White Paper Application

Project Title: Comparative genome analysis of enterotoxigenic *E. coli* strains isolates from infections of different clinical severity.

Authors: Fleckenstein JM, Qadri F, Rasko DR
Primary Investigator Contact:

Name	James Michael Fleckenstein
Position	Associate Professor, Medicine and Molecular Sciences
Institution	University of Tennessee Health Science Center
Address	1030 Jefferson Avenue, Research (151), Memphis
State	TN
ZIP Code	38104
Telephone	901-270-8796
Fax	901-577-7273
E-Mail	jflecke1@tennessee.edu

Name	Firdausi Qadri
Position	Director, Immunology Laboratory
Institution	International Center for Diarrhoeal Disease Research, (ICDDR,B)
Address	Mohakhali, Dhaka 1212 Bangladesh
Telephone	880-2-8860523-32 x2431
Fax	880-2-8812529
E-Mail	fqadri@icddrb.org

Name	David Allan Rasko
Position	Assistant Professor, IGS and Dept of Microbiology and Immunology
Institution	University of Maryland School of Medicine
Address	801 W. Baltimore Street, Room 619, Baltimore
State	MD
ZIP Code	21201
Telephone	410-706-6774
Fax	410-706-1482
E-Mail	drasko@umaryland.edu

1. Executive Summary ***(Please limit to 500 words.)***

Enterotoxigenic *E. coli* (ETEC) are the most common bacterial etiology of diarrheal illness in developing countries[1], causing millions of infections and hundreds of thousands of deaths annually. ETEC infections range in severity from mild asymptomatic colonization to severe cholera-like illness[2-4] associated with rapid diarrheal fluid losses resulting in profound dehydration and death due to hypovolemic shock. Cholera-like illness in patients without detectable *Vibrio cholerae* prompted early discovery of *E. coli* that produced toxin[5-9] (the heat-labile toxin, LT) very similar to cholera toxin, and subsequently to the detection of heat-stable toxin (ST)[10] in similar isolates. In its most severe form, diarrheal disease caused by ETEC is clinically indistinguishable from cholera [4, 11, 12], and like disease due to *Vibrio cholerae*, it may occur in epidemic form[12, 13].

A genetically heterogeneous collection of organisms that share the ability to produce and

effectively deliver the LT and/or ST enterotoxin to effector sites in the small intestine constitute the ETEC pathotype[14]. Elucidation of particular virulence factors or biomarkers associated with severe disease could permit a rational approach to vaccines specifically designed to prevent deaths due to ETEC. Unfortunately, while strains exhibit clear phenotypic differences in the ability of to cause more severe forms of diarrhea, the genetic/genomic reasons for this are not clear. The genomes of eight ETEC strains are currently available (seven of which were generated by investigators on this project). However, overall plasticity of *E. coli* genomes in general and incomplete accompanying clinical information for these strains precludes assignment of any particular genetic feature with disease outcome.

Another problem confounding rational vaccine design is that frequently, a single isolate from a patient/subject identified as ETEC (by identification of LT and/or ST in fecal *E. coli* colonies) is typically stored. Yet, ETEC infections typically follow ingestion of heavily contaminated food or water,and neither the inoculum nor the strains emerging in diarrheal stools are necessarily clonal.

This proposal seeks to exploit a wealth of existing ETEC strains for which there accompanying detailed clinical data. The majority of these isolates are from the International Centre for Diarrhoeal Disease Research (ICDDR,B) in Dhaka, Bangladesh, where multiple isolates have been carefully catalogued during many clinical studies on ETEC [13, 15-18].

The strains in this collection include isolates from:

(a) patients hospitalized at ICDDR,B with severe infection,

(b) patients with mild infection not requiring hospitalization,

(c) asymptomatically colonized children and adults.

(d) In a subset of isolates from severe disease we plan to examine the diversity of the diarrheal isolates to examine the population variability during a single infection. Also included in the list of proposed isolates are a limited number of well-characterized strains (non-ICDDR, B) of established provenance collected from patients with severe infection.

Altogether, DNA from a total of 141 ETEC strains linked to various clinical outcomes will be examined. Comparisons of the genomes of these strains will facilitate the identification of specific features associated with severe disease, aid in the identification of highly conserved vaccine targets, and address intra-infection population variability. We anticipate that these data will greatly accelerate the pace of biomarker identification, pathogenesis studies and vaccine development for these pathogens of great importance.

2. Justification

Provide a succinct justification for the sequencing or genotyping study by describing the significance of the problem and providing other relevant background information.

Relevance of ETEC to infectious diseases

The enterotoxigenic *Escherichia coli* are a principal cause of diarrheal illness in developing countries[1], where they account millions of infections and result in hundreds of thousands of deaths in young children each year[16]. In addition, these genetically and phenotypically heterogeneous pathogens, which share the ability to produce and effectively deliver heat-labile (LT) and/or heat-stable (ST) enterotoxins, are perennially the major cause of traveller's diarrhea[19], and an emerging cause of diarrheal illness in the United States as exemplified by multiple recent large-scale outbreaks[20-28].

ETEC, *Shigella,* and rotavirus, the three major causes of death due to infectious diarrhea in developing countries, each account for approximately one third of the estimated 1.5-2 million deaths, largely among children less than five years of age, that occur each year[29]. Furthermore, studies suggest that ETEC and cholera are among the organisms most likely to require hospitalization for diarrheal illness in older children and adults in developing countries[30]. Therefore, these organisms are essential targets for vaccine and therapeutic development[31, 32]. **Notably, while the overall death rate from diarrheal**

illnesses has in general declined since the widespread implementation of oral rehydration therapy, morbidity attributable to ETEC remains largely unchanged. Indeed, ETEC has been implicated in recent studies to delayed growth and malnutrition that accompany repeated bouts of diarrhea[16, 33].

Despite the global importance of these infections, there remains no effective vaccine for ETEC[31]. Until very recently, most pathogenesis studies, as well as virtually all vaccinology efforts for ETEC had focused very narrowly on a small subset of antigens[14], including the toxins, and fimbrial colonization factors (CFs). Unfortunately, these antigens have not been sufficient to formulate a broadly-protective vaccine because of substantial antigenic heterogeneity[34] in the case of the CFs, and the insufficient un-sustained protection afforded by anti-toxin based immunity[35, 36]. More recent studies of ETEC support the notion that there are other virulence determinants (potentially conserved in the ETEC pathotype[37]), including the recently discovered EtpBAC two partner secretion (TPS) adhesin locus[38, 39], or the EatA autotransporter protein[40] which are recognized by the host during the course of infection[41] and may serve as protective antigens[42, 43]. However, the remarkable plasticity of *E. coli* genomes predicts that multiple conserved antigens will likely need to be targeted to achieve broad-based protection.

Antimicrobial resistance:

In addition to the virulence factors more than half of all ETEC strains in some developing countries are now resistant to agents commonly employed in treatment of infectious diarrhea, including ampicillin, trimethoprim-sulfamethoxazole, and tetracyclines[44]. Moreover, recent reports suggest that ETEC isolated from travelers to developing countries are increasingly resistant to newer antimicrobial agents, including fluoroquinolones[45, 46]. This established and emerging proliferation of antibiotic resistance among enteropathogens in general, and particularly in the ETEC pathotype provides a strong impetus for developing a vaccine to these infections. The rapid spread of antimicrobial resistance mechanisms may be linked to the identification of large plasmids in the majority of the sequenced members of the ETEC pathovar [37, 47, 48].

Current status of genome data for ETEC

What is the status of other sequencing / genotyping projects on the same organism including current and past projects of the NIAID GSC? Provide information on other characteristics (genome size, GC content, repetitive DNA, pre-existing arrays etc.) relevant to the proposed study. Have analyses been performed on the raw data already generated/published? If additional strains are proposed for a species, please provide a justification for additional strains?

A total of eight ETEC strains have now been sequenced and published. These include the prototype H10407 strain [47], 5 isolates from Guinea-Bissau representing diverse MLST types [37], and E24377A, and B7A[48] (table 1).

table 1. summary of current completed ETEC genome sequencing projects

strain designation	origin	genome size (x 10^6 bp)	GC (%)	toxins	Colonization Factors	other virulence genes	
						etpBAC	*eatA*
H10407	Bangladesh	5.326 [(a)]	50.8	LT, ST	CFA/I	+	+
B7A	Vietnam	5.300 [(b)]	50.7	LT, ST	CS6	-	-
E24377A	Egypt	5.136 [(c)]	50.6	LT, ST	CS1,CS3	+	+
TW10598	Guinea-Bissau		50.7	LT, ST	CS2,CS3,CS21	+	+
TW10722	Guinea-Bissau		50.7	ST	CS5,CS6	-	+
TW11681	Guinea-Bissau		50.7	ST	CFA/I, CS21	+	+
TW10828	Guinea-Bissau		50.7	ST	CS7	+	+
TW14425	Guinea-Bissau		50.5	LT	CS14	+	-

(a) includes chromosome of 5,153,435 bp and 4 plasmids
(b) draft contigs
(c) includes chromosome of 4,979,619 bp and 6 plasmids

If analyses have been conducted, briefly describe utility of the new sequencing or genotyping information with an explanation of how the proposed study to generate additional data will advance diagnostics, therapeutics, epidemiology, vaccines, or basic knowledge such as species diversity, evolution, virulence, etc. of the proposed organism to be studied.

The existing sequencing data from these eight strains has already provided useful information. For example these data demonstrated that two novel putative vaccine candidates, EtpA and EatA are each present in 75% of a diverse clonal population of

strains[37], potentially enhancing their value as molecular targets for further study. However, recent immunoproteomic studies[41] suggest that these are among many antigens that may contribute to immune responses that develop following infection, and the precise nature of the protection that develops following natural infections remains unclear[49]. A detailed study of organisms isolated from individuals with various forms of diarrhea due to ETEC could highlight the most relevant and highly conserved antigens to target in the development of an ETEC vaccine.

With the exception of the H10407, there is little if any accompanying clinical data regarding the severity of the clinical illness from which these recently sequenced strains were originally derived. Therefore, the current small number of sequenced isolates is insufficient to permit any detailed correlation of genes or genomic regions with specific disease manifestations. A systematic comparison of ETEC from severe, cholera-like illness to organisms from mild disease or asymptomatic colonization could identify specific antigenic targets to prevent the most lethal forms of infection. Consequently, a collection of strains with established provenance from different disease categories that has not been subjected to serial passage would be invaluable in defining potential associations.

Identification of virulence determinants associated with severe disease

Because ETEC is among the bacterial pathogens most likely to be associated with diarrhea leading to hospitalization and death in developing countries, targeting specific antigens associated with more serve forms of disease seems paramount. Yet, much remains to be learned with respect to the differential capacity of strains to cause severe illness. Clearly, however there are significant phenotypic differences among ETEC strains that might be exploited in the discovery of particularly aggressive virulence traits. Some strains such as H10407 (LT^+, ST^+, CFA/I), originally isolated from a Bangladeshi patient with severe cholera-like diarrheal illness[50], clearly cause more severe illness in volunteers than other ETEC strains such as B7A (LT^+, ST^+, CS6)[51], yet the reasons at the genetic level for this are not clear. Interestingly, B7A does lack two recently characterized virulence factors, the EatA autotransporter and a novel two-partner secretion locus, which encodes the EtpA adhesin. These loci are present in H10407 and other isolates[40, 38] associated with severe cholera-like diarrheal illness. Nonetheless, a systematic unbiased investigation of strains isolated from patients with more acute forms of diarrheal illness in comparison with isolates from milder diarrhea or asymptomatically colonized subjects is essential before attribution to any particular genetic feature(s). Altogether, *identification of virulence factors contributing to severe illness could highlight antigens that need to be targeted to prevent deaths due to diarrheal disease in developing countries.*

Molecular epidemiology of ETEC virulence factors and implications for vaccine development General plasticity with attendant loss and acquisition of potential virulence factors characterizes genomes of all *E. coli* pathotypes including ETEC[48]. Therefore, conservation of putative antigens across a broad array of isolates is likely to be a critical determinant of the utility of any particular molecule for a vaccine. Indeed, after nearly four decades of vaccine development focused almost exclusively on colonization factors (CFs), we now understand that these antigens are not sufficiently conserved nor appreciably cross-protective[35]. *Expansion of the existing dataset to include a broad collection of well-validated isolates will be essential for the rational selection of highly conserved target antigens to incorporate into a vaccine.*

Examination of ETEC population diversity from individual patients For many years, the standard for detecting pathogenic *E. coli* in fecal specimens has been to test (up to) six *E. coli* colonies[18, 52] using PCR, genetic hybridization for toxins or other assays to identify the particular pathotype involved. However, the scientific validity of this approach has recently been questioned[53]. Indeed, prior methodology clearly misses cases of ETEC by testing an insufficient number of colonies. These findings suggest that the actual number of ETEC

infections and related morbidity is actually under-represented. This also highlights another problem in that neither the ingested inoculum, nor the population that emerges in diarrheal stools is necessarily clonal[17]. Moreover, secreted ETEC virulence proteins can potentially complement strains deficient in these factors rendering them virulent[39], therefore selection of a limited number of colonies from an individual stool could underestimate the relevance of particular virulence traits. Moreover, ETEC are known to lose[54, 55] potential virulence factors as plasmids or other mobile elements are lost in the absence of selection. Not surprisingly then, serologic data obtained following natural infection occasionally demonstrate immune responses to one or more proteins that do not appear to be present in the corresponding isolate. *Therefore, understanding the nature of the ETEC population from a given infection has critical implications for vaccine development.*

3. Rationale for Strain Selection

1. *Provide the rationale behind the selection of strains and the number of strains proposed in the study. The focus of the program is on potential agents of bioterrorism or organisms responsible for emerging or re-emerging infectious diseases. Non-select agents or non-pathogenic organisms will be considered when they can provide insight into these scientific areas.*

Source of strains

Each of the ETEC strains proposed for sequencing here meet several criteria:

(1) isolates have been carefully maintained without serial passage between multiple laboratories,

(2) the clinical characteristics of the corresponding ETEC infection, including the severity of the illness, are well-documented,

(3) the provenance of each isolate is well-established.

The majority of isolates are derived from the extensive collection of ETEC isolates obtained from multiple clinical studies at ICDDR,B, that have been maintained by Dr. Qadri and her colleagues in Dhaka. Many of these strains have associated serologic specimens collected during convalescence that can be used to assess immune responses to selected antigens. Where possible we have included additional isolates from outside Bangladesh for which there are solid supporting clinical data, and a clearly defined clinical outcome.

Strains from severe illness:

For these studies we have included 50 isolates from patients with severe clinical illness, with the majority (n=45) of these coming from subjects hospitalized at ICDDR,B. These strains have been carefully catalogued and include isolates from various time periods (range in years). Complementing these isolates are strains from outbreaks of cholera-like ETEC diarrhea that occurred in 1998 in two villages of the Amazonian rainforest in Brazil, with fatal outcomes in both adults and children[12]. These isolates were kindly provided by Dr. Ana Vicente of the Oswaldo Cruz Institute in Rio de Janeiro. Also included in the collection of isolates submitted for sequencing is ThroopD, isolated in 1973 from a 36 year old female patient who presented with severe cholera-like disease, hypovolemic shock, and diarrhea in excess of 50 liters over 3 days[3]. The isolate was originally obtained from "rice-water" stool, and was provided by Dr. Richard Finklestein, who originally isolated the strain in Dallas.

Strains from asymptomatic colonization, and mild disease

These strains were acquired during the course of a number of community ETEC studies performed by Dr. Qadri and her laboratory at ICDDR,B. We plan to include 20 isolates from asymptomatically colonized subjects, and 10 strains from mild disease not requiring hospitalization for comparison.

Multiple ETEC strains isolated simultaneously from individual patients

To assess genomic diversity of isolates obtained from individual patients we plan to examine 10 isolates that have already been collected from 2 different patients with severe diarrhea (5 isolates each, see isolate numbers 82-91 on the attached spreadsheet). Already, these isolates suggest that we will find significant phenotypic and genotypic diversity in these isolates as 3 different CF types (and one CF-negative isolate) are represented in the isolates from just 2 subjects. In addition to these isolates, we plan to prospectively collect 10 strains obtained from five individual subjects hospitalized with severe watery diarrhea (total of n=50, isolate numbers 92-141 on the attached spreadsheet). These isolates will be obtained prospectively from the diarrheal stools of 5 individuals with severe illness. Briefly stools will be freshly inoculated onto the surface of MacConkey agar plates, and incubated for 18 hours at 37°C. Then, 100 *E. coli* colonies will be picked per patient and tested for LT and ST by ELISA as previously described. Each of the LT^+/ST^+, LT^+, or ST^+ isolates will be stored at -80°C, and 10 isolates will be chosen from each patient for sequencing. Dr. Fleckenstein is traveling to Dhaka in early 2011 during the peak of ETEC diarrhea season, to aid in the collection and characterization of these additional isolates. These isolates will then be typed and sent back to the United States for sequencing and further characterization. The abundance of patients during the time Dr. Fleckenstein will be in Dhaka will ensure the successful collection of these additional 50 isolates from 5 patients.

4a. Approach to Data Production: **Data Generation**

2. *State the data and resources planned to be generated. (e.g draft genome sequences, finished sequence data, SNPs, DNA/protein arrays generation, clone generation etc.)*

Sequencing strategy. Since sufficient amounts of high-quality genomic DNA (>20 µg) can be obtained from clonal cultures (in single passage) of all strains suggested for sequencing, we propose to apply a hybrid sequencing strategy that combines the advantages of Roche's 454 GS FLX Titanium with Illumina's HiSeq2000 sequencing platforms. Using 3kb paired-end sequencing libraries for the 454 platform with a moderate overall genome coverage (~20- fold), large continuous contig scaffolds can be generated that will span the vast majority of each genome sequence. These genome scaffolds can be combined with high-coverage (at least 100-fold) short read (100 bp) sequence data from the Illumina platform to generate hybrid assemblies which should leave few to no physical gaps in each genome sequence. Since one of the goals in this proposal is new gene discovery and comparative genomics, we feel as though this hybrid approach will provide the most complete datasets for the successful identification of these genes and potential vaccine candidates. Additionally, IGS is scheduled to take delivery of a Pacific Biosciences RS (PacBioRS) machine early in 2011. Considering the wealth of data on *E. coli* genomes this would make this proposal an ideal high-throughput test case for this new technology.

Sequence assembly and annotation and whole-genome comparative analysis.

Genome assembly. Draft genome sequences generated under this proposal will be assembled using Newbler and Celera Assemblers. The resulting contigs will be filtered for sequences of the indigenous virulence plasmids using the existing ETEC plasmids as a reference.

Genome annotation. All individual contigs will be submitted to the IGS annotation pipelines and evaluated in Manatee (http://ae.igs.umaryland.edu/cgi/manatee_intro.cgi). The annotated contigs will be released in the whole genome sequencing (WGS) section of

GenBank (http://www.ncbi.nlm.nih.gov/projects/WGS/WGSprojectlist.cgi). Manatee is a freely available, open source, web-based gene evaluation and genome annotation tool (http://manatee.sourceforge.net/). The Manatee interface allows users to quickly identify genes and make high quality functional assignments, such as functional classifications, using search data, paralogous families, and annotation suggestions generated from automated analysis. The analyses will then be presented in a graphical user interface that will be available to the scientific community over the National Center for Biotechnology information (NCBI) (http://www.ncbi.nlm.nih.gov/). We will use the two reference ETEC genomes H10407 and E24377A as reference molecules for the annotation. Other regions that are unique will be assigned automated annotation using appropriate references.

Whole genome comparative analysis. All genomes will be examined using Mugsy, a whole genome aligner developed at IGS by Sam Anguioli [56] and first implemented in the comparison of ETEC genomes. This tool will allow for the identification of novel regions in any genome, or set of genomes as well as the generation of a conserved core genomic region. These applications will provide the ability to rapidly identify regions that are conserved or unique and thus speeding the pace of development of molecular diagnostics or identification of vaccine and therapeutic targets.

4b. Approach to Data Production: **Data Analysis**

3. *Briefly describe the analysis (value-add) envisioned to be performed subsequently by the community and the potential to develop hypotheses driven proposals given the datasets and resources produced by this work.*

Enterotoxigenic *E. coli* remain a persistent threat to the health of both children and adults in developing countries, as well as an emerging pathogen in more developed regions. Yet, at present there is no broadly protective vaccines on the immediate horizon, and despite decades of investigation the pathogenesis of these organisms remains relatively poorly understood. The data generated in this proposal should greatly accelerate vaccine development for these important pathogens as well as provide new avenues for investigation of the pathogenesis of ETEC. In preparing this proposal, we have focused only on strains of established provenance with clearly defined clinical outcomes to enhance our ability to link specific disease patterns to particular virulence genes. The data generated by this proposal, made publicly available, should be invaluable to the ETEC research community at large in:

- aiding the rational selection of vaccine targets by identifying highly conserved antigens in a broad population of isolates
- promoting investigation of ETEC pathogenesis by highlighting specific genes or loci that appear in strains with severe clinical outcomes. These genetic features can be examined for their ability to contribute to pathogenesis using both *in vivo* and *in vitro* models of ETEC infection.
- developing our understanding of the nature of the protective immune responses that develop following ETEC infection, something which is currently poorly understood. Because patient serologic samples (including antibody in lymphocyte supernatants, ALS) exist for many of the ICDDR,B isolates included in this proposal, immune responses to specific antigens (including otherwise hypothetical features), can be assessed following infection with ETEC. The ultimate development of *in vitro* tests that correlate with protective immunity is a high priority for the ETEC research community, and will likely be crucial in the development of an ETEC vaccine [32].
- developing advanced rapid diagnostic tests for ETEC infections, another priority for the community[32]. Despite the ubiquitous nature of these infections, diagnostic capability currently rests in highly specialized reference laboratories. Development of a rapid test would speed trials of vaccine efficacy and ultimately allow more effective deployment of vaccine to high-risk populations.

5. Community Support and Collaborator Roles:

4. *Provide evidence of the relevant scientific community's size and depth of interest in the proposed sequencing or genotyping data for this organism or group of organisms. Please provide specific examples.*

Diarrheal diseases are second only to pneumonia as a cause of childhood mortality in developing countries. Three pathogens, ETEC, rotavirus, and *Shigella* collectively account for the majority of these deaths. With the advent of effective rotovirus vaccines the remaining bacterial pathogens, ETEC and *Shigella* are now the principle targets for vaccine development[32, 57]. These pathogens are also high priority targets for the U.S. military as they are the pathogens that commonly affect soldiers deployed to endemic regions. The proposed data to be generated here, and made freely available would greatly accelerate rational selection of target antigens. Multiple organizations in addition to the NIAID, are invested in development, testing, and eventual deployment of vaccines for ETEC and Shigella. These include:

- the Gates Foundation (http://www.gatesfoundation.org/global-health/Documents/enteric-and-diarrheal-diseases-strategy.pdf
- PATH, through its enteric vaccine initiative (EVI) (http://www.path.org/projects/enteric_vaccine.php)
- EntVac, an enteric vaccine initiative funded by the Research Council of Norway.
- the U. S. military (http://www.med.navy.mil/sites/nmrc/pages/id_ed.htm)
- the WHO (http://www.who.int/vaccine_research/diseases/diarrhoeal/en/index4.html)

Initiatives already underway involve hundreds of individual investigators that would soon have access to an invaluable dataset to direct future investigations as outlined above. We believe that open dissemination of data as proposed here would greatly facilitate these collective efforts.

5. *List all project collaborators and their roles in the project*

James M. Fleckenstein, M.D. (PI) Dr. Fleckenstein's laboratory at the University of Tennessee has amassed considerable experience in the identification and characterization of putative ETEC virulence factors, and exploring their potential as vaccine candidates in an animal model of ETEC infection. Dr. Fleckenstein will coordinate the acquisition, maintenance, and deposition of the ETEC strains in the BEI archive. He will also oversee prospective collection of multiple isolates from individual patients at the Institute for Diarrheal Diseases Research in Dhaka, Bangladesh (ICDDR,B) and be responsible for isolation of DNA in accord with GSC protocols. Dr. Fleckenstein's lab will ultimately clone and express antigens of interest identified in these studies for subsequent characterization of immune responses in infected patients, and test the protective efficacy of individual proteins in a murine model of ETEC infection. In addition, his laboratory plans to investigate the role of candidate virulence molecules identified in these studies by constructing mutant strains to assess contribution of individual genes in his *in vitro* and *in vivo* models.

Firdausi Qadri, Ph.D. (co-investigator) Dr. Qadri heads the Immunology laboratory at ICDDR,B where she has conducted multiple seminal studies on ETEC. Dr. Qadri and her laboratory routinely isolate and characterize ETEC strains obtained from patients and maintain a vast collection of bacterial specimens and corresponding serologic samples. Her laboratory is uniquely positioned to investigate immune responses to ETEC during the course of natural infections.

David Rasko, Ph.D. (co-investigator) Dr. Rasko is a faculty member of the University of Maryland School of Medicine and the Institute for Genome Sciences (IGS). Dr. Rasko and his laboratory have sequenced 7/8 currently available ETEC genomes, and has

considerable experience in bioinformatic analysis of pathogen genomes, expression profiling with DNA microarrays, and real-time PCR.

6. *List availability of other funding sources for the project.*

Dr Fleckenstein is the PI for project 1R01AI089894-01 that includes both Dr. Rasko and Dr. Qadri as co-investigators. The major focus of this project is the discovery and testing of novel candidate antigens that might be incorporated into a multivalent, broadly protective vaccine for ETEC. These studies involve high-throughput proteomic platforms, comparative genome hybridizations (CGH), and transcriptome analysis to identify genes that conserved, expressed, and recognized immunologically during the course of infection. Dr. Fleckenstein is also the PI on a separate ETEC pathogenesis grant that is funded by the Department of Veterans Affairs (Merit Review, Molecular Pathogenesis of Enterotoxigenic *Escherichia coli* infections). These studies focus on the role of novel adhesins or other surface expressed molecules of ETEC, and their role in bacterial adherence, toxin delivery, and colonization.

Dr. Rasko is involved in a number of *E. coli* sequencing projects, namely an additional *E. coli* and *Shigella* GSCID project that more broadly covers these pathogenic species (http://gscid.igs.umaryland.edu/emerging_diarrheal_pathogens.php). Additionally Dr. Rasko is part of the UMB team that has a funded U19- 1U19AI090873-01 “Severe Enteric Disease: Pathogenesis and Response”. This program examines the genomics of *Shigella* species and enteropathogenic *E. coli*.

6. Availability & Information of Strains:

7. *Indicate availability of relevant laboratory strains and clinical isolates. Are the strains/isolates of interest retrospectively collected, prepared and ready to ship?*

Note: If samples are prospectively prepared the GSC can provide protocols and recommendation based on the Centers past experiences. The samples must however meet minimum quality standards as established by the Center for the optimal technology platform (sequencing/ genotyping) to be used in the study.

All of the strains in this proposal have been collected and categorized with the exception of samples for testing the genotypic population diversity of strains from individual patients with severe diarrhea. (n=50 strains, samples 92-141 in the accompanying Excel file). Collection of these later strains is scheduled to begin in March, 2011 when Dr. Fleckenstein will travel to Dr. Qadri’s laboratory at ICDDR,B.

8. *Attach relevant information, if available in an excel spreadsheet for multiple samples: e.g*

see attached Excel Spreadsheet

9. *What supporting metadata and clinical data have been collected or are planned on being collected that could be made available for community use?*

The isolates in these proposed studies are confined to strains for which there are at a minimum accompanying data regarding the severity of the clinical illness. Each strain also comes linked to metadata that includes origin, date of isolation, serotype, toxin profile, and if published, the accompanying reference with PMID number referring to the unique PubMed ID. Metadata made available with the strains at the completion of the studies, will include each of these parameters as well as a finite list of antigens of interest.

7. Compliance Requirements:

7a. Review NIAID’s Reagent, Data & Software Release Policy:

NIAID supports rapid data and reagent release to the scientific community for all sequencing and genotyping projects funded by NIAID GSC. It is expected that projects will adhere to the data and reagent release policy described in the following web sites.

http://www3.niaid.nih.gov/LabsAndResources/resources/mscs/data.htm
http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-013.html
Once a white paper project is approved, NIAID GSC will develop with the collaborators a detailed data and reagent release plan to be reviewed and approved by NIAID.

Accept ☒ Decline ☐

7b. Public Access to Reagents, Data, Software and Other Materials:

10. State plans for deposit of starting materials as well as resulting reagents, resources, and datasets in NIAID approved repositories. Sequencing projects will not begin until the strain is deposited into NIAID funded BEI repository (http://www.beiresources.org/). This includes web based forms are completed by the collaborator and received by the NIAID BEI (http://www.beiresources.org/).

Following approval by NIAID, all of the strains in this proposal will be deposited in the Biodefense and Emerging infections Research Resources Repository (http://www.beiresources.org/) prior to initiation of sequencing.

All of the genomic data will be deposited in the NCBI GenBank upon verification by the GSCID. All of the sequence assemblies, and annotations will likewise be submitted within 45 (calendar) days of generation and GSCID verification. All accompanying metadata will also be submitted using tools available via the NIAID Bioinformatic Resource Centers, including the Meta Genome Rapid Annotation Subsystem Technology (MG-RAST) platform available (http://metagenomics.nmpdr.org/) via the NIAID-sponsored National Microbial Pathogen Data Resource (NMPDR) http://metagenomics.nmpdr.org/.

7c. Research Compliance Requirements

Upon project approval, NIAID review of relevant IRB/IACUC documentation is required prior to commencement of work. Please contact the GSC Principal Investigator(s) to ensure necessary documentation are filed for / made available for timely start of the project.

Investigator Signature:

Investigator Name: **James M. Fleckenstein, M.D.**

Date **01.24.2011**

references

1. Qadri, F., Svennerholm, A.M., Faruque, A.S. & Sack, R.B. Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin Microbiol Rev* **18**, 465-483 (2005).
2. Sack, R.B. et al. Enterotoxigenic Escherichia coli isolated from patients with severe cholera-like disease. *J Infect Dis* **123**, 378-385 (1971).

3. Finkelstein, R.A., Vasil, M.L., Jones, J.R., Anderson, R.A. & Barnard, T. Clinical cholera caused by enterotoxigenic Escherichia coli. *J Clin Microbiol* **3**, 382-384 (1976).
4. Taneja, N., Rao, P., Rao, D.S., Singh, M. & Sharma, M. Enterotoxigenic Escherichia coli causing cholerogenic syndrome during an interepidemic period of cholera in North India. *Jpn J Infect Dis* **59**, 245-248 (2006).
5. Nalin, D.R. & Richardson, S.H. Diarrhea resembling cholera induced by Escherichia coli culture filtrate. *Lancet* **1**, 677-678 (1973).
6. De, S.N., Bhattacharya, K. & Sarkar, J.K. A study of the pathogenicity of strains of Bacterium coli from acute and chronic enteritis. *J Pathol Bacteriol* **71**, 201-209 (1956).
7. Nalin, D.R., Bhattacharjee, A.K. & Richardson, S.H. Cholera-like toxic effect of culture filtrates of Escherichia coli. *J Infect Dis* **130**, 595-607 (1974).
8. Etkin, S. & Gorbach, S.L. Studies on enterotoxin from Escherichia coli associated with acute diarrhea in man. *J Lab Clin Med* **78**, 81-87 (1971).
9. Nalin, D.R., McLaughlin, J.C., Rahaman, M., Yunus, M. & Curlin, G. Enterotoxigenic Escherichia coli and idiopathic diarrhoea in Bangladesh. *Lancet* **2**, 1116-1119 (1975).
10. Sack, D.A., Merson, M.H., Wells, J.G., Sack, R.B. & Morris, G.K. Diarrhoea associated with heat-stable enterotoxin-producing strains of Escherichia coli. *Lancet* **2**, 239-241 (1975).
11. Sack, D.A., McLaughlin, J.C., Sack, R.B., Orskov, F. & Orskov, I. Enterotoxigenic Escherichia coli isolated from patients at a hospital in Dacca. *J Infect Dis* **135**, 275-280 (1977).
12. Vicente, A.C. et al. Outbreaks of cholera-like diarrhoea caused by enterotoxigenic Escherichia coli in the Brazilian Amazon Rainforest. *Trans R Soc Trop Med Hyg* **99**, 669-674 (2005).
13. Qadri, F. et al. Enterotoxigenic Escherichia coli and Vibrio cholerae diarrhea, Bangladesh, 2004. *Emerg Infect Dis* **11**, 1104-1107 (2005).
14. Fleckenstein, J.M. et al. Molecular mechanisms of enterotoxigenic Escherichia coli infection. *Microbes Infect* **12**, 89-98 (2010).
15. Ahmed, T. et al. Children with the Le(a+b-) blood group have increased susceptibility to diarrhea caused by enterotoxigenic Escherichia coli expressing colonization factor I group fimbriae. *Infect Immun* **77**, 2059-2064 (2009).
16. Qadri, F. et al. Disease burden due to enterotoxigenic Escherichia coli in the first 2 years of life in an urban community in Bangladesh. *Infect Immun* **75**, 3961-3968 (2007).
17. Chowdhury, F. et al. Concomitant enterotoxigenic Escherichia coli infection induces increased immune responses to Vibrio cholerae O1 antigens in patients with cholera in Bangladesh. *Infect Immun* **78**, 2117-2124 (2010).
18. Qadri, F. et al. Prevalence of Toxin Types and Colonization Factors in Enterotoxigenic Escherichia coli Isolated during a 2-Year Period from Diarrheal Patients in Bangladesh. *J. Clin. Microbiol.* **38**, 27-31 (2000).
19. Shah, N., DuPont, H.L. & Ramsey, D.J. Global etiology of travelers' diarrhea: systematic review from 1973 to the present. *Am J Trop Med Hyg* **80**, 609-614 (2009).
20. Roels, T.H. et al. Clinical features of infections due to Escherichia coli producing heat-stable toxin during an outbreak in Wisconsin: a rarely suspected cause of diarrhea in the United States. *Clin Infect Dis* **26**, 898-902 (1998).
21. Beatty, M.E. et al. Enterotoxin-producing Escherichia coli O169:H41, United States. *Emerg Infect Dis* **10**, 518-521 (2004).
22. Devasia, R.A. et al. Endemically acquired foodborne outbreak of enterotoxin-producing Escherichia coli serotype O169:H41. *Am J Med* **119**, 168 e167-110 (2006).
23. Jain, S. et al. An outbreak of enterotoxigenic Escherichia coli associated with sushi restaurants in Nevada, 2004. *Clin Infect Dis* **47**, 1-7 (2008).
24. Mintz, E.D. Enterotoxigenic Escherichia coli: outbreak surveillance and molecular testing. *Clin Infect Dis* **42**, 1518-1520 (2006).
25. Dalton, C.B., Mintz, E.D., Wells, J.G., Bopp, C.A. & Tauxe, R.V. Outbreaks of enterotoxigenic Escherichia coli infection in American adults: a clinical and epidemiologic profile. *Epidemiol Infect* **123**, 9-16 (1999).
26. Daniels, N.A. et al. Traveler's diarrhea at sea: three outbreaks of waterborne enterotoxigenic Escherichia coli on cruise ships. *J Infect Dis* **181**, 1491-1495 (2000).

27. Naimi, T.S. et al. Concurrent outbreaks of Shigella sonnei and enterotoxigenic Escherichia coli infections associated with parsley: implications for surveillance and control of foodborne illness. *J Food Prot* **66**, 535-541 (2003).
28. Beatty, M.E. et al. Epidemic diarrhea due to enterotoxigenic Escherichia coli. *Clin Infect Dis* **42**, 329-334 (2006).
29. Boschi-Pinto, C., Velebit, L. & Shibuya, K. Estimating child mortality due to diarrhoea in developing countries. *Bull World Health Organ* **86**, 710-717 (2008).
30. Fischer Walker, C.L., Sack, D. & Black, R.E. Etiology of diarrhea in older children, adolescents and adults: a systematic review. *PLoS Negl Trop Dis* **4**, e768 (2010).
31. WHO Future directions for research on enterotoxigenic Escherichia coli vaccines for developing countries. *Wkly Epidemiol Rec* **81**, 97-104 (2006).
32. Enterotoxigenic Escherichia coli: advances in technical and laboratory aspects of research and development of vaccines. *Wkly Epidemiol Rec* **83**, 92-95 (2008).
33. Petri, W.A., Jr. et al. Enteric infections, diarrhea, and their impact on function and development. *J Clin Invest* **118**, 1277-1290 (2008).
34. Peruski, L.F., Jr. et al. Phenotypic Diversity of Enterotoxigenic Escherichia coli Strains from a Community-Based Study of Pediatric Diarrhea in Periurban Egypt. *J. Clin. Microbiol.* **37**, 2974-2978 (1999).
35. Boedeker, E.C. Vaccines for enterotoxigenic Escherichia coli: current status. *Curr Opin Gastroenterol* **21**, 15-19 (2005).
36. Clemens, J.D. et al. Cross-protection by B subunit-whole cell cholera vaccine against diarrhea associated with heat-labile toxin-producing enterotoxigenic Escherichia coli: results of a large-scale field trial. *J Infect Dis* **158**, 372-377 (1988).
37. Sahl, J.W. et al. A comparative genomic analysis of diverse clonal types of enterotoxigenic Escherichia coli reveals pathovar-specific conservation. *Infect Immun* (2010).
38. Fleckenstein, J.M., Roy, K., Fischer, J.F. & Burkitt, M. Identification of a two-partner secretion locus of enterotoxigenic Escherichia coli. *Infect Immun* **74**, 2245-2258 (2006).
39. Roy, K. et al. Enterotoxigenic Escherichia coli EtpA mediates adhesion between flagella and host cells. *Nature* **457**, 594-598 (2009).
40. Patel, S.K., Dotson, J., Allen, K.P. & Fleckenstein, J.M. Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic Escherichia coli. *Infect Immun* **72**, 1786-1794 (2004).
41. Roy, K., Bartels, S., Qadri, F. & Fleckenstein, J.M. Enterotoxigenic Escherichia coli elicits immune responses to multiple surface proteins. *Infect Immun* **78**, 3027-3035 (2010).
42. Roy, K., Hamilton, D., Allen, K.P., Randolph, M.P. & Fleckenstein, J.M. The EtpA exoprotein of enterotoxigenic Escherichia coli promotes intestinal colonization and is a protective antigen in an experimental model of murine infection. *Infect Immun* **76**, 2106-2112 (2008).
43. Roy, K., Hamilton, D., Ostmann, M.M. & Fleckenstein, J.M. Vaccination with EtpA glycoprotein or flagellin protects against colonization with enterotoxigenic Escherichia coli in a murine model. *Vaccine* (2009).
44. Vila, J. et al. Quinolone resistance in enterotoxigenic Escherichia coli causing diarrhea in travelers to India in comparison with other geographical areas. *Antimicrob Agents Chemother* **44**, 1731-1733 (2000).
45. Ouyang-Latimer, J. et al. In Vitro Antimicrobial Susceptibility of Bacterial Enteropathogens Isolated from International Travelers to Mexico, Guatemala, and India, 2006-2008. *Antimicrob Agents Chemother* (2010).
46. Mendez Arancibia, E. et al. Evolution of antimicrobial resistance in enteroaggregative Escherichia coli and enterotoxigenic Escherichia coli causing traveller's diarrhoea. *J Antimicrob Chemother* **64**, 343-347 (2009).
47. Crossman, L.C. et al. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407. *J Bacteriol* (2010).
48. Rasko, D.A. et al. The pan-genome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. *J Bacteriol* (2008).

49. Steinsland, H. et al. Protection from natural infections with enterotoxigenic Escherichia coli: longitudinal study. *Lancet* **362**, 286-291 (2003).
50. Evans, D.J., Jr. & Evans, D.G. Three characteristics associated with enterotoxigenic Escherichia coli isolated from man. *Infect Immun* **8**, 322-328 (1973).
51. Coster, T.S. et al. Immune response, ciprofloxacin activity, and gender differences after human experimental challenge by two strains of enterotoxigenic Escherichia coli. *Infect Immun* **75**, 252-259 (2007).
52. Shaheen, H.I. et al. Phenotypic and genotypic analysis of enterotoxigenic Escherichia coli in samples obtained from Egyptian children presenting to referral hospitals. *J Clin Microbiol* **47**, 189-197 (2009).
53. Galbadage, T., Jiang, Z.D. & DuPont, H.L. Improvement in detection of enterotoxigenic Escherichia coli in patients with travelers' diarrhea by increasing the number of E. coli colonies tested. *Am J Trop Med Hyg* **80**, 20-23 (2009).
54. Ram, S., Khurana, S., Singh, R.P. & Khurana, S.B. Loss of some virulence factors of enterotoxigenic Escherichia coli on repeated subcultures. *Indian J Med Res* **95**, 284-287 (1992).
55. Evans, D.G., Evans, D.J., Jr., Tjoa, W.S. & DuPont, H.L. Detection and characterization of colonization factor of enterotoxigenic Escherichia coli isolated from adults with diarrhea. *Infect Immun* **19**, 727-736 (1978).
56. Angiuoli, S.V. & Salzberg, S.L. Mugsy: Fast multiple alignment of closely related whole genomes. *Bioinformatics* (2010).
57. Svennerholm, A.M. & Tobias, J. Vaccines against enterotoxigenic Escherichia coli. *Expert review of vaccines* **7**, 795-804 (2008).