THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF FOOD SCIENCE

IDENTIFICATION OF PUTATIVE VIRULENCE BIOMARKERS IN PATHOGENIC SHIGA TOXIN-PRODUCING *ESCHERICHIA COLI*

EMILY A. McINTYRE
Spring 2012

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biology
with honors in Food Science

Reviewed and approved* by the following:

Edward Dudley
Assistant Professor of Food Science
Thesis Supervisor

Stephen Knabel
Professor of Food Science
Honors Advisor

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Shiga toxin-producing *E. coli* (STEC) are zoonotic pathogens that produce Shiga toxins, causing illness in humans ranging from mild diarrhea to bloody diarrhea, hemorrhagic colitis, and hemolytic uremic syndrome. Enterohemorrhagic *E. coli* (EHEC), such as *E. coli* O157:H7, are a subset of STEC that are capable of causing the most severe disease, such as hemorrhagic colitis and the hemolytic uremic syndrome, in humans. *E. coli* O157:H7 is the STEC most frequently isolated from clinical outbreaks, with raw ground beef being the most common vehicle for transmission. While the unique biochemical properties of *E. coli* O157:H7 have made detecting this serotype fairly straightforward, other STEC that are capable of causing disease lack distinguishing biochemical characteristics and have been underrecognized as pathogens. The increased recognition of non-O157 STEC as a cause of disease and outbreaks has made detecting them in food and in clinical samples of great importance. Also, because not all STEC are capable of causing disease, there is need for a method of differentiating highly virulent from less harmful strains. A recent study identified 87 genes that are conserved among fully sequenced strains of O157, O26, O111, and O103 serogroups but absent in other *E. coli.* The objective of this study was to determine if these 87 genes are present in a larger collection of EHEC strains and to determine the prevalence of a subset of these genes in STEC. The *nleB* gene, encoded on OI-122 (a genome segment found in *E. coli* O157:H7 but not *E. coli* K-12), is an effector gene of the type III secretion system. This gene was found in all isolates of O26, O103, O111, O157, O145, and O121, but was less prevalent in other STEC serogroups. The results of this study indicate that *nleB* is the most reliable marker of virulence among the STEC isolates screened.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am very grateful for the help and support I received as I learned lab techniques, carried out my project, and wrote this thesis. The assistance of the following individuals is greatly appreciated:

Dr. Edward Dudley

Dr. Stephen Knabel

Dr. Chobi DebRoy

Shuang Yin

Chun Chen

Carrie Lewis

Kakolie Goswami

I would especially like to acknowledge and sincerely thank Dr. Edward Dudley for his endless guidance throughout this process. Also, thank you to lab-mate Shuang Yin for helping me begin this project and analyze my results. Thank you to lab-mate Chun Chen for his continuous help and encouragement. Finally, I am very grateful for the funding I received from the Penn State Center for Food Manufacturing.

LITERATURE REVIEW

Escherichia coli

Escherichia coli are a genetically diverse group of gram negative, rod-shaped bacteria. Many *E. coli* are commensal organisms, living in humans and warm-blooded animals as part of the normal gut microflora. These bacteria are generally harmless to the host, causing disease only in rare cases or in a host that is immunocompromised. Other *E. coli* have acquired virulence factors that enable them to cause a broad spectrum of disease, such as diarrheal disease, urinary tract infections, sepsis and meningitis (39).

Serotyping is used to classify microorganisms based on antigen-antibody reactions. *E. coli* isolates can be differentiated by serotyping their O and H antigens. The O antigen derives from the lipopolysaccharde of the outer membrane and identifies the serogroup while the H antigen derives from the flagellar protein and, along with the O antigen, identifies the serotype (39). Serotyping is used for tracing outbreaks and for determining risk factors during infection. Certain serogroups, such as O157, O26, and O103 are associated with human disease.

Pathogenic *E. coli* that cause gastroenteritis are divided into six different classes. These are enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), diffuse-adhering *E. coli*, enteroaggregative *E. coli* (EAEC), and enterohemorrhagic *E. coli* (EHEC) (5). This review will focus on EHEC and related organisms.

Shiga Toxin-Producing *Escherichia coli*

Shiga toxin-producing *E. coli* (STEC) are zoonotic pathogens that produce Shiga toxins, of which the two main types are Shiga toxin 1 (Stx 1) and Shiga toxin 2 (Stx2; 29). These toxins

contribute to pathology by inhibiting host protein synthesis and damaging endothelial cells (30, 35, 49). STEC are capable of causing illness in humans ranging from mild diarrhea to bloody diarrhea, hemorrhagic colitis and the hemolytic uremic syndrome (4). EHEC, such as *E. coli* O157:H7, are a subset of STEC that cause the most severe disease, such as hemorrhagic colitis and the hemolytic uremic syndrome, in humans. Hemorrhagic colitis (HC) is a syndrome characterized by bloody diarrhea, severe abdominal pain, and little or no fever (50). The hemolytic uremic syndrome (HUS), which mainly affects children, is characterized by acute renal failure, thrombocytopenia, and microangiopathic hemolytic anemia (34, 40, 48). HUS, caused by *E. coli* O157:H7 and other STEC, is estimated to cause death or end-stage renal disease in about 12% of patients and about 25% of these survivors demonstrate long-term renal sequelae (24).

More than 600 STEC serotypes have been described (21) but only a few have been associated with human illness. *E. coli* O157:H7 is the most recognized STEC, estimated to cause 2,238 hospitalizations and about 20 deaths each year in the United States (54). This human pathogen is a commensal organism in the intestines of cattle and other ruminants, which serve as reservoir hosts (14). The most frequent mode of transmission to humans is through the consumption of contaminated food and water (4) and the infectious dose may be fewer than 50 organisms (60). Up to 30% of healthy cattle carry this organism in their intestines (4, 18, 29, 52, 59, 63) and undercooked ground beef is the most recognized food vehicle for *E. coli* O157:H7 transmission (21). Outbreaks are most common during the warmer months of the year (8, 15) and infections are more frequently seen in the young and the elderly (26).

The term “typical EHEC” is used here to designate isolates of *E. coli* O157:H7 that lack

the ability to ferment sorbitol within 24 hours and the ability to produce β-glucuronidase (47). These distinguishing biochemical characteristics make detection of *E. coli* O157:H7 in environmental, food, and clinical samples straightforward and inexpensive. The ease in identification of this serotype has contributed to its recognition as a human pathogen.

The term non-O157 STEC refers to *E. coli* strains that produce Shiga toxins but belong to serogroups other than O157. The Center for Disease Control and Prevention estimates that non-O157 STEC causes 112,752 annual foodborne illnesses in the U.S. while O157 causes 63,153 (54). In July of 2011, STEC O104 linked to raw sprouts was responsible for a large-scale outbreak in Germany, resulting in 852 HUS cases and 32 associated deaths (http://www.cdc.gov/outbreaknet/outbreaks.html). Other recent non-O157 STEC outbreaks in the U.S. include one in May of 2011 traced to *E. coli* O145 from Romaine lettuce, and another in March of 2012 linked to *E. coli* O26 isolated from raw clover sprouts at Jimmy John's restaurants (http://www.cdc.gov/outbreaknet/outbreaks.html).

Although non-O157 STEC have been implicated in clinical outbreaks and human disease, they have been underrecognized due to limited diagnostics and surveillance capabilities. Unlike *E. coli* O157:H7, these strains lack distinguishing biochemical properties that would assist in their detection. The most common non-O157 STEC serogroups associated with human outbreaks include O26, O111, O103, O121, O45, and O145 (10). These 6 serogroups accounted for 71% of non-O157 STEC isolates recovered in the United States from 1983 to 2002 (10). Although a large number of non-O157 STEC strains have been identified, most have rarely, if ever, been associated with human disease. A means of differentiating between pathogenic non-O157 STEC strains and STEC that apparently do not transmit to humans or cause disease is

currently lacking.

STEC and the Food Supply

Although many different STEC vehicles have been identified, the majority of *E. coli* O157:H7 infections have been linked to the consumption of undercooked ground beef (39). Cattle can harbor these organisms and other STEC serotypes in their GI tract (63) without becoming ill and shed the bacteria in their feces. Contamination of commercial meat with STEC usually occurs in beef cattle processing plants. One study determined a non-O157 STEC prevalence of greater than 50% on beef carcasses prior to in-plant antimicrobial interventions in U.S. processing plants (2). When beef cattle are slaughtered, contact with feces or transfer of STEC from the cattle hides to the carcasses can lead to the contamination of meat. Also, when needle tendered beef is processed, the needles can transfer bacteria present on the outside of the meat into the center of a product where it is less likely to be killed by high temperatures during food preparation. Cooking beef products to an internal temperature of 68.3 °C will inactivate *E. coli* O157:H7 (32), however, this is not always accomplished by consumers or restaurants.

Outbreaks involving ground beef can be difficult to trace because meat processing plants obtain cattle and cuts of meat from many different sources (4). These industries distribute their products over large distances to many different areas. As a result, a product contaminated with EHEC can cause widespread transmission and disease. It is therefore important to detect EHEC strains in ground beef before a contaminated product leaves the manufacturer.

Due to its high virulence and implication in human disease, in 1994 the USDA's Food Safety and Inspection Service declared *E. coli* O157:H7 an adulterant in raw ground beef (46). Six non-O157 STEC serogroups (O26, O103, O111, O145, O121, and O45) that are capable of causing severe human illness were recently proposed to be added to the list of adulterants (1). Because of the damage caused by Shiga toxins, it has also been argued that all STEC be considered dangerous and kept out of the food supply. While it is easy for meat processing plants and the USDA to detect *E. coli* O157:H7, non-O157 STEC lack unique biochemical properties that would aid in their detection. A means by which these strains can be identified and characterized is therefore needed. Also, because not all STEC are capable of causing human disease, a method of differentiating between pathogenic STEC and those that are less harmful must be developed. **The main goal of this study was to identify genes that are more prevalent in highly virulent STEC and that could be used as molecular markers in detecting them. This information could then be used to assess the risks posed by these potentially highly virulent STEC strains.**

Virulence Factors

EHEC have acquired a number of virulence factors that contribute to their pathogenicity. These factors include toxins, attachment proteins, capsules, and other proteins used to avoid host immune systems, adhere to host cell membranes, and colonize the intestines (55). The most well-known virulence factors of EHEC are summarized below. Many of these are encoded within bacteriophages or on other mobile genetic elements that have been obtained by the cell through horizontal gene transfer.

Shiga toxin

The Shiga toxins produced by STEC are encoded within the genome of lambdoid bacteriophages (38). These toxins are absorbed through the intestine and can inhibit protein synthesis once internalized by host cells (31). They may also damage endothelial cells in the kidney and trigger blood clot formation (58). These clots can block small blood vessels, causing HUS and decreased urine output, leading to kidney failure and death (58). EHEC strains produce one or both of two allelic variants of Shiga toxin, designated Stx1 and Stx2, with Stx2 production more frequently linked to severe disease (7). *E. coli* O157:H7 strains isolated from HUS patients usually produce both Stx1 and Stx2, or Stx2 only (42).

The LEE Pathogenicity Island

Pathogenicity islands are horizontally-acquired genetic elements found in pathogenic bacteria that encode virulence functions (20). Many of these are known as O-islands, genome segments found in *E. coli* O157:H7 but not *E. coli* K-12, an avirulent laboratory strain (45). EHEC carry the locus of enterocyte effacement (LEE), a chromosomal pathogenicity island that encodes proteins associated with attaching and effacing (A/E) lesion formation in the large intestine (27). Some of these proteins include components of a type III secretion system, the adhesion molecule intimin (encoded by the *eae* gene), Tir, the translocated intimin receptor, and other effector proteins that are secreted into host cells where they disrupt various cellular processes (25, 64).

Type III secretion systems (T3SS) directly inject proteins from the cytoplasm of a bacterial pathogen into a eukaryotic host cell, translocating virulence factors in a single step (25). EspA filaments from the bacterial cell extend and adhere to a host cell, forming a translocation pore through which effector proteins can be delivered (22). Tir is injected into the host cell and inserted into the host cell membrane where it binds intimin expressed on the surface of the bacterial cell (19). The T3SS-mediated A/E lesions cause the destruction of the host intestinal mucosal epithelium and microvilli (25).

EHEC also carry genes for a variety of non-LEE encoded (*nle*) effector genes that encode proteins secreted by the T3SS but that are not part of the LEE pathogenicity island (16). Studies have shown that several of these genes, including *nleA, nleB, nleC, nleE, nleF, nleG2, nleG5, nleG6, nleH1-2* and *ent/espL2,* are potential virulence factors (11, 12, 16, 33). The presence of O-Island 122 has shown the strongest correlation with highly virulent STEC (13, 33) and O-Islands 57 and 71 are prevalent in strains causing HUS and outbreaks (16).

Plasmid-Encoded

Other virulence factors of EHEC may be encoded on pO157, a 60 MDa plasmid found in several EHEC serotypes (36). While its specific role in virulence is unknown, the potential virulence genes encoded on this plasmid include *toxB*, EHEC-hemolysin, and *katP* (14). The *toxB* gene is important in adherence because it affects the production of the adhesion proteins EspA, EspB, and Tir (57). EHEC-hemolysin enhances intestinal cell

permeability by forming small pores in the plasma membrane (56). KatP, a catalase peroxidase, protects the bacterium against oxidative stress (61).

While many of the factors summarized above have been suggested to play roles in the pathogenicity of EHEC, the presence of one or more of these genes does not necessarily imply high virulence. For example, identification of Shiga toxin alone does not predict that an *E. coli* strain will cause human disease. Instead, it is likely that the prevalence of several different genes determines virulence. **It is important to identify these genes and develop a gene-typing system by which potentially high virulent STEC strains can be detected.**

Previous Research

The increased concern about STEC has made the development of a rapid, specific, and low-cost method for their identification a priority. In recent years, a number of different studies have focused on identifying virulence factors in STEC. In an analysis by Bugarel *et al.*, a micro-array-based method was used to screen a collection of STEC isolates for O-antigens, H-types, and virulence factors, simultaneously determining serotypes and potential virulence of the strains (12). In addition to genes encoded on virulence plasmids, the virulence factors screened for were the *nle* genes encoded on O-island 122 and O-island 71. The enterohemolysin (*ehxA*) gene was more prevalent in EHEC serotypes O104, O113, O91, O103, O118, O123, O15, O165, O172, O177, O45, O5, O55, O111, O121, O157, O145, and O26 (85%) than STEC (43%) and the intimin (*eae*) gene was detected in all typical EHEC strains. The

ent/espL2, nleB and *nleE* genes encoded on OI-122 were detected in all EHEC while *nleF*, *nleH1-2*, and *nleA* genes encoded on OI-71 were less common in EHEC strains. This suggests that OI-122 genes are associated with highly virulent strains. The catalase peroxidase (*katP*) and serine protease (*espP*) genes were variably present but more commonly found in EHEC than in STEC strains. The gene *etpD* of the T3SS was detected in 29% of the EHEC strains in this study but was not found in other STEC strains, making it the only unique marker for EHEC. The results of this study suggested that no single gene or virulence factor distinguishes EHEC from less virulent STEC strains.

In another study by Bugarel *et al.*, a low-density macroarray was used to screen for the *stx1*, *stx2*, *eae*, and *ehxA* genes, along with 6 *nle* genes encoded on genomic islands OI-122 (*ent, nleB*, and *nleE*) and OI-71 (*nleF, nleH1-2,* and *nleA*), in a collection of *E. coli* isolates (11). The *ehxA* gene was found more prevalently in EHEC (90%) than in STEC (42.66%) and *eae* was present in all typical EHEC strains. The *nle* genes detected in some EHEC strains had various distributions and seemed to be associated with certain serotypes and intimin genotypes. The *nle* genes encoded on OI-122 were detected in all typical EHEC while those encoded on OI-71 were not. Of the OI-71 *nle* genes, *nleH1-2* was found in 99.5 % of EHEC strains while *nleF* and *nleA* were found in 72.8% and 79%, respectively. From their results, the authors identified the genes *eae, ent/espL2, nleB, nleE,* and *nleH1-2* as possible virulence determinants that associate with *E. coli* O157:H7 and other EHEC strains of high virulence (11).

In a molecular analysis study by Coombes *et al.*, researchers analyzed the distribution of 16 *nle* genes, encoded in O-islands 36, 57, and 71 as part of the T3SS, among O157 and non-O157 STEC isolates (16). A complete O-Island 57 was found in 41% of HUS-associated and 46%

of outbreak-associated strains while a complete O-Island 71 was found in 45% of HUS-associated and 67% outbreak-associated strains. The increased prevalence of these O-Islands in strains that cause HUS and outbreaks suggests their contribution to virulence. Fourteen *nle* genes (*nleA, nleB, nleC, nleE, nleF, nleG, nleG2-1, nleG2-3, nleG6-2, nleG9, nleH1, nleH2,* and *ent/espL2*) were more prevalent in isolates associated with HUS and outbreaks than in serogroups considered to be less harmful. The *nleA, nleB,* and *ent/espL2* genes were found in 100% of outbreak-associated and 69% of HUS-associated STEC strains but only in 33% of non-outbreak associated and 32% of non-HUS associated STEC strains. The results of this study also showed that non-O157 STEC strains associated with severe human disease and linked to outbreaks contain significantly more *nle* genes and more complete O-islands (OI- 57, OI-71, and OI-122 but not OI-36) than STEC strains not associated with HUS or outbreaks. These results suggest that the presence of *nle* genes may predict, and contribute additively to, the virulence of STEC strains.

PCR binary typing (P-BIT) may also be useful in the risk assessment of STEC isolates due to its speed and low cost (17). This method involves screening for the presence or absence of virulence genes (binary typing) in *E. coli* to produce a genetic fingerprint for each isolate (9). A recent analysis by Brandt *et al.* used a 24-gene target P-BIT system to produce "virulence bar codes" for O157 and non-O157 STEC isolates that could be used in risk assessment and epidemiological studies (9). The genes detected in that study have been associated with outbreaks and severe disease and include the LEE, OI-36, OI-57, OI-71, OI-122, pO157, the long polar fimbriae operon, the urease gene cluster, *stx1,* and *stx2.* Bar codes with higher numbers indicated the detection of more virulence genes in the strain and, therefore, a greater potential

to cause disease in humans. The P-BIT system results were used to separate isolates into clusters based on their association with outbreaks and severe disease. The clusters were compared to seropathotype (SPT) classification of each isolate, a system of identifying the pathogenicity of STEC serotypes linked to outbreaks or disease (33). In the P-BIT cluster dendogram that was developed, two clusters, designated 7 and 8, contained non-O157 serotypes associated with severe disease and outbreaks. The strains in these clusters contained the genes *pic, espC, iha, eaeA, pagC, sen, nleG2-3, nleG, agn43*$_{EDL933}$*, paa, ureC, toxB, espP,* *lpfA*$_{O26}$, and *lpfA*$_{O113}$. Because their P-BIT clustering corresponds well with SPT classification, the authors suggest the use of P-BIT bar codes and clustering in risk assessment of STEC isolates.

The results of the studies summarized above suggest that the prevalence of a combination of genes may predict virulence of STEC. Differentiating between highly virulent STEC and those not known to cause disease or outbreaks requires the identification of genetic elements, coding for virulence factors, which are found more commonly in pathogenic strains.

Objective

Full genome sequencing has been used to identify genes conserved in O157 EHEC and other EHEC serogroups that may contribute to virulence. A recent report describes a genomic comparison of fully sequenced strains from serogroups O157, O26, O111, and O103 (41). This study reported that 87 genes were conserved amongst these EHEC strains and absent in other *E. coli*. These results suggest that a set of these genes may associate with EHEC and other high risk STEC strains. Determining whether any of these genes could be used to identify highly virulent strains of EHEC would be useful for the development of novel diagnostics.

The present study aims to determine whether these 87 genes are present within a larger collection of non-EHEC strains, and to determine their prevalence in STEC by screening a library of clinical and non-clinical isolates. **This research will help in identifying genes that are more prevalent in high risk STEC than in strains that have not been associated with disease or outbreaks. The results of this study will contribute to the development of methods by which potentially highly virulent STEC can be screened for in environmental, food and clinical samples.**

MATERIALS AND METHODS

Genes

Eighty-seven genes (Table S1) found to be conserved in fully sequenced EHEC serotypes O157, O26, O111, and O103 (41) were researched using GenBank (3) to determine putative gene function and length. Using BLAST, each gene sequence was analyzed for its presence in O157, O55, O111, O103, and O26 strains, along with the mouse pathogen *Citrobacter rodentium*. When using this database to search a gene, results suggesting its putative function and presence in an organism were considered a "hit." Genes found to be present in O157, O111, O103, and O26 serogroups but absent in other *E. coli* were considered for use in this study.

The Shiga toxin genes used in the genetic screening were not included in this computational analysis because they are not consistently found in all members of the O157, O111, O103, and O26 serogroups. While the *nleB* gene was chosen for screening because of previous reports suggesting its usefulness as an EHEC marker, the other 11 genes used in this study were chosen at random from the list of 87. Because some STEC strains have multiple copies of a gene, this selection was based on the ease with which primers for a given gene target could be designed

Bacterial Strains

A total of 62 STEC strains were used in this project (Table 1; obtained from Dr. Chobi DebRoy, Director of the *E. coli* Reference Center, Penn State University). In addition, *E. coli* strain Sakai (28) was obtained from Dr. Wei Zhang at the Illinois Institute of Technology and the

K-12 strain (6) was obtained from laboratory stocks (Dr. Edward Dudley, Food Science Department, Penn State University). Stocks for each strain were made in a 10% glycerol solution and were stored at -80°C. All strains were streaked onto Lysogeny Broth (LB) medium (53) and plates were incubated at 37°C overnight. Single isolated colonies were then grown in LB overnight to form a pure culture for subsequent gene typing.

Table 1: Collection of Clinical and Non-Clinical STEC Strains Used for Genetic Screening

Accession Number	O-Type	H-Type	Sources
7.3964	26	11	Unknown
81.0211	26	11	Antelope
0.1302	26	N/A	Cow
82.0219	26	11	Cow
95.1144	26	N/A	Cow
6.1592	26	11	Test strain
8.0176	26	30	Unknown
5.2217	26	11	Human
93.0494	26	2	Human
99.0849	26	N/A	Human
99.0850	26	N/A	Human
99.0869	26	N/A	Human
99.1761	26	7	Human
99.1773	26	N/A	Human
10.2529	103		Cow
90.0107	103	12	Cow
90.1764	103	2	Cow
0.1623	103	N/A	Deer
10.0941	103	2	Food
87.1368	103	2	Goat
3.2605	103	2	Horse
9.0108	103	2	Test strain
93.0626	103	6	Human
93.1685	103	2	Human
99.1791	103	N/A	Human
99.1792	103	N/A	Human
99.1806	103	N/A	Human
99.1807	103	N/A	Human
99.1822	103	N/A	Human
0.2981	111	N/A	Unknown
4.0005	111	N/A	Cow

Accession Number	O-Type	H-Type	Sources
8.0288	111	8	Cow
87.1377	111	11	Cow
10.0815	111	8	Food
7.1639	111	8	Test strain
11.0247	111		Human
11.0284	111		Human
96.1529	111	N/A	Human
96.1530	111	N/A	Human
99.1769	111	N/A	Human
93.0525	111	N/A	Human
93.1707	111	N/A	Human
6.1192	157	12	Pig
6.1194	157	12	Pig
6.1195	157	12	Pig
9.0538	145		Food, Ground Beef
0.2732	121		Pig
5.0959	121	19	Other
7.1636	121	19	Test strain
1.2805	2	42	Cow
2.0665	5		Food, Dairy Products
96.1898	50		Food
96.0876	75		Gazelle, Curvier's
1.2265	76		Goat, Siberian Ibex
97.1571	146		Okapi
10.1083	4	7	Food, Spinach
10.0939	8	16	Food
96.1534	5	11	Human
96.0611	84		Human
95.3644	88		Human
93.0602	85		Human
95.4080	6	16	Human

N/A, Unknown H-type

DNA Isolation and Quantification

The Wizard® Genomic DNA Purification Kit (Promega Corporation, Madison, Wisconsin) was used to isolate DNA from the *E. coli* strains. The solutions used in this protocol were obtained from this kit. From each cell culture, 1.5 mL was centrifuged at 16,873 ×*g* for 2.5 minutes and the supernatant was poured off. This step was repeated so that a total of 3 mL of cell culture was used for each strain. The cells were re-suspended in Nuclei Lysis solution and incubated at 80°C for 5 minutes. After cooling the samples to room temperature, 3 µL of RNase solution was added to the cell lysates and the samples were incubated at 37°C for 15 minutes. The samples were vortexed after adding 200 µL of Protein Precipitation Solution to the RNase-treated lycate; samples were incubated on ice for 5 minutes. The samples were then centrifuged at 16,873 ×*g* for 3 minutes; the supernatants were transferred to new tubes containing 600 µL of isopropanol. The tubes were mixed by inversion until threadlike strands of DNA had formed, then centrifuged at 16,873 ×*g* for 2 minutes. The supernatants were poured off and 600 µL of 70% ethanol was added to wash each sample. After centrifuging at 16,873 ×*g* for 2 minutes, the tubes were drained and the samples were subjected to vacuum centrifugation for 30 minutes. Finally, 200 µL of distilled water was added to each tube and samples were incubated at 65 °C for 1 hour.

The concentration of double-stranded DNA (µg/mL) present in each sample was determined using the Qubit® Fluorometer dsDNA High Sensitivity Assay (Invitrogen Corporation, Life Technologies Corporation, Grand Island, New York). Samples were diluted to 4 ng/µL and stored at -20°C; these were used as template DNA in PCR reactions.

Primer Design

Primer sequences were determined using GenBank (3; Table 2); primers for each gene were designed using Primer3 (51). Primer sequences had been previously reported for both Shiga toxin 1 (23) and Shiga toxin 2 (44) genes.

Table 2: Primers and PCR Conditions for Gene Targets

Primer Name	Product Size (bps)	Type	Sequence[a]	Annealing Temperature (°C)	Extension Time (sec.)
Shiga toxin 1	600	Forward	ACACTGGATGATCTCAGTGG	58	40
		Reverse	CTGAATCCCCCTCCATTATG		
Shiga toxin 2	255	Forward	GGCACTGTCTGAAACTGCTCC	58	40
		Reverse	TCGCCAGTTATCTGACATTCT		
ECs848- Hypothetical Protein	398	Forward	AGAAACCTGACTTCGCCTGA	60	40
		Reverse	CTTTTGTTGCATCCTCAGCA		
ECs1322- *ureA* urease subunit γ	199	Forward	CGCGGCCTGAAACTTAACTA	60	40
		Reverse	CGGATTATGGACGGTAACCA		
ECs1323- *ureB* urease subunit β	242	Forward	CCGGGCTACCTGCAGTATTA	66	30
		Reverse	ATCATTTGCCTCCAGCTCAC		
ECs1326- *ureF*	334	Forward	GTAACAGCCTCCCTGTTGGA	60	40
		Reverse	TACCGGTTAACTGGCTTTGC		
ECs1561- Hypothetical Protein	284	Forward	TTTCTTTCGTGACGCTGATG	60	40
		Reverse	AGCTTGAATACCGGTTGTGC		
ECs1568- Hypothetical Protein	334	Forward	CTCCCTCGCGAAAGTGTAAG	60	40
		Reverse	GTAGCGCCACAGACAGCATA		
ECs2226- Hypothetical Protein	302	Forward	TGCTGTTTCTGTTGGTCTGC	56	30
		Reverse	CCGAAAACGCCCTTAAAAAT		
ECs3857- *nleB*	782	Forward	GCCAGAGCGATACGAAAAAG	60	40
		Reverse	TAAAATGCCGCTTGATACCC		
ECs3858- Hypothetical Protein	582	Forward	AGGGCGTGTCCCCTATAAAT	48	40
		Reverse	TATTTCCCCAGGCATGTAGC		
ECs4552- *escF*	179	Forward	CGGTTAGAAATGGTTGAGACC	65	40
		Reverse	GAAATTACTCAACAAATGGGTGAA		
ECs4553- Hypothetical Protein	188	Forward	ATTTCATCGCGCTGATTTTC	48	20
		Reverse	CTTTGTGGCAGTGCCTTTTT		
ECs4557- *sepL*	687	Forward	CAATCGATACCCGAGAAGGA	64	40
		Reverse	CAAAGGTAGCGCAAGGAAAG		

[a] Primer sequences were determined using GenBank (3); primers for each gene were designed using Primer3 (51)

PCR

For each PCR reaction, Sakai (28) and K12 (6) template DNA were used as positive and negative controls, respectively. The concentration of template DNA used was 4 ng/µL and the concentration of each primer was 10 µM. While the annealing time was constant (30 sec.) for each PCR reaction, the annealing temperature varied for each set of primers (Table 2). Also, the extension temperature was constant (72°C) for each reaction but the extension time varied (1 minute per 1 kb of DNA; Table 2). Two different PCR reaction mixtures were used throughout this study (Tables 3, 4). Genes ECs0848, ECs1322, and ECs1326 were screened for using reaction mixture #1 (Table 3) while the remaining genes were screened for using reaction mixture #2 (Table 4). The thermocycler (iCycler and C1000 Touch™ Thermal Cyclers, Bio-Rad Laboratories, Hercules, CA 94547) conditions used for the PCR reactions are presented in Table 5.

Table 3: PCR Reaction Mixture #1

Reagent	Volume (µL)
Buffer	2.5
dNTP	0.5
Forward Primer	1
Reverse Primer	1
Template	3
Taq	0.4
Water	16.6
Total Volume	25

Table 4: PCR Reaction Mixture #2

Reagent	Volume (µL)
Buffer	2
dNTP	0.2
Forward Primer	1
Reverse Primer	1
Template	0.5
Taq	0.1
Water	15.2
Total Volume	20

Table 5: Thermocycler Conditions Used for PCR Reactions

Step	Temperature (°C)	Time	Repeats
Initialization	94	2 min.	1x
Denaturation	95	30 sec.	30x
Annealing	48-66[a]	30 sec.	30x
Extension	72	40 sec.[b]	30x
Final Extension	72	5 min.	1x
Hold	4	∞	1x

[a] Varied according to specific primer conditions
[b] Average time was 40 sec. but time varied from 20 sec. to 40 sec. based on primer conditions

Following the PCR reactions, 6 x gel-loading buffer (53) was added to each product. The products were loaded onto a 1.2% agarose gel which was run at 160 volts for 30-50 minutes in TAE buffer (53). Ethidium bromide was used to stain the DNA fragments and the gels were visualized using an EC3 Bioimaging System (UVP, Upland, California).

Statistical Analysis

The results of each PCR reaction were reported as positive, faint positive, or negative for each gene. The Fisher exact test (Minitab 16 statistical software) was used to determine the difference in prevalence between genes found in EHEC serogroups O26, O103, O111, O157, O145, and O121 and those of other STEC serogroups. Differences in gene distributions with *p* values of less than 0.05 were considered significant.

RESULTS

Computational Analysis

The 87 genes previously reported to be conserved in 4 fully sequenced EHEC strains of serogroups O157, O26, O111, and O103 (41) were obtained using GenBank (3). In addition to determining the length and putative function of the genes, the presence of each in the O157, O55, O111, O103, O26, and *Citrobacter rodentium* genomes was determined using BLAST (Table S1). Genes found to be present in O157, O111, O103, and O26 serogroups but absent in other *E. coli* were considered for use in this study. The results for 12 of the genes used in the genetic screening are presented in Table 6.

Genetic Screening

The two Shiga toxin genes were used in the genetic screening to ensure that all isolates were STEC. The *nleB* gene was chosen because of previous reports suggesting its increased prevalence in highly virulent STEC. The other 11 gene targets used in this study were chosen at random from the list of 87 based on the ease with which primers could be designed.

A total of 62 clinical and non-clinical STEC strains were screened for the 14 gene targets using PCR. The screening results for 57 strains are shown in Table S2 and incomplete results for the remaining 5 strains are reported in Table S3. Of the 57 strains with complete data sets, 23 were isolated from clinical sources while 34 were from other animals, environmental, or other sources. There were 44 strains of EHEC serogroups (14 O26, 13 O103, 12 O111, 2 O157, 1 O145, and 2 O121) and 13 strains of serogroups not known to be associated with severe disease.

The percentage of EHEC and other STEC isolates that were positive for each gene is reported in Table 7. The *stx1* gene was present in 93% of STEC isolates (91.7% of clinical and 94.1% of non-clinical isolates) and was found in 95.5% of the isolates from EHEC serogroups and in 84.6% of the other isolates. The *stx2* gene was detected in 38.6% of all isolates (29.2% clinical; 44.1% non-clinical). EHEC serogroups were 31.8% positive for the *stx2* gene while 61.5% of the other isolates tested positive. Every gene except ECs2226, a hypothetical protein, was detected in all O157 STEC isolates. Also, ECs3857 (*nleB*) was detected in all of the EHEC seroroups but only in 53.9% of isolates belonging to serogroups not associated with severe human disease. All 44 isolates belonging to the EHEC serogroups screened positive for the *nleB* gene.

Six of the gene targets in this study encode hypothetical proteins (Table 6). These genes were found in 35-42 of the 44 total O26, O103, O111, O157, O145, and O121 EHEC isolates and in 5-6 of the 13 other STEC isolates (Table 8). Three of the gene targets encode urease subunits (*urea, ureB, ureF*) and another encodes EscF. These genes were found in 35-40 of the 44 total EHEC isolates and in 5-6 of the other STEC isolates.

Statistical Analysis

The Fisher exact test was used to determine the significance of gene distribution differences observed between the EHEC serogroups and other STEC (Table 8). There was no significant difference between the prevalence of *stx1* ($p = 0.22$) and *stx2* ($p = 0.102$) in these two groups. Also, the distribution of ECs4557 (*sepL*) in EHEC serogroups did not differ from that in the other serogroups ($p = 0.319$). The prevalence of the 11 other genes (ECs0848, ECs1322,

ECs1323, ECs1326, ECs1561, ECs1568, ECs2226, ECs3857, ECs3858, ECs4552, and ECs4553) differed significantly between the EHEC and other STEC serogroups ($p \leq 0.033$; Table 8).

Table 6: Computational Analysis of the 12 non-Stx Gene Targets Used in Screening STEC Isolates

Gene	Putative Function	Size	O157	O55	O111	O103	O26	*C. rodentium*
ECs0848	Hypothetical Protein	882 bps	+	+	+	+	+	+
ECs1322	*ureA* (urease subunit gamma)	303 bps	+	-	+	+	+	-
ECs1323	*ureB* (urease subunit beta)	321 bps	+	-	+	+	+	-
ECs1326	*ureF*	675 bps	+	-	+	+	+	-
ECs1561	Hypothetical Protein	3.4 kb	+	-	+	+	+	+
ECs1568	Hypothetical Protein	1.36 kb	+	+	+	+	+	+
ECs2226	Hypothetical Protein	372 bps	+	+	+	+	+	-
ECs3857	*nleB*	990 bps	+	+	+	+	+	+
ECs3858	Hypothetical Protein	675 bps	+	+	+	+	+	+
ECs4552	*escF*	222 bps	+	+	+	+	+	-
ECs4553	Hypothetical Protein	408 bps	+	+	+	+	+	+
ECs4557	*sepL*	1.06 kb	+	+	+	+	+	+

+, Present
-, Absent

Table 7: Percentage of EHEC and Non-EHEC Strains with Positive Gene Screenings Using PCR

Gene	% of EHEC Strains[a] Screening (+)	% of Other STEC Strains[b] Screening (+)
stx1	95.5	84.6
stx2	31.8	61.5
ECs848	95.5	46.2
ECs1322	88.6	38.5
ECs1326	88.6	38.5
ECs1323	79.5	38.5
ECs4557	13.6	0.0
ECs4552	90.9	46.2
ECs3858	93.2	38.5
ECs4553	95.5	46.2
ECs3857	100.0	53.8
ECs1568	79.5	23.1
ECs1561	90.9	38.5
ECs2226	93.2	46.2

[a] EHEC refers to strains of serogroups O26, O103, O111, O157, O145, and O121

[b] Other STEC strains include those of serogroups other than O26, O103, O111, O157, O145, and O121

Table 8: Prevalence of genes in STEC Collection and Variation between EHEC and non-EHEC

Gene	No. of STEC Strains[a] (n=44) Screening (+)	No. of Other STEC Strains[b] (n=13) Screening (+)	*p* [c]
stx1	42	11	0.22
stx2	14	8	0.102
ECs848	42	6	<0.001
ECs1322	39	5	0.001
ECs1323	35	5	0.012
ECs1326	39	5	0.001
ECs1561	40	5	<0.001
ECs1568	35	6	0.033
ECs2226	41	6	0.001
ECs3857	44	7	<0.001
ECs3858	41	5	<0.001
ECs4552	40	6	0.001
ECs4553	42	6	<0.001
ECs4557	6	0	0.319

[a] EHEC strains refers to isolates of serogroups O26, O103, O111, O157, O145, and O121

[b] Other STEC strains refers to isolates of serogroups other than O26, O103, O111, O157, O145, and O121

[c] The Fisher exact test was used to determine the statistical difference between the distribution of genes in EHEC and non-EHEC serotypes used in this study; a significant difference is seen with a *p* value less than 0.05

DISCUSSION

The increased recognition of non-O157 Shiga toxin-producing *E. coli* (STEC) as a cause of disease and outbreaks has made detecting them in food and in clinical samples of great importance. Although many different STEC strains have been identified, few are known to cause human disease. Enterohemorrhagic *Escherichia coli* (EHEC) are a subset of STEC that are capable of causing the most severe disease in humans, such as hemorrhagic colitis and the hemorrhagic uremic syndrome. At present, there is no method of distinguishing between EHEC and STEC that have not been linked to disease.

Full genome sequencing has been used to identify genes conserved in O157 EHEC and other EHEC serogroups that may contribute to virulence. A recent study reported 87 genes that are conserved among fully sequenced strains of O157, O26, O111, and O103 serogroups, but absent in other *E. coli* (41). The purpose of the present study was to determine if these 87 genes are present in a larger collection of EHEC strains and to determine their prevalence in STEC. Determining if any of these genes could be used in identifying EHEC would help in developing novel diagnostics.

Using GenBank (3), a computational analysis was done on the 87 genes (Table S1). Each gene was noted for its presence in *E. coli* serogroups O157, O55, O111, O103, and O26, along with its presence in other organisms and in *C. rodentium*, a mouse pathogen. The size and putative function of each gene was also determined. From this list, 12 genes, in addition to *stx1* and *stx2* genes, were chosen as targets for screening (Table 6). Although all 87 genes were previously reported as being conserved in O157, O26, O111, and O103 serogroups, only 3 (*stx1, stx2,* and ECs3857) of the 14 genes used in this study have been screened for in other reports of

gene prevalence in STEC. The remaining 11 genes (ECs0848, ECs1322, ECs1323, ECs1326, ECs1561, ECs1568, ECs2226, ECs3858, ECs4552, ECs4553, and ECs4557) were chosen because of the lack of knowledge surrounding their possible roles in the virulence of STEC. Also, while 8 of the 14 genes have established putative functions, the other 6 (ECs0848, ECs1561, ECs1568, ECs2226, ECs3858, ECs4552, and ECs4553) are thought to code for hypothetical proteins. These 6 genes were chosen at random from the list of 87 to identify a possible correlation between their prevalence in STEC and virulence.

A collection of 62 clinical and non-clinical STEC isolates were used in this study (Table 1); complete screening data was only obtained for 57 of these isolates (Table S2). Of these 57 strains, 44 were of EHEC serogroups (O26, O103, O111, O157, O145, and O121) and 13 belonged to serogroups other than these. The strains in this collection were chosen because of their various serogroups and sources. Twenty-three of the 57 strains were isolated from clinical sources while the other 34 were isolated from non-human animals, the environment, or other sources.

When using PCR to amplify DNA in a large collection of isolates, single nucleotide polymorphisms in primer binding sites may exist, resulting in false negatives. Because of this uncertainty, slot-blot analysis was initially considered as the method for screening our strain collection. Slot-blot was thought to offer an advantage over PCR because the probe used is typically much longer than a PCR primer, and, therefore, this technique is less affected by naturally-occurring allelic variations. Early in this study, all of the STEC isolates were screened for *stx1* and *stx2* genes using slot-blot, however, we found several issues with this method that gave us reason to use PCR to perform the genetic screening. The process was labor intensive

because conditions had to be specifically altered to accommodate the probe for each gene. Also, because of non-specific binding, true positive bands on the membrane were difficult to differentiate from false positives. PCR was therefore chosen as the method used to screen our strain collection for the presence of the 14 gene targets. The advantages it offers include its ease of use, low-cost, and wide availability.

The *stx* genes, of allelic variants designated Stx1 and Stx2, were used in the screening to confirm that all isolates were STEC. The *stx1* gene was found to be more prevalent in the 57 isolates than the *stx2* gene (Table 7); there was no variation between the prevalence of these genes in EHEC versus other STEC (Table 8). In addition to these gene targets, the distribution of ECs4557 (*sepL*) did not differ between EHEC and other STEC serogroups. SepL is a T3SS protein that may bind to Tir or LEE-encoded SepD and control the timing of effector protein secretion during translocon assembly (43, 62). The low prevalence of *sepL* in the highly virulent isolates of our STEC collection suggests its insignificance in T3SS function and, therefore, virulence.

The distribution of 11 of the gene targets (ECs0848, ECs1322, ECs1323, ECs1326, ECs1561, ECs1568, ECs2226, ECs3857, ECs3858, ECs4552, and ECs4553) differed significantly between EHEC strains of serogroups O26, O103, O111, O157, O145, and O121 and STEC belonging to other serogroups ($p \leq 0.033$). All 6 genes encoding hypothetical proteins were found to be more prevalent in the EHEC serogroups, suggesting their possible role in the virulence of these strains. Genes encoding UreA, UreB, UreF, and EscF were also more prevalent in EHEC. The urease proteins function as enzymes that cleave urea (37). The ammonia produced in this process neutralizes the highly acidic pH of the gut allowing survival of the bacterium (37). The prevalence of these ureases in isolates of EHEC serogroups suggests

their role in the virulence of STEC. More research is needed to determine the exact function of EscF and its possible contribution to virulence.

ECs3857 (*nleB*) was found in 100% of isolates belonging to EHEC serogroups and in only 53.8% of isolates belonging to other isolates (Table 7). There was a significant difference in its distribution between EHEC serogroups and the other isolates in our collection ($p < 0.001$; Table 8). The prevalence of this gene in highly virulent strains suggests its usefulness in identifying EHEC and other STEC associated with human illness.

CONCLUSIONS AND FUTURE DIRECTIONS

Of the genes screened for in this study, *nleB* is the most reliable marker of EHEC and can be used to identify highly virulent STEC. Also, the prevalence of urease genes *ureA, ureB,* and *ureF* in isolates of EHEC serogroups suggest their contribution to virulence. More research needs to be conducted to determine the specific functions of the EscF protein and the hypothetical proteins used in this study, and to assess their possible role in virulence.

Possible areas for future research include completing the genetic screening for the data set in Table S3 and determining whether or not the 12 non-Stx genes used in this study actually play a role in STEC virulence. Also, similar studies could be done to investigate the prevalence of the remaining genes from the list of 87 in a larger collection of STEC.

REFERENCES

1. **Almanza, A. V.** 2012. Shiga Toxin-Producing *Escherichia coli* in Certain Raw Beef Products. Federal Register **77**:1-2.

2. **Arthur, T. M., G. A. Barkocy-Gallagher, M. Rivera-Betancourt, and M. Koohmaraie**. 2002. Prevalence and Characterization of *Escherichia coli* on Carcasses in Commercial Beef Cattle Processing Plants. Applied and Environmental Microbiology **68**:4847-4852.

3. **Benson, D. a, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers**. 2009. GenBank. Nucleic Acids Research **37**:D26-31.

4. **Besser, R. E., P. M. Griffin, and L. Slutsker**. 1999. *Escherichia coli* O157:H7 Gastroenteritis and the Hemolytic Uremic Syndrome: An Emerging Infectious Disease. Annual Review of Medicine **50**:355-67.

5. **Bettelheim, K. A.** 2007. The non-O157 Shiga-toxigenic (Verocytotoxigenic) *Escherichia coli*; Under-Rated Pathogens. Critical Reviews in Microbiology **33**:67-87.

6. **Blattner, F. R., G. Plunkett Iii, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, and G. F. Mayhew**. 1997. The Complete Genome Sequence of *Escherichia coli* K-12. Science **277**:1453.

7. **Boerlin, P., S. a McEwen, F. Boerlin-Petzold, J. B. Wilson, R. P. Johnson, and C. L. Gyles**. 1999. Associations between Virulence Factors of Shiga Toxin-Producing *Escherichia coli* and Disease in Humans. Journal of Clinical Microbiology **37**:497-503.

8. **Bonardi, S., E. Maggi, a Bottarelli, M. L. Pacciarini, a Ansuini, G. Vellini, S. Morabito, and a Caprioli**. 1999. Isolation of Verocytotoxin-producing *Escherichia coli* O157:H7 from cattle at slaughter in Italy. Veterinary Microbiology **67**:203-11.

9. **Brandt, S. M., N. King, A. J. Cornelius, A. Premaratne, T. E. Besser, and S. L. W. On**. 2011. Molecular Risk Assessment and Epidemiological Typing of Shiga Toxin-Producing *Escherichia coli* by Using a Novel PCR Binary Typing System. Applied and Environmental Microbiology **77**:2458-70.

10. **Brooks, J. T., E. G. Sowers, J. G. Wells, K. D. Greene, P. M. Griffin, R. M. Hoekstra, and N. A. Strockbine**. 2005. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983-2002. The Journal of Infectious Diseases **192**:1422-1429.

11. **Bugarel, M., L. Beutin, and P. Fach**. 2010. Low-Density Macroarray Targeting Non-Locus of Enterocyte Effacement Effectors (nle genes) and Major Virulence Factors of Shiga Toxin-Producing *Escherichia coli* (STEC): A New Approach for Molecular Risk assessment of STEC isolates. Applied and Environmental Microbiology **76**:203-11.

12. **Bugarel, M., L. Beutin, A. Martin, A. Gill, and P. Fach**. 2010. Micro-array for the Identification of Shiga Toxin-Producing *Escherichia coli* (STEC) Seropathotypes Associated with Hemorrhagic Colitis and Hemolytic Uremic Syndrome in Humans. International Journal of Food Microbiology. Elsevier B.V. **142**:318-29.

13. **Bugarel, M., A. Martin, P. Fach, and L. Beutin**. 2011. Virulence gene profiling of enterohemorrhagic (EHEC) and enteropathogenic (EPEC) *Escherichia coli* strains: a basis for molecular risk assessment of typical and atypical EPEC strains. BMC microbiology. BioMed Central Ltd **11**:142.

14. **Caprioli, A., S. Morabito, H. Brugère, and E. Oswald**. 2005. Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. Veterinary Research **36**:289-311.

15. **Chapman, P. A., C. A. Siddons, A. T. Gerdan Malo, and M. A. Harkin**. 1997. A 1-year study of *Escherichia coli* O157 in cattle, sheep, pigs and poultry. Epidemiology and Infection **119**:245-50.

16. **Coombes, B. K., M. E. Wickham, M. Mascarenhas, S. Gruenheid, B. B. Finlay, and M. A. Karmali**. 2008. Molecular Analysis as an Aid to Assess the Public Health Risk of Non-O157 Shiga Toxin-Producing *Escherichia coli*. Applied and Environmental Microbiology.

17. **Cornelius, A. J., B. Gilpin, P. Carter, C. Nicol, and S. L. W. On**. 2010. Comparison of PCR Binary Typing (P-BIT), a New Approach to Epidemiological Subtyping of *Campylobacter jejuni*, with Serotyping, Pulsed-Field Gel Electrophoresis, and Multilocus Sequence Typing Methods. Applied and Environmental Microbiology **76**:1533-44.

18. **Cray, W. C., L. a Thomas, R. a Schneider, and H. W. Moon**. 1996. Virulence attributes of *Escherichia coli* isolated from dairy heifer feces. Veterinary Microbiology **53**:369-74.

19. **Delahay, R. M., G. Frankel, and S. Knutton**. 2001. Intimate interactions of enteropathogenic *Escherichia coli* at the host cell surface. Current Opinion in Infectious Diseases **14**:559.

20. **Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker**. 2004. Genomic islands in pathogenic and environmental microorganisms. Nature Reviews Microbiology **2**:414-424.

21. **Doyle, M. P., L. R. Beuchat, and T. J. Montville**. 2007. Food Microbiology: Fundamentals and Frontiers. ASM Press Washington DC.

22. **Frankel, G., A. D. Phillips, I. Rosenshine, G. Dougan, J. B. Kaper, and S. Knutton**. 1998. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. Molecular Microbiology **30**:911-921.

23. **Gannon, V. P., R. K. King, J. Y. Kim, and E. J. Thomas**. 1992. Rapid and Sensitive Method for Detection of Shiga-Like Toxin-Producing *Escherichia coli* in Ground Beef Using the Polymerase Chain Reaction. Applied and Environmental Microbiology **58**:3809-15.

24. **Garg, A. X., R. S. Suri, M. P. Rosas-arellano, R. B. Haynes, W. F. Clark, N. Barrowman, F. Rehman, D. Matsell, and M. Salvadori**. 2003. Long-term Renal Prognosis of Diarrhea-Associated Hemolytic Uremic Syndrome. Journal of the American Medical Association **290**.

25. **Garmendia, J., G. Frankel, and V. F. Crepin**. 2005. Minireview: Enteropathogenic and Enterohemorrhagic *Escherichia coli* Infections: Translocation, Translocation, Translocation. Infection and Immunity **73**.

26. **Griffin, P. M., and R. V. Tauxe**. 1991. The Epidemiology of Infections Caused by *Escherichia coli* O157: H7, Other Enterohemorrhagic *E. coli*, and the Associated Hemolytic Uremic Syndrome. Epidemiologic Reviews **13**:60.

27. **Hansen, A.-M., and J. B. Kaper**. 2009. Hfq affects the expression of the LEE pathogenicity island in enterohaemorrhagic *Escherichia coli*. Molecular Microbiology **73**:446-65.

28. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, and T. Murata**. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157: H7 and genomic comparison with a laboratory strain K-12. DNA Research **8**:11.

29. **Hussein, H. S.** 2007. Prevalence and pathogenicity of Shiga toxin-producing *Escherichia coli* in beef cattle and their products. Journal of Animal Science **85**:E63-72.

30. **Jacewicz, M. S., D. W. K. Acheson, D. G. Binion, G. A. West, L. L. Lincicome, C. Fiocchi, and G. T. Keusch**. 1999. Responses of Human Intestinal Microvascular Endothelial Cells to Shiga Toxins 1 and 2 and Pathogenesis of Hemorrhagic Colitis. Infection and Immunity **67**:1439-1444.

31. **Jaeger, J., and D. Acheson**. 2000. Shiga Toxin-Producing *Escherichia coli*. Current Infectious Disease Reports **2**:61-67.

32. **Juneja, V. K., O. P. Snyder, A. C. Williams, and B. S. Marmer**. 1997. Thermal Destruction of *Escherichia coli* 0157:H7 in Hamburger. Journal of Food Protection **60**:1163-1166.

33. **Karmali, M. A., M. Mascarenhas, S. Shen, K. Ziebell, S. Johnson, R. Reid-smith, C. Clark, K. Rahn, J. B. Kaper, and J. Isaac-renton**. 2003. Association of Genomic O Island 122 of *Escherichia coli* EDL 933 with Verocytotoxin-Producing *Escherichia coli* Seropathotypes that are Linked to Epidemic and / or Serious Disease. Journal of Clinical Microbiology **41**:4930-4940.

34. **Karmali, M. A., M. Petric, C. Lim, P. C. Fleming, G. S. Arbus, and H. Lior**. 1985. The Association Between Idiopathic Hemolytic Uremic Syndrome and Infection by Verotoxin-Producing *Escherichia coli*. The Journal of Infectious Diseases **151**.

35. **Lentz, E. K., D. Leyva-Illades, M.-S. Lee, R. P. Cherla, and V. L. Tesh**. 2011. Differential Response of the Human Renal Proximal Tubular Epithelial Cell Line HK-2 to Shiga Toxin Types 1 and 2. Infection and Immunity **79**:3527-40.

36. **Levine, M. M., J.-guo Xu, J. B. Kaper, H. Lior, V. Prado, J. Nataro, H. Karch, K. Wachsmuth, and B. Tall**. 1987. A DNA Probe to Identify Enterohemorrhagic *Escherichia coli* of O157:H7 and Other Serotypes That Cause Hemorrhagic Colitis and Hemolytic Uremic Syndrome. The Journal of Infectious Diseases **156**:175-182.

37. **Mobley, H. L., M. D. Island, and R. P. Hausinger**. 1995. Molecular Biology of Microbial Ureases. Microbiological Reviews **59**:451-80.

38. **Murphy, K. C., J. M. Ritchie, M. K. Waldor, A. Løbner-Olesen, and M. G. Marinus**. 2008. Dam Methyltransferase is Required for Stable Lysogeny of the Shiga Toxin (Stx2)-Encoding Bacteriophage 933W of Enterohemorrhagic *Escherichia coli* O157:H7. Journal of Bacteriology **190**:438-41.

39. **Nataro, J. P., and J. B. Kaper**. 1998. Diarrheagenic *Escherichia coli*. Clinical Microbiology Reviews **11**:142.

40. **Noris, M., and G. Remuzzi**. 2005. Hemolytic Uremic Syndrome. Journal of the American Society of Nephrology: JASN **16**:1035-50.

41. **Ogura, Y., T. Ooka, A. Iguchi, H. Toh, M. Asadulghani, K. Oshima, T. Kodama, H. Abe, K. Nakayama, and K. Kurokawa**. 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proceedings of the National Academy of Sciences **106**:17939.

42. **Ostroff, S. M., P. I. Tarr, M. A. Neill, J. H. Lewis, J. M. Kobayashi, A. Neill, and N. Hargrett-bean**. 1989. Toxin Genotypes and Plasmid Profiles as Determinants of Systemic Sequelae in *Escherichia coli* 0157:H7 Infections. The Journal of Infectious Diseases **160**.

43. **O'Connell, C. B., E. a Creasey, S. Knutton, S. Elliott, L. J. Crowther, W. Luo, M. J. Albert, J. B. Kaper, G. Frankel, and M. S. Donnenberg**. 2004. SepL, a protein required for enteropathogenic *Escherichia coli* type III translocation, interacts with secretion component SepD. Molecular Microbiology **52**:1613-25.

44. **Paton, A. W., and J. C. Paton**. 1998. Detection and Characterization of Shiga Toxigenic *Escherichia coli* by Using Multiplex PCR Assays for stx1, stx2 , eaeA , Enterohemorrhagic *E . coli* hlyA , rfb O111 , and rfb O157. Journal of Clinical Microbiology **36**:598-602.

45. **Perna, N. T., G. Plunkett, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Pósfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner**. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409**:529-33.

46. **Pihkala, N., N. Bauer, D. Eblen, P. Evans, R. Johnson, J. Webb, and C. Williams**. 2011. Risk Profile for Pathogenic Non-O157 Shiga Toxin-Producing *Escherichia coli*.

47. **Ratnam, S., S. B. March, R. Ahmed, and G. S. Bezanson**. 1988. Characterization of *Escherichia coli* Serotype 0157:H7. Journal of Clinical Microbiology **26**:2006-2012.

48. **Remuzzi, and Ruggenenti**. 1995. The hemolytic uremic syndrome. Kidney International **47**:2-19.

49. **Richardson, S. E., M. a Karmali, L. E. Becker, and C. R. Smith**. 1988. The Histopathology of the Hemolytic Uremic Syndrome Associated With Verocytotoxin-Producing *Escherichia coli* Infections. Human Pathology **19**:1102-8.

50. **Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, E. S. Olcott, L. M. Johnson, and N. T. Hargrett**. 1983. Hemorrhagic Colitis Associated With a Rare *Escherichia coli* Serotype. New England Journal of Medicine **308**:681-685.

51. **Rozen, S., and H. Skaletsky**. 2000. Primer3 on the WWW for General Users and for Biologist Programmers. Methods in Molecular Biology **132**.

52. **Samadpour, M., M. Kubler, F. C. Buck, G. a Depavia, E. Mazengia, J. Stewart, P. Yang, and D. Alfi**. 2002. Prevalence of Shiga Toxin-Producing *Escherichia coli* in Ground Beef and Cattle Feces from King County, Washington. Journal of Food Protection **65**:1322-5.

53. **Sambrook, J., and D. W. Russell**. 2001. Molecular Cloning: A Laboratory Manual. CSHL press.

54. **Scallan, E.** 2011. Foodborne Illness Acquired in the United States—Major Pathogens. Emerging Infectious Diseases **17**:7-15.

55. **Slonczewski, J., and J. W. Foster**. 2008. Microbiology: An Evolving Science. WW Norton.

56. **Suttorp, N., B. Flöer, H. Schnittler, W. Seeger, and S. Bhakdi**. 1990. Effects of *Escherichia coli* Hemolysin on Endothelial Cell Function. Infection and Immunity **58**:3796-801.

57. **Tatsuno, I., M. Horie, H. Abe, T. Miki, K. Makino, H. Shinagawa, H. Taguchi, S. Kamiya, T. Hayashi, and C. Sasakawa**. 2001. toxB Gene on pO157 of Enterohemorrhagic *Escherichia coli* O157:H7 Is Required for Full Epithelial Cell Adherence Phenotype. Infection and Immunity **69**:6660-6669.

58. **Tesh, V. L., and a D. O'Brien**. 1991. The pathogenic mechanisms of Shiga toxin and the Shiga-like toxins. Molecular Microbiology **5**:1817-22.

59. **Thran, B. H., H. S. Hussein, M. R. Hall, and S. F. Khaiboullina**. 2001. Shiga Toxin-Producing *Escherichia coli* in Beef Heifers Grazing an Irrigated Pasture. Journal of Food Protection **64**:1613-6.

60. **Tilden, J., W. Young, a M. McNamara, C. Custer, B. Boesel, M. a Lambert-Fair, J. Majkowski, D. Vugia, S. B. Werner, J. Hollingsworth, and J. G. Morris**. 1996. A New Route of Transmission for *Escherichia coli*: Infection from Dry Fermented Salami. American Journal of Public Health **86**:1142-45.

61. **Uhlich, G. a**. 2009. KatP contributes to OxyR-regulated hydrogen peroxide resistance in *Escherichia coli* serotype O157:H7. Microbiology **155**:3589-98.

62. **Wang, D., A. J. Roe, S. McAteer, M. J. Shipston, and D. L. Gally**. 2008. Hierarchal type III secretion of translocators and effectors from *Escherichia coli* O157:H7 requires the carboxy terminus of SepL that binds to Tir. Molecular Microbiology **69**:1499-512.

63. **Wells, J. G., L. D. Shipman, K. D. Greene, E. G. Sowers, J. H. Green, D. N. Cameron, and W. E. T. Al**. 1991. Isolation of *Escherichia coli* Serotype 0157:H7 and Other Shiga- Like-Toxin-Producing *E . coli* from Dairy Cattle. Journal of Clinical Microbiology **29**:985-989.

64. **Wong, A. R. C., J. S. Pearson, M. D. Bright, D. Munera, K. S. Robinson, S. F. Lee, G. Frankel, and E. L. Hartland**. 2011. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. Molecular Microbiology **80**:1420-38.

APPENDIX

Table S1: Computational Analysis of 87 Genes Conserved in EHEC strains of Serogroups O157, O26, O111, and O103

Gene	Size	O157	O55	O111	O103	O26	*C. rodentium*	Others	Function
ECs814	960 bps	+	+	+	+	+	-	*E. coli* O127:H6 E2348/69	Outer Membrane Protein
ECs848	882 bps	+	+	+	+	+	+	*E. coli* O127:H6 E2348/69; Phage cdtI DNA; *E. coli* cell cycle inhibiting factor (cif), hypothetical protein	Hypothetical Protein
ECs1091	315 bps	+	+	+	+	+	-	None	Transcriptional Regulator
ECs2182	315 bps	+	+	+	+	+	-	None	Transcriptional Regulator
ECs2737	315 bps	+	+	+	+	+	-	None	Transcriptional Regulator
ECs1322	303 bps	+	-	+	+	+	-	*E. coli* serogroup O149 urease operon; *Klebsiella variicola* At-22; *Klebsiella pneumoniae* 342; *Variovorax paradoxus* EPS	*ureA* (urease subunit gamma)
ECs1323	321 bps	+	-	+	+	+	-	*E. coli* serogroup O149 urease operon; *Klebsiella variicola* At-22*; Klebsiella pneumoniae* 342; *Variovorax paradoxus* EPS	*ureB* (urease subunit beta)
ECs1324	1.71 kb	+	-	+	+	+	-	*E. coli* serogroup O149 urease operon; *Klebsiella variicola* At-22; *Klebsiella pneumoniae* 342; *Klebsiella pneumoniae* subsp. Pneumoniae MGH 78578; *Klebsiella pneumoniae* NTUH-K2044 DNA; *K. aerogenes* urease subunits A (ureA), B (ureB), C (ureC); *Enterobacter cloacae* subsp. cloacae ATCC 13047; *Pseudomonas putida* GB-1; Uncultured organism clone 100203CB100S_C_18 urea amidohydrolase	*ureC* (urease subunit alpha)
ECs1325	465 bps	+	-	+	+	+	-	*E. coli* serogroup O149 urease operon	*ureE* (urease accessory protein UreE)
ECs1326	675 bps	+	-	+	+	+	-	*E. coli* serogroup O149 urease operon	UreF

ECs1327	618 bps	+	-	+	+	+	-	*E. coli* serogroup O149 urease operon; ureG {insertion site of IS3411}; *Bradyrhizobium japonicum* USDA 110 DNA	UreG
ECs1560	2.35 kb	+	-	+	+	+	+	None	Secreted effector protein
ECs1561	3.4 kb	+	-	+	+	+	+	None	Hypothetical Protein
ECs1568	1.36 kb	+	+	+	+	+	+	None	Hypothetical Protein
ECs1772	759 bps	+	+	+	+	+	-	*E. coli* porcine attaching-effacing associated protein (paa); *E. coli* strain ECL1060/4677/3126/1062/8389/8559/1071/8394 porcine attaching-effacing associated protein; *E. coli* plasmid pTENT2 porcine attaching effacing associated protein	Colonization Factor
ECs1995	630 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008; *E. coli* potB, trcA, ORF2, ORF3, ORF4 genes, complete cds	Hypothetical Protein
ECs2155	630 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008; *E. coli* potB, trcA, ORF2, ORF3, ORF4 genes, complete cds	Hypothetical Protein
ECs2226	372 bps	+	+	+	+	+	-	*E. coli* ORF1 (partial), ORF2, ORF3, and rorf1 gene, multiple strains; Stx2-converting phage 1717; Enterobacteria phage 2851	Hypothetical Protein
ECs2715	1.01 kb	+	+	+	+	+	-	*E. coli* strain ICC199 tir-cytoskeleton coupling protein; *E. coli* NT:H19 tccP2 gene for tir-cytoskeleton coupling protein; *E. coli* 051:H- / O109:H9/ O13:H11/O154:H9 tccP2 gene for tir-cytoskeleton coupling protein; *E. coli* NT:H40 tccP2 gene; *E. coli* O104:H12 pseudogene; *E. coli* O21:H- tccP2 gene; *E. coli* O115:K+ tccP2 gene	EspF-like protein
ECs3486	651 bps	+	+	+	+	+	-	*E. coli* potB, trcA, ORF2, ORF3, ORF4 genes, complete cds	Hypothetical Protein
ECs3487	1.06 kb	+	+	+	+	+	-	*E. coli* potB, trcA, ORF2, ORF3, ORF4 genes, complete cds	Hypothetical Protein
ECs3855	1.65 kb	+	+	+	+	+	+	*E. coli* DNA for locus of enterocyte effacement (LEE); *E. coli* pathogenicity island I; *E. coli* CadC gene, partial cds, transposase 2; *E. coli* O127:H6 E2348/69; multiple LEE strains	Enterotoxin

ECs3857	990 bps	+	+	+	+	+	+	*E. coli* DNA for locus of enterocyte effacement (LEE); *E. coli* pathogenicity island I; *E. coli* CadC gene, partial cds, transposase 2; *E. coli* O127:H6 E2348/69; multiple LEE strains	*nleB*
ECs3858	675 bps	+	+	+	+	+	+	*E. coli* DNA for locus of enterocyte effacement (LEE); *E. coli* pathogenicity island I; *E. coli* CadC gene, partial cds, transposase 2; *E. coli* O127:H6 E2348/69; multiple LEE strains	Hypothetical Protein
ECs4552	222 bps	+	+	+	+	+	-	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain 33264/EDS-58 enterocyte effacement gene locus; locus of enterocyte effacement, strain 0181-6/86; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* strain E2348/69 pathogenicity island; *E. coli* unknown gene, partial cds; multiple LEE strains; *E. coli* strain C74 EscF gene; *E. coli* serovar O11:H0 putative EscF gene; *E. coli* pathogenicity island I; *E. coli* Esp B gene, CadC gene; *E. coli* Ler, EscR, EscS, EscT, SepL gene	Hypothetical Protein
ECs4553	408 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain 33264/EDS-58 enterocyte effacement gene locus; locus of enterocyte effacement, strain 0181-6/86; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* O127:H6 E2348/69 pathogenicity island; *E. coli* strain E2348/69 pathogenicity island; *E. coli* unknown gene, partial cds; multiple LEE strains; *E. coli* pathogenicity island I; *E. coli* Esp B gene, CadC gene; *E. coli* Ler, EscR, EscS, EscT, SepL gene; *E. coli* strain 83/39 EspB gene, E65/56 EspB gene, 84/110-1 EspB gene; *E. coli* espB gene, strain E2348/69	Hypothetical Protein

ECs4557	1.06 kb	+	+	+	+	+	+	*E. coli* strain 71074/33264 enterocyte effacement ; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* pas, sepL, espA, espD, espB genes; *E. coli* strain EDS-58 enterocyte effacement gene locus; *E. coli* locus of enterocyte effacement, strains 0181-6/86 and 3431-4/86; *E. coli* O127:H6 E2348/69; *E. coli* strain E2348/69 pathogenicity island; CadC gene; Ler, EscR, EscS, EscT; multiple LEE strains; *E. coli* (STEC) escD, sepL, espA, espD and espB genes; *E. coli* SepL gene, partial cds; *E. coli* translocated intimin receptor	Type III Secretion System Protein
ECs4558	1.22 kb	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity; strain EDS-58 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* pas, sepL, espA, espD, espB genes; *E. coli* strain 33264 enterocyte effacement gene locus; *E. coli* locus of enterocyte effacement, strain 0181-6/86; *E. coli* O127:H6 E2348/69 genome and pathogenicity island; multiple LEE strains; *E. coli* CadC gene, Ler, EscR, EscS, EscT; *E. coli* (STEC) escD, sepL, espA, espD and espB genes; *E. coli* translocated intimin receptor; *E. coli* strain REPEC RDEC-1, intimin (eaeA) gene; enteropathogenic strain REPEC 84/110/1; partial escD gene for secreted protein D, multiple strains	Type III Secretion System Protein
ECs4560	471 bps	+	+	+	-	-	-	*E. coli* strain 71074 enterocyte effacement pathogenicity; *E. coli* strain 33264 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene; strain EDS-58 enterocyte effacement gene locus; *E. coli* strain G5101, 493/89, 2886-75, 86-24, OK-1, 93-111, 16581, 16940, 15533, 17004, 18257, and E32511 intimin (eae) genes; *E. coli* strain aEPEC BA320, aEPEC 9100-83, and aEPEC EC292/84 CesT genes; multiple LEE strains; *E. coli* O119:UT cesT gene for CesT; *E. coli* O127:H6 E2348/69 complete genome and cesT gene	CesT protein

ECs4568	1.34 kb	+	+	+	+	+	+	*E. coli* strain 71074/EDS-58/33264 enterocyte effacement genes; L0001 (yicJ) gene; multiple LEE strains; *E. coli* O127:H6 E2348/69; strain E2348/69 pathogenicity island; *E. coli* O2:H4 putative EscN gene; multiple LEE strains; *E. coli* CadC gene, Ler, EscR, EscS, EscT; E.coli sepA and sepB genes; E.coli isolate 69.5 putative EscN protein gene	Type III Secretion System Protein
ECs4569	2.03 kb	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* strain EDS-58 enterocyte effacement gene locus; multiple LEE strains; *E. coli* O127:H6 E2348/69; strain E2348/69 pathogenicity island; strain 33264 enterocyte effacement gene locus; *E. coli* sepA and sepB genes; CadC gene, partial cds; Ler, EscR, EscS, EscT; pathogenicity island I; *E. coli* strain aEPEC EC292/84 EscV gene; *E. coli* serotype O11:H0 EPEC EscV-like protein gene; strain aEPEC BA4013, aEPEC BA320, and aEPEC 9100-83 EscV genes	Hypothetical Protein
ECs4570	354 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity; L0001 (yicJ) gene; strain 33264 enterocyte effacement gene locus; multiple LEE strains; *E. coli* O127:H6 E2348/69 complete genome; *E. coli* locus of enterocyte effacement, strain 0181-6/86; *E. coli* strain E2348/69 pathogenicity island; strain EDS-58 enterocyte effacement gene locus; CadC gene, partial cds; Ler, EscR, EscS, EscT; pathogenicity island I; *E. coli* locus of enterocyte effacement, strain 3431-4/86	Hypothetical Protein

ECs4572	429 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity; *E. coli* strain 33264 enterocyte effacement gene locus; strain EDS-58 enterocyte effacement gene locus; L0001 (yicJ) gene; *E. coli* locus of enterocyte effacement, strain 3431-4/86; multiple LEE strains; *E. coli* O127;H6 E2348/69 complete genome; *E. coli* strain E2348/69 pathogenicity island; *E. coli* locus of enterocyte effacement, strain 0181-6/86; *E. coli* CadC gene, partial cds; Ler, EscR, EscS, EscT; *E. coli* pathogenicity island I; *E. coli* enterohemorrhagic strain 86-24 serotyp O157:H7; enteropathogenic strain O55:H7 SepZ gene; enterohemorrhagic strain EDL 933 serotype O157:H7	Hypothetical Protein
ECs4573	573 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* strain 33264 and EDS-58 enterocyte effacement gene locus; multiple LEE strains; *E. coli* O127:H6 E2348/69 complete genome; *E. coli* strain E2348/69 pathogenicity island; *E. coli* isolate 75.1 putative EscJa protein gene; *E. coli* locus of enterocyte effacement, strains 0181-6/86 and 3431-4/86; pathogenicity island I; CadC gene, partial cds; DNA for locus of enterocyte effacement II; Ler, EscR, EscS, EscT; *E. coli* serotype O11:H0 putative EscJ protein gene; *E. coli* isolate 190.5 putative EscJ protein gene	Type III Secretion System Protein
ECs4574	456 bps	+	+	+	+	+	+	*E. coli* strain 33264 and EDS-58 enterocyte effacement gene locus; L0001 (yicJ) gene, partial cds; *E. coli* strain 71074 enterocyte effacement pathogenicity..; multiple LEE strains; *E. coli* O127;H6 E2348/69; *E. coli* locus of enterocyte effacement, strain 0181-6/86; *E. coli* strain E2348/69 pathogenicity island; sepC and sepD genes; *E. coli* locus of enterocyte effacement, strain 3431-4/86; pathogenicity island I; CadC gene, partial cds; transposase 2; Ler, EscR, EscS, EscT; DNA for locus of enterocyte effacement II	Type III Secretion System Protein

ECs4575	1.54 kb	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; L0001 (yicJ) gene, partial cds; *E. coli* strains 33264 and EDS-58 enterocyte effacement gene locus; *E. coli* O127:H6 E2348/69; multiple LEE strains; *E. coli* locus of enterocyte effacement, strain 0181-6/86; pathogenicity island I; CadC gene, partial cds; DNA for locus of enterocyte effacement II; Ler, EscR, EscS, EscT	Type III Secretion System Protein
ECs4576	456 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; L0001 (yicJ) gene, partial cds; *E. coli* strains 33264 and EDS-58 enterocyte effacement gene locus; *E. coli* O127:H6 E2348/69; multiple LEE strains; *E. coli* locus of enterocyte effacement, strain 0181-6/86; pathogenicity island I; CadC gene, partial cds; DNA for locus of enterocyte effacement II; Ler, EscR, EscS, EscT	Type III Secretion System Protein
ECs4577	414 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain EDS-58 and 33264 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* O127:H6 E2348/69; *E. coli* locus of enterocyte effacement, strain 0181-6/86; strain E2348/69 pathogenicity island; multiple LEE strains; pathogenicity island I; DNA for locus of enterocyte effacement II; CadC gene, partial cds; transposase 2; Ler, EscR, EscS, EscT; *E. coli* locus of enterocyte effacement, strain 3431-4/86	Hypothetical Protein
ECs4578	372 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain EDS-58 and 33264 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* O127:H6 E2348/69; *E. coli* locus of enterocyte effacement, strain 0181-6/86; strain E2348/69 pathogenicity island; multiple LEE strains; pathogenicity island I; DNA for locus of enterocyte effacement II; CadC gene, partial cds; transposase 2; Ler, EscR, EscS, EscT; *E. coli* locus of enterocyte effacement, strain 3431-4/86	Negative Regulator GrlR

ECs4580	1.04 kb	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain EDS-58 and 33264 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* O127:H6 E2348/69; *E. coli* locus of enterocyte effacement, strain 0181-6/86; strain E2348/69 pathogenicity island; multiple LEE strains; pathogenicity island I; DNA for locus of enterocyte effacement II; CadC gene, partial cds; transposase 2; Ler, EscR, EscS, EscT; *E. coli* locus of enterocyte effacement, strain 3431-4/86	secretion system apparatus protein SsaU
ECs4581	777 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain EDS-58 and 33264 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* O127:H6 E2348/69; *E. coli* locus of enterocyte effacement, strain 0181-6/86; strain E2348/69 pathogenicity island; multiple LEE strains; pathogenicity island I; DNA for locus of enterocyte effacement II; CadC gene, partial cds; transposase 2; Ler, EscR, EscS, EscT; *E. coli* locus of enterocyte effacement, strain 3431-4/86	Type III Secretion System Protein
ECs4582	270 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain EDS-58 and 33264 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* O127:H6 E2348/69; *E. coli* locus of enterocyte effacement, strain 0181-6/86; strain E2348/69 pathogenicity island; multiple LEE strains; pathogenicity island I; DNA for locus of enterocyte effacement II; CadC gene, partial cds; transposase 2; Ler, EscR, EscS, EscT; *E. coli* locus of enterocyte effacement, strain 3431-4/86	Type III Secretion System Protein

ECs4583	654 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain EDS-58 and 33264 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* O127:H6 E2348/69; *E. coli* locus of enterocyte effacement, strain 0181-6/86; strain E2348/69 pathogenicity island; multiple LEE strains; pathogenicity island I; DNA for locus of enterocyte effacement II; CadC gene, partial cds; transposase 2; Ler, EscR, EscS, EscT; *E. coli* locus of enterocyte effacement, strain 3431-4/86; *E. coli* serovar O11:H0 putative EscR gene, complete cds	Type III Secretion System Protein
ECs4584	696 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strains EDS-58 and 33264 enterocyte effacement gene locus; L0001 (yicJ) gene; *E. coli* locus of enterocyte effacement, strain 0181-6/86; multiple LEE strains; *E. coli* O127:H6 E2348/69 complete genome; strain E2348/69 pathogenicity island; pathogenicity island I; *E. coli* transposase 2, transposase 1, EspG; Cad C gene, partial cds; DNA locus of enterocyte effacement II; Ler, EscR, EscS, EscT; *E. coli* locus of enterocyte effacement, strain 3431-4/86	Hypothetical Protein
ECs4585	600 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strains EDS-58 and 33264 enterocyte effacement gene locus; L0001 (yicJ) gene; *E. coli* locus of enterocyte effacement, strain 0181-6/86; multiple LEE strains; *E. coli* O127:H6 E2348/69 complete genome; strain E2348/69 pathogenicity island; pathogenicity island I; *E. coli* transposase 2, transposase 1, EspG; Cad C gene, partial cds; DNA locus of enterocyte effacement II; Ler, EscR, EscS, EscT; *E. coli* locus of enterocyte effacement, strain 3431-4/86	Hypothetical Protein
ECs4587	219 bps	+	+	+	-	-	-	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain 33264 and EDS-58 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; multiple LEE strains; *E. coli* O127:H6 E2348/69 complete genome; *E. coli* strain E2348/69 pathogenicity island; *E. coli* locus of enterocyte effacement, strain 0181-6/86	Hypothetical Protein

ECs4588	390 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* strain 33264 and EDS-58 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene; multiple LEE strains; *E. coli* O127:H6 E2348/69 complete genome; *E. coli* strain E2348/69 pathogenicity island; *E. coli* locus of enterocyte effacement, strain 0181-6/86; *E. coli* transposase 2, transposase1 , EspG; *E. coli* Cadc gene, partial cds; DNA for locus of enterocyte effacement II; Ler, EscR, EscS, EscT; *E. coli* strain aEPEC 9100-83, aEPEC BA320, aEPEC BA4013, and aEPEC EC292/84 Ler genes; *E. coli* LEE encoded regulator (ler) gene, complete cds; *E. coli* pathogenicity island I; *E. coli* locus of enterocyte effacement, strain 3431-4/86	Ler
ECsL17/ pO157p17	492 bps	+	-	+	+	+	-	*E. coli* strain 83-75 plasmid hemolysin C; *E. coli* strain H30 plasmid pO26-Vir, complete sequence; *E. coli* plasmid pO103 DNA for partial HlyA gene; *E. coli* plasmid-DNA for EHEC-hemolysin operon; *E. coli* plasmid DNA hlyC gene; *E. coli* EHEC-hlyC, A, B, D genes for hemolysin C, A, B, D; *E. coli* strain EH41 plasmid pO113, complete sequence; *E. coli* plasmid pO157 DNA	Hemolysin C
ECsL18/ pO157p18	3.0 kb	+	-	+	+	+	+	*E. coli* strain AGR270 enterohemolysin (*ehxA*) gene; *E. coli* strain ER03/4238 enterohemolysin gene; *E. coli* serotype O157:H7 plasmid pO157 hemolysin toxin; *E. coli* strains AGR047, AGR119, AGR374, AGR053, AGR158, AGR340, AGR670, AGR674 enterohemolysin (*ehxA*) genes; *E. coli* plasmid-DNA for EHEC-hemolysin operon; *E. coli* strain 83-75 plasmid hemolysin C; *E. coli* strain H30 plasmid pO26-Vir; *E. coli* EHEC-hlyA gene; *E. coli* EHEC-hlyC, A, B, D genes for hemolysin C, A, B, D; *E. coli* hemolysin (*ehxA*) gene; *E. coli* strain EH41 plasmid pO113, complete sequences; *E. coli* EHEC HlyA gene, partial cds; *E. coli* strains G5101, 493/89, 2886-75, 86-24, OK-1, 93-111, 18257, 15533, 17004 + more hemolysin A genes.. REFER TO GENE LIST FOR OTHERS	Hemolysin A

ECsL19/ pO157p19	2.12 kb	+	-	+	+	+	-	*E. coli* plasmid-DNA for EHEC-hemolysin operon; *E. coli* strain 83-75 plasmid hemolysin C; *E. coli* strain H30 plasmid pO26-Vir; *E. coli* EHEC-hlyC, A, B, D genes for hemolysin C, A, B, D; *E. coli* strain EH41 plasmid pO113; *E. coli* EHEC HlyA (hlyA) gene, partial cds; *E. coli* strain CH014 transport protein E-HlyB gene; *E. coli* UM146 and UTI89 , complete genomes; *E. coli* hemolysin (hlyB) gene; *E. coli* CFT073, complete genome	Hemolysin B
ECsL20/ pO157p20	1.44 kb	+	-	+	+	+	-	*E. coli* strain 83-75 plasmid hemolysin C; *E. coli* strain H30 plasmid pO26-Vir, complete sequence; E.coli plasmid-DNA for EHEC-hemolysin operon, *E. coli* EHEC-hlyC, A, B, D genes for hemolysin C, A, B, D; *E. coli* strain EH41 plasmid pO113, complete sequence	Hemolysin D
ECsL88/ pO157p79	438 bps	+	-	+	+	+	-	*E. coli* strain H30 plasmid pO26-Vir; *E. coli* 7.4 kb DNA from plasmid pO157; *E. coli* plasmid pO157 ecf4 gene, partial cds	Hypothetical Protein
ECsL89/ pO157p80	1.11 kb	+	-	+	+	+	-	*E. coli* 7.4 kb DNA from plasmid pO157; *E. coli* plasmid pO157 ecf4 gene, partial cds; *E. coli* strain H30 plasmid pO26-Vir	Hypothetical Protein
ECsL91/ pO157p82	999 bps	+	-	+	+	+	-	*E. coli* 7.4 kb DNA from plasmid pO157; *E. coli* strain H30 plasmid pO26-Vir; *E. coli* plasmid pO157 ecf4 gene, partial cds; *Photorhabdus asymbiotica* ATCC43949 complete genome	Lipid A Biosynthesis
ECs0815	690 bps	+	+	+	+	+	-	*E. coli* O127:H6 E2348/69 complete genome; Bacteriophage 82 orf33, orf151, orf56, orf96, rus, orf45, and Q gene; Bacteriophage 82 antitermination protein (Q) gene, partial cds	Anti-termination protein
ECs1084	372 bps	+	+	+	+	+	-	Shigella sonnei Ss046, complete genome	Anti-termination protein
ECs1778	531 bps	+	-	+	+	+	-	*E. coli* O91:H21 Shiga toxin 2 variant d A subunit (stx2d2A); *E. coli* ED1a chromosome; *E. coli* ETEC H10407, complete genome; *E. coli* O127:H6 E2348/69	Hypothetical Protein

ECs2183	186 bps	+	+	+	+	+	-	Phage BP-4795; Enterobacteria phage YYZ-2008; *Shigella boydii* CDC 3083-94; Bacteriophage CP-1639 and chromosomal integration site; *E. coli* q gene (partial), stx1A gene, stx1B gene, ORF101; Enterobacteria phage VT1-Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Stx2 converting phage II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic DNA; Bacteriophage 933W; Enterobacteria phage Min27; *E. coli* ED1a chromosome; *E. coli* ABU 83972, complete genome; *E.coli* O127:H6 E2348/69 complete genome; *E. coli* SMS-3-5; Phage cdtI DNA; *Shigella boydii* Sb227, complete genome	Putative lipoprotein Rz1 precursor
ECs2738	186 bps	+	+	+	+	+	-	Phage BP-4795; Enterobacteria phage YYZ-2008; *Shigella boydii* CDC 3083-94; Bacteriophage CP-1639 and chromosomal integration site; *E. coli* q gene (partial), stx1A gene, stx1B gene, ORF101; Enterobacteria phage VT1-Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Stx2 converting phage II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic DNA; Bacteriophage 933W; Enterobacteria phage Min27; *E. coli* ED1a chromosome; *E. coli* ABU 83972, complete genome; E.col O127:H6 E2348/69 complete genome; *E. coli* SMS-3-5; Phage cdtI DNA; *Shigella boydii* Sb227, complete genome	Putative lipoprotein precursor

ECs2166	2.61 kb	+	+	+	+	+	-	Bacteriophage CP-1639 and chromosomal integration site; Enterobacteria phage VT1-Sakai genomic DNA; *Shigella boydii* CDC 3083-94; *E. coli* ED1a, 55989, S88, UMN026, IAI39; IAI1, LF82 chromosomes; *E. coli* 042, APEC 01, UTI89, IHE3034, UM146, CFT073, SE11 DNA, Ss046, ABU83972, ETEC H10407, 536, O83:H1 str. NRG 857C, W, BL21(DE3), 'BL21-Gold(DE3)pLysS AG', BW2952 complete genomes; Enterobacteria phage DE3, complete genome; Cloning vector lambdaS2775; *Monosiqa brevicollis* clone JGIACYI-5B3 and JGIACYI-5A5, complete sequences; Uncultured bacterium clone zdt-9n2; Enterobacteria phage lambda, complete genome; Cloning vectors TLF97-3, TLF97-2, TLF97-1, phage lambda lacZ transitional fusion vector; Cloning vector lambda EMBL3 SP6/T7, left arm; *E. coli* ATCC 8739, complete genome; *E. coli* SE15 DNA; *E. coli* O127:H6 E2348/69; *Homo sapiens* genomic sequence surrounding NotI site, clones NR5-KJ and NR1-W; TSA: *Linepithema humile* isotiq02319.Lihu mRNA sequence	tail length tape measure protein precursor
ECs2725	765 bps	+	+	+	+	+	-	Bacteriophage CP-1639 and chromosomal integration site; Enterobacteria phage VT1-Sakai genomic DNA; *Shigella boydii* CDC 3083-94; *E. coli* ED1a, 55989, S88, UMN026, IAI39; IAI1, LF82 chromosomes; *E. coli* 042, APEC 01, UTI89, IHE3034, UM146, CFT073, SE11 DNA, Ss046, ABU83972, ETEC H10407, 536, KO11, O83:H1 str. NRG 857C, W, BL21(DE3), 'BL21-Gold(DE3)pLysS AG', BW2952 complete genomes; Enterobacteria phage DE3, complete genome; Cloning vector lambdaS2775; *Monosiqa brevicollis* clone JGIACYI-5B3 and JGIACYI-5A5, complete sequences; Uncultured bacterium clone zdt-9n2; Enterobacteria phage lambda, complete genome; Cloning vectors TLF97-3, TLF97-2, TLF97-1, phage lambda lacZ transitional fusion vector; Cloning vector lambda TXF97, lacZ; Cloning vector lambda EMBL3 SP6/T7, left arm; *E. coli* ATCC 8739, complete genome; *E. coli* SE15 DNA; *E. coli* O127:H6 E2348/69; *Homo sapiens* genomic sequence surrounding NotI site, clones NR5-KJ and NR1-W	putative tail length tape measure protein precursor

ECs2949	2.65 kb	+	+	+	+	+	-	Bacteriophage CP-1639 and chromosomal integration site; Enterobacteria phage VT1-Sakai genomic DNA; *Shigella boydii* CDC 3083-94; *E. coli* ED1a, 55989, S88, UMN026, IAI39; IAI1, LF82 chromosomes; *E. coli* 042, APEC 01, UTI89, IHE3034, UM146, CFT073, SE11 DNA, Ss046, ABU83972, ETEC H10407, 536, KO11, O83:H1 str. NRG 857C, W, BL21(DE3), 'BL21-Gold(DE3)pLysS AG', BW2952 complete genomes; *Shigella sonnei* Ss046, complete genome; Enterobacteria phage DE3, complete genome; Cloning vector lambdaS2775; *Monosiqa brevicollis* clone JGIACYI-5B3 and JGIACYI-5A5, complete sequences; Uncultured bacterium clone zdt-9n2; Enterobacteria phage lambda, complete genome; Cloning vectors TLF97-3, TLF97-2, TLF97-1, phage lambda lacZ transitional fusion vector; Cloning vector lambda TXF97, lacZ; Cloning vector lambda EMBL3 SP6/T7, left arm; *E. coli* ATCC 8739, complete genome; *E. coli* SE15 DNA; *E. coli* O127:H6 E2348/69; Homo sapiens genomic sequence surrounding NotI site, clones NR5-KJ and NR1-W; TSA: *Linepithema humile* isotiq02319.Lihu mRNA sequence	putative tail length tape measure protein precursor
ECs2168	423 bps	+	+	+	+	+	-	Bacteriophage CP-1639 and chromosomal integration site; Enterobacteria phage VT1-Sakai genomic DNA; *Shigella boydii* CDC 3083-94; *E. coli* UM146, IHE3034; 042, APEC O1, UTI89, CFT073 complete genomes; *E. coli* S88, 55989, UMN026, ED1a, IAI39 chromosomes, complete genomes; *E. coli* 0127:H6 E2348/69 complete genome	minor tail protein
ECs2951	423 bps	+	+	+	+	+	-	Bacteriophage CP-1639 and chromosomal integration site; Enterobacteria phage VT1-Sakai genomic DNA; Shigella boydii CDC 3083-94; *E. coli* UM146, IHE3034; 042, APEC O1, UTI89, CFT073 complete genomes; *E. coli* S88, 55989, UMN026, ED1a, IAI39 chromosomes, complete genomes; *E. coli* 0127:H6 E2348/69 complete genome	minor tail protein

ECs1092	186 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008; *Shigella boydii* CDC 3083-94; Bacteriophage CP-1639 and chromosomal integration site; *E. coli* q gene (partial), stx1A gene, stx1B gene; Enterobacteria phage VT1-Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Stx2 converting phage II DNA; Stx1 converting phage DNA; Phage BP-4795; Enterobacteria phage VT2-Sakai genomic DNA; Bacteriophage 933W; Enterobacteria phage Min27; *E. coli* ED1a, 55989, UMN026 chromosomes; *E. coli* 0127:H6 E2348/69, ABU 83972, Sb227, SMS-3-5, APEC O1, IHE3034, BL21(DE3), O83:H1 str. NRG 857C, UM146, complete genomes; Phage cdtI DNA; *Shigella boydii* Sb227; Enterobacteria phage VT2phi_272; Stx2-converting phage 86 and I DNA; *Shigella sonnei* bacteriophage 7888 stxA, q(partial), stxB; *Shigella sonnei* Ss046, complete genome; refer to gene list for more	putative lipoprotein Rz1 protein precursor

ECs1216	186 bps	+	+	+	+	+	-	Stx2 converting phage II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic DNA; Bacteriophage 933W; Enterobacteria phage Min27 and YYZ-2008, complete genomes; *Shigella boydii* CDC 3083-94; Bacteriophage CP-1639 and chromosomal integration site; *E. coli* q gene (partial), stx1A, stx1B genes; Enterobacteria phage VT1-Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Phage BP-4795; *E. coli* ED1a, 55989, UMN026, LF82, S88, IAI1, SE11, 536, UTI89, CFT073, DH1 complete chromosomes; *E. coli* 0127:H6 E2348/69, ABU 83972, SMS-3-5, APEC O1; IHE3034; BL21(DE3), O83:H1 str. NRG 857C, UM146, B str. REL606, 'BL21-Gold(DE3)pLysS AG', SE11 DNA, BW2952, 042, ETEC H10407, complete genomes; Enterobacteria phage VT2phi_272; Stx2-converting phage 86 and I DNA; *Shigella boydii* Sb227; *Shigella sonnei* bacteriophage 7888 stxA, q (partial); *Shigella sonnei* Ss046; Bacillus thuringiensis CT43 plasmid pBMB0558; Enterobacteria phage DE3; Cloning vector lambdaS2775; *E. coli* str. K12 substr. DH10B, W3110 DNA, and MG1655 complete genomes; Mus musculus activated spleen cDNA; Paracoccidioides brasiliensis endopeptidase mRNA; Enterobacteria phage lambda; Bacteriophage lambda Rz1 protein precursor; Cloning vector TLF97-3,2,1 phage lambda lacZ translational fusion vector; Cloning vector lambda TXF97 and EMBL3, right arm; Bacteriophage 21 lysis genes S, R, and Rz; *E. coli* fergusonii ATCC 35469 chromosome	putative lipoprotein Rz1 protein precursor

ECs1787	186 bps	+	+	+	+	+	-	Stx2 converting phage II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic DNA; Bacteriophage 933W; Enterobacteria phage Min27 and YYZ-2008, complete genomes; *Shigella boydii* CDC 3083-94; Bacteriophage CP-1639 and chromosomal integration site; *E. coli* q gene (partial), stx1A, stx1B genes; Enterobacteria phage VT1-Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Phage BP-4795; *E. coli* ED1a, 55989, UMN026, LF82, S88, IAI1, SE11, 536, UTI89, CFT073, DH1 complete chromosomes; *E. coli* 0127:H6 E2348/69, ABU 83972, SMS-3-5, APEC O1; IHE3034; BL21(DE3), O83:H1 str. NRG 857C, UM146, B str. REL606, 'BL21-Gold(DE3)pLysS AG', SE11 DNA, BW2952, 042, ETEC H10407, complete genomes; Enterobacteria phage VT2phi_272; Stx2-converting phage 86 and I DNA; *Shigella boydii* Sb227; *Shigella sonnei* bacteriophage 7888 stxA, q (partial); *Shigella sonnei* Ss046; Bacillus thuringiensis CT43 plasmid pBMB0558; Enterobacteria phage DE3; Cloning vector lambdaS2775; *E. coli* str. K12 substr. DH10B, W3110 DNA, and MG1655 complete genomes; Mus musculus activated spleen cDNA; Paracoccidioides brasiliensis endopeptidase mRNA; Enterobacteria phage lambda; Bacteriophage lambda Rz1 protein precursor; Cloning vector TLF97-3,2,1 phage lambda lacZ translational fusion vector; Cloning vector lambda TXF97 and EMBL3, right arm; Bacteriophage 21 lysis genes S, R, and Rz; *E. coli* fergusonii ATCC 35469 chromosome	lipoprotein Rz1 precursor

ECs2256	186 bps	+	+	+	+	+	-	Stx2 converting phage II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic DNA; Bacteriophage 933W; Enterobacteria phage Min27 and YYZ-2008, complete genomes; *Shigella boydii* CDC 3083-94; Bacteriophage CP-1639 and chromosomal integration site; *E. coli* q gene (partial), stx1A, stx1B genes; Enterobacteria phage VT1-Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Phage BP-4795; *E. coli* ED1a, 55989, UMN026, LF82, S88, IAI1, SE11, 536, UTI89, CFT073, DH1 complete chromosomes; *E. coli* 0127:H6 E2348/69, ABU 83972, SMS-3-5, APEC O1; IHE3034; BL21(DE3), O83:H1 str. NRG 857C, UM146, B str. REL606, 'BL21-Gold(DE3)pLysS AG', SE11 DNA, BW2952, 042, ETEC H10407, complete genomes; Enterobacteria phage VT2phi_272; Stx2-converting phage 86 and I DNA; *Shigella boydii* Sb227; *Shigella sonnei* bacteriophage 7888 stxA, q (partial); *Shigella sonnei* Ss046; Bacillus thuringiensis CT43 plasmid pBMB0558; Enterobacteria phage DE3; Cloning vector lambdaS2775; *E. coli* str. K12 substr. DH10B, W3110 DNA, and MG1655 complete genomes; Mus musculus activated spleen cDNA; Paracoccidioides brasiliensis endopeptidase mRNA; Enterobacteria phage lambda; Bacteriophage lambda Rz1 protein precursor; Cloning vector TLF97-3,2,1 phage lambda lacZ translational fusion vector; Cloning vector lambda TXF97 and EMBL3, right arm; Bacteriophage 21 lysis genes S, R, and Rz; *E. coli* fergusonii ATCC 35469 chromosome	lipoprotein Rz1 precursor

ECs2965	186 bps	+	+	+	+	+	-	Stx2 converting phage II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic DNA; Bacteriophage 933W; Enterobacteria phage Min27 and YYZ-2008, complete genomes; *Shigella boydii* CDC 3083-94; Bacteriophage CP-1639 and chromosomal integration site; *E. coli* q gene (partial), stx1A, stx1B genes; Enterobacteria phage VT1-Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Phage BP-4795; *E. coli* ED1a, 55989, UMN026, LF82, S88, IAI1, SE11, 536, UTI89, CFT073, DH1 complete chromosomes; *E. coli* 0127:H6 E2348/69, ABU 83972, SMS-3-5, APEC O1; IHE3034; BL21(DE3), O83:H1 str. NRG 857C, UM146, B str. REL606, 'BL21-Gold(DE3)pLysS AG', SE11 DNA, BW2952, 042, ETEC H10407, complete genomes; Enterobacteria phage VT2phi_272; Stx2-converting phage 86 and I DNA; *Shigella boydii* Sb227; *Shigella sonnei* bacteriophage 7888 stxA, q (partial); *Shigella sonnei* Ss046; Bacillus thuringiensis CT43 plasmid pBMB0558; Enterobacteria phage DE3; Cloning vector lambdaS2775; *E. coli* str. K12 substr. DH10B, W3110 DNA, and MG1655 complete genomes; Mus musculus activated spleen cDNA; Paracoccidioides brasiliensis endopeptidase mRNA; Enterobacteria phage lambda; Bacteriophage lambda Rz1 protein precursor; Cloning vector TLF97-3,2,1 phage lambda lacZ translational fusion vector; Cloning vector lambda TXF97 and EMBL3, right arm; Bacteriophage 21 lysis genes S, R, and Rz; *E. coli* fergusonii ATCC 35469 chromosome	lipoprotein Rz1 precursor
ECs5623	No Hit	-	-	-	-	-	-	None	Unknown
ECs1214	570 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008; Stx2 converting phage II DNA; Stx1 converting phage DNA; Phage BP-4795; Enterobacteria phage VT1 and VT2 -Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Bacteriophage 933W; Enterobacteria phage Min27; Stx2-converting phage 86 and I DNA; *Shigella sonnei* bacteriophage 7888 stxA, q (partial), stxB; Enterobacteria phage VT2phi_272, complete sequence	anti-repressor protein

ECs1533	570 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008; Stx2 converting phage II DNA; Stx1 converting phage DNA; Phage BP-4795; Enterobacteria phage VT1 and VT2 -Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Bacteriophage 933W; Enterobacteria phage Min27; Stx2-converting phage 86 and I DNA; *Shigella sonnei* bacteriophage 7888 stxA, q (partial), stxB; Enterobacteria phage VT2phi_272, complete sequence	putative tail length tape measure protein
ECs1785	570 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008; Stx2 converting phage II DNA; Stx1 converting phage DNA; Phage BP-4795; Enterobacteria phage VT1 and VT2 -Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Bacteriophage 933W; Enterobacteria phage Min27; Stx2-converting phage 86 and I DNA; *Shigella sonnei* bacteriophage 7888 stxA, q (partial), stxB; Enterobacteria phage VT2phi_272, complete sequence	antirepressor protein
ECs1965	570 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008; Stx2 converting phage II DNA; Stx1 converting phage DNA; Phage BP-4795; Enterobacteria phage VT1 and VT2 -Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Bacteriophage 933W; Enterobacteria phage Min27; Stx2-converting phage 86 and I DNA; *Shigella sonnei* bacteriophage 7888 stxA, q (partial), stxB; Enterobacteria phage VT2phi_272, complete sequence	putative antirepressor protein
ECs2258	570 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008; Stx2 converting phage II DNA; Stx1 converting phage DNA; Phage BP-4795; Enterobacteria phage VT1 and VT2 -Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Bacteriophage 933W; Enterobacteria phage Min27; Stx2-converting phage 86 and I DNA; *Shigella sonnei* bacteriophage 7888 stxA, q (partial), stxB; Enterobacteria phage VT2phi_272, complete sequence	putative antirepressor protein

ECs2967	570 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008; Stx2 converting phage II DNA; Stx1 converting phage DNA; Phage BP-4795; Enterobacteria phage VT1 and VT2 -Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Bacteriophage 933W; Enterobacteria phage Min27; Stx2-converting phage 86 and I DNA; *Shigella sonnei* bacteriophage 7888 stxA, q (partial), stxB; Enterobacteria phage VT2phi_272, complete sequence	putative antirepressor protein
ECs1316	360 bps	+	-	+	+	+	-	*E. coli* ED1a chromosome, complete genome	putative diacylglycerol kinase
ECs1332	342 bps	+	-	+	+	+	-	*E. coli* ED1a chromosome, complete genome	putative colicin immunity protein
ECs2153	243 bps	+	+	+	+	+	-	Bacteriophage CP-1639 and chromosomal integration site; Enterobacteria phage VT1-Sakai genomic DNA; *E. coli* ETEC H10407; Shigella flexneri 2002017; *E. coli* cdtB, cdtA, cdtC genes; Salmonella enterica subsp. Enterica serovar Schwarzengrund str. CV; Phage cdtI DNA; Shigella flexneri 5 str. 8401, 2a str. 301, 2a str. 2457T complete genomes; *Shigella boydii* Sb227; *Shigella sonnei* Ss046; *E. coli* S88 chromosome; E. fergusonii ATCC 35469 chromosome; *Shigella boydii* CDC 3083-94; *E. coli* APEC O1, CFT073, IHE3034, and UTI89 complete genomes; *E. coli* ED1a chromosome; Bacteriophage P-EibC ORF-191C immunoglobin binding..; Phage BP-4795	putative damage-inducible protein

ECs2218	255 bps	+	+	+	+	+	-	*E. coli* ED1a chromosome; Bacteriophage P-EibC ORF-191C immunoglobin binding..; Phage BP-4795; *E. coli* SE11 DNA, complete genome; Shigella flexneria 2a str. 301 and 2a str. 2457T complete genoms; Shigella flexneri aerobactin pathogenicity and SHI-2 pathogenicity islands; *E. coli* plasmid pO86A1 DNA; Salmonella enterica subsp. enterica serovar Typhimurium; *E. coli* plasmid pEC_L46 and pEC_L8 complete sequences; Shigella flexneri 2002017 plasmid pSFxv_1; *E. coli* strain A plasmid pEK499; Shigella flexneri 2a str. 301 virulence plasmid pCP301; Stx2 converting phage II DNA; Shigella flexneri virulence plasmid pWR100; Uncultured bacterium pRSB107 plasmid; *E. coli* pathogenicity island I; *E. coli* plasmid pO103 DNA for partial HlyA gene; Shigella flexneri 5a plasmid virulence plasmid pWR501; *E. coli* plasmid pO157, espP natural mutant; *E. coli* 8.6 kb DNA from plasmid p0157; Enterobacteria phage VT2-Sakai genomic DNA; *E. coli* 0127:H6 E2348/69 plasmid pMAR2; *E. coli* strain E2348/69 plasmid pMAR7; *E. coli* plasmid pB171 DNA	Hypothetical Protein
ECs0813	666 bps	+	+	+	0	+	-	Bacteriophage LC159 DNA-binding protein Roi; *E. coli* strain R1388, R834, EC970520, F1081, and LS68 stx2ca upstream region genomic sequences; Stx2-converting phage 1717; Enterobacteria phage 2851; Bacteriophage 2851 roi gene, ORF2, ORF3, Q gene, stx2A gene; Bacteriophage Nil2 proviral n gene, ORF22, ci gene, cro gene; ENterobacteria phage Sf6; E. fergusonii ATCC 35469 chromosome; *E. coli* BW2952; Enterobacteria phage lambda; *E. coli* BL21(DE3); Enterobacteria phage DE3; *E. coli* 'BL21-Gold(DE3)pLysS AG'; Cloning vector lambda TXF97, lacZ transcriptional fusion vector; Cloning vector lambda EMBL3, right arm; Cloning vector lambdaS2775; Mus musculus melanocyte cDNA, RIKEN; TSA Arachis hypogaea CL10403Contig1.Arhy mRNA sequence	serine/threonin protein phosphatase

ECs3502	666 bps	+	+	+	+	+	-	Bacteriophage LC159 DNA-binding protein Roi; *E. coli* strain R1388, R834, EC970520, F1081, and LS68 stx2ca upstream region genomic sequences; Stx2-converting phage 1717; Enterobacteria phage 2851; Bacteriophage 2851 roi gene, ORF2, ORF3, Q gene, stx2A gene; Bacteriophage Nil2 proviral n gene, ORF22, ci gene, cro gene; ENterobacteria phage Sf6; E. fergusonii ATCC 35469 chromosome; *E. coli* BW2952; Enterobacteria phage lambda; *E. coli* BL21(DE3); Enterobacteria phage DE3; *E. coli* 'BL21-Gold(DE3)pLysS AG'; Cloning vector lambda TXF97, lacZ transcriptional fusion vector; Cloning vector lambda EMBL3, right arm; Cloning vector lambdaS2775; Mus musculus melanocyte cDNA, RIKEN; TSA Arachis hypogaea CL10403Contig1.Arhy mRNA sequence	serine/threonine protein phosphatase
ECs1085	819 bps	+	+	+	+	+	-	None	Hypothetical Protein
ECs1097	237 bps	+	-	+	+	+	-	None	Hypothetical Protein
ECs1315	201 bps	+	-	+	+	+	-	*E. coli* ED1a chromosome, complete genome	Hypothetical Protein
ECs1319	276 bps	+	-	+	+	+	-	*E. coli* serogroup O149 urease operon; *E. coli* ED1a chromosome, complete genome	Hypothetical Protein
ECs1320	204 bps	+	-	+	+	+	-	*E. coli* serogroup O149 urease operon, complete genome	Hypothetical Protein
ECs1331	282 bps	+	-	+	+	+	-	*E. coli* ED1a chromosome, complete genome	Hypothetical Protein
ECs1333	672 bps	+	-	+	+	+	-	*E. coli* ED1a chromosome, complete genome	Hypothetical Protein
ECs1342	330 bps	+	-	+	+	+	-	*E. coli* W plasmid pRK1, complete sequence; Gryllus bimaculatus mRNA, GBcontig28810	Hypothetical Protein
ECs1774	258 bps	+	+	+	+	+	-	*E. coli* SMS-3-5; *E. coli* porcine attaching-effacing associated protein; *Shigella sonnei* Ss046; *E. coli* ABU 83972 and 042 complete genomes	Hypothetical Protein
ECs2196	273 bps	+	-	+	+	+	-	None	Hypothetical protein

ECs1244	345 bps	+	-	+	+	+	-	Stx2 converting phage II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Enterobacteria phage Min27; Stx2 converting phage I DNA	Hypothetical Protein
ECs2197	345 bps	+	-	+	+	+	-	Stx2 converting phage II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic DNA; Enterobacteria phage VT2-Sakai proviral DNA; Enterobacteria phage Min27; Stx2 converting phage I DNA; Bacteriophage 933W	Hypothetical Protein
ECs2199	558 bps	+	+	+	+	+	-	*E. coli* SE11 DNA; Mus musculus adult male kidney cDNA; *E. coli* ED1a chromosome; *E. coli* 042; Salmonella enterica subsp. Arizonae serovar 62:z4,z23; *E. coli* IAI39 chromosome; Enterobacteria phage Min27; Stx2 converting phage II and I DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W; *E. coli* UTI89	Hypothetical Protein
ECs2200	219 bps	+	+	+	+	+	-	*Shigella sonnei* Ss046; Shigella flexneri 2002017, 5 str. 8401, and 2a str. 301, complete genomes; *E. coli* SE11 DNA, ABU 83972, and 042, complete genomes	Hypothetical Protein
ECs2202	228 bps	+	-	+	+	+	-	None	Hypothetical Protein
ECs2203	282 bps	+	-	+	+	+	-	None	Hypothetical Protein

ECs2213	261 bps	+	+	+	+	+	-	*E. coli* E24377A, SMS-3-5, 042, IHE3034, UM146, 0127:H6 E2348/69, APEC O1; UTI89, CFT073, W, ETEC H10407, ABU 83972, ATCC 8739, SE11 DNA, DH1 (ME8569) DNA, DH1, BW2952, str. K12 substr. DH10B, str. K12 substr. W3110 DNA, str. K-12 substr. MG1655, complete genomes; *E. coli* 55989, LF82, S88, ED1a, UMN026, IAI1 chromosomes; *Shigella boydii* CDC 3083-94; *E. coli* genes dicA, dicB, dicC and dicF; *E. coli* O83:H1 str. NRG 857C, BL21(DE3), B str. REL606; 'BL21-Gold(DE3)pLysS AG', complete genomes; Shigella flexneri 2002017, 2a str. 301, 2a str. 2457T, complete genomes; S. flexneri insertion sequence IS911 DNA; Phage cdtI DNA; Gryllus bimaculatus mRNA, GBcontig05773; Salmonella enterica subsp. enterica serovar Enteritidis, Dublin, Choleraesuis str.; Bacteriophage P27	Hypothetical Protein
ECs1816	258 bps	+	+	+	+	+	-	*E. coli* 0127:H6 E2348/69; Edwardsiella tarda EIB202; Edwardsiella ictaluri 93-146, D,D -heptose 1/7-bisphosphate phosphotase, transposase-like protein gene, and putative transposase and heat stabel enterotoxins; Salmonella enterica insertion sequence IS1414; *E. coli* plasmid pAPEC-O2-ColV; Primary endosymbiont of Sitophilus zeamais SZPEC circle 6; Salmonella enterica J pseudogene; *E. coli* ETEC H10407 p948 plasmid and ETEC H10407, complete genome; *E. coli* str. 55989 plasmid 55989p; *E. coli* putative plasmid pSC1 EtpB; *E. coli* strain 55989 plasmid pAA-like agg3 gene cluster; *E. coli* insertion sequence IS1414 transposase-like protein; Primary endosymbiont of Sitophilus oryzae insertion sequence, Sitophilus zeamais SZPE circle 1, Sitophilus oryzae SOPE circles 1 and 2; Primary endosymbiont of Sitophilus zeamais clone 04 and 03 insertion sequences; Primary endosymbiont of Sitophilus oryzae insertion sequence; Primary endosymbiont of Sitophilus zeamais clones and insertion sequence (see gene list); Shigella flexneri 2a she pathogenicity island; *E. coli* SMS-3-5, complete genome; *E. coli* SE11 plasmid pSE11-3 DNA; *E. coli* 0127:H6 E2348/69 plasmid pMAR2, strain E2348/69; *E. coli* strain E2348/69 plasmid pMAR7; *E. coli* perA, perB, perC and perD genes; *E. coli* plasmid pB171 DNA; *E. coli* bfpT, bfpV, bfpW and transposase genes; Dickeya zeae Ech1591; *E. coli* cell cycle inhibiting factor (cif), hypothetical protein	Hypothetical Protein

ECs2221	870 bps	+	+	+	-	+	-	Edwardsiella ictaluri 93-146, transposase-like protein gene, and D,D-heptose 1,7-bisphosphate phosphatase; Edwardsiella tarda EIB202; Enterobacter cloacae SCF1; *E. coli* ETEC 1392/75 plasmid p1018; *E. coli* E24377A plasmid pETEC_80; Primary endosymbiont of Sitophilus oryzae insertion sequence and of multiple Sitophilus zeamais clones; Salmonella enterica J pseudogene, partial sequence; *E. coli* ETEC H10407 p948 plasmid along with complete ETEC H10407 genome; *E. coli* str. 55989 plasmid 55989p; *E. coli* putative plasmid pSC1 EtpB, EtpA; *E. coli* strain 55989 plasmid pAA-like agg3 gene cluster; *E. coli* insertion sequence IS1414 transposase-like protein; *E. coli* 042 plasmid pAA; *E. coli* plasmid pAPEC-O2-ColV; Candidatus Serratia symbiotica insertion sequence IS256 transposase; Edwardsiella ictaluri putative transposase and heat stable enterotoxin; Primary endosymbiont of Sitophilus zeamais SZPE circle 6, 1, 2, 3, 4, 1, 9, 10, 5; *E. coli* SMS-3-5 and CFT073, O83:H1 str. NRG 857C, ABU83972, LF82 chromosome, CFT073, complete genomes; Salmonella enterica subsp. enterica serovar Enteritidis str. P125109; EAST1=enterotoxin 1; Salmonella enterica subsp. enterica serovar Gallinarum strain 287/91; Shigella dysenteriae Sd197; Cronobacter sakazakii thermoresistance gene cluster; Dickeya zeae Ech1591; *E. coli* 47IIR, 47IIM, 47IR genes; Salmonella enterica subsp. enterica serrovar Typhi Ty2;Salmonella enterica subsp. enterica serovar Typhimurium (multiple strains), Heidelberg str. SL476, Paratyphi C, A, and B (multiple strains), and Dublin str. CT_02021853; *E. coli* 0127:H6 E2348/69 plasmid pMAR2, strain E2348/69; *E. coli* strain E2348/69 plasmid pMAR7; *E. coli* plasmid pB171 DNA; Shigella dysenteriae Sd197 plasmid pSD1_197; *E. coli* 042, complete genome; Shigella flexneri 2002017; *E. coli* 55989 chromosome	Transposase

ECs2255	228 bps	+	+	+	+	+	-	Enterobacteria phage YYZ-2008, 2851; Phage BP-4795; Stx2-converting phage 1717; *E. coli* S88 chromosome, complete genome; *Shigella boydii* CDC 3083-94 and Sb227, complete genomes; *E. coli* DH1 (ME8569) DNA, W, 042, DH1, BW2952, IAI39 chromosome, IAI1 chromosome, str. K12 subst. DH10B, str K12 substr. W3110 DNA, K-12 substr. MG1655, BL21(DE3), UM146, IHE3034, 'BL21-Gold(DE3)pLysS AG', 55989 chromosome, 0127:H6 E2348/69, SE11 DNA, ATCC 8739, E24377A, APEC O1, UTI89, CFT073, complete genomes; *E. coli* DNA, ecos1 cleavage site region by lambda terminase; Mus musculus transgenic locus rearrangement example 4; Enterobacteria phage DE3; Enterobacteria phage lambda; Cloning vector pAL-F insertion sequence IS1 galactokinase; *Shigella sonnei* Ss046; Cloning vector pAL-Z galactokinase, TLF97-3,2,1 phage lambda lacZ translational fusion vector; Cloning vector lambda TXF97; Cloning vectors pDO193, pDO192, pDO6, pDO2, pDO19, pDO184, pDO17, complete sequences; Cloning vector lambda EMBL2, right arm; Sequencing vector pAA113M DNA; *E. coli* UMN026 chromosome, B str. REL606, complete genomes; Bacillus thuringiensis CT43 plasmid pMBM0558; Mus musculus trasgenic locus rearrangement example 2	Hypothetical Protein

ECs1210	180 bps	+	+	+	+	+	-	Enterobacteria phage VT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Bacteriophage 933W ileX, stx2A, stx2B genes; Enterobacteria phage VT1-Sakai genomic DNA; *Shigella sonnei* bacteriophage 7888 stxA, q, stxB, s, r, ant; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W; *E. coli* stx2cB gene, ORF645, ORF59, s gene, stx2c, ORF107, ORF59 DNA, stx2B; Stx2-converting phage 1717, Enterobacteria phage 2851, Bacteriophage 2851 roi gene, ORF2, ORF3, Q gne; *E. coli* O91:H21 Shiga toxin 2 variant d A subunity; Phage BP-4795; Stx1-converting phage phi-O153; Phage 258-320 lysogen T4, T3, T2 shiga toxin 2 subunit B; *E. coli* Shiga toxin 2 subunit B; *Shigella boydii* CDC 3083-94, Sb227, ABU83972, ED1a chromosome, S88 chromosome, complete genomes; Shigella dysenteriae Sd197, shiga toxin A gene, shiga toxin B gene, SapF gene	Hypothetical Protein
ECs2971	180 bps	+	+	+	+	+	-	Enterobacteria phage VT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Bacteriophage 933W ileX, stx2A, stx2B genes; Enterobacteria phage VT1-Sakai genomic DNA; *Shigella sonnei* bacteriophage 7888 stxA, q, stxB, s, r, ant; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W; *E. coli* stx2cB gene, ORF645, ORF59, s gene, stx2c, ORF107, ORF59 DNA, stx2B; Stx2-converting phage 1717, Enterobacteria phage 2851, Bacteriophage 2851 roi gene, ORF2, ORF3, Q gne; *E. coli* O91:H21 Shiga toxin 2 variant d A subunity; Phage BP-4795; Stx1-converting phage phi-O153; Phage 258-320 lysogen T4, T3, T2 shiga toxin 2 subunit B; *E. coli* Shiga toxin 2 subunit B; *Shigella boydii* CDC 3083-94, Sb227, ABU83972, ED1a chromosome, S88 chromosome, complete genomes; Shigella dysenteriae Sd197, shiga toxin A gene, shiga toxin B gene, SapF gene	Hypothetical Protein

ECs3489	390 bps	+	+	+	+	+	+	Enterobacteria phage YYZ-2008; Bacteriophage CP-1639 and chromosomal integration site; *E. coli* ORF1, ORF2, ORF3 and rorf1 gene; Phage BP-4795; Enterobacteria phage BT1-Sakai genomic DNA; *E. coli* potB, trcA, ORF4; Enterobacteria phage 2851 and BT2phi_272; Stx2-converting phage 1717; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W; Phage cdtI DNA	Hypothetical Protein
ECs2716	270 bps	+	+	+	+	+	+	Enterobacteria phage YYZ-2008; Enterobacteria phage VT1-Sakai genomic DNA; *E. coli* ORF1, ORF2, ORF3, rorf1 gne; E.coli 0127:H6 E2348/69; Bacteriophage CP-1639 and chromosomal integration site; Phage BP-4795; Stx2-converting phage 1717; Enterobacteria phage 2851, complete genome; *E. coli* potB, trcA, ORF4 genes; Phage cdtI DNA; Enterobacteria phage BT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W	Hypothetical Protein
ECs1809	270 bps	+	+	+	+	+	+	Enterobacteria phage YYZ-2008; Enterobacteria phage VT1-Sakai genomic DNA; *E. coli* ORF1, ORF2, ORF3 and rorf1 gene; Phage BP-4795; Bacteriophage CP-1639 and chromosomal integration site; Stx2-converting phage 1717; Enterobacteria phage 2851; *E. coli* potB, trcA, ORF4 genes; Phage cdtI DNA; Enterobacteria phage VT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT1-Sakai genomic and proviral DNA; Bacteriophage 933W	Hypothetical Protein

ECs2230	270 bps	+	+	+	+	+	+	Enterobacteria phage YYZ-2008; Enterobacteria phage VT1-Sakai genomic DNA; *E. coli* ORF1, ORF2, ORF3 and rorf1 gene; Phage BP-4795; Bacteriophage CP-1639 and chromosomal integration site; Stx2-converting phage 1717; *E. coli* 0127:H6 E2348/69; Enterobacteria phage 2851; *E. coli* potB, trcA, ORF4 genes; Phage cdtI DNA; Enterobacteria phage VT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W	Hypothetical Protein
ECs2157	270 bps	+	+	+	+	+	+	Enterobacteria phage YYZ-2008; Enterobacteria phage VT1-Sakai genomic DNA; *E. coli* ORF1, ORF2, ORF3 and rorf1 gene; Phage BP-4795; Bacteriophage CP-1639 and chromosomal integration site; Stx2-converting phage 1717; *E. coli* 0127:H6 E2348/69; Enterobacteria phage 2851; *E. coli* potB, trcA, ORF4 genes; Phage cdtI DNA; Enterobacteria phage VT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W	Hypothetical Protein
ECs2940	309 bps	+	+	+	+	+	+	Enterobacteria phage YYZ-2008; Enterobacteria phage VT1-Sakai genomic DNA; *E. coli* ORF1, ORF2, ORF3 and rorf1 gene; Phage BP-4795; Bacteriophage CP-1639 and chromosomal integration site; Stx2-converting phage 1717; *E. coli* 0127:H6 E2348/69; Enterobacteria phage 2851; *E. coli* potB, trcA, ORF4 genes; Phage cdtI DNA; Enterobacteria phage VT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W	Hypothetical Protein

ECs1229	270 bps	+	+	+	+	+	+	Enterobacteria phage YYZ-2008; Enterobacteria phage VT1-Sakai genomic DNA; *E. coli* ORF1, ORF2, ORF3 and rorf1 gene; Phage BP-4795; Bacteriophage CP-1639 and chromosomal integration site; Stx2-converting phage 1717; *E. coli* 0127:H6 E2348/69; Enterobacteria phage 2851; *E. coli* potB, trcA, ORF4 genes; Phage cdtI DNA; Enterobacteria phage VT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W	Hypothetical Protein
ECs0845	309 bps	+	+	+	+	+	+	Enterobacteria phage YYZ-2008; Enterobacteria phage VT1-Sakai genomic DNA; *E. coli* ORF1, ORF2, ORF3 and rorf1 gene; Phage BP-4795; Bacteriophage CP-1639 and chromosomal integration site; Stx2-converting phage 1717; *E. coli* 0127:H6 E2348/69; Enterobacteria phage 2851; *E. coli* potB, trcA, ORF4 genes; Phage cdtI DNA; Enterobacteria phage VT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W	Hypothetical Protein
ECs1993	342 bps	+	+	+	+	+	+	Enterobacteria phage YYZ-2008; Enterobacteria phage VT1-Sakai genomic DNA; *E. coli* ORF1, ORF2, ORF3 and rorf1 gene; Phage BP-4795; Bacteriophage CP-1639 and chromosomal integration site; Stx2-converting phage 1717; *E. coli* 0127:H6 E2348/69; Enterobacteria phage 2851; *E. coli* potB, trcA, ORF4 genes; Phage cdtI DNA; Enterobacteria phage VT2phi_272; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W	Hypothetical Protein

ECs1124	270 bps	+	+	+	+	+	+	Stx2-converting phage 1717; Enterobacteria phage 2851; Bacteriophage CP-1639 and chromosomal integration site; Enterobacteria phage VT2phi_272; *E. coli* ORF1, ORF2, ORF3 and rorf1 gene; Enterobacteria phage Min27; Stx2 converting phage I and II DNA; Stx1 converting phage DNA; Enterobacteria phage VT2-Sakai genomic and proviral DNA; Bacteriophage 933W; *E. coli* 0127:H6 E2348/69; Enterobacteria phage VT1-Sakai genomic DNA; Phage BP-4795; Enterobacteria phage YYZ-2008; Phage cdtI DNA; *E. coli* potB, trcA, ORF4 genes	Hypothetical Protein
ECs3500	717 bps	+	+	+	+	+	-	*E. coli* 0127:H6 E2348/69 complete genome; Shigella dysenteriae Sd197 plasmid pSD1_197; *Shigella boydii* CDC 3083-94 plasmid pBS512_211; Shigella flexneri plasmid pINV_F6_M1382 ORF137 pseudogene; Shigella flexneri virulence plasmid pWR100; Shigella flexneri 5a plasmid virulence plasmid pWR501; Shigella flexneri 2002017 plasmid pSFxv_1; Shigella flexneri 2a str. 301 virulence plasmid pCP301	Hypothetical Protein
ECs0816	717 bps	+	+	+	+	+	-	*E. coli* 0127:H6 E2348/69 complete genome; Shigella dysenteriae Sd197 plasmid pSD1_197; *Shigella boydii* CDC 3083-94 plasmid pBS512_211; Shigella flexneri plasmid pINV_F6_M1382 ORF137 pseudogene; Shigella flexneri virulence plasmid pWR100; Shigella flexneri 5a plasmid virulence plasmid pWR501; Shigella flexneri 2002017 plasmid pSFxv_1; Shigella flexneri 2a str. 301 virulence plasmid pCP301	Hypothetical Protein
ECs4567	378 bps	+	+	+	+	+	+	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* locus of enterocyte effacement (LEE) strains; *E. coli* strain 33264, EDS-58 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* 0127:H6 E2348/69 complete genome and pathogenicity island; *E. coli* pathogenicity island I; *E. coli* CadC gene, partial cds; transposase 2; *E. coli* Ler, EscR, EscS, EscT	Hypothetical Protein

ECs4579	459 bps	+	+	+	+	+	-	*E. coli* strain 71074 enterocyte effacement pathogenicity..; *E. coli* locus of enterocyte effacement (LEE) strains; *E. coli* strain 33264, EDS-58 enterocyte effacement gene locus; *E. coli* L0001 (yicJ) gene, partial cds; *E. coli* 0127:H6 E2348/69 complete genome and pathogenicity island; *E. coli* pathogenicity island I; *E. coli* CadC gene, partial cds; transposase 2; *E. coli* Ler, EscR, EscS, EscT; *E. coli* DNA for locus of enterocyte effacement II; *E. coli* locus of enterocyte effacement, strain 3431-4/86	Hypothetical Protein
ECs5637	No Hit	-	-	-	-	-	-	None	Unknown
ECs5627	No Hit	-	-	-	-	-	-	None	Unknown
ECs5625	No Hit	-	-	-	-	-	-	None	Unknown
ECs1960	168 bps	+	+	+	+	+	-	*E. coli* ABU 83972; *E. coli* ED1a chromosome; *E. coli* 0127:H6 E 2348/69, complete genome	Hypothetical Protein
ECs2748	168 bps	+	+	+	+	+	-	*E. coli* ABU 83972; *E. coli* ED1a chromosome	Hypothetical Protein
ECs5614	No Hit	-	-	-	-	-	-	None	Unknown
ECsL90	No Hit	-	-	-	-	-	-	None	Unknown

Table S2: Genetic Screening Results for Clinical and Non-Clinical STEC strains

Strain	*stx1*	*stx2*	ECs848	ECs1322	ECs1323	ECs1326	ECs1561[a]	ECs1568	ECs2226	ECs3857	ECs3858	ECs4552	ECs4553	ECs4557
7.3964	+	-	+	+	+	+	+	+	+	+	+	+	+	-
81.0211	+	-	+	+	+	+	-	+	+	+	+	+	+	-
0.1302	+	-	+	+	+	+	+	+	+	+	+	+	+	-
82.0219	+	+	+	+	+	+	-	+	+	+	+	+	+	-
95.1144	+	-	+	+	+	+	+	+	+	+	+	+	+	+[b]
6.1592	+	+	+	+	+	+	+	+	+	+	+	+	+	+
8.0176	+	-	+	+	-	+	+	+	+	+	+	+	+	-
5.2217	+	-	+	-	-	-	+	-	+	+	+	+	+	-
93.0494	+	-	+	+	+	+	+	+	+	+	+	+	+	-
99.0849	+	-	+	+	+	+	+	+	+	+	+	+	+	-
99.0850	+	-	+	+	+	+	+	+	+	+	+	+	+	-
99.0869	+	-	+	+	+[b]	+	+	+	+	+	+	+	+	-
99.1761	+	-	+	+	-	+	+	+	+	+	+	+	+	-
99.1773	+	-	+	+	+[b]	+	+	+	+	+	+	+	-	-
10.2529	+	-	+	+	+	+	+	-	-	+	+	+	+[b]	-
90.0107	+	-	+	+	+	+	-	-	+	+	+	+	-	-
90.1764	+	-	+	+	+	+	+	+	+	+	+	+	+	-
0.1623	+	-	+	+	+	+	+	+	+	+	+	+	+	-
10.0941	+	-	+	+	+	+	+	-	+	+	-	+	+	-
87.1368	+	+	+	+	+	+	+	+	-	+	-	+	+	+
3.2605	+	-	+	+	+	+	-	-	+	+	+	+	+	-
9.0108	+	-	-	-	-	-	+	-	+	+	+	-	+	-
93.1685	+	-	+	-	-	-	+	-	+	+	+	+	+	-
99.1791	+	-	+	-	-	-	+	-	+	+	-	+	+	-
99.1792	+	-	+	+	+	+	+	-	+	+	+	+	+	-
99.1806	+	+	+	+	-	+[b]	+	+	+	+	+	-	+	-
99.1822	+	+	+	+	-	+	+	+	+	+	+	+	+	-
0.2981	+	-	+	+	+	+	+	+	+	+	+	-	+	-
4.0005	+	+	+	+	+	+	+	+	+	+	+	+	+	-
8.0288	+	-	+	+	+	+	+	+	+	+	+	+	+	-
87.1377	+	-	+	+	+	+	+	+	+	+	+	-	+	-
10.0815	+	+	+	+	+	+	+	+	+	+	+	+	+	-
7.1639	+	+	+	+	+	+	+	+	+	+	+	+	+	-
11.0247	+	-	+	+	+	+	+	+	+	+	+	+	+	-

11.0284	+	-	+	+	+	+	+	+	+	+	+	+	+	-
96.1530	+	-	+	+	+	+	+	+	+	+	+	+	+	-
99.1769	+	-	+	+	+	+	+	+	+	+	+	+	+	-
93.0525	+	+	+	+	+	+	+	+	+	+	+	+	+	-
93.1707	+	+	+	+	+	+	+	+	+	+	+	+	+	-
6.1194	+	+	+	+	+	+	+	+	+	+	+	+	+	+
6.1195	+	+	+	+	+	+	+	+	-	+	+	+	+	+
9.0538	+	-	-	+	+	+	+	+	+	+	+	+	+	+
5.0959	-	+	+	+	+	+	+	+	+	+	+	+	+	-
7.1636	-	+	+	-	-	-	+	+	+	+	+	+	+	-
1.2805	+	+	-	-	-	-	-	-	-	-	-	-	-	-
2.0665	+	+	+	+	+	+	+	+	+	+	+	+	+	-
96.1898	+	+	-	-	-	-	+	-	-	-	-	-	-	-
96.0876	+	+	-	-	-	-	-	-	-	-	-	-	-	-
1.2265	+	+	-	-	-	-	-	-	-	-	-	-	-	-
97.1571	+	-	-	-	-	-	-	-	-	-	-	-	-	-
10.1083	+	-	-	-	-	-	-	-	-	-	-	-	-	-
10.0939	+	-	-	-	-	-	-	-	-	+[b]	-	-	-	-
96.1534	-	+	+	+	+	+	-	+	+	+	+	+	+	-
96.0611	-	+	+	-	-	-	+	+	+	+	+	+	+	-
95.3644	+	+	+	+	+	+	-	+	+	+	+	+	+	-
93.0602	+	-	+	+	+	+	+	+	+	+	-	+	+	-
95.4080	+	-	+	+	+	+	+	+	+	+	+	+	+	-

[a] The expected DNA product size for this gene was 284 bps, however, all strains produced products of 326 bps in length with the exception of strain 96.1898, which was 294 bps in length

[b] Strain gave a positive result for presence of gene but band on DNA gel was weak

Table S3: Additional Genetic Screening Results for 5 STEC Strains

Strain	Stx1	Stx2	ECs848	ECs1322	ECs1323	ECs1326	ECs1561	ECs1568	ECs2226	ECs3857	ECs3858	ECs4552	ECs4553	ECs4557
93.0626	+	-		-	-	-	+	-	+	+	+	+	+	-
99.1807	+	-	+		-	-	+	+	+	+	+	+		-
96.1529	+	-	-	+	+	+	+	+	+	+	+	+		-
6.1192	+	-	+	+		+	+	+	+	+	+	+		-
0.2732	-	+	-	-		-	-	+	-	-	-	-		-

Emily A. McIntyre

eam5221@gmail.com

EDUCATION

The Pennsylvania State University University Park, PA
Bachelor of Science; Biology, Vertebrate Physiology Anticipated 05/2012
Schreyer Honors College

Honors Thesis: Identification of Virulence Biomarkers in Pathogenic Shiga Toxin-Producing *Escherichia coli*

EXPERIENCE

Penn State University Department of Food Science University Park, PA
Undergraduate Researcher *09/2012-Present*

- Use PCR, gel electrophoresis, and aseptic technique to screen Shiga toxin-producing *E. coli* for genes that may contribute to virulence
- Analyze data using GenBank, BLAST, and Excel

Family Dentistry, Dr. Kent H. Landin, D.M.D. Warren, PA
Volunteer *05/2010-Present*

- Observe and assist with restorative procedures, and crowns, dentures and partials fittings and placements
- Clean equipment, prepare rooms, develop x-rays, and communicate with patients and staff

HONORS AND AWARDS

- Golden Key International Honour Society *2010-Present*
- Behrend Honors Program, Penn State Erie *2008-2010*
- NSF STEM Scholarship Recipient *2008-2010*
- Freshman of the Year in Chemistry, Penn State Erie *04/2009*
- NETS SCITECH Scholarship Recipient *2011-2012*

ACTIVITIES

- Pre-Dental Society, Penn State University *08/2010-Present*
- Operation Smile, Penn State University Chapter *04/2011-Present*
- Intramural Sports, Penn State University *08/2008-Present*
- Red Cross Club Volunteer, Penn State University *11/2010-Present*