Distribution of Pathogenicity Islands OI-122, OI-43/48, OI-57 and High Pathogenicity Island (HPI) in Shiga Toxin-producing Escherichia coli

Wenting Ju¹, Jinling Shen¹,³, Magaly Toro¹, Shaohua Zhao⁴, and Jianghong Meng¹,²*

¹Department of Nutrition and Food Science, and ²Joint Institute of Food Safety and Applied Nutrition, and, University of Maryland College Park, MD 20740, USA
³College of Food Science and Engineering, Northwest A&F University, Yangling, Shaanxi Province 712100, China
⁴Division of Animal and Food Microbiology, Office of Research, Center for Veterinary Medicine, Food and Drug Administration, Laurel, MD 20708, USA

* Corresponding author. Present address: 0112 Skinner Building, Department of Nutrition and Food Science, University of Maryland, College Park, MD 20742, USA. Tel.: +1 301 405 8382; fax: +1 301 314 3313. E-mail address: jmeng@umd.edu.

Key words: STEC, PAIs, eae, iha.
ABSTRACT

Pathogenicity islands (PAIs) play an important role in STEC pathogenicity. The distribution of PAIs OI-122, OI-43/48, OI-57 and high pathogenicity island (HPI) were determined among 98 STEC strains assigned to seropathotypes (SPT) A to E. PCR and PCR-RFLP assays were used to identify 14 virulence genes that belonged to the four PAIs, and to subtype eae and stx genes, respectively. Phylogenetic trees were constructed based on the sequences of pagC among 34 STEC strains and iha among 67 diverse pathogenic E. coli, respectively. Statistical analysis demonstrated that the prevalence of OI-122 (55.82%) and OI-57 (82.35%) was significantly greater in seropathotypes (SPT A, B and C) that are frequently associated with severe disease than in other seropathotypes. terC (62.5%) and ureC (62.5%) in OI-43/48 were also significantly more prevalent in SPT A, B and C than in SPT D and E. Additionally, OI-122, OI-57, OI-43/48 and their associated virulence genes (except iha) were found to be primarily associated with eae-positive STEC, whereas HPI occurred independently of the eae presence. The strong association of OI-122, OI-43/48 and OI-57 with eae-positive STEC suggests in part different pathogenic mechanisms exist between eae-positive and eae-negative STEC. Virulence genes in PAIs that are associated with severe diseases can be used as potential markers to aid in identifying highly virulent STEC.
Shiga toxin-producing *Escherichia coli* can cause human illnesses ranging from self-limiting diarrhea to life-threatening diseases such as hemolytic uremic syndrome (HUS), a leading cause of kidney failure in children (1). *E. coli* O157:H7 is the single serotype that causes most STEC outbreaks and HUS cases. Like O157, non-O157 STEC can also cause severe diseases and foodborne outbreaks (1). More than 470 non-O157 STEC serotypes have been associated with human illness (2) and the public health concerns of non-O157 STEC are increasing (1). Estimations indicate that non-O157 STEC cause 112,752 illness each year in the United States, almost double the number of O157:H7 illnesses (63,153) (3). While some non-O157 STEC have been associated with disease symptoms indistinguishable from O157:H7, not all STEC can cause HUS and outbreaks, and some STEC serotypes have never been reported related to any human illness (4). The scientific basis for this difference, however, is poorly understood. Increasing evidence shows that differences in virulence between pathogenic and nonpathogenic bacterial strains can be attributed in part to virulence genes located in pathogenicity islands (PAIs) (5). PAIs usually contain blocks of virulence genes and are greater than 10 Kb (6). Several PAIs have been identified and characterized in STEC. A chromosomal pathogenicity island termed the locus of enterocyte effacement (LEE) was identified in *E. coli* O157:H7 strain EDL933, which encodes a type III secretion system (TTSS) as well as virulence genes (*eae* and *tir*) associated with the intimate attachment of bacteria to intestinal epithelial cells (4). LEE appears to confer enhanced virulence, since LEE-positive STEC are much more commonly associated with HUS and outbreaks than LEE-negative STEC (5). However, some LEE-positive STEC serotypes have never been associated with disease, and some LEE-negative STEC can
cause HUS and outbreaks, indicating that virulence factors other than those in LEE may contribute to pathogenesis of STEC (5).

Pathogenicity island OI-122 is also well characterized in O157:H7 (5, 7). OI-122 is a 23 kb PAI consisting of three modules (5, 8, 9). Z4321 is located in module one and encodes a protein sharing 46% similarity with the phoP-activated gene C (pagC) of Salmonella Typhimurium (5, 9). Z4326, Z4328, and Z4329 are located in module two. Z4326 (sen) encodes a protein that shares 38.2% similarity to Shigella flexneri enterotoxin 2 (5), whereas Z4328 and Z4329 encode proteins that have 89% and 86% similarity to non-LEE encoded effectors NleB and NleE, respectively (9). The enterohemorrhagic E. coli factor for adherence (Efa), which is involved in epithelial cell adhesion and inhibiting the proliferation of bovine peripheral blood lymphocytes, is located in module three (10).

OI-43 and OI-48 are duplicate genomic islands found in EDL933 (8). OI-43/48 genes are divided into three functional groups: a seven-gene cluster ureDABCEFG that encodes urease and accessory proteins hydrolyzing urea to ammonia and carbon dioxide; telluride resistance genes terZABCDEF (11); and two putative adhesion genes, iha (iron-regulated gene A) and aidA-1 (autotransporter adhesin involved in diffuse adherence) (12).

In EDL933, OI-57 contains non-LEE encoded effector genes nleG2-3, nleG6-2, and nleG 5-2 (13, 14). NleG proteins are E3 ubiquitin ligases analogous to RING finger and U-box enzymes in eukaryotes. Although the exact functions of NleG2-3, NleG6-2, and NleG 5-2 are still unclear, similar proteins have been identified as effectors that suppress immune response from the host (15).
High pathogenicity island (HPI) was first detected in *Yersinia pestis* and other highly virulent *Yersinia* species, and encodes a siderophore (yersiniabactin) mediated iron-uptake system (16). HPI is required for full virulence expression in *Yersinia* (16), and contains two main virulence genes, *fyuA* and *irp2*. FyuA is an outer membrane protein acting as a receptor for ferric-yersiniabactin uptake and for bacteriocin pesticin, while Irp2 is involved in yersiniabactin synthesis (17). An orthologous and highly conserved HPI is widely distributed among different species and genera of the family *Enterobacteriaceae* (16).

Few studies have investigated PAIs other than LEE in STEC to date. Since PAIs are normally absent in non-pathogenic strains of the same or closely related species, they may serve as useful markers to distinguish highly virulent from less virulent or harmless strains (5, 6). In addition, PAIs can be used to identify new and emerging pathogenic bacteria. In this study, we reported the distribution of OI-122, OI-43/48, OI-57, and HPI and their virulence genes in STEC, and evaluated the association of the PAIs and individual virulence genes with STEC seropathotypes linked to severe diseases and outbreaks. In addition, the association of the four PAIs with LEE was determined.
MATERIALS AND METHODS

Bacterial strains. A total of 98 STEC strains from humans, animals and food were used in this study (Table S1). Strains were classified into seropathotypes A to E according to the criteria described by Karmali et al. (5). The assignment of seropathotypes was based on published references (5, 14, 18) and a large online database on non-O157 STEC (http://www.lugo.usc.es/ecoli/SEROTIPOSHUM.htm).

stx and eae subtyping. stx and eae subtypes were determined using PCR-RFLP analysis (19, 20), and stx2dact was confirmed by PCR as previously described (21). Genomic DNA was extracted using boiling method as previously described (22, 23). STEC strains S1191 (stx2e), EDL933 (stx1a and stx2a), E32511 (stx2c), EH250 (stx2b), B2F1 (stx2dact) and N15018 (stx1c) were used as positive controls for the stx subtyping; STEC strains 86-24 (gamma 1), EDL933 (gamma 1), TW06584 (kappa), E2348-69 (alpha), TW07920 (epsilon), RDEC-1 (beta), TW10366 (rho), TW03501 (iota), TW07892 (eta), and TW01387 (gamma 2/theta) were used as positive controls for the eae subtyping. E. coli K12 was used as a negative control strain for both stx and eae subtyping.

Presence of OI-122, OI-43/48, O-57 and HPI. PCR assays were used to determine the presence of 14 virulence genes in STEC OI-122, OI-43/48, O-57 and HPI as described (5, 11, 13, 24, 25). The presence of a PAI was determined by several marker genes located in different regions of the island, including pagC, sen, nleB, efa-1, and efa-2 for OI-122; terC, ureC, iha and aidA-1 for OI-43/48; nle2-3, nleG6-2 and nleG5-2 for OI-57; and irp2 and fyuA for HPI. PCR was performed in a 25 µl reaction mixture, containing 2 µl of DNA template, 2.5 µl of 10x PCR buffer, 2 µl of 25 mmol l⁻¹ MgCl₂, 2 µl of 1.25 mmol l⁻¹ dNTP mix, 0.125 µl of 5 U µl⁻¹ AmpliTaq Gold DNA polymerase mix (Applied Biosystems, Branchburg, NJ) and 0.2 µl of 50
pmol µl⁻¹ of each primer. *E. coli* O157:H7 EDL 933 was used as a positive control for the virulence genes of OI-122, OI-43/48 and OI-57, and *E. coli* O26:H11 SJ-13 for virulence genes of HPI. *E. coli* K12 was used as a negative control for all PCR assays.

**Phylogenetic and sequences analysis.** *iha* and *pagC* were the only two genes that were highly prevalent in both *eae*-positive and *eae*-negative STEC. To determine the evolutionary relationship between the two groups of STEC, *iha* and *pagC* were selected for phylogenetic analysis studies. *iha* sequences from 67 *E. coli* and *Shigella* were obtained from the GenBank. A multiple sequence alignment of *iha* was performed using ClustalW in MEGA 5.05, and a maximum likelihood phylogenetic tree was generated using the General Time Reversible model (26). A bootstrapping of 2,000 replicates was used to estimate the confidence of the branching patterns of the phylogenetic tree using *iha* of *E. coli* SMS-3-5 as the phylogenetic tree’s root.

Additionally, PCR was used to amplify *pagC* of OI-122 from 12 selected STEC strains representing different serotypes as described by Konczy et al. (9). PCR products were sequenced by GeneWiz (Germantown, MD). Twenty two *pagC* sequences representing different STEC serotypes and one *Citrobacter* were downloaded from the GenBank. The *pagC* sequences were cropped to 446 bp prior alignment. Phylogenetic analysis was performed using ClustalW within MEGA 5.05 (26). A phylogenetic tree based on *pagC* sequences was constructed using maximum likelihood methods by MEGA 5.05 with bootstrapping of 2,000 replicates using *pagC* of *Citrobacter rodentium* IC168 as the tree’s root.

**Statistical analysis.** Chi square or Fisher’s exact test was used for data analysis using SAS9.2 (SAS Institute, Cary, N.C.). A P-value of < 0.01 was considered statistically significant.
RESULTS

Distribution of OI-122, OI-43/48, OI-57 and HPI in STEC. The 98 STEC strains were classified into seropathotypes A to E (Table S1). Overall, the prevalence of OI-122 and OI-57 decreased progressively from seropathotype A (SPT A) to seropathotype E (SPT E) (Fig. S1). The prevalence of OI-122 and OI-57 was significantly higher in seropathotypes associated with severe diseases (SPT A, B, and C) and outbreaks (SPT A and B) than in other seropathotypes (P<0.0001) (Table 1). Although the prevalence of OI-43/48 was greater in seropathotypes associated with HUS (SPT A, B and C) and outbreaks (SPT A and B) than in other seropathotypes, the difference was not statistically significant (P= 0.1356 and 0.02, respectively).

HPI was not found in SPT A (O157), but was almost evenly distributed from SPT B to SPT E (Fig. S1).

ten, nleB, efa-1, efa-2, terC, ureC, nleG2-1, nleG5-2 and nleG6-2 were significantly more prevalent in SPT A and B than in SPT C, D and E (Table 1). pagC, sen, nleB, efa-1, efa-2, terC, ureC, nleG2-1, nleG5-2 and nleG6-2 were statistically more prevalent in SPT A, B and C than in SPT D and E (Table 1). Although aidA-1 was more prevalent in SPT A and B than in SPT C, D and E, the difference was not statistically significant (P=0.27). iha, fyuA, and irp2 were less prevalent in SPT A, B and C than in SPT D and E, but the differences were not statistically significant.

Distribution of OI-122, OI-43/48, OI-57 and HPI in EHEC O157. PAIs showed three patterns of distribution in EHEC O157 (Table S1). As to β-glucuronidase (GUD)-negative O157:H7, four strains all contained marker genes for OI-122, OI-57, and OI-43/48. In GUD-positive O157: H7,
none of the five strains carried efa-1 and efa-2 (located at the third module of OI-122) or aidA-1 (located at the end of OI-43/48). Sorbitol fermenting O157:NM strains contained all virulence genes of OI-122 and OI-57 but were negative for all the OI-43/48 virulence marker genes, indicating the absence of OI-43/48 in O157:NM. Additionally, none of the O157:H7 and O157:NM strains were positive of HPI virulence genes.

**Association of OI-122, OI-43/48, OI-57 and HPI with eae.** We compared the distribution of virulence genes of OI-122, OI-43/48, OI-57, and HPI between eae-positive and eae-negative STEC. All virulence genes of OI-122 and OI-57 (pagC, sen, nleB, efa-1, efa-2, nleG2-3, nleG5-2 and nleG6-2) were highly prevalent in eae-positive strains (Table 1 and Table S1). However, these genes, with the exception of pagC, were less prevalent in eae-negative STEC (Table 1 and Table S1). Although 38.6% of eae-negative STEC strains were positive for pagC, its prevalence was significantly higher (64.8%) in eae-positive STEC (P=0.005). There was no apparent physical or functional relationship identified between OI-43/48 and LEE, but three OI-43/48 virulence genes (ureC, terC and aidA-1) were mainly associated with the presence of eae (P<0.0001). On the other hand, iha was more prevalent in eae-negative than in eae-positive STEC strains (P=0.007). As for HPI, there were no significant differences in the distribution of fyuA or irp2 between eae-positive and eae-negative STEC (P=0.36).

**Phylogenetic analysis of iha from diverse pathogenic E. coli.** A phylogenetic tree based on iha separated eae-positive and eae-negative STEC strains into two distinct clades (Fig. 1). In clade I, two subgroups, Ia and Ib, shared at least 98.0% sequence similarity. eae-positive EHEC
serotypes highly associated with outbreaks and severe diseases were located in clade I (O157:H7, O26:H11, O103:H2, O111:NM, O145:H28). Those sequences shared at least 99% similarity and clustered together with iha from other pathogenic *E. coli* including enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), enterotoxigenic *E. coli* (ETEC), uropathogenic *E. coli* (UPEC), neonatal meningitis *E. coli* (NMEC), and Shiga toxin-producing EAEC O104:H4 (from a German outbreak in 2011). *iha* sequences from the O26:H11, O111:H11, and O111:NM strains formed subgroup Ib and shared at least 98.0% sequence similarity with subgroup Ia. Interestingly, strains DEC10A (O26:H11), DEC10C (O26:H11), 11368 (O26:H11) and DEC8C (O111:NM) carried two *iha* that clustered separately in Ia and Ib.

All 15 *iha* sequences from *eae*-negative STEC clustered together to form clade II. Multiple sequence alignments demonstrated that *iha* from *eae*-negative STEC shared only 91.1-93.6% sequence similarity with *iha* from clade I. *iha* from subgroups IIa and IIb shared only 93.8-94.3% sequence similarity. As in some *eae*-positive strains, *eae*-negative STEC strains CL-3 (O113:H21), 96.0497 (O91:H21) and B2F1 (O91:H21) also carried two *iha* genes that clustered separately in subgroups IIa and IIb.

**Phylogenetic and sequence analysis of pagC.** The *pagC* phylogenetic tree showed four clades (Fig. 2). The *eae*-positive *E. coli* STEC formed a single clade with EPEC and ETEC, whereas the *eae*-negative STEC strains formed two clades, along with one strain that clustered with a *Citrobacter rodentium* strain. We identified 15 single nucleotide polymorphisms (SNPs) and one indel among the 35 *pagC* sequences. Sequence analysis revealed that an insertion of adenine at...
nucleotide 388 in two O103:H25, two O45:H2, one O103:H2 and one O103:H6 strains led to a frame shift mutation, and that a premature stop codon truncated the protein at the third loop resulting in the loss of the fourth and last loops.
DISCUSSION

In the present study, PAIs OI-122 and OI-57 of STEC were found highly associated with seropathotypes that can cause severe disease and outbreaks, as previously demonstrated (7, 13, 14). Several OI-122 virulence factors play important roles in bacterial pathogenesis. For example, PagC can promote the survival of *Salmonella* within macrophages (5, 9). Efa is an adhesion protein originally described in some EHEC strains (27). The *efa-1* gene is almost identical to *lifA*, an EPEC gene encoding lymphostatin (LifA) (28), which inhibits the proliferation of mitogen-activated lymphocytes and the synthesis of proinflammatory cytokines (28). Efa1/LifA also contributes to EPEC adherence to epithelial cells and is critical for the intestinal colonization by *C. rodentium* (29). NleB was required for full colonization and colonic hyperplasia in mice and a mutation of *nleB* abolished lethality of *C. rodentium* in C3H/HeJ mice (7, 30).

Whereas OI-122 is highly related to colonization and suppression of the host immune system, the function of OI-57 is largely unknown. Wu *et al.* (15) determined that NleG like proteins and U-box enzymes in eukaryotes. Although the targets of the OI-57 Nle effectors are unknown, several similar effectors are primarily involved in suppressing host immune response by degrading immune-related host proteins (15). Thus, it is possible that OI-57, similar to OI-122, would be also related to suppression of host immune system.

In addition to the virulence genes in OI-122 and OI-57, *ureC* and *terC* located on OI-43/48 were also highly associated with seropathotypes related to severe disease and outbreaks. Urease has been confirmed as an important virulence factor in several bacterial species, such as *Helicobacter pylori, Yersinia enterocolitica, Proteus mirabilis, Brucella* species, and *Klebsiella pneumonia* (31). Mutation of *ureC* has led to a reduced adherence of EHEC O157:H7 in ligated...
pig intestine (12). A recent study by Steyert et al. (32) revealed that strains with non-functional urease were two times unlikely to survive passage through the stomach and had a reduced ability to colonize the mouse intestinal tract compared with urease-positive strains. These data demonstrate that urease can help STEC strains survive in the stomach and enhance its competitiveness in colonization in calf and human intestinal tracts. The role of tellurite resistance genes (terZABCDEF) in STEC is still not well understood. Yin et al. (12) showed that mutation of the ter cluster in O157:H7 led to fewer adherence to epithelial cells and smaller bacterial clusters compared with wild-type strains. Therefore, ter genes might encode an adhesin or a gene product that promotes the function of adhesion(s). In addition, tellurite salts are strong oxidative agents, and it is possible that ter genes might offer a selective advantage in the host environment and aid STEC in general stress response (12).

Interestingly, ureC has been more frequently found in eae-positive STEC (113/132) than in eae-negative strains (4/70) although no physical linkage of ureC and eae has been identified (33). The prevalence of ureC in eae-positive STEC (45/55) was significantly higher than in eae-negative STEC (2/44) (P < 0.0001). Similarly, terC was also more prevalent in eae-positive STEC (45/55) than eae-negative strains (5/44) (P < 0.0001). Even though OI-43/48 and LEE are physically distant, our observations indicated that there might be a functional relationship between them.

The arrangement of OI-122 genes was found to be serotype dependent and all O157:H7 strains have a complete OI-122 (5, 9). However, we found that two patterns of OI-122 existed in O157:H7. An incomplete OI-122 lacking the third module was identified in all GUD-positive O157:H7 strains. Additionally, aidA-1 of OI-43/48 was absent in GUD-positive O157:H7.
Most OI-122, OI-43/48 and OI-57 virulence genes (pagC, sen, nleB, efa-1, efa-2, terC, ureC, iha, aidA-1, nleG2-3, nleG6-2, and nleG5-2) were highly prevalent in eae-positive STEC. However, they were largely absent in eae-negative STEC with the exception of pagC and iha. Phylogenetic analysis revealed that iha genes from eae-positive STEC had high similarity (99.6%), whereas they had lower sequence similarity (91.1-93.6%) with iha genes from eae-negative STEC, indicating that iha from eae-positive and eae-negative STEC may have evolved independently or have different origins. Such a difference also existed in pagC between eae-positive and eae-negative STEC. Schmidt et al. (34) reported that iha was carried by a 33,014 bp PAI in STEC serotype O91:H- strains (eae-negative). In addition, iha was found in pO113 plasmid of STEC serotype O113:H21 (eae-negative) (35). Moreover, Shen et al. (36, 37) reported that pagC was identified within a mosaic PAI from STEC0113:H21 strain CL-3 (eae-negative). Thus, the higher prevalence of iha and pagC in the eae-negative STEC strains, as compared with other virulence marker genes in this study, is likely due to the presence of the same or similar PAIs and/or plasmids as previously described. The similar prevalence of iha genes in the seropathotypes highly associated with severe diseases and other seropathotypes indicates that iha is not related with severe clinical outcomes, but the significantly higher prevalence of pagC in the seropathotypes associated with severe diseases indicates that this gene has some association with severe clinical outcome whether a strain carries the gene in OI-122 or in some other PAIs. The distribution of PAI virulence genes and the phylogenetic analysis of iha and pagC support the hypothesis that OI-122, OI-43/48 and OI-57 are primarily associated with eae-positive strains in STEC. However, some eae-negative STEC serotypes, for example, O113:H21 and O91:H21, are also associated with life threatening diseases such as HUS (5). Virulence factors such as subtilase cytotoxin AB5 (subAB5) and Saa (STEC autoagglutinating adhesion) are more
commonly associated with eae-negative STEC. Moreover, it has been shown that some LEE-negative STEC, especially O113:H21, can invade tissue culture cells (38). Whole genome comparison between nine eae-negative and five eae-positive STEC strains revealed that eae-negative strains did not carry any LEE or other phage encoded non-LEE effectors (39). These observations indicate that some differences in pathogenesis mechanisms may exist between eae-positive and eae-negative STEC. Additional studies, especially genomics and proteomics, are needed to unveil the difference in the pathogenicity mechanisms between eae-negative and eae-positive STEC.

The strong association of O1-122, O1-57 and O1-43 with eae-positive STEC offers an important basis for STEC molecular risk assessment (MRA). The MRA, which uses 14 non-LEE encoded virulence factors to distinguish high risk from low risk non-O157 STEC, was proposed by Coombes et al. in 2008 (13). Other researchers adopted this concept and applied it to their own studies (40-43). However, the current work demonstrated that some of non-LEE encoded effectors (nleB, nle2-3, nleG5-2 and nleG6-2) were primarily associated with eae-positive STEC strains. In addition, Mundy et al. (44) reported that nleA was present in 37 out of 43 (86%) eae-positive STEC, but absent in 50 eae-negative STEC clinical strains. Konczy et al. (9) reported that nleB and nleE of O1-122 were highly correlated with LEE. Moreover, comparative genomics analysis demonstrated that all known phage encoded non-LEE effector genes were absent in eae-negative STEC (39). Based on the MRA framework, which uses non-LEE effector genes as sole markers, all eae-negative virulence STEC, including HUS associated O113:H21, O91:H21 and O104:H21, would be grouped as harmless STEC; other serotypes, for example, O103:H11 and O119:H25, which have not been reported as associated with severe disease or outbreaks but carry similar non-LEE encoded virulence effectors as O157 EHEC, would be considered as
outbreak- and severe disease-associated serotypes. Therefore, additional markers or methods of assessment, especially for eae-negative STEC, are needed to accurately distinguish highly pathogenic STEC from low virulent or harmless STEC.

In summary, O-122 and OI-57, and their virulence genes were highly associated with seropathotypes that cause severe diseases and outbreaks. In addition, ureC and terC located at OI-43/48 were also identified as markers related to high risk seropathotypes. Virulence genes in PAIs that are associated with severe diseases can be used as markers to identify potentially highly virulent STEC. Furthermore, this study demonstrated that OI-122, OI-43/48, and OI-57 are highly associated with eae-positive STEC, which offers an important basis for STEC MRA.
ACKNOWLEDGEMENTS

The study was supported in part by the Joint Institute for Food Safety & Applied Nutrition (JIFSAN), University of Maryland, College Park, Maryland. The authors thank Dr. Julie Kase FDA/CFSAN for offering DNA of positive controls for \textit{stx} and \textit{eae} subtyping.


10. Abu-Median AB, van Diemen PM, Dziva F, Vlisidou I, Wallis TS, Stevens MP.


11. Taylor DE, Rooker M, Keelan M, Ng LK, Martin I, Perna NT, Burland NT, Blattner FR.


13. Coombes BK, Wickham ME, Mascarenhas M, Gruenheid S, Finlay BB, Karmali MA.


integrated at selC in locus of enterocyte effacement-negative, Shiga toxin-producing


Table 1. Association of pathogenicity islands (PAIs) and its virulence genes with seropathotypes associated outbreak (SPT A and B), severe disease, (SPT A, B and C) and LEE.

<table>
<thead>
<tr>
<th>PAI Gene</th>
<th>Prevalence (%) related to outbreak in</th>
<th>Prevalence (%) in</th>
<th>Prevalence (%) in</th>
<th>Prevalence (%) in</th>
<th>Prevalence (%) in</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPT A and B (n=34)</td>
<td>SPT C, D and E (n=64)</td>
<td>STEC SPT A, B and C (n=56)</td>
<td>STEC in SPT D and E (n=42)</td>
<td>eae-positive STEC (n=54)</td>
</tr>
<tr>
<td>OI-122</td>
<td>55.82 * 17.18 * 46.43 * 9.52 *</td>
<td>55.56 * 0 *</td>
<td>55.56 * 0 *</td>
<td>55.56 * 0 *</td>
<td>55.56 * 0 *</td>
</tr>
<tr>
<td>pagC</td>
<td>70.59 * 46.88 * 69.64 * 35.71 *</td>
<td>64.81 * 38.64 *</td>
<td>64.81 * 38.64 *</td>
<td>64.81 * 38.64 *</td>
<td>64.81 * 38.64 *</td>
</tr>
<tr>
<td>sen</td>
<td>100.00 * 31.25 * 76.79 * 26.45 *</td>
<td>100.00 * 2.27 *</td>
<td>100.00 * 2.27 *</td>
<td>100.00 * 2.27 *</td>
<td>100.00 * 2.27 *</td>
</tr>
<tr>
<td>efa-1</td>
<td>82.35 * 31.25 * 66.07 * 30.95 *</td>
<td>88.89 * 6.82 *</td>
<td>88.89 * 6.82 *</td>
<td>88.89 * 6.82 *</td>
<td>88.89 * 6.82 *</td>
</tr>
<tr>
<td>efa-2</td>
<td>82.35 * 31.25 * 66.07 * 26.45 *</td>
<td>88.89 * 2.27 *</td>
<td>88.89 * 2.27 *</td>
<td>88.89 * 2.27 *</td>
<td>88.89 * 2.27 *</td>
</tr>
<tr>
<td>nleB</td>
<td>100.00 * 31.25 * 76.79 * 26.45 *</td>
<td>100.00 * 2.27 *</td>
<td>100.00 * 2.27 *</td>
<td>100.00 * 2.27 *</td>
<td>100.00 * 2.27 *</td>
</tr>
<tr>
<td>OI-43/48</td>
<td>32.35 * 12.50 * 26.79 * 14.28 *</td>
<td>37.03 * 2.27 *</td>
<td>37.03 * 2.27 *</td>
<td>37.03 * 2.27 *</td>
<td>37.03 * 2.27 *</td>
</tr>
<tr>
<td>terC</td>
<td>76.47 * 35.94 * 62.50 * 33.33 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
</tr>
<tr>
<td>ureC</td>
<td>76.47 * 35.94 * 62.50 * 33.33 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
</tr>
<tr>
<td>iha</td>
<td>64.70 * 70.31 * 64.29 * 73.80 *</td>
<td>57.41 * 84.09 *</td>
<td>57.41 * 84.09 *</td>
<td>57.41 * 84.09 *</td>
<td>57.41 * 84.09 *</td>
</tr>
<tr>
<td>OI-57</td>
<td>38.23 * 32.81 * 39.28 * 28.57 *</td>
<td>65.90 * 18.18 *</td>
<td>65.90 * 18.18 *</td>
<td>65.90 * 18.18 *</td>
<td>65.90 * 18.18 *</td>
</tr>
<tr>
<td>terC</td>
<td>76.47 * 35.94 * 62.50 * 33.33 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
</tr>
<tr>
<td>ureC</td>
<td>76.47 * 35.94 * 62.50 * 33.33 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
<td>81.48 * 11.36 *</td>
</tr>
<tr>
<td>iha</td>
<td>64.70 * 70.31 * 64.29 * 73.80 *</td>
<td>57.41 * 84.09 *</td>
<td>57.41 * 84.09 *</td>
<td>57.41 * 84.09 *</td>
<td>57.41 * 84.09 *</td>
</tr>
<tr>
<td>OI-57</td>
<td>38.23 * 32.81 * 39.28 * 28.57 *</td>
<td>65.90 * 18.18 *</td>
<td>65.90 * 18.18 *</td>
<td>65.90 * 18.18 *</td>
<td>65.90 * 18.18 *</td>
</tr>
<tr>
<td>nleG2-3</td>
<td>97.06 * 31.25 * 73.21 * 28.57 *</td>
<td>94.44 * 6.82 *</td>
<td>94.44 * 6.82 *</td>
<td>94.44 * 6.82 *</td>
<td>94.44 * 6.82 *</td>
</tr>
<tr>
<td>nleG5-2</td>
<td>82.35 * 21.88 * 60.71 * 19.04 *</td>
<td>85.19 * 2.27 *</td>
<td>85.19 * 2.27 *</td>
<td>85.19 * 2.27 *</td>
<td>85.19 * 2.27 *</td>
</tr>
<tr>
<td>nleG6-2</td>
<td>97.18 * 28.13 * 66.07 * 28.57 *</td>
<td>75.93 * 4.56 *</td>
<td>75.93 * 4.56 *</td>
<td>75.93 * 4.56 *</td>
<td>75.93 * 4.56 *</td>
</tr>
<tr>
<td>HPI</td>
<td>17.65 * 25.00 * 17.86 * 28.57 *</td>
<td>25.92 * 18.18 *</td>
<td>25.92 * 18.18 *</td>
<td>25.92 * 18.18 *</td>
<td>25.92 * 18.18 *</td>
</tr>
<tr>
<td>fyuA</td>
<td>17.65 * 23.44 * 16.07 * 28.57 *</td>
<td>25.92 * 18.18 *</td>
<td>25.92 * 18.18 *</td>
<td>25.92 * 18.18 *</td>
<td>25.92 * 18.18 *</td>
</tr>
<tr>
<td>ihaA1</td>
<td>17.65 * 23.44 * 16.07 * 28.57 *</td>
<td>25.92 * 18.18 *</td>
<td>25.92 * 18.18 *</td>
<td>25.92 * 18.18 *</td>
<td>25.92 * 18.18 *</td>
</tr>
</tbody>
</table>

* Statistical difference between SPT A and B with SPT C, D and E.
+: Statistical difference between SPT A, B and C with SPT D and E.

^: Statistical difference between eae-positive and eae-negative STEC.

A P-value <0.01 was considered as statistically significant.
Figure Legends

Fig. 1. Phylogenetic tree based on *iha* sequences from 67 *E. coli* and *Shigella* strains. *iha* sequences were aligned and a tree was constructed using the maximum likelihood method with 2,000 iterations utilizing MEGA 5.05 (26). *iha* sequences from *eae*-positive and *eae*-negative STEC strains segregated into two distinct clades, clades I (with subgroup Ia and Ib) and II (with subgroups IIa and IIb). *iha* sequences from *eae*-negative STEC were marked in bold and italic, and *eae*-positive STEC strains were marked in bold. EPEC: enteropathogenic *E. coli*; EIEC: enteroinvasive *E. coli*; EAEC: enteroaggregative *E. coli*; stx-producing EAEC: Shiga toxin-producing EAEC; ETEC: enterotoxigenic *E. coli*; UPEC: uropathogenic *E. coli*; NMEC: neonatal meningitis *E. coli*; and ABU: asymptomatic bacteriuria *E. coli*.

Fig. 2. Phylogenetic tree based on *pagC* sequences from 35 pathogenic *E. coli* strains. *pagC* sequences were aligned and a tree was constructed using the maximum likelihood method with 2,000 iterations utilizing MEGA 5.05 (26). *pagC* sequences from *eae*-negative STEC were marked in bold and italic, and *eae*-positive STEC strains were marked in bold. *pagC* genes sequenced by this study were marked by black frames. EPEC: enteropathogenic *E. coli*, ETEC: enterotoxigenic *E. coli*.