n  
d  
i  
d  
a  
t  
e  
g  
e  
n  
e  
s  
t  
h  
a  
t

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

a  
r  
e  
n  
o  
t  
f  
o  
u  
n  
d  
i  
n  
a  
n  
y  
o  
f  
t  
h  
e  
c  
o  
m  
m  
e  
n  
s  
a  
l  
s  
o  
r  
*E*  
·  
*c*  
*o*  
*I*  
*i*  
K  
-  
1  
2  
·  
I

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

n  
c  
l  
u  
d  
e  
d  
i  
n  
t  
h  
i  
s  
l  
i  
s  
t  
a  
r  
e  
g  
e  
n  
e  
s  
i  
n  
v  
o  
l  
v  
e  
d  
i  
n  
s  
y  
n  
t  
h  
e  
s  
i  
s  
o

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

f  
t  
h  
e  
c  
o  
l  
o  
n  
i  
z  
a  
t  
i  
o  
n  
f  
a  
c  
t  
o  
r  
a  
n  
t  
i  
g  
e  
n  
C  
F  
A  
/  
I  
,  
t  
h  
e  
*E*  
*t*  
*p*  
*B*  
*A*  
*C*  
o

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

p  
e  
r  
o  
n  
,  
a  
n  
d  
t  
h  
e  
E  
a  
t  
A  
s  
e  
r  
i  
n  
e  
p  
r  
o  
t  
e  
a  
s  
e  
a  
u  
t  
o  
t  
r  
a  
n  
s  
p  
o  
r  
t  
e

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

r  
.  
B  
L  
A  
S  
T  
-  
P  
(  
h  
t  
t  
p  
:  
/  
/  
b  
l  
a  
s  
t  
.  
n  
c  
b  
i  
.  
n  
l  
m  
.  
n  
i  
h  
.  
g  
o  
v  
)  
a  
n  
a  
l  
y

•

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

s  
i  
s  
o  
f  
E  
t  
p  
A  
p  
e  
p  
t  
i  
d  
e  
s  
e  
q  
u  
e  
n  
c  
e  
(  
G  
e  
n  
B  
a  
n  
k  
a  
c  
c  
e  
s  
s  
i  
o  
n  
n  
u  
m  
b

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

e  
r  
A  
A  
X  
1  
3  
5  
0  
9  
)  
i  
d  
e  
n  
t  
i  
f  
i  
e  
s  
m  
a  
n  
y  
c  
l  
o  
s  
e  
h  
o  
m  
o  
l  
o  
g  
u  
e  
s  
o  
f  
E  
t  
P



A  
·  
H  
o  
w  
e  
v  
e  
r  
,  
t  
h  
e  
s  
e  
a  
r  
e  
s  
t  
r  
i  
c  
t  
l  
y  
c  
o  
n  
f  
i  
n  
e  
d  
t  
o  
E  
T  
E  
C  
s  
t  
r  
a  
i

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

n  
s  
t  
h  
a  
t  
h  
a  
v  
e  
b  
e  
e  
n  
s  
e  
q  
u  
e  
n  
c  
e  
d  
t  
o  
d  
a  
t  
e  
.  
I  
n  
f  
a  
c  
t  
,  
r  
e  
c  
e  
n  
t  
s  
t

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

u  
d  
i  
e  
s  
s  
u  
g  
g  
e  
s  
t  
t  
h  
a  
t  
t  
h  
i  
s  
a  
n  
t  
i  
g  
e  
n  
i  
s  
r  
e  
l  
a  
t  
i  
v  
e  
l  
y  
c  
o  
n  
s  
e  
r

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

v  
e  
d  
i  
n  
t  
h  
e  
E  
T  
E  
C  
P  
a  
t  
h  
o  
v  
a  
r  
[  
3  
1  
].  
C  
o  
n  
v  
e  
r  
s  
e  
l  
y  
,  
B  
L  
A  
S  
T  
-  
P  
a  
g

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

a  
i  
n  
s  
t  
c  
o  
m  
m  
e  
n  
s  
a  
l  
s  
H  
S  
,  
S  
E  
l  
l  
,  
S  
E  
l  
5  
,  
N  
i  
s  
s  
l  
e  
l  
9  
l  
7  
,  
o  
r  
t  
h  
e  
l

a  
b  
o  
r  
a  
t  
o  
r  
y  
i  
s  
o  
l  
a  
t  
e  
M  
G  
1  
6  
5  
5  
f  
a  
i  
l  
s  
t  
o  
r  
e  
v  
e  
a  
l  
a  
n  
y  
s  
i  
g  
n  
i  
f  
i

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

c  
a  
n  
t  
h  
o  
m  
o  
l  
o  
g  
y  
,  
f  
u  
r  
t  
h  
e  
r  
s  
u  
g  
g  
e  
s  
t  
i  
n  
g  
t  
h  
a  
t  
t  
h  
i  
s  
p  
a  
r  
t  
i  
c  
u

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

l  
a  
r  
p  
r  
o  
t  
e  
i  
n  
i  
s  
“  
p  
a  
t  
h  
o  
v  
a  
r  
s  
p  
e  
c  
i  
f  
i  
c  
.”

**b.**

Examination of potential surface expression: Analysis of the EtpA peptide sequence with each of the cell localization and domain characterization algorithms outlined in Table 1 yields results provided in Table 4. Without prior knowledge of EtpA function, these algorithms would have (correctly) predicted that this protein shares a number of features with filamentous



hemagglutinin (FHA), a component of the acellular pertussis vaccine and that similar to FHA it belongs to the two-partner family of secretion molecules, which feature atypical extended signal peptides. Like FHA, EtpA is correctly predicted to function as an extracellular adhesin [47] molecule. Collectively, these results suggest that when applied to ETEC, reverse vaccinology approaches have the potential to select novel candidate antigens for downstream validation in vitro and ultimately for vaccine testing in vivo.

**3. Refining antigen selection.**

**a.**

Using immunoproteomics to narrow antigen selection: While the bioinformatics approaches above can offer a list of candidates, the list may be extensive, and additional criteria will likely be needed to refine selection of molecules for further testing as vaccine candidates. Therefore a number of additional modalities have recently been used to highlight key antigens that can be exploited in a vaccine. Previous efforts have combined an examination of the proteome of ETEC with immune response s generated during experimental infection of animals or natural infections

in humans to identify novel antigens [48]. A number of antigens that are not currently targeted in ETEC vaccine approaches were identified including the secreted proteins EtpA, EatA, and YghJ, and antigen 43, an autotransporter protein. To date, three of these proteins EtpA, EatA, and antigen 43 have been shown to offer protection against ETEC infection in an animal model [40, 49, 50]. Nevertheless, this approach has a number of very important limitations: (1) it requires fractionation of bacterial samples to separate proteins which are secreted or which localized to the outer membrane; (2) only one strain can be examined at a time; (3) laboratory culture conditions may not reflect those in vivo, impeding identification of proteins which are not optimally expressed or present in low abundance; (4) it is laborious requiring 2D separation of proteins, subsequent identification of immunoreactive spots by immunoblotting, and extraction of the corresponding protein from a parallel sample which is then identified using mass spectrometry.

**b.**

Protein microarrays: It is now possible to overcome many of the limitations inherent in the approach

outlined above through the use of protein microarrays [51]. These arrays, which have been applied in the investigation of immune responses to diverse pathogens [52–55], offer a number of theoretical advantages. (1) First, they can incorporate features from a number of isolates, thereby coupling bioinformatic analysis of many strains to the printing of key conserved antigens onto the array. (2) It is now possible to synthesize sufficient protein by *in vitro* transcription-translation to accomplish high-throughput antigen synthesis needed to construct hundreds of arrays. (3) The relatively small format of these arrays greatly reduces the sample volumes required for analysis of thousands of candidate antigens simultaneously. Several unpublished projects using ETEC-specific protein microarrays show significant promise in profiling protective immune responses to candidate ETEC vaccines, and in assessing responses that follow natural infections. While experimental and natural ETEC infections offer protection against subsequent disease, the mechanistic correlates of protection have not been established. Protein

microarrays potentially afford an unbiased approach to finding immunologic signatures associated with protection that can then be mined to prioritize antigens for subsequent vaccine testing.

#### 4. Summary.

Since the discovery of ETEC now more than 40 years ago in individuals with severe cholera-like diarrheal illness [56, 57], vaccine development efforts have largely focused on a small group of plasmid-encoded antigens, namely the colonization factors (CFs). With time, investigators have gained an increased appreciation for the complex valency requirements for vaccines based exclusively on CFs [9, 23] and/or heat-labile toxin [58], stimulating the discovery of additional antigens that could complement existing approaches. A compilation of more than 100 sequenced ETEC genomes provides a very rich dataset to interrogate in pursuit of additional vaccine targets.

## References

1. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, 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; 382:209–222. [PubMed: 23680352]
2. Fleckenstein JM, Hardwidge PR, Munson GP, Rasko DA, Sommerfelt H, Steinsland H. Molecular mechanisms of enterotoxigenic *Escherichia coli* infection. *Microbes Infect*. 2010; 12:89–98. [PubMed: 19883790]
3. Cheng SH, Rich DP, Marshall J, Gregory RJ, Welsh MJ, Smith AE. Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel. *Cell*. 1991; 66:1027–1036. [PubMed: 1716180]
4. Vaandrager AB, Smolenski A, Tilly BC, Houtsmuller AB, et al. Membrane targeting of cGMP-dependent protein kinase is required for cystic fibrosis transmembrane conductance regulator Cl<sup>-</sup> channel activation. *Proc Natl Acad Sci U S A*. 1998; 95:1466–1471. [PubMed: 9465038]
5. Golin-Bisello F, Bradbury N, Ameen N. STa and cGMP stimulate CFTR translocation to the surface of villus enterocytes in rat jejunum and is regulated by protein kinase G. *Am J Physiol Cell Physiol*. 2005; 289:C708–C716. [PubMed: 15872007]
6. Chao AC, de Sauvage FJ, Dong YJ, Wagner JA, Goeddel DV, Gardner P. Activation of intestinal CFTR Cl<sup>-</sup> channel by heat-stable enterotoxin and guanylin via cAMP-dependent protein kinase. *EMBO J*. 1994; 13(5):1065–1072. [PubMed: 7510634]
7. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, et al. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *J Bacteriol*. 2010; 192:5822–5831. [PubMed: 20802035]
8. Wolf MK. Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic *Escherichia coli*. *Clin Microbiol Rev*. 1997; 10:569–584. [PubMed: 9336662]

9. Peruski LF Jr, Kay BA, El-Yazeed RA, El-Etr SH, et al. Phenotypic diversity of enterotoxigenic *Escherichia coli* strains from a community-based study of pediatric diarrhea in periurban Egypt. *J Clin Microbiol.* 1999; 37:2974–2978. [PubMed: 10449484]
10. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, et al. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol.* 2008; 190:6881–6893. [PubMed: 18676672]
11. von Mentzer A, Connor TR, Wieler LH, Semmler T, et al. Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nat Genet.* 2014; 46:1321–1326. [PubMed: 25383970]
12. Touchon M, Hoede C, Tenailon O, Barbe V, et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* 2009; 5(1):e1000344. [PubMed: 19165319]
13. Dorsey FC, Fischer JF, Fleckenstein JM. Directed delivery of heat-labile enterotoxin by enterotoxigenic *Escherichia coli*. *Cell Microbiol.* 2006; 8:1516–1527. [PubMed: 16922869]
14. Tacket CO, Losonsky G, Link H, Hoang Y, Guesry P, Hilpert H, Levine MM. Protection by milk immunoglobulin concentrate against oral challenge with enterotoxigenic *Escherichia coli*. *N Engl J Med.* 1988; 318:1240–1243. [PubMed: 3283555]
15. Freedman DJ, Tacket CO, Delehanty A, Maneval DR, Nataro J, Crabb JH. Milk immunoglobulin with specific activity against purified colonization factor antigens can protect against oral challenge with enterotoxigenic *Escherichia coli*. *J Infect Dis.* 1998; 177:662–667. [PubMed: 9498445]
16. McKenzie R, Darsley M, Thomas N, Randall R, et al. A double-blind, placebo-controlled trial to evaluate the efficacy of PTL-003, an attenuated enterotoxigenic *E. coli* (ETEC) vaccine strain, in protecting against challenge with virulent ETEC. *Vaccine.* 2008; 26:4731–4739. [PubMed: 18602960]
17. Darsley MJ, Chakraborty S, Denearing B, Sack DA, et al. ACE527 oral, live attenuated ETEC vaccine reduces the incidence and severity of diarrhea in a human challenge model of diarrheal disease. *Clin Vaccine Immunol.* 2012; 19:1921–1931. [PubMed: 23035175]
18. Sack DA, Shimko J, Torres O, Bourgeois AL, et al. Randomised, double-blind, safety and efficacy of a killed oral vaccine for enterotoxigenic *E. coli* diarrhoea of travellers to Guatemala and Mexico. *Vaccine.* 2007; 25:4392–4400. [PubMed: 17448578]
19. Li YF, Poole S, Rasulova F, McVeigh AL, Savarino SJ, Xia D. A receptor-binding site as revealed by the crystal structure of CfaE, the colonization factor antigen I fimbrial adhesin of enterotoxigenic *Escherichia coli*. *J Biol Chem.* 2007; 282:23970–23980. [PubMed: 17569668]
20. Li YF, Poole S, Nishio K, Jang K, et al. Structure of CFA/I fimbriae from enterotoxigenic *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2009; 106:10793–10798. [PubMed: 19515814]
21. Giron JA, Levine MM, Kaper JB. Longus: a long pilus ultrastructure produced by human enterotoxigenic *Escherichia coli*. *Mol Microbiol.* 1994; 12:71–82. [PubMed: 7914665]
22. Roy SP, Rahman MM, Yu XD, Tuittila M, Knight SD, Zavialov AV. Crystal structure of enterotoxigenic *Escherichia coli* colonization factor CS6 reveals a novel type of functional assembly. *Mol Microbiol.* 2012; 86:1100–1115. [PubMed: 23046340]
23. Isidean SD, Riddle MS, Savarino SJ, Porter CK. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine.* 2011; 29:6167–6178. [PubMed: 21723899]
24. Del Canto F, Botkin DJ, Valenzuela P, Popov V, et al. Identification of the Coli Surface Antigen 23 (CS23), a novel adhesin of enterotoxigenic *Escherichia coli*. *Infect Immun.* 2012; 80:2791–2801. [PubMed: 22645287]
25. Rockabrand DM, Shaheen HI, Khalil SB, Peruski LF Jr, Rozmajzl PJ, et al. Enterotoxigenic *Escherichia coli* colonization factor types collected from 1997 to 2001 in US military personnel during operation Bright Star in northern Egypt. *Diagn Microbiol Infect Dis.* 2006; 55:9–12. [PubMed: 16542813]
26. Peruski LF Jr, Kay BA, El-Yazeed RA, El-Etr SH, Cravioto A, et al. Phenotypic diversity of enterotoxigenic *Escherichia coli* strains from a community-based study of pediatric diarrhea in periurban Egypt. *J Clin Microbiol.* 1999; 37:2974–2978. [PubMed: 10449484]

27. Manoil C, Beckwith J. TnpA: a transposon probe for protein export signals. *Proc Natl Acad Sci U S A*. 1985; 82:8129–8133. [PubMed: 2999794]
28. Patel SK, Dotson J, Allen KP, Fleckenstein JM. Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic *Escherichia coli*. *Infect Immun*. 2004; 72:1786–1794. [PubMed: 14977988]
29. Fleckenstein JM, Roy K, Fischer JF, Burkitt M. Identification of a two-partner secretion locus of enterotoxigenic *Escherichia coli*. *Infect Immun*. 2006; 74:2245–2258. [PubMed: 16552055]
30. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Rd Science*. 1995; 269:496–512. [PubMed: 7542800]
31. 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. [PubMed: 25629897]
32. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012; 486:207–214. [PubMed: 22699609]
33. De Gregorio E, Rappuoli R. From empiricism to rational design: a personal perspective of the evolution of vaccine development. *Nat Rev Immunol*. 2014; 14:505–514. [PubMed: 24925139]
34. Rinaldo CD, Telford JL, Rappuoli R, Seib KL. Vaccinology in the genome era. *J Clin Invest*. 2009; 119:2515–2525. [PubMed: 19729849]
35. Sahl JW, Steinsland H, Redman JC, Angiuoli SV, Nataro JP, Sommerfelt H, Rasko DA. A comparative genomic analysis of diverse clonal types of enterotoxigenic *Escherichia coli* reveals pathovar-specific conservation. *Infect Immun*. 2011; 79:950–960. [PubMed: 21078854]
36. Allen KP, Randolph MM, Fleckenstein JM. Importance of heat-labile enterotoxin in colonization of the adult mouse small intestine by human enterotoxigenic *Escherichia coli* strains. *Infect Immun*. 2006; 74:869–875. [PubMed: 16428729]
37. Sheikh A, Luo Q, Roy K, Shabaan S, Kumar P, Qadri F, Fleckenstein JM. Contribution of the highly conserved EaeH surface protein to enterotoxigenic *Escherichia coli* pathogenesis. *Infect Immun*. 2014; 82:3657–3666. [PubMed: 24935979]
38. Roy K, Hamilton D, Allen KP, Randolph MP, Fleckenstein JM. The EtpA exoprotein of enterotoxigenic *Escherichia coli* promotes intestinal colonization and is a protective antigen in an experimental model of murine infection. *Infect Immun*. 2008; 76:2106–2112. [PubMed: 18285493]
39. Luo Q, Kumar P, Vickers TJ, Sheikh A, Lewis WG, Rasko DA, Sistrunk J, Fleckenstein JM. Enterotoxigenic *Escherichia coli* secretes a highly conserved mucin-degrading metalloprotease to effectively engage intestinal epithelial cells. *Infect Immun*. 2014; 82:509–521. [PubMed: 24478067]
40. 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; 82:500–508. [PubMed: 24478066]
41. Fleckenstein JM, Holland JT, Hasty DL. Interaction of an outer membrane protein of enterotoxigenic *Escherichia coli* with cell surface heparan sulfate proteoglycans. *Infect Immun*. 2002; 70:1530–1537. [PubMed: 11854241]
42. 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; 286:29771–29779. [PubMed: 21757737]
43. Fleckenstein JM, Kopecko DJ, Warren RL, Elsinghorst EA. Molecular characterization of the tia invasion locus from enterotoxigenic *Escherichia coli*. *Infect Immun*. 1996; 64:2256–2265. [PubMed: 8675335]
44. Roy K, Hilliard GM, Hamilton DJ, Luo J, Ostmann MM, Fleckenstein JM. Enterotoxigenic *Escherichia coli* EtpA mediates adhesion between flagella and host cells. *Nature*. 2009; 457:594–598. [PubMed: 19060885]
45. Sheikh A, Luo Q, Roy K, Shaaban S, Kumar P, Fleckenstein JM. Contribution of the highly conserved EaeH surface protein to enterotoxigenic *Escherichia coli* pathogenesis. *Infect Immun*. 2014; 82:3657–3666. [PubMed: 24935979]

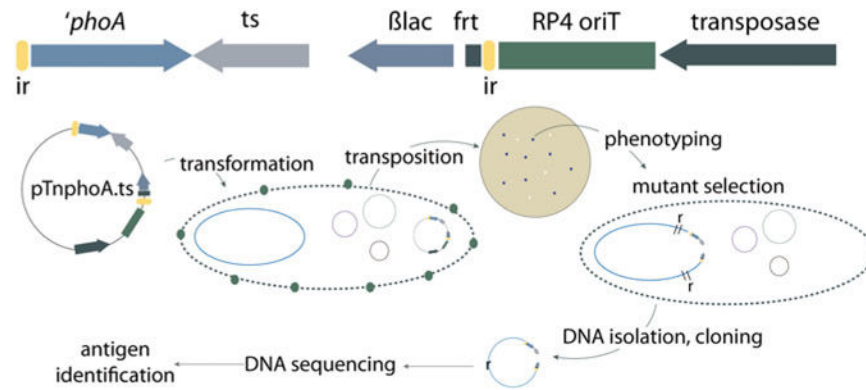
46. Roy K, Hamilton DJ, Fleckenstein JM. Cooperative role of antibodies against heat-labile toxin and the EtpA Adhesin in preventing toxin delivery and intestinal colonization by enterotoxigenic *Escherichia coli*. *Clin Vaccine Immunol*. 2012; 19:1603–1608. [PubMed: 22875600]
47. Sachdeva G, Kumar K, Jain P, Ramachandran S. SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. *Bioinformatics*. 2005; 21:483–491. [PubMed: 15374866]
48. Roy K, Bartels S, Qadri F, Fleckenstein JM. Enterotoxigenic *Escherichia coli* elicits immune responses to multiple surface proteins. *Infect Immun*. 2010; 78:3027–3035. [PubMed: 20457787]
49. 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; 27:4601–4608. [PubMed: 19523914]
50. Harris JA, Roy K, Woo-Rasberry V, Hamilton DJ, Kansal R, Qadri F, Fleckenstein JM. Directed evaluation of enterotoxigenic *Escherichia coli* autotransporter proteins as putative vaccine candidates. *PLoS Negl Trop Dis*. 2011; 5(12):e1428. [PubMed: 22163060]
51. Davies DH, Liang X, Hernandez JE, Randall A, et al. Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. *Proc Natl Acad Sci U S A*. 2005; 102:547–552. [PubMed: 15647345]
52. Crompton PD, Kayala MA, Traore B, Kayentao K, et al. A prospective analysis of the Ab response to *Plasmodium falciparum* before and after a malaria season by protein microarray. *Proc Natl Acad Sci U S A*. 2010; 107:6958–6963. [PubMed: 20351286]
53. Liang L, Juarez S, Nga TV, Dunstan S, et al. Immune profiling with a *Salmonella Typhi* antigen microarray identifies new diagnostic biomarkers of human typhoid. *Sci Rep*. 2013; 3:1043. [PubMed: 23304434]
54. Lee SJ, Liang L, Juarez S, Nanton MR, et al. Identification of a common immune signature in murine and human systemic Salmonellosis. *Proc Natl Acad Sci U S A*. 2012; 109:4998–5003. [PubMed: 22331879]
55. Felgner PL, Kayala MA, Vigil A, Burk C, Nakajima-Sasaki R, et al. A *Burkholderia pseudomallei* protein microarray reveals serodiagnostic and cross-reactive antigens. *Proc Natl Acad Sci U S A*. 2009; 106:13499–13504. [PubMed: 19666533]
56. Sack RB. The discovery of cholera-like enterotoxins produced by *Escherichia coli* causing secretory diarrhoea in humans. *Indian J Med Res*. 2011; 133:171–180. [PubMed: 21415491]
57. Sack RB, Gorbach SL, Banwell JG, Jacobs B, Chatterjee BD, Mitra RC. Enterotoxigenic *Escherichia coli* isolated from patients with severe cholera-like disease. *J Infect Dis*. 1971; 123:378–385. [PubMed: 4938945]
58. Riddle MS, Savarino SJ. Moving beyond a heat-labile enterotoxin-based vaccine against enterotoxigenic *Escherichia coli*. *Lancet Infect Dis*. 2013; 14:174–175. [PubMed: 24291167]
59. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, et al. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res*. 2014; 42:W337–W343. Web Server issue. [PubMed: 24799431]
60. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011; 8:785–786. [PubMed: 21959131]
61. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel*. 2004; 17:349–356. [PubMed: 15115854]
62. Bendtsen JD, Kiemer L, Fausboll A, Brunak S. Non-classical protein secretion in bacteria. *BMC Microbiol*. 2005; 5:58. [PubMed: 16212653]
63. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*. 2010; 26:1608–1615. [PubMed: 20472543]
64. Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci*. 2004; 13:1402–1406. [PubMed: 15096640]

65. He Y, Xiang Z, Mobley HL. Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J Biomed Biotechnol.* 2010; 2010:297505. [PubMed: 20671958]
66. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics.* 2008; 9:75. [PubMed: 18261238]
67. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep.* 2015; 5:8365. [PubMed: 25666585]
68. Blom J, Albaum SP, Doppmeier D, Puhler A, Vorholter FJ, Zakrzewski M, Goesmann A. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics.* 2009; 10:154. [PubMed: 19457249]
69. Minkin I, Pham H, Starostina E, Vyahhi N, Pham S. C-Sibelia: an easy-to-use and highly accurate tool for bacterial genome comparison. *F1000Res.* 2013; 2:258. [PubMed: 25110578]
70. Finn RD, Bateman A, Clements J, Coggill P, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2014; 42:D222–D230. Database issue. [PubMed: 24288371]
71. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, et al. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* 2013; 41:D348–D352. Database issue. [PubMed: 23197659]
72. Grozdanov L, Raasch C, Schulze J, Sonnenborn U, et al. Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *J Bacteriol.* 2004; 186:5432–5441. [PubMed: 15292145]
73. Oshima K, Toh H, Ogura Y, Sasamoto H, Morita H, et al. Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. *DNA Res.* 2008; 15:375–386. [PubMed: 18931093]
74. Toh H, Oshima K, Toyoda A, Ogura Y, Ooka T, et al. Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. *J Bacteriol.* 2010; 192:1165–1166. [PubMed: 20008064]
75. de Muinck EJ, Lagesen K, Afset JE, Didelot X, et al. Comparisons of infant *Escherichia coli* isolates link genomic profiles with adaptation to the ecological niche. *BMC Genomics.* 2013; 14:81. [PubMed: 23384204]
76. Peterson MD, Mooseker MS. Characterization of the enterocyte-like brush border cytoskeleton of the C2BBE clones of the human intestinal cell line, Caco-2. *J Cell Sci.* 1992; 102(Pt 3):581–600. [PubMed: 1506435]
77. Tom BH, Rutzky LP, Jakstys MM, Oyasu R, Kaye CI, Kahan BD. Human colonic adenocarcinoma cells. I. Establishment and description of a new line. *In Vitro.* 1976; 12:180–191. [PubMed: 1262041]
78. Dharmasathaphorn K, McRoberts JA, Mandel KG, Tisdale LD, Masui H. A human colonic tumor cell line that maintains vectorial electrolyte transport. *Am J Physiol.* 1984; 246(2 Pt 1):G204–G208. [PubMed: 6141741]



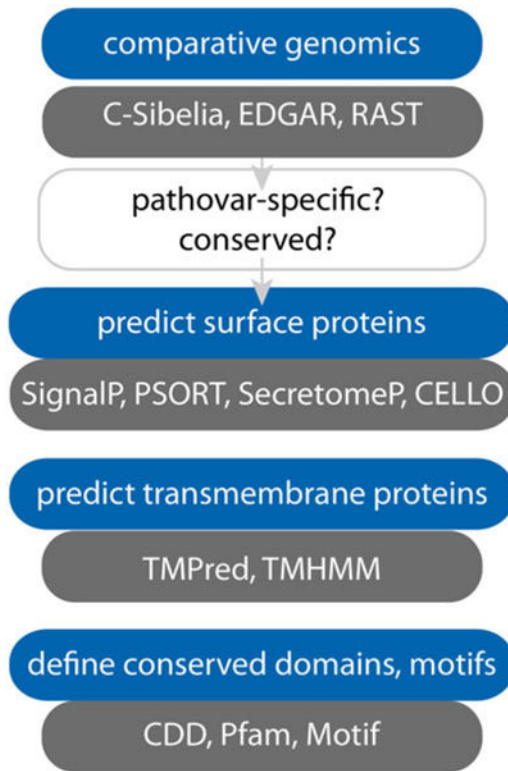
	advantages	limitations
classical genetics	once completed the targets are well identified and characterized	time consuming and labor intensive - requires specialized knowledge
comparative genomics	easy to generate large amounts of data from clinically relevant samples	gene presence and absence provides no information about expression
transcriptomics	global and unbiased methods have been developed that are easy to employ	cost to generate data and use of relevant conditions for virulence not always clear
proteomics	protein data demonstrates that the gene is active and produced	global, high-throughput tools are still developing
immunoproteomics	direct evidence for immune reactivity	global, high-throughput tools are still developing

**Fig. 1.** Methodologies employed to identify and characterize potential vaccine antigens for enterotoxigenic *Escherichia coli*



**Fig. 2.**

TnpA.ts mutagenesis strategy: (*top*) linear plasmid map of pTnpA.ts. The transposon is shown between the two inverted repeat elements (*ir*, *yellow*). Within the transposon are a truncated version of the alkaline phosphatase gene (*phoA*), a temperature-sensitive origin of replication (*ts*), beta lactamase gene encoding ampicillin resistance (*βlac*), and a flp recombination recognition site (*frt*). Encoded on the plasmid backbone outside the repeat element are the RP4 transfer locus (*RP4 oriT*), and the Tn5 transposase gene. Shown below the map are major steps in transposon mutant generation starting with transformation of the recipient strain at permissive temperature (30 °C). Selection for transposition events with productive fusions then takes place at higher temperature under antibiotic and XP (*blue colony*) selection. Resulting colonies can then be tested to identify mutants phenotypically different than the parental ETEC strain. Total genomic DNA (plasmid and chromosome) is then digested with a restriction enzyme that cuts outside of the transposon, religated, and used to transform a laboratory cloning strain to ampicillin resistance at 30 °C. The resulting plasmid can be sequenced from to identify the gene interrupted by transposition, and the corresponding protein can be identified in GenBank



**Fig. 3.** Algorithms useful in identification and in silico characterization of candidate vaccine antigens



**Table 2**Commensal *E. coli* strains with sequenced genomes

Strain designation	NCBI/EMBL accession number(s)	Origin	Reference
HS	NC_009800.1	Healthy adult, USA	[10]
Nissle	1917 CP007799.1	Healthy adult, Germany	[72]
SE11	AP009240.1 (chromosome) AP009241.1 (pSE11-1) AP009242.1 (pSE11-2) AP009243.1 (pSE11-3) AP009244.1 (pSE11-4) AP009245.1 (pSE11-5) AP009246.1 (pSE11-6)	Healthy adult, Japan	[73]
SE15	NC_013654.1 (chromosome) NC_013655.1 (plasmid)	Healthy adult, Japan	[74]
EDM1c <sup>a</sup>	ERS155053	Healthy child, Norway	[75]
EDM3c <sup>a</sup>	ERS155049	Healthy child, Norway	
EDM16c <sup>a</sup>	ERS155051	Healthy child, Norway	
EDM70c <sup>a</sup>	ERS155055	Healthy child, Norway	
EDM49c <sup>a</sup>	ERS155056	Healthy child, Norway	
EDM101c <sup>a</sup>	ERS155057	Healthy child, Norway	
EDM106c <sup>a</sup>	ERS155058	Healthy child, Norway	
EDM116c <sup>a</sup>	ERS155052	Healthy child, Norway	
EDM123c <sup>a</sup>	ERS155054	Healthy child, Norway	
EDM530c <sup>a</sup>	ERS155050	Healthy child, Norway	

<sup>a</sup>Draft genomes

**Table 3**

Gastrointestinal cell lines used to examine ETEC pathogen-host interactions

Cell line	Distinguishing characteristics	ATCC	Toxin response			Reference(s)
			LT	ST	ST	
Caco-2	Polarize	HTB-37	+	+	+/-	
C2BBel	Form brush borders	CRL-2102	+	+	+/-	[76]
LS-174 T	Goblet cell like; abundant MUC2 production	CL-188	-	-	-	[77]
T84	Polarity	CCL-248	+	+	+	[78]
HT-29	Blood group A, MUC2	HTB-38	+	+	-	

ATCC American Type Culture Collection

**Table 4**

EtpA as a prototype molecule in reverse vaccinology algorithms

Program	Result/prediction	Notes/score
Signal-P	No signal peptide identified	–
PredictProtein	Predicted location: secreted	–
Secretome	Secreted	(SecP) 0.939918
Psortb	Extracellular	10.00
CELLO	Extracellular	2.777
Vaxign	Adhesin	via SPAAN [47]
Motif <sup>a</sup>	Multiple domains identified: CDD130956, adhes_NPXG CDD214908, haemagg_act CDD257462, ESPR	Domain description Filamentous hemagglutinin family N-terminal domain Hemagglutination activity domain Extended signal peptide of type V secretion system

<sup>a</sup> Motif searches both CDD and Pfam databases

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript