

## MPB FINAL RESEARCH GRANT REPORT

**I. Project Title and NPB project identification number:** Reverse Vaccinology to Identify Antigens for Control of ETEC-caused Post-Weaning Diarrhea in Swine

**Principal Investigators:** Timothy Johnson and Richard Isaacson

**Institution:** University of Minnesota

**Date Report Submitted:** 12/19/2009

### II. Industry Summary.

Post-weaning diarrhea (PWD) in pigs is a continuing problem for the swine industry. The goal of this study was to utilize a concept known as reverse vaccinology towards the identification of novel vaccine candidates to be used towards the future development of a vaccine providing heterologous protection against *E. coli* causing PWD in swine. The objectives used here to achieve this goal included sequencing the genome of a K88<sup>+</sup> porcine *E. coli* strain implicated in PWD and the use of this sequence to identify a subset of proteins with the greatest antigenic potential. This subset was identified using computational approaches to predict surface exposed proteins, and subsequent assessment of their prevalence among porcine diarrheagenic and commensal isolate collections. Using this approach, we were able to identify a subset of approximately 50 predicted proteins (from greater than 5,000 genes encoded by this genome) with antigenic potential. We sought the presence of the genes encoding the most promising of these proteins among 300 porcine *E. coli* isolates, and narrowed this subset to eight novel proteins that are significantly associated with *E. coli* causing PWD. This promising subset represents proteins that will be further characterized and assessed for their ability to elicit a mucosal immune response in weaned pigs. The incorporation of multiple antigens into a

recombinant attenuated *Salmonella* strain would enable us to invoke heterologous protection against PWD-associated *E. coli*, which would ultimately enable a cost effective and reasonable *E. coli* PWD vaccine for the swine industry.

### **III. Scientific Abstract.**

Enterotoxigenic *Escherichia coli* (ETEC) is the causative agent of post-weaning diarrhea (PWD) in production pigs, which is a disease that continues to be a major cause of morbidity for the swine industry. The primary ETEC strains implicated in PWD are well known and characterized. However, traditional approaches at a vaccine providing protection against PWD have not proven effective. Here, we utilized a global approach known as reverse vaccinology towards the identification of novel vaccine candidates. This approach involved generation of the first completed genome sequence of a K88<sup>+</sup> porcine ETEC isolate, subsequent mining of the predicted proteins within this genome for suitable antigenic candidates, and assessment of the prevalence of the gene sequences encoding for suitable antigenic candidates among porcine *E. coli* collections. Using this approach, we identified approximately 50 antigenic candidates from the K88 genome sequence. We have screened a collection of 300 porcine *E. coli* isolates for the presence of these candidates, narrowing this subset to eight genes that are significantly associated with PWD ETEC isolates and possess antigenic potential. This subset of isolates will be used in future studies aimed at developing a recombinant attenuated *Salmonella* vaccine that will elicit a mucosal immune response and protect against ETEC strains implicated in PWD.

#### **IV. Introduction.**

PWD is perhaps the most regularly-occurring disease of commercial pig farms (Fairbrother et al., 2005). This disease most often affects pigs that are weaning at approximately 3-4 weeks of age, and the disease starts a few days after lacteal protection stops. Protection of pigs and other animals from *E. coli* pathogens is a constant challenge, as these strains have proven highly adaptable under environmental pressures. This is especially true for ETEC and other types harboring transmissible plasmids encoding virulence factors (Nagy and Fekete, 1999). In fact, there are reports of recent surges in the occurrence of PWD in commercial farms around the world, and the ETEC implicated in these cases are often K88<sup>+</sup> or F18<sup>+</sup> multidrug resistant strains. Because of the genome diversity of ETEC, it has been challenging to respond to outbreaks involving these strains, and to develop strategies to prevent and treat diseases caused by them.

Great effort has been put forth towards controlling PWD in pigs. Potential control measures sought have included treatment with antibiotics, bacteriophage, probiotics, and vaccines. Of these, vaccination currently holds the greatest promise for a long-term solution. Unfortunately, active immunization against PWD in pigs has been met with very mixed results (Haesebrouck et al., 2004). The primary reasons for this include an inability to stimulate mucosal immunity, the complexities of PWD itself, and hypersensitivity or tolerance to antigens encountered earlier in life. Most of the vaccines against ETEC-caused PWD have targeted the K88 and F18 fimbrial components. This has proven ineffective towards providing comprehensive protection against PWD. Therefore, we sought to use a global approach allowing us to identify novel vaccine candidates that are universal among PWD strains. This approach is coined “reverse vaccinology.” Reverse

vaccinology is based on the concept that secreted or extracellular proteins are more accessible to antibodies and are therefore the best vaccine candidates (Mora et al., 2006). With the availability of advanced computer prediction programs that predict the destination of proteins based upon sequence, the availability of whole genome sequences means that vaccine candidates can be identified using a global, reverse approach.

In organisms such as *E. coli*, it is desirable to avoid targeting proteins possessed by commensal strains because they would invoke a response against the normal flora in the host. By screening both pathogen and commensal populations using PCR, the subset of vaccine candidates can be further refined to include only pathogen-specific sequences. The resulting subset of genes encodes predicted proteins are those with the greatest potential as vaccine candidates, are pathogen-associated, and can be further evaluated (Mora et al., 2006). Here, modern reverse vaccinology was applied to porcine ETEC in our initial efforts to identify alternative approaches towards a vaccine protecting against PWD in pigs.

## **V. Objectives.**

- 1. Genome sequencing of a representative ETEC strain implicated in porcine PWD.**
- 2. *in silico* identification of ETEC-specific proteins with vaccine potential.**
- 3. Determination of the prevalence of genes with vaccine potential among 300 porcine ETEC and porcine commensal isolates.**

## **VI. Materials and Methods.**

*Bacterial strain.* Our choice of a K88<sup>+</sup> strain to be sequenced utilized a large collection of pathogenic *E. coli* characterized for serotype, genotype, phylogenetic type. From these

collections and based on current literature, we selected a representative strain for sequencing, UMNK88. This strain belongs to the O149 serogroup, which is the most frequently occurring PWD ETEC serogroup. It possesses the F4ac (K88) fimbrial antigen, the EAST1 heat stable enterotoxin, STb, and LT. This strain was selected from recent Midwestern U.S. field isolates available within our collections and from the University of Minnesota Veterinary Diagnostic Lab.

Genome sequencing. UMNK88 was subjected to whole genome draft sequencing using 454 GS FLX technology with Titanium chemistry. We sequenced two libraries: a draft shotgun library and an 8-kb paired-end mate library. For each library, 20 micrograms of total genomic DNA was isolated from UMNK88 and diluted to a concentration of 500 ng/uL. Library preparation and draft sequencing was performed at the University of Minnesota's Biomedical Genomics Center. In total, ½ of a picotitre plate was sequenced, resulting in about 200,000,000 total bases sequenced. This corresponded to approximately 35-fold coverage of the genome. The sequencing reads were assembled using the Newbler Assembler from 454 Life Sciences, converted into an .ace file, and imported into SeqManPro software from Lasergene (Madison, WI). The Lasergene package was used to finish and edit the sequence, predict genes, and annotate the final gene predictions.

in silico identification of putative antigens. Each predicted protein was assessed for its potential as a vaccine candidate. P-SORT was used to predict the subcellular locations of each of the ETEC-specific proteins in the two sequenced strains. Signal peptide prediction was carried out on each protein using SignalP 3.0 (Bendtsen et al., 2004).  $\alpha$ -helix transmembrane topology prediction was carried out using TMHMM (Krogh et al., 2001). BOMP was used to predict  $\beta$ -barrel outer membrane proteins (Berven et al., 2004).

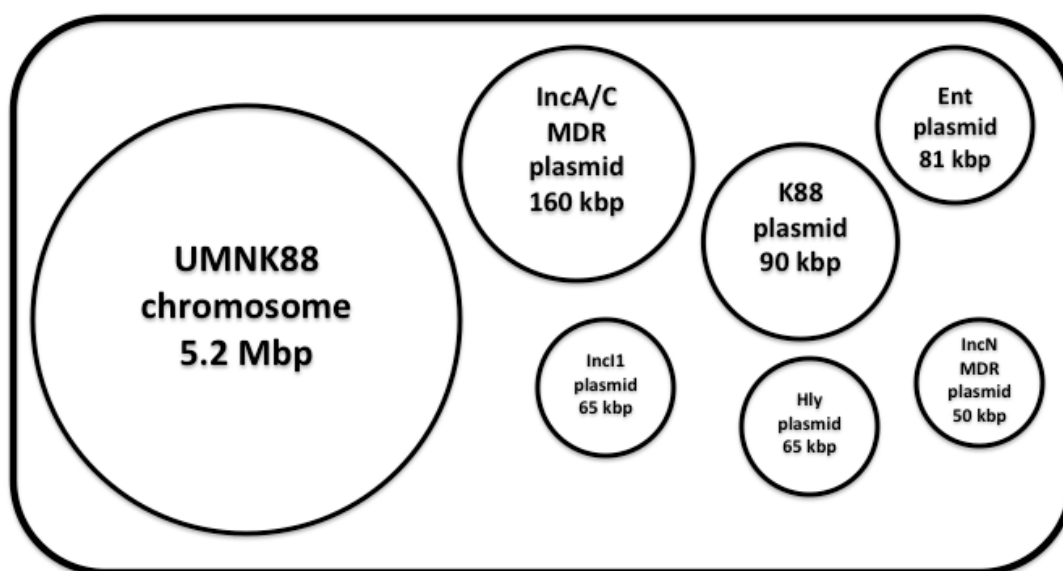
Putative lipoproteins were predicted by SpLiP (Setubal et al., 2006). From the analyses above, the best putative antigens were selected as those that 1) were surface exposed, and 2) possessed at least one of the following: a predicted signal peptide, no more than four  $\alpha$ -helices with transmembrane topology,  $\beta$ -barrel topology (indicating an outer membrane protein), or a predicted lipoprotein signal peptide.

*Prevalence of vaccine candidates.* The presence of identified vaccine candidates were sought among populations of 100 porcine ETEC involved in PWD, 100 porcine ETEC involved in neonatal diarrhea, 100 and porcine commensal *E. coli*. Total DNA was extracted using a boiled lysis preparation from an overnight 1 mL culture to serve as template DNA (Rodriguez-Siek et al., 2005). Template DNA from each isolate was denatured and spotted onto nylon membranes. Each membrane was probed with a PCR product amplifying one of the vaccine candidates. PCR products were DIG-labeled, hybridized to the array, stringency washed, and chemically detected.

## **VI. Results.**

*UMNK88 genome sequence.* The chromosome of UMNK88 was closed at approximately 5.2 Mbp. Remarkably, UMNK88 possessed 6 plasmids totaling approximately 511 kb, accounting for 9% of the strain's genome (Fig. 1). These plasmids included the IncF-type virulence plasmids (K88, Ent, and Hly) encoding K88 fimbriae, LT and ST enterotoxins, and hemolysin, respectively. Also included with UMNK88's plasmid complement were two multidrug resistance (MDR)-encoding plasmids containing the IncA/C and IncN replicon types. Encoded by the virulence plasmids were a number of novel predicted proteins with antigenic potential. In addition to the plasmid complement of UMNK88, 44

chromosomal genomic islands (GIs) greater than 2 kb were identified that were present in UMNK88 but absent in the laboratory strain K-12 MG1655 (Table 1). The remainder of the vaccine candidates identified were located in these genomic islands. Despite the presence of 44 GIs totaling 793 kb, the UMNK88 chromosome was most phylogenetically related to those of human commensal and laboratory *E. coli* strains (Fig. 2).



**Figure 1 (above).** Overview of the *E. coli* UMNK88 genome containing 6 plasmids.

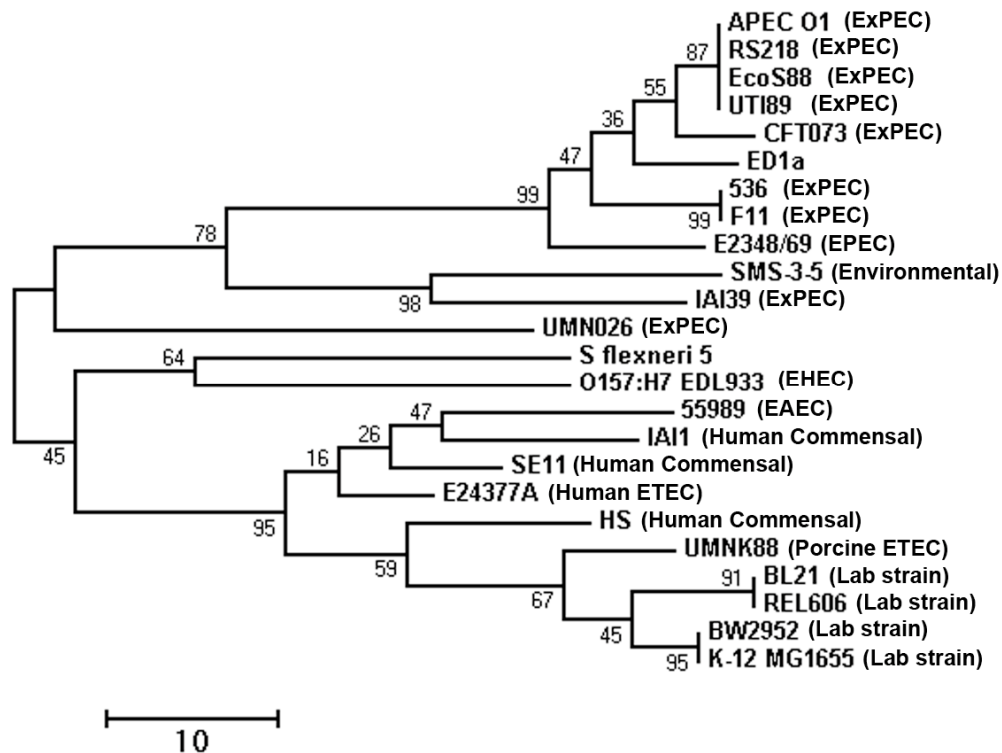
**Table 1.** Chromosomal GIs of porcine ETEC strain UMNK88.

GI #	Start	Stop	Length (bp)	Description
1	50185	75281	25096	Unknown GI
2	100379	102441	2062	Unknown GI
3	164818	170629	5811	Putative nutrient utilization and drug efflux GI
4	170777	175916	5139	Putative ribitol metabolism and transport operon
5	460372	499708	39336	Prophage GI
6	649495	687601	38106	Prophage GI

7	812602	849173	36571	Prophage GI
8	1067592	1075796	8204	Ett2 type 3 secretion system (incomplete)
9	1188623	1227474	38851	Putative pathogenicity island
10	1229124	1237125	8001	GI containing a type II secretion system
11	1255187	1263596	8409	Putative sugar metabolism GI
12	1412077	1414381	2304	Putative nutrient utilization GI
13	1695574	1699017	3443	Unknown GI
14	1930283	1934072	3789	O-antigen region
15	1948649	1965121	16472	Prophage GI
16	2162948	2171499	8551	Putative nutrient utilization GI
17	2173322	2178916	5594	Unknown GI
18	2372386	2376665	4279	GI conserved in pathogenic <i>E. coli</i>
19	2392712	2399029	6317	Putative sugar utilization GI
20	2530532	2596129	65597	Putative pathogenicity island
21	2642989	2648123	5134	Putative nutrient utilization GI
22	2730492	2740595	10103	Putative virulence-associated GI
23	2751281	2757975	6694	Unknown GI
24	2850871	2854920	4049	Unknown GI
25	3073007	3103024	30017	Unknown GI
26	3126039	3164558	38519	Prophage GI
27	3182661	3186377	3716	GI conserved in pathogenic <i>E. coli</i>
28	3439102	3449599	10497	GI conserved in pathogenic <i>E. coli</i>
29	3544061	3546215	2154	GI conserved in pathogenic <i>E. coli</i>
30	3744785	3780241	35456	Prophage GI
31	3834410	3871853	37443	Prophage GI
32	4016547	4109393	92846	Putative pathogenicity island
33	4195479	4237064	41585	Prophage GI
34	4348104	4359886	11782	Prophage GI
35	4447949	4503753	55804	Prophage GI

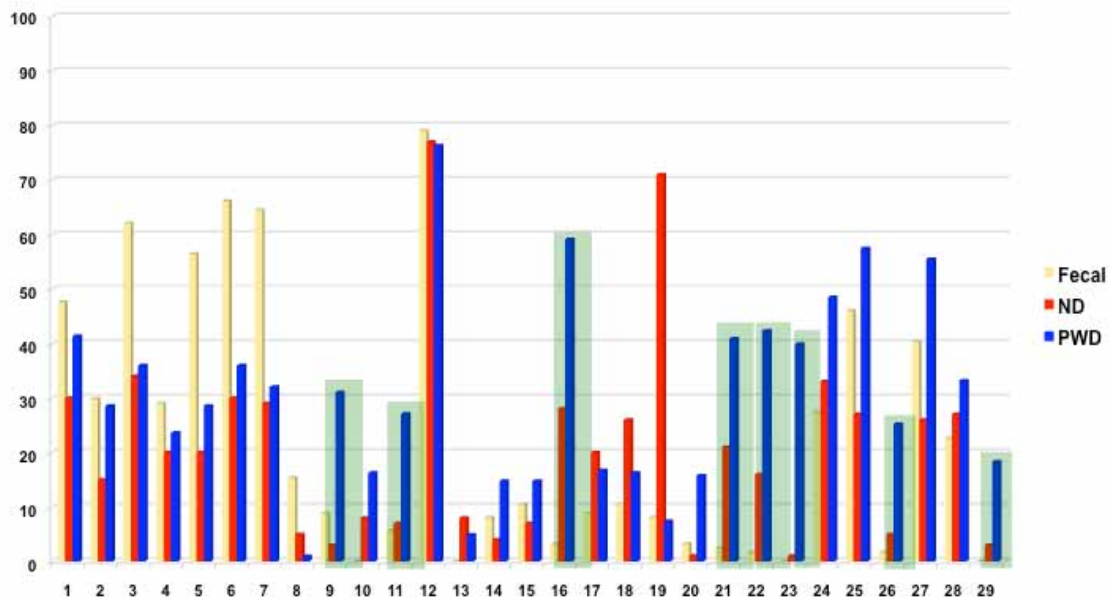
36	4599104	4606199	7095	GI conserved in pathogenic <i>E. coli</i>
37	4665100	4685274	20174	Unknown GI
38	4715231	4721516	6285	Putative sugar utilization GI
39	4736835	4758255	21420	Prophage GI
40	4769776	4774576	4800	Prophage GI
41	4778066	4780237	2171	GI conserved in pathogenic <i>E. coli</i>
42	4787892	4796961	9069	Unknown GI with matches in <i>Shigella</i> spp.
43	5146301	5148883	2582	Mobile elements
44	5173522	5176002	2480	GI conserved in pathogenic <i>E. coli</i>
		Total GIs	793,807	

**Figure 2** (below). *in silico* sequence alignment of housekeeping genes from UMNK88 and other sequenced *E. coli* strains.



From the UMNK88 finished sequence, nearly 50 predicted proteins were identified with vaccine candidate potential. We selected 29 of these putative antigens because they were

novel targets and subsequently screened 300 porcine *E. coli* isolates for their presence (Fig. 3). Eight of these targets were significantly associated with PWD isolates, as compared to those isolated from cases of neonatal diarrhea and commensal isolates from healthy pigs.



**Figure 3 (above).** Screening of 300 porcine *E. coli* isolates (100 fecal, 100 neonatal diarrhea - ND, and 100 PWD) for the presence of 29 genes encoding predicted proteins with antigenic potential (X-axis = antigen number; Y-axis = percent prevalence). From this group, eight were identified (highlighted in light green) that are significantly associated with PWD ( $p < 0.05$ ).

## VIII. Discussion.

Control of PWD in pigs has been problematic because of the complexities of the disease, the challenge of stimulating mucosal immunity in weaned pigs, and the diversity of *E. coli* strains causing PWD. Since classical approaches has been ineffective, we opted to take a global approach towards the identification of novel antigenic candidates to be used in future studies developing an effective vaccine against PWD. As expected, we identified a number of promising targets here for future study. However, we also obtained interesting results when we analyzed the UMNK88 genome. We expected the UMNK88 chromosome

to be most closely related to other intestinal pathogenic strains from a phylogeny standpoint. However, UMNK88 is most closely related to laboratory and commensal *E. coli* strains. This is not necessarily a surprising finding, since we would expect strains that are adapted as intestinal pathogens to be extremely efficient at colonizing the gut. Certainly, porcine strains with a commensal chromosomal background that acquire ETEC virulence-associated plasmids might represent the best combination required to cause disease. What is currently unknown is if all porcine ETEC are “commensal” clones with virulence plasmids, or if these plasmids have been acquired by multiple *E. coli* lineages, as demonstrated in human ETEC (Turner et al., 2006). If porcine ETEC do belong to commensal lineages, is this because they are the most efficient host-adapted strains or is it instead the result of a greater propensity for these strains to acquire the ETEC virulence plasmids? A better understanding of these questions will aid in our overall understanding of ETEC evolution and the risks posed by virulence plasmid transfer in production animal environments.

Because vaccines targeting the fimbrial components of common PWD-associated strains have not provided comprehensive protection against the full repertoire of porcine ETEC causing this disease, we hypothesize that a more effective approach is to utilize reverse genomics to identify a subset of proteins that will elicit an immune response against most of these strains. Usually, these subsets of proteins are subsequently used to design protein subunit vaccines. This approach has worked well for many types of bacterial disease. However, since a primary problem with PWD protection involves the stimulation of a mucosal immune response, logic would suggest that a protein subunit vaccine won't work well towards this endeavor. Therefore, we believe a more effective approach will be

the integration and expression of these proteins in a bacterium known to effectively stimulate mucosal immunity. Certainly, one such candidate for this approach is *S. Choleraesuis* strain CS54, which is licensed for use as a live vaccine for pigs in the United States. Recently, it has been shown that *E. coli* antigens can be incorporated into the chromosome of this strain and ultimately expressed. In the future, we hope to use such an approach with our subset of proteins to ultimately protect against the diverse collection of strains causing PWD in weaned pigs.

## IX. References.

**Bendtsen, J. D., H. Nielsen, H. G. von, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**:783-795.

**Berven, F. S., K. Flikka, H. B. Jensen, and I. Eidhammer.** 2004. BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Res.* **32**:W394-W399.

**Fairbrother, J. M., E. Nadeau, and C. L. Gyles.** 2005. *Escherichia coli* in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies. *Anim. Health Res. Rev.* **6**:17-39.

**Haesebrouck, F., F. Pasmans, K. Chiers, D. Maes, R. Ducatelle, and A. Decostere.** 2004. Efficacy of vaccines against bacterial diseases in swine: what can we expect? *Vet. Microbiol.* **100**:255-268.

**Krogh, A., B. Larsson, H. G. von, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**:567-580.

**Lee, E., R. Platt, S. Kang, J.A. Roth, and G.J. Phillips.** 2001. Chromosomal integration and expression of the *Escherichia coli* K88 gene cluster in *Salmonella enterica* ser. *Choleraesuis* strain 54 (SC54). *Vet Microbiol.* **83**:177-183.

**Mora, M., C. Donati, D. Medini, A. Covacci, and R. Rappuoli.** 2006. Microbial genomes and vaccine design: refinements to the classical reverse vaccinology approach. *Curr. Opin. Microbiol* **9**:532-536.

**Nagy, B. and P. Z. Fekete.** 1999. Enterotoxigenic *Escherichia coli* (ETEC) in farm animals. *Vet. Res.* **30**:259-284.

**Rodriguez-Siek, K. E., C. W. Giddings, C. Doetkott, T. J. Johnson, M. K. Fakhr, and L. K. Nolan.** 2005. Comparison of *Escherichia coli* isolates implicated in human urinary tract infection and avian colibacillosis. *Microbiol.* **151**:2097-2110.

**Setubal, J. C., M. Reis, J. Matsunaga, and D. A. Haake.** 2006. Lipoprotein computational prediction in spirochaetal genomes. *Microbiol.* **152**:113-121.

**Turner, S.M., R.R. Chaudhuri, Z. Jiang, H. DuPont, C. Gyles, C.W. Penn, M.J. Pallen, and I.R. Henderson.** 2006. Phylogenetic Comparisons Reveal Multiple Acquisitions of the Toxin Genes by Enterotoxigenic *Escherichia coli* Strains of Different Evolutionary Lineages. *J. Clin. Microbiol.* 44:4528-4536.