Virulence gene profiles and phylogeny of Shiga toxin-positive *Escherichia coli* strains isolated from FDA regulated foods during 2010-2017

Narjol Gonzalez-Escalona[1,*], and Julie Ann Kase[1]

[1]Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, MD, USA

Key words: Shiga toxin-producing *Escherichia choli*, Whole genome sequencing, virulence genes, phylogeny, foods

Running title: Whole genome sequencing of STECs FDA regulated foods

*Corresponding author. Mailing address: FDA, CFSAN, 5100 Paint Branch Parkway HFS-712, College Park, MD 20740. Phone: (240) 436-1937. Fax: (301) 436-2644. E-mail: narjol.gonzalez-escalona@fda.hhs.gov

Abstract:

Illnesses caused by Shiga toxin-producing *Escherichia* coli (STECs) can be life threatening, such as hemolytic uremic syndrome (HUS). The STECs most frequently identified by USDA's Microbiological Data Program (MDP) carried toxin gene subtypes *stx1a* and/or *stx2a*. Here we describe the genome sequences of 331 STECs isolated from foods regulated by the FDA 2010-2017, determining their genomic identity, serotype, sequence type, virulence potential, and prevalence of antimicrobial resistance. Isolates were selected from the MDP archive, routine food testing by field labs (ORA), food testing by a contract company, and our laboratory (ORS). Only 276 (83%) were confirmed as STECs by *in silico* analysis. Foods from which STECs were recovered included cilantro (6%), spinach (25%), lettuce (11%), and flour (9%). Phylogenetic analysis using core genome MLST revealed these STEC genomes were highly variable, with some clustering associated with ST types and serotypes. We detected 95 different sequence types (ST); several ST were previously associated with HUS: ST21 and ST29 (O26:H11), ST11 (O157:H7), ST33 (O91:H14), ST17 (O103:H2), and ST16 (O111:H-). *in silico* virulome analyses showed ~ 51% of these strains were potentially pathogenic [besides *stx* gene they also carried *eae* (25%) or 26% *subA* (26%)]. Virulence gene prevalence was also determined: *stx*1 only (19%) -variants a and c; *stx*2 only (66%) – variants a, b, c, d, e, and g; and *stx*1/*sxt*2 (15%). Our data form a new WGS database that can be used to support food safety investigations and monitor the recurrence/emergence of *E. coli* in foods.

47 **Importance**

48

49 Shiga toxin-producing *Escherichia* coli (STECs) are associated with foodborne outbreaks
50 worldwide; however, surveillance has not previously included genomic analyses for
51 phylogenetics, prevalence, or potential virulence. We constructed the first genomic database of
52 isolates from FDA-regulated foods to help monitor the emergence of new pathogenic STECs.
53 Although only ~30 STECs were isolated per year, 50% of these carried markers associated with
54 pathogenesis either a combination of *eae* plus *stx*, or *subA* plus *stx*. Moreover, those strains
55 also carried virulence genes associated with severe illnesses. Here we showed that WGS
56 enabled comparisons across isolates to establish phylogeny, help in identification of antibiotic
57 resistance by monitoring the presence of antimicrobial resistance genes, and determined the
58 presence of known virulence genes that have been linked with illnesses. Future food safety
59 investigations will benefit from improved source tracking and risk assessments made possible
60 by these analyses and new WGS database.

61

62

Introduction

Shiga toxin-producing *Escherichia* coli (STECs) have the potential to cause infections, from mild to life threatening outcomes such as hemolytic uremic syndrome (HUS). STECs causing HUS, hemorrhagic colitis and bloody diarrhea are known as enterohemorrhagic *E. coli* (EHEC). Among the most common EHECs are O157:H7, O26, O121, O103, O111, and O145. O157:H7 strains are responsible for most foodborne outbreaks in the last two decades (1) while non-O157 serogroups, O26, O121, O103, O111, and O145 are the second most common cause of EHEC foodborne infections in the US (2,3) and worldwide (4-7). Each year in the US, O157:H7 cause an approximately 95,000 cases with 2,150 hospitalizations, with non-O157 STECs responsible for an estimated 170,000 cases (3). These serotypes carry Shiga toxin genes (*stx1* and/or *stx2*) and there are at least 130 EHEC serotypes that have been recovered from human patients. The US Department of Agriculture Food Safety and Inspection Services (USDA FSIS) in 2011, declared O26 and five other non-O157 serogroups, O45, O103, O111, O121, and O145 as adulterants in ground beef and non-intact beef products, and in mid-2012 began testing for these pathogens in both domestic and imported beef trimmings (8).

In order to cause illness, STEC strains need a set of genes that allow them to attach, colonize, and produce and secrete Shiga toxin protein (9-12). Genes described for attachment and colonization include *eae* (intimin), other proteins present in the locus of enterocyte effacement (LEE) locus, T3SS effectors, as well as biofilm production, and other virulence genes that are usually located in a plasmid (e.g. *ehxA*), referred as the virulence plasmid to differentiate it from other possible plasmids that can be carried by the same strain as well (9,11,13). Although the precise role of *ehxA* in STEC pathogenesis remains to be elucidated, several studies indicate an association of *ehxA* in clinical disease since 1) *ehxA* was found to be produced by many STEC associated with diarrheal disease and HUS (14-16), and 2) serum samples from

89 HUS patients have been shown to react specifically to *ehxA* (17). Some STECs do not carry
90 *eae*, however they possess other genes believed to compensated for the lack of *eae* (e.g. *subA,*
91 *saa* or *sat*) (18,19).
92
93 STECs can be transmitted by various means with food remaining the predominant transmission
94 route (1). Among the illnesses caused by STECs in FDA regulated food products (FRFDA),
95 fresh produce has been implicated in several outbreaks, as well as some other atypical
96 commodities, such as flour (20). Leafy greens and other agricultural food crops are particularly
97 susceptible to contamination since they are grown in close contact with the ground where
98 runoff from livestock areas, particularly cattle, contaminated irrigation water, manure used as
99 fertilizer, and the intrusion of wildlife into growing fields can occur (21). Many of these same
100 items are consumed raw and possibly with little cleaning. Some noteworthy *E. coli* outbreaks
101 reported by the Center of Disease and Control (CDC) in the US in the last 10 years are: 2009 -
102 beef (O157:H7) and prepackaged cookie dough (O157:H7); 2010 cheese (O157:H7), romaine
103 lettuce (O145) and beef (O157:H7); 2011 - romaine lettuce (O157:H7), Lebanon bologna
104 (O157:H7), and in-shell hazelnuts (O157:H7); 2012 - spinach and spring mix blend (O157:H7),
105 unknown source (O145), and raw clover sprouts (O26); 2013 ready-to-eat salads (O157:H7),
106 and frozen food products (O121); 2014 - raw clover sprouts (O121), and ground beef
107 (O157:H7); 2015 - rotisserie chicken salad (O157:H7), and Mexican-style restaurant chain
108 (O26); 2016 - flour (O121 and O26), and alfalfa sprouts (O157); 2017 - leafy greens (O157:H7),
109 and soy nut butter (O157:H7); in this year, 2018, there has been an outbreak link to romaine
110 lettuce caused by O157:H7 (https://www.cdc.gov/ecoli/outbreaks.html).
111
112 Beyond the noted outbreaks, there have been several reports on STECs found in FRFDA
113 (22,23). The most comprehensive survey was the USDA Microbiological Data Program (MDP)
114 that collected domestic and imported fresh fruit and vegetable samples from primarily terminal

markets and wholesale distribution centers from 2001-2012 (https://www.ams.usda.gov/datasets/mdp/mdp-program-data-and-reports). This program tested approximately up to 15,000 samples annually, and tested for the presence of *Salmonella*, *E. coli O157:*H7, and other STECs. STEC were found most frequently from spinach samples (0.5%), and of the 132 STECs isolated, 9% were found to carry *eae*. The most prevalent Shiga toxin variants found were *stx*1a (22%) and/or *stx*2a (56%) (23). However, little other information about the genome content of those strains is publicly available.

Whole genome sequencing (WGS) technology is reshaping food safety and food-borne illness investigations (24). The use of WGS is becoming more useful as the cost of bacterial genome sequencing decreases every year. WGS cost per bacterial sequence is now comparable to PFGE. There are many attractive attributes with regards to the use of WGS for analyzing food samples including the potential to identify all pathogens present in that sample (25). Among other applications of the use of WGS are: it can help in identifying genes that allow for resistance/survival or virulence of certain bacterial strains (26-28), can help in establishing phylogenetic relationships among old strains of STECs isolated from either clinical cases or environmental samples (7,29,30), and can further help in Identifying matches between environmental and outbreak strains during outbreaks scenarios (26,30-32). Furthermore, using WGS can help in identifying matches among bacterial strains isolated from environmental samples in production facilities and may help locate contamination sources (33). It can also be extremely helpful in establishing mechanism of evolution among pathogens (34). For example the 2011 outbreak in Germany linked to fenugreek seeds caused by an *E. coli* strain with a genomic backbone and virulence traits of entero-aggregative *E. coli* (EAEC) but had acquired a *stx* phage (*stx*2a gene variant) and caused a more aggressive disease with high HUS rate cases (35,36). This event highlighted the high plasticity of the *E. coli* genomes and it constitutes a warning of the possible arise of more of this new “hybrid pathotype” strains.

Therefore, we wanted to further characterize and catalog historical strains of STECs isolated from FRFDA by performing WGS analysis of every STEC strains isolated by the MDP and other FDA surveillance programs, as well as, some FDA historical isolates. This work establishes is the first genomic database of FRFDA isolates which in turn, will allow improved surveillance for both recurrence and the emergence of new strains that might be impacting our food supply. A total of 296 presumptive STECs were isolated during 2010-2017, and 35 additional STECs were historical isolates from our collection. The 331 presumptive STEC strains were analyzed for virulence genes [encompassing all *E. coli* virulent types - STEC, entero-pathogenic *E. coli* (EPEC), entero-toxigenic *E. coli* (ETEC), entero-invasive *E. coli* (EIEC), and EAEC], *in silico* MLST, and antibiotic resistance genes. Finally, their phylogenetic relationships and diversity were determined by whole genome phylogeny analysis using an allele-based whole genome multilocus sequence analysis (MLST) or core genome MLST analysis (cgMLST).

Materials and Methods

Bacterial strains and media. *E. coli* (n = 331) presumptive Shiga toxin-positive strains used in this study are listed in supplementary table 1 (Table S1). Each strain was assigned a CFSAN number for future tracking. The FRFDA strains were isolated by us (n = 196), FDA Office of Regulatory Affairs (ORA) laboratories (n = 74), and a contracting lab (n = 63) during 2012-2017 in the US.

DNA preparation. Genomic DNA from each strain was isolated from overnight cultures using the DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA), following the manufacturer's instructions. The resultant DNA extract was stored at -20°C until used as a template for whole genome sequencing. The concentration was determined using a Qubit double-stranded DNA HS assay kit and a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA), according to manufacturer's instructions.

Whole genome sequencing, contig assembly and annotation. The genomes of the strains were sequenced, using an Illumina MiSeq sequencer (Illumina, San Diego, CA), with the 2x250 bp pair-end chemistry according to manufacturer's instructions, at approximately 80X average coverage. The genome libraries were constructed using the Nextera XT DNA sample prep kit (Illumina). Genomic sequence contigs were *de novo* assembled using default settings within CLC Genomics Workbench v9.5.2 (QIAGEN) with a minimum contig size threshold of 500 bp in length.

***in silico* serotyping.** The serotype of each strain analyzed in this study was confirmed using the genes deposited in the Center for Genomic Epidemiology (http://www.genomicepidemiology.org) for *E. coli* as part of their web-based serotyping tool

(SerotypeFinder 1.1 - https://cge.cbs.dtu.dk/services/SerotypeFinder) (37,37,38). We used Ridom for performing batch screening of the genomes analyzed. Briefly, all the genes were divided into O-type (*wzx* and *wzy*) and H-type (*fli*C) genes in FASTA format (ex. All *wzx* alleles were in a single FASTA file), and used as task template. For the virulence screening, a project was created using all three task templates and each whole genome sequence was screened for the presence of each gene type (O-type or H-type gene). Results were similar to SerotypeFinder, and as done for the virulence genes previously, the data was now in a database and new alleles (if found) could be added to the task templates.

***in silico* MLST phylogenetic analysis.** The initial analysis and identification of the strains were performed using an *in silico E. coli* MLST approach, based on the information available at the *E. coli* MLST website (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli) and using Ridom SeqSphere+ software v2.4.0 (Ridom; Münster, Germany) (http://www.ridom.com/seqsphere). Seven housekeeping genes (*dnaE, gyrB, recA, dtdS, pntA, pyrC,* and *tnaA*), described previously for *E. coli* (39), were used for MLST analysis. The same *E. coli* MLST database was also used to assign numbers for alleles and STs.

***in silico* determination of virulence genes.** Virulence genes were determined using the genes deposited in the Center for Genomic Epidemiology (http://www.genomicepidemiology.org) for *E. coli* as part of their VirulenceFinder 1.5 web-based tool (https://cge.cbs.dtu.dk/services/VirulenceFinder) (38), except that we used Ridom for performing batch screening of the genomes analyzed. Briefly, all the genes were divided into classes or groups by homology in FASTA format (e.g. All *ast*A alleles were in a single FASTA file), and used as a task template. Afterwards a project was created using all these task templates, and each WGS was screened for the presence of each gene class (virulence gene). We tested for 95 virulence genes previously reported here (27). These 95 virulence genes include different *E.*

206 *coli* pathotypes (ETEC, STEC, EAEC, and EPEC) in order to detect any possible E. coli hybrid if
207 present as observed for the O104:H4 Germany. The *stx* gene variants analyzed are available at
208 https://cge.cbs.dtu.dk/services/VirulenceFinder. The result was very similar to the one displayed
209 at VirulenceFinder, except that the data are now in a database and new alleles (if found) could
210 be added to the task templates.
211
212 ***in silico* antimicrobial resistance genes identification.** Antimicrobial resistance genes
213 present in sequenced genomes as well as in those retrieved from GenBank (Table 2) were
214 identified by using the genes deposited in the Center for Genomic Epidemiology
215 (http://www.genomicepidemiology.org) as part of their Resfinder 2.1 web-based tool
216 (https://cge.cbs.dtu.dk/services/ResFinder) (40), except that we used Ridom for performing
217 batch screening of the genomes analyzed. Briefly, all the genes were divided into classes or
218 groups by homology in fasta format (e.g. All *bla*TM alleles were located in a single fasta file),
219 and used as task template. Later a project was created using all these task templates, and each
220 WGS was screened for the presence of each gene class (antimicrobial resistance gene). The
221 result was very similar to the one displayed at ResFinder, except that the data are now in a
222 database and new alleles (if found) could be added to the task templates.
223
224 **Phylogenetic relationship of the strains by cgMLST analysis.** The phylogenetic relationship
225 of the strains was assessed by a core genome multilocus sequence typing (cgMLST) analysis
226 using Ridom SeqSphere+ software v2.4.0. The genome of O26:H11 strain 11368
227 (NC_013361.1) was used as a reference. After eliminating loci that were missing from the
228 genome of any strain used in our analyses, we performed a cgMLST analysis. These remaining
229 loci were considered the core genome shared by the analyzed strains .We used Nei's DNA
230 distance method (41) for calculating the matrix of genetic distance, taking only the number of
231 same/different alleles in the core genes into consideration. A Neighbor-Joining (NJ) tree using

the appropriate genetic distances was built after the cgMLST analysis. The discriminatory index was calculated with the Ridom software using the Simpson's discriminatory index as described (42); cgMLST uses alleles number of each loci for determining the genetic distance and build the phylogenetic tree. The use of allele numbers reduces the influence of recombination in the dataset studied and allow for fast clustering determination of genomes

Nucleotide sequence accession numbers. The draft genome sequences of 196 *E. coli* strains used in our study are available in GenBank under the accession numbers listed in Table S1.

RESULTS

Presence of STEC in FDA regulated foods

Among 331 suspected STEC strains isolated from FRFDA between 2003 – 2017 and sequenced by several labs and deposited at NCBI, only 276 were confirmed to be STECs by *in silico* analysis for the presence of either *stx1* or *stx2* (Table S1). Of the 196 identified and sequenced by our lab, 92% carried either stx1 or 2 (181/196). From the 74 strains which genomes were retrieved from NCBI and were initially isolated and sequenced by FDA ORA, 94% carried either stx1 or 2 (70/74). Of the 63 *E. coli* strains isolated and sequenced by a FDA contracting laboratory, 43% carried either stx1 or 2 (25/61). The frequency of isolation of STECs from foods are listed in Table 1. STECs were isolated from 22 food commodities. Most of these STECs were isolated from spinach (32%), flour (21%), lettuce (13%), and cilantro (12%).

The frequency of STEC isolation per year, their sequence type, food commodity and state of isolation (if available) is listed in Table 2. A median of 30 STECs were recovered from FDA regulated foods per year. We used 2010 as our starting year, since the number of strains before that year were sporadically found and came from our STEC historical collection. The STEC strains analyzed were isolated in 22 states.

Characterization of STEC strains by *in silico* MLST

Among the 276 STECs analyzed in this study we identified 95 different sequence types (STs) (35%) by *in silico* MLST. Strains belonging to a ST were isolated between 1 to 12 times in the period studied (2010-2017) (Table 3 and Table S2). The majority of the STs identified were observed only one time (45%).

Characterization of STEC strains by serotyping, and virulence gene profiles

However, belonging to a known ST that caused HC is not enough to predict the probability of the strain to cause disease illness. Therefore, we further characterized these STECs by in silico virulence determination as well as their predicted serotype (Table 4). The detailed in silico analysis for presence of virulence genes and serotype is listed in Supplementary Table 2. Table 4 lists only the serotype and some of the most known virulence genes for each strain: *stx1* type*, stx2* type*, eae* type*, ehxA, espP, etpD, toxB, katP, subA, saa*, and *sab*. We identified at least 81 different serotypes among the 276 STECs sequenced (Table 4). Many of the O types were not present in our O types database and were listed as unknown. Among those some of the most common clinical STECs serotypes were identified, such as: O157:H7, O26:H11, O113:H21, O121:H19, O91:H21, O103:H2, and O111:H8.

Adherence factors *eae* and *subA* genes were found in 67 (24%) and 72 (26%), respectively (Table 4). Shiga toxin genes were present as follows: *stx*1- 53 (19%) (variants a and c), *stx*2- 184 (67%) (variants a, b, c**,** d, d/e, e, and g), while *stx*1+*stx*2 – 39 (15%). The other virulence genes were more sporadically found: *exhA* gene was present in 169 (61%), *espP* was present in 118 (43%), *katP* in 24 (9%), *etpD* in 4 (2%), and finally *toxB* was present in 10 (4%).

Presence of antimicrobial resistance genes

Thirty-three of the 276 STEC strains (12%) carried antimicrobial resistance genes (Table 5). Thirty of them carried multiple antibiotic resistance genes while the remainder three carried a single gene (*tetA*- IEH-NGS-ECO-00231, FDA00011218, and CFSAN051521). Among the

antimicrobial classes observed were genes resistant to aminoglycosides, beta-lactamases, macrolides, phenicols, quinolones, sulphonamides, tetracyclines, and trimethoprim.

Phylogenetic relationship of the STEC strains by cgMLST analysis

The phylogenetic relationships among the 276 STECs from this study determined by cgMLST analysis is shown in Figure 1. The genome of O157:H7 strain Sakai (NC_002695.1) was used as the reference for the cgMLST. This *E. coli* strain has 5,204 genes, of which 3,860 genes (core genes) were present in the six genomes used as comparison to generate the cgMLST scheme (NC_011353.1 – O157:H7 strain EC4115, NC_002655.2 - O157:H7 strain EDL933, NC_013008.1 – O157:H7 strain TW14359, NC_013941.1 - O55:H7 strain CB9615, NC_017656.1 - O55:H7 strain RM12579, and NC_017906.1 – O157:H7 strain Xuzhou21). While 791 genes were found in some of the compared genomes. The remainder of the genes were eliminated from the analysis for several reasons (genes were paralogous, or pseudogenes). Therefore, a total of 4,651 genes were used as templates for the analysis of the STECs from this study.

The initial phylogenetic analysis [Neighbor-Joining (NJ) tree] based on gene differences (allele based) among these 276 STECs (Figure 1) revealed a complex evolutionary history with the existence of multiple, highly diverse genomic variants of strains isolated from RFFDA. Some of these genomes formed discrete groups and clustering was consistent with their ST (ex. all ST655 strains clustered together). A further analysis by a minimum spanning tree allows visualization of alleles differences between strains with the same ST that was not seen with the NJ tree (Figure 2).

***eae* positive Non-STEC strains virulence gene profiles**

Among the 55 non-STECs (lacking either *stx* gene by *in silico* analysis) strains isolated from FDA regulated foods, we found 35 that were positive for the *eae* gene (Table 6). Most of them were classified as atypical EPEC (aEPEC) *eae*$^{+}$ and *bfpA*$^{-}$. Even though two of them (IEH-NGS-ECO-00094, and IEH-NGS-ECO-00100) carried *bfpA* (*eae*$^{+}$ and *bfpA*$^{+}$) they were missing most of the common genes found in typical EPEC (Table 6, typical EPEC lineage 1 strain E2348/69). Therefore, we classified them as aEPEC. Virulence genes for ETEC, EIEC, and EAEC were not detected among the 331 sequenced *E. coli* strains (results not shown).

Discussion

STECs are the most dangerous among diarrheagenic *E. coli* for public health worldwide (5,7,23,43-45). Usually the most threatening STEC are those of O157:H7 serotype (46,47). However, in recent years there has been an increase in the occurrence of many non-O157 serotypes in humans associated with consumption of contaminated food, including produce and other FDA regulated products (2,48,49). Some studies have characterized STECs presence and their virulence potential from FDA regulated products (20,23,50). Most of the STEC isolated from those products have been only initially screened for the presence of some virulence genes using PCR (23,51). In the present study, we performed an in-depth analysis by whole genome sequencing of 331 presumptive STEC strains isolated from FDA regulated foods recovered during a period of 2003-2017 by two surveillance programs (FDA ORA, and MDP USDA) and other sources. STECs were isolated from 22 food commodities. It is worth mentioning that even though the sampling occurred in not all states, the food commodities had nationwide (or at least multistate) distribution.

The STEC analyzed in this study were isolated from a wide variety of foods (Table 1), with the majority isolated from spinach (32%), flour (21%), lettuce (13%), and cilantro (12%) samples during the period 2010-2017. The actual frequency of flour STECs should be assessed at a lower frequency of 9%; the spike observed in their frequency was due to the outbreak in flour in 2016, where most STECs (37 strains (66%) of total flour STECs) were isolated. A better reflection of the frequency of STEC isolated per food commodity, specifically produce, can be found in Feng and Reddy (2013) (23). Nevertheless, the presence of STECs in FRFDA per year remained relatively low, with a median of 30 isolates per year. As pointed previously, these variations in frequency of isolation can be due to seasonal variations, geographical variations, or

even due to sporadic outbreaks as was observed for O121:H19 STEC strains isolated from flour in 2016 (23).

WGS revealed that these STECs were highly variable with the existence of 95 different sequence types (STs) and belonging to at least 81 different serotypes. Some serotypes could not be predicted and might be due to that the O type and H type genes were not present in the database used which includes the most frequent serotypes found in clinical cases. Most STs were observed only once while some others were observed more frequently. ST655 was observed up to 38 times among the STECs analyzed and it was because 37 of those STEC strains were recovered during the flour outbreak in 2016 (20). The majority of the STEC STs observed in this study [69/95 -73%] had been reported as causing disease in humans, according to what was found in Enterobase (http://enterobase.warwick.ac.uk). Furthermore, of these potential human pathogenic STECs, strains belonging to 18 of those STs (19%) were additionally associated with strains causing EHEC-related illnesses (Table 3). Among the known ST associated with causing HC illnesses or HUS cases we found: ST21 and ST29 (O26:H11), ST11 (O157:H7), ST33 (O91:H14), ST17 (O103:H2), and ST16 (O111:H-), among others (44,52-54).

Some samples have the same ST however they show differences in their virulence profile as well as their Shiga toxin gene content. For example, there were 5 strains that were ST10 and from these only 2 were classified as STECs, with one carrying *stx*1a while the other carried *stx*2a, both were negative for any of the attaching genes (*eae*, *saa,* and *subA* genes), therefore considered as low risk of causing infection in a healthy individual. This is an example that a single characteristic (e.g. ST or serotype) is not enough to make an inference of the potential pathogenic trait of any STECs (http://www.fao.org/documents/card/en/c/CA0032EN). The better way is to take all the information into consideration (ST, *stx* type, attaching genes, serotype, etc)

in order to make a more informed prediction of the pathogenic potential of any STEC in conjunction with historical available data on clinical cases. For example, a strain of O113:H21 *stx*2a positive and that doesn't possess *eae* but it has *subA, saa*, and *sab* genes, and has been linked to HUS cases (19), therefore, we can foresee that this strain might be harmful to humans. A similar analysis could be done in the case of any STEC that have all those attributes but that has not been linked to any human cases. Even though we cannot predict the actual outcome of an infection with this strain, it still warrants a warning about its presence in foods that are consumed raw as it is the case of fresh produce.

We tested for 95 known virulence genes (27) found in the most common *E. coli* pathotypes and did not find any genes present that would characterize the strains as STEC/EAEC//ETEC/EIEC hybrids. Among the adherence factors, *eae* and *subA* genes were found in 24% and 26% of the STEC strains, respectively. Strains that carry *eae* did not carry *subA* or *saa* genes, and viceversa, as perviously observed for STEC isolated from fresh produce (23). Regarding the presence of Shiga toxin type, there was great variation with most strains (67%) carrying only stx type 2, 19% carrying only stx type 1 while 15% carried both stx types. Among the stx type 2 there were 144 that were either a,d, or c, which are the stx2 types found among clinical cases (31,32,55-57) and that have specific trophism for humans (56). The remaining 40 STECs carrying *stx* type 2 alone were stx type (e, d/e and g) which have been described in animal reservoirs (58). The ones carrying both stx types were all stx1a. The remaining virulence genes were sporadically found with the most common *exhA* gene was found in 61% of the STECs, while *espP* was found in 43% of the STECs. These two genes can be found in the virulence plasmid and appears to participate in STECs infection in humans (9,11,13). In summary, 46 of the STECs analyzed in this study carried both *stx2*a and *eae* gene which is considered of elevated risk to human health (22,56).

We also confirmed the presence of antimicrobial resistance (AMR) genes using the DTU database with our modified protocol (Ridom) and find that their prevalence to be low at only 12%, although carrying multiple antimicrobial resistance genes. The presence of strains carrying multiple AMR genes is worrisome since they can be shared amongst other *E. coli* and could possible participate in the dissemination of AMR in their environments, as has been observed occurring for tetracycline genes in *E. coli* isolates from beef cattle (59), for colistin resistance (*mcr*-1 gene) through plasmid-mediated transfer (60), and for ampicillin resistance genes in *E. coli* in an infant treated with antibiotics (61).

Phylogenetic analysis by a custom cgMLST analysis of these 276 STECs confirmed the MLST *in silico* analysis, with many different defined clades among these STECs isolated from FRFDA. The cgMLST analysis is a fast method of analysis and provides an initial visualization of the relationships among the strains analyzed. Comparable results have been observed for establishing fast relationships among genomes from diverse bacterial pathogens (29,43,62-67). A further analysis using only the genomes of strains that are located within each individual or among selected clades can produce a more detailed evolutionary history, using single nucleotide analyses, which can help in determining to understand the potential source, phylogenetic nature, lineage, and timeline of transmission of each group, as has been shown for the ST36 lineage of *Vibrio parahaemolyticus* (68).

EPECs are the leading cause of infantile diarrhea in developing countries (70,71). Typical EPECs (tEPEC) have *eae* and *bfp* genes, and their main reservoir is humans (72). The *eae* gene is located in the chromosome, in the LEE operon, while the *bfp* operon is typically located in the large EPEC adherence factor (EAF) virulence plasmid (72). These tEPEC also carry the *perA* gene, which increases the expression of LEE elements (72,73). Interestingly, 17% of our presumptive STECs were shown by *in silico* analysis to be atypical EPECs

(aEPEC). Among their unusual features are the absence of the EAF plasmid, and their reservoirs can be animals or humans (72). It is possible these aEPECs might had lost their phages upon culturing, as this pattern has been observed in clinical isolates of *E. coli* upon sub-cultivation (69). The aEPEC we observed in this study may have the capacity to produce A/E lesions, since they carried both the *eae* and *tir* gene, which are the effector and receptor necessary for the formation of the A/E lesion (74).

Our results suggest that finding aEPECs in food could be of particular concern, as these strains have the potential for acquiring the *stx* phage, as observed in the *E. coli* O104:H4 strain found in Germany (75). That strain was an entero-aggregative *E. coli* (EAEC) that had acquired an *stx*2a phage, and human illnesses that resulted during 2011 became the largest known HUS outbreak of STEC-related illness in the world (75). Similarly, anO26:H11 strain 21765, isolated in 2005 during a milk cheese outbreak in France (76) was shown to be an EPEC strain that had probably acquired a *stx*2a phage (27). In Gonzalez-Escalona et al (2016), the authors demonstrated that some strains of *E. coli* O26:H11 isolated from US cattle were phylogenetically more closely related to ST29 O26:H11 EHECs but t because these did not carry the *stx* phage, they would have been classified as EHEC-like by previous methods (77). Over the last five years, the analyses of thousands of *E. coli* genomes have revealed that so-called *E. coli* “hybrid strains” – strains that belong to one pathotype but acquire virulence markers, such as *stx* genes, from another pathotype – could be more common than previously believed. If this is the case, this suggests that environmental *E. coli* strains currently considered harmless could acquire the potential to pose risks to human health; for example, both aEPEC and STEC strains were isolated from foods such as flour, cilantro, lettuce, and kale (Table S1).

We are heading to a new phase in surveillance of STECs in the US by using a genomic monitoring approach and our STEC sequences from FRFDA provides a solid foundation to build

upon (78)(https://www.cdc.gov/pulsenet/pathogens/wgs.html). There already exists a database that achieve the first goal of source tracking by using core genome information (NCBI pathogen detection tool), but there is a need for improved databases that allow for fast analysis of the WGS data for detecting virulence genes, phages and plasmids content, as well as antimicrobial resistance genes.

In conclusion, STECs were isolated from diverse FRFDA food sources during the period study. The contamination frequency was relatively low (average 30 STEC strains isolated per year). However, fifty percent of the STECs analyzed in this study carried either a combination of *eae* plus *stx*, or *subA* plus *stx*, therefore being potentially pathogenic to humans. Moreover, those STECs carried most of the virulence genes described for STECs causing infections with a diverse range from HC (e.g. ST655 O111:H19 strains) to HUS (e.g. ST21 O26:H11 strains) (20,44). Some others have not been described as causing disease in humans but have the potential to do so (e.g. ST342 O5:H-unknown strains) since they carried all virulence genes described in pathogenic strains (*stx1a*, *eae*-beta1, *exhA*, *tir*, and many of the T3SS effectors and non-LEE effectors) (Table 4). Nonetheless, the determination of the presence of STECs in FRFDA with potential to cause disease in humans reinforce the need to continue surveillance for this important pathogen which is of importance for food safety and public health. Furthermore, the availability of these genomes could provide early warnings of food contamination from cattle or other animals, since some of the STEC isolated were carrying stx2e that have been usually observed causing edema in pigs (79) and are considered as probably non-pathogenic to humans (56). Here we showed that WGS enabled comparisons across isolates to establish phylogeny, help in identification of antibiotic resistance by monitoring the presence of antimicrobial resistance genes, and determined the presence of known virulence genes that have been linked with illnesses. A freely accessible database of high-quality reference genome sequences of FRFDA was previously unavailable. Future food

482 safety investigations will benefit from the comparisons made possible by WGS databases like
483 ours as it allows for the monitoring of the recurrence and emergence of strains in the food
484 supply. It is our goal to help develop databases that will allow for fast source tracking and
485 accurate categorization (low risk or high risk) of STECs food isolates in a more comprehensive
486 manner.
487

ACKNOWLEDGMENTS

This work was supported by the FDA Foods Science and Research Intramural Program. We also want to thank Sabina Lindley for assistance in sequencing some of these *E. coli* strains. The authors thank Lili Fox Vélez for her helpful comments on this manuscript.

FIGURES

Figure 1. Phylogenetic relationships among the 276 STEC genomes of *E. coli* sequenced in this study by cgMLST analysis. Ridom SeqSphere+ (v5.0.0) identified 4,651core genes. The evolutionary history was inferred by using the Neighbor-joining (NJ) tree built using the genetic distance and showing the existence of many diverse clades with a complex evolutionary history. Strains are colored based on different STs as labeled.

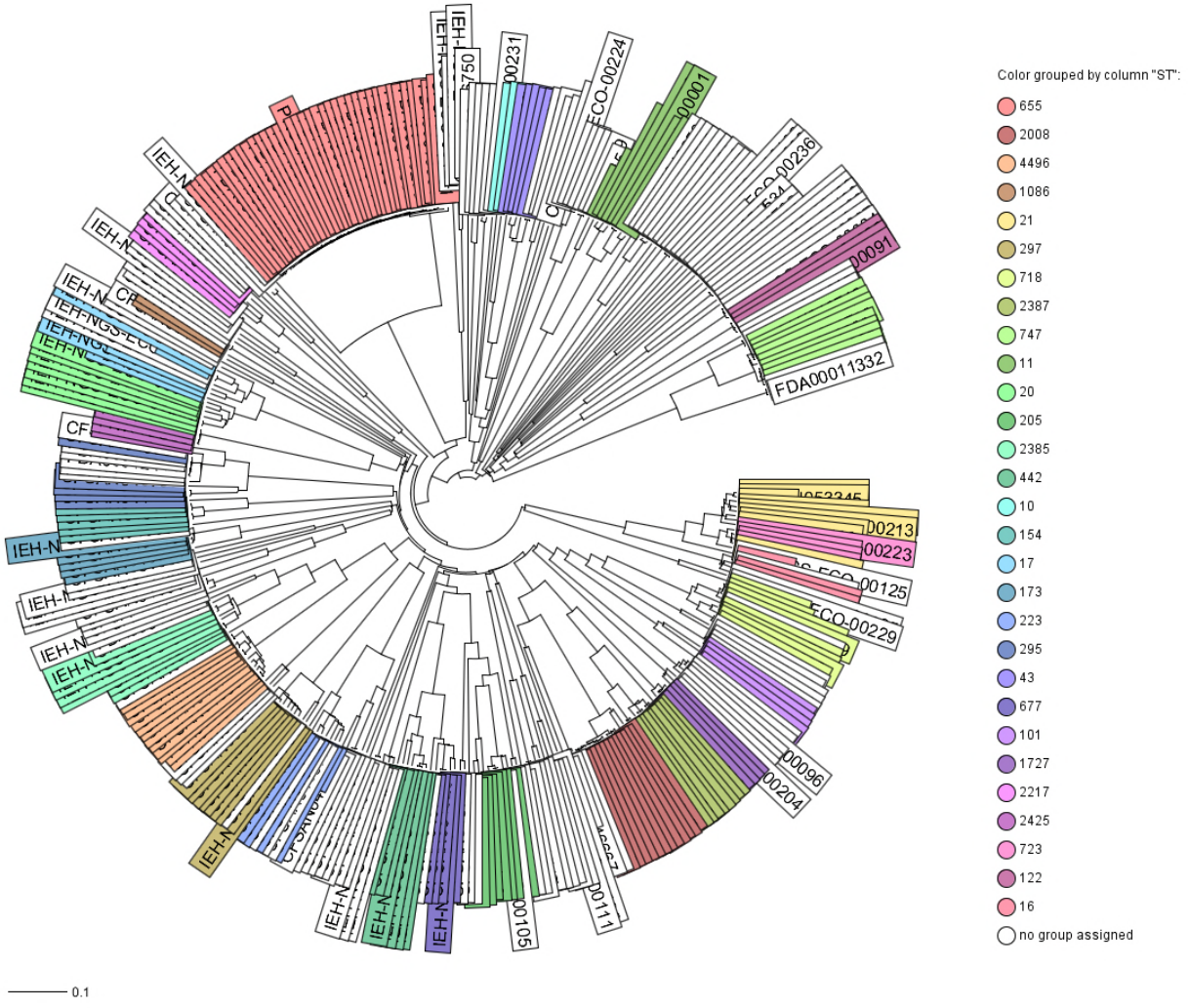
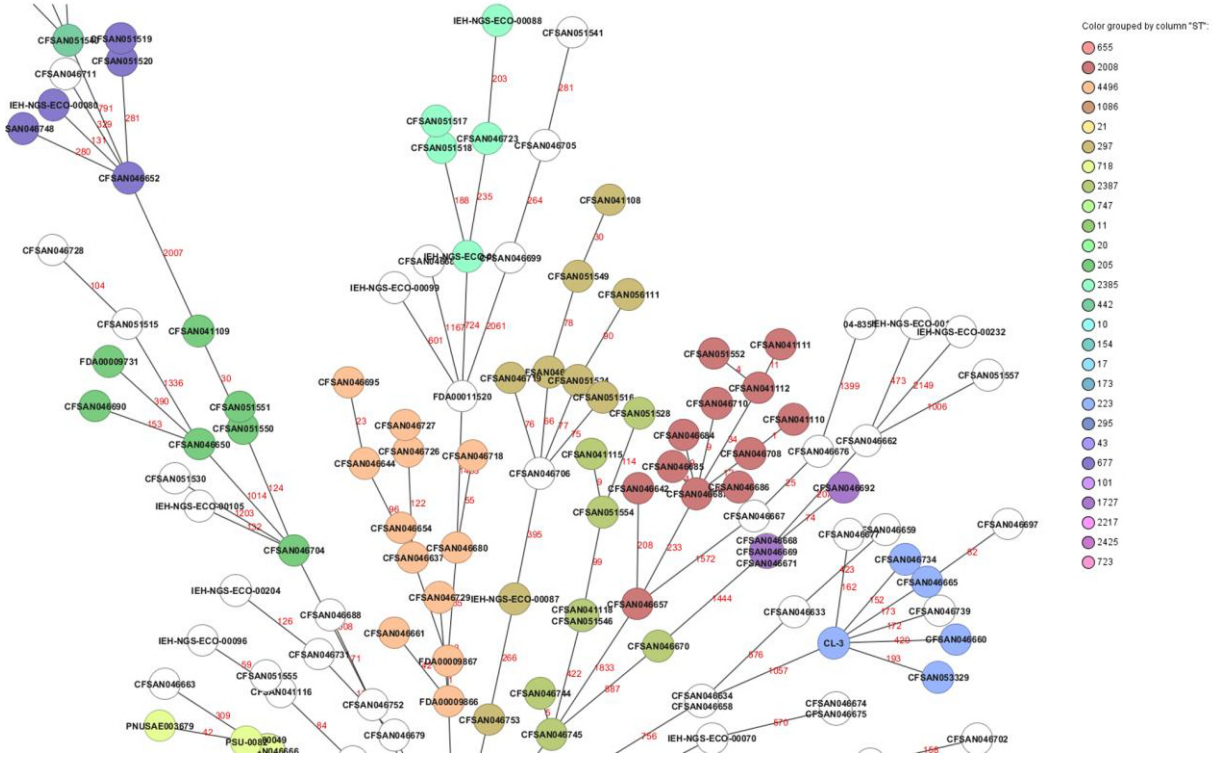


Figure 2. Snapshot of a minimum spanning tree showing the relationships among all different STECs. The numbers above the connected lines (not to scale) represent allele differences between strains belonging to the same ST. The isolates are colored based on different STs as labeled.



References

1. **Rangel, J. M., P. H. Sparling, C. Crowe, P. M. Griffin, and D. L. Swerdlow**. 2005. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982-2002. Emerg. Infect Dis. **11**:603-609.
2. **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. J. Infect Dis. **192**:1422-1429.
3. **Scallan, E., R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M. A. Widdowson, S. L. Roy, J. L. Jones, and P. M. Griffin**. 2011. Foodborne illness acquired in the United States--major pathogens. Emerg. Infect. Dis. **17**:7-15.
4. **Beutin, L., G. Krause, S. Zimmermann, S. Kaulfuss, and K. Gleier**. 2004. Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from human patients in Germany over a 3-year period. J Clin Microbiol **42**:1099-1108.
5. **Kuehne, A., M. Bouwknegt, A. Havelaar, A. Gilsdorf, P. Hoyer, K. Stark, and D. Werber**. 2016. Estimating true incidence of O157 and non-O157 Shiga toxin-producing Escherichia coli illness in Germany based on notification data of haemolytic uraemic syndrome. Epidemiol. Infect. **144**:3305-3315.
6. **Byrne, L., C. Jenkins, N. Launders, R. Elson, and G. K. Adak**. 2015. The epidemiology, microbiology and clinical impact of Shiga toxin-producing Escherichia coli in England, 2009-2012. Epidemiol. Infect. **143**:3475-3487.

7. **McAllister, L. J., S. J. Bent, N. K. Petty, E. Skippington, S. A. Beatson, J. C. Paton, and A. W. Paton**. 2016. Genomic Comparison of Two O111:H- Enterohemorrhagic Escherichia coli Isolates from a Historic Hemolytic-Uremic Syndrome Outbreak in Australia. Infect. Immun. **84**:775-781.
8. **U.S.Department of Agriculture, F. S. a. I. S.** 2012. Shiga Toxin-Producing *Escherichia coli* in Certain Raw Beef Products. Fed. Regist. **77**:31975-31981.
9. **Kaper JB, Nataro JP, and Mobley HL**. 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol **2**:123-140.
10. **Karmali, M. A., M. Mascarenhas, S. Shen, K. Ziebell, S. Johnson, R. Reid-Smith, J. Isaac-Renton, C. Clark, K. Rahn, and J. B. Kaper**. 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. J. Clin. Microbiol. **41**:4930-4940.
11. **Nataro JP and Kaper JB**. 1998. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev **11**:142-201.
12. **Garmendia, J., G. Frankel, and V. F. Crepin**. 2005. Enteropathogenic and enterohemorrhagic Escherichia coli infections: translocation, translocation, translocation. Infect. Immun. **73**:2573-2585.
13. **Garmendia, J., Z. Ren, S. Tennant, M. A. Midolli Viera, Y. Chong, A. Whale, K. Azzopardi, S. Dahan, M. P. Sircili, M. R. Franzolin, L. R. Trabulsi, A. Phillips, T. A. Gomes, J. Xu, R. Robins-Browne, and G. Frankel**. 2005. Distribution of tccP in clinical enterohemorrhagic and enteropathogenic Escherichia coli isolates. J Clin Microbiol. **43**:5715-5720.

14. **Beutin, L., J. Prada, S. Zimmermann, R. Stephan, I. Orskov, and F. Orskov**. 1988. Enterohemolysin, a new type of hemolysin produced by some strains of enteropathogenic E. coli (EPEC). Zentralbl. Bakteriol. Mikrobiol. Hyg. A **267**:576-588.

15. **Schmidt, H., C. Kernbach, and H. Karch**. 1996. Analysis of the EHEC hly operon and its location in the physical map of the large plasmid of enterohaemorrhagic Escherichia coli O157:h7. Microbiology **142 (Pt 4)**:907-914.

16. **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. J Clin Microbiol. **37**:497-503.

17. **Schmidt, H., L. Beutin, and H. Karch**. 1995. Molecular analysis of the plasmid-encoded hemolysin of *Escherichia coli* O157:H7 strain EDL 933. Infect. Immun. **63**:1055-1061.

18. **Khaitan, A., D. M. Jandhyala, C. M. Thorpe, J. M. Ritchie, and A. W. Paton**. 2007. The operon encoding SubAB, a novel cytotoxin, is present in shiga toxin-producing Escherichia coli isolates from the United States. J Clin Microbiol. **45**:1374-1375.

19. **Paton, A. W., P. Srimanote, M. C. Woodrow, and J. C. Paton**. 2001. Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxigenic Escherichia coli strains that are virulent for humans. Infect. Immun. **69**:6999-7009.

20. **Crowe, S. J., L. Bottichio, L. N. Shade, B. M. Whitney, N. Corral, B. Melius, K. D. Arends, D. Donovan, J. Stone, K. Allen, J. Rosner, J. Beal, L. Whitlock,**

A. Blackstock, J. Wetherington, L. A. Newberry, M. N. Schroeder, D. Wagner, E. Trees, S. Viazis, M. E. Wise, and K. P. Neil. 2017. Shiga Toxin-Producing E. coli Infections Associated with Flour. N Engl J Med **377**:2036-2043.

21. **Erickson, M. C. and M. P. Doyle**. 2007. Food as a vehicle for transmission of Shiga toxin-producing Escherichia coli. J Food Prot. **70**:2426-2449.
22. **Feng, P. C. and S. P. Reddy**. 2014. Prevalence and diversity of enterotoxigenic Escherichia coli strains in fresh produce. J Food Prot. **77**:820-823.
23. **Feng, P. C. and S. Reddy**. 2013. Prevalences of Shiga toxin subtypes and selected other virulence factors among Shiga-toxigenic Escherichia coli strains isolated from fresh produce. Appl. Environ. Microbiol. **79**:6917-6923.
24. **Allard, M. W., R. Bell, C. M. Ferreira, N. Gonzalez-Escalona, M. Hoffmann, T. Muruvanda, A. Ottesen, P. Ramachandran, E. Reed, S. Sharma, E. Stevens, R. Timme, J. Zheng, and E. W. Brown**. 2018. Genomics of foodborne pathogens for microbial food safety. Curr. Opin. Biotechnol **49**:224-229.
25. **Bergholz, T. M., A. I. Moreno Switt, and M. Wiedmann**. 2014. Omics approaches in food safety: fulfilling the promise? Trends Microbiol. **22**:275-281.
26. **Hoffmann, M., Y. Luo, S. R. Monday, N. Gonzalez-Escalona, A. R. Ottesen, T. Muruvanda, C. Wang, G. Kastanis, C. Keys, D. Janies, I. F. Senturk, U. V. Catalyurek, H. Wang, T. S. Hammack, W. J. Wolfgang, D. Schoonmaker-Bopp, A. Chu, R. Myers, J. Haendiges, P. S. Evans, J. Meng, E. A. Strain, M. W. Allard, and E. W. Brown**. 2016. Tracing Origins of the *Salmonella* Bareilly Strain Causing a Food-borne Outbreak in the United States. J. Infect Dis. **213**:502-508.

27. **Gonzalez-Escalona, N., M. Toro, L. V. Rump, G. Cao, T. G. Nagaraja, and J. Meng**. 2016. Virulence Gene Profiles and Clonal Relationships of Escherichia coli O26:H11 Isolates from Feedlot Cattle as Determined by Whole-Genome Sequencing. Appl. Environ. Microbiol. **82**:3900-3912.

28. **Maiden, M. C. and O. B. Harrison**. 2016. The population and functional genomics of the Neisseria revealed with gene-by-gene approaches. J. Clin. Microbiol.

29. **Lorenz, S. C., M. L. Kotewicz, M. Hoffmann, N. Gonzalez-Escalona, M. Fischer, and J. A. Kase**. 2016. Complete Genome Sequences of Four Enterohemolysin-Positive (ehxA) Enterocyte Effacement-Negative Shiga Toxin-Producing Escherichia coli Strains. Genome Announc. **4**.

30. **Dallman, T. J., L. Byrne, N. Launders, K. Glen, K. A. Grant, and C. Jenkins**. 2015. The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing *Escherichia coli* serogroup O26:H11. Epidemiol. Infect **143**:1672-1680.

31. **Beutin, L. and A. Martin**. 2012. Outbreak of Shiga toxin-producing *Escherichia coli* (STEC) O104:H4 infection in Germany causes a paradigm shift with regard to human pathogenicity of STEC strains. J. Food Prot. **75**:408-418.

32. **Pennington, H.** 2011. Escherichia coli O104, Germany 2011. Lancet Infect. Dis.

33. **Li, Z., A. Perez-Osorio, Y. Wang, K. Eckmann, W. A. Glover, M. W. Allard, E. W. Brown, and Y. Chen**. 2017. Whole genome sequencing analyses of Listeria monocytogenes that persisted in a milkshake machine for a year and caused illnesses in Washington State. BMC Microbiol. **17**:134.

624 34. **Ogura, Y., T. Ooka, A. Iguchi, H. Toh, M. Asadulghani, K. Oshima, T.**
625 **Kodama, H. Abe, K. Nakayama, K. Kurokawa, T. Tobe, M. Hattori, and T.**
626 **Hayashi**. 2009. Comparative genomics reveal the mechanism of the parallel
627 evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc. Natl.
628 Acad. Sci. U. S A **106**:17939-17944.
629 35. **Beutin, L. and A. Martin**. 2012. Outbreak of Shiga toxin-producing Escherichia
630 coli (STEC) O104:H4 infection in Germany causes a paradigm shift with regard
631 to human pathogenicity of STEC strains. J. Food Prot. **75**:408-418.
632 36. **Rasko, D. A., D. R. Webster, J. W. Sahl, A. Bashir, N. Boisen, F. Scheutz, E.**
633 **E. Paxinos, R. Sebra, C. S. Chin, D. Iliopoulos, A. Klammer, P. Peluso, L.**
634 **Lee, A. O. Kislyuk, J. Bullard, A. Kasarskis, S. Wang, J. Eid, D. Rank, J. C.**
635 **Redman, S. R. Steyert, J. Frimodt-Moller, C. Struve, A. M. Petersen, K. A.**
636 **Krogfelt, J. P. Nataro, E. E. Schadt, and M. K. Waldor**. 2011. Origins of the *E.*
637 *coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. N
638 Engl J Med **365**:709-717.
639 37. **Joensen, K. G., A. M. Tetzschner, A. Iguchi, F. M. Aarestrup, and F. Scheutz**.
640 2015. Rapid and Easy In Silico Serotyping of *Escherichia coli* Isolates by Use of
641 Whole-Genome Sequencing Data. J. Clin. Microbiol. **53**:2410-2426.
642 38. **Joensen, K. G., F. Scheutz, O. Lund, H. Hasman, R. S. Kaas, E. M. Nielsen,**
643 **and F. M. Aarestrup**. 2014. Real-time whole-genome sequencing for routine
644 typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J.
645 Clin. Microbiol. **52**:1501-1510.

646 39. **Wirth, T., D. Falush, R. Lan, F. Colles, P. Mensa, L. H. Wieler, H. Karch, P. R.**
647 **Reeves, M. C. Maiden, H. Ochman, and M. Achtman**. 2006. Sex and virulence
648 in *Escherichia coli*: an evolutionary perspective. Mol. Microbiol **60**:1136-1151.
649 40. **Zankari, E., H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O.**
650 **Lund, F. M. Aarestrup, and M. V. Larsen**. 2012. Identification of acquired
651 antimicrobial resistance genes. J. Antimicrob. Chemother. **67**:2640-2644.
652 41. **Nei, M., F. Tajima, and Y. Tateno**. 1983. Accuracy of estimated phylogenetic
653 trees from molecular data. II. Gene frequency data. J. Mol. Evol. **19**:153-170.
654 42. **Hunter, P. R. and M. A. Gaston**. 1988. Numerical index of the discriminatory
655 ability of typing systems: an application of Simpson's index of diversity. J. Clin.
656 Microbiol. **26**:2465-2466.
657 43. **Gonzalez-Escalona, N., M. Toro, L. V. Rump, G. Cao, T. G. Nagaraja, and J.**
658 **Meng**. 2016. Virulence Gene Profiles and Clonal Relationships of Escherichia
659 coli O26:H11 Isolates from Feedlot Cattle as Determined by Whole-Genome
660 Sequencing. Appl. Environ. Microbiol. **82**:3900-3912.
661 44. **Bielaszewska, M., A. Mellmann, S. Bletz, W. Zhang, R. Kock, A. Kossow, R.**
662 **Prager, A. Fruth, D. Orth-Holler, M. Marejkova, S. Morabito, A. Caprioli, D.**
663 **Pierard, G. Smith, C. Jenkins, K. Curova, and H. Karch**. 2013.
664 Enterohemorrhagic *Escherichia coli* O26:H11/H-: a new virulent clone emerges in
665 Europe. Clin. Infect Dis. **56**:1373-1381.
666 45. **Taylor, E. V., T. A. Nguyen, K. D. Machesky, E. Koch, M. J. Sotir, S. R.**
667 **Bohm, J. P. Folster, R. Bokanyi, A. Kupper, S. A. Bidol, A. Emanuel, K. D.**
668 **Arends, S. A. Johnson, J. Dunn, S. Stroika, M. K. Patel, and I. Williams**.

669 2013. Multistate outbreak of *Escherichia coli* O145 infections associated with
670 romaine lettuce consumption, 2010. J. Food Prot. **76**:939-944.
671 46. **Watanabe, H., A. Wada, Y. Inagaki, K. Itoh, and K. Tamura**. 1996. Outbreaks
672 of enterohaemorrhagic *Escherichia coli* O157:H7 infection by two different
673 genotype strains in Japan, 1996. Lancet **348**:831-832.
674 47. **Bell, B. P., M. Goldoft, P. M. Griffin, M. A. Davis, D. C. Gordon, P. I. Tarr, C.**
675 **A. Bartleson, J. H. Lewis, T. J. Barrett, J. G. Wells, and .** 1994. A multistate
676 outbreak of *Escherichia coli* O157:H7-associated bloody diarrhea and hemolytic
677 uremic syndrome from hamburgers. The Washington experience. JAMA
678 **272**:1349-1353.
679 48. **Ethelberg, S., B. Smith, M. Torpdahl, M. Lisby, J. Boel, T. Jensen, E. M.**
680 **Nielsen, and K. Molbak**. 2009. Outbreak of non-O157 Shiga toxin-producing
681 *Escherichia coli* infection from consumption of beef sausage. Clin. Infect Dis.
682 **48**:e78-e81.
683 49. **Fey, P. D., R. S. Wickert, M. E. Rupp, T. J. Safranek, and S. H. Hinrichs**.
684 2000. Prevalence of non-O157:H7 shiga toxin-producing *Escherichia coli* in
685 diarrheal stool samples from Nebraska. Emerg. Infect. Dis. **6**:530-533.
686 50. **Cooper, K. K., R. E. Mandrell, J. W. Louie, J. Korlach, T. A. Clark, C. T.**
687 **Parker, S. Huynh, P. S. Chain, S. Ahmed, and M. Q. Carter**. 2014.
688 Comparative genomics of enterohemorrhagic *Escherichia coli* O145:H28
689 demonstrates a common evolutionary lineage with *Escherichia coli* O157:H7.
690 BMC. Genomics **15**:17.

691 51. **Feng, P. C., S. Delannoy, D. W. Lacher, L. F. Dos Santos, L. Beutin, P. Fach,**
692 **M. Rivas, E. L. Hartland, A. W. Paton, and B. E. Guth**. 2014. Genetic diversity
693 and virulence potential of shiga toxin-producing Escherichia coli O113:H21
694 strains isolated from clinical, environmental, and food sources. Appl. Environ.
695 Microbiol. **80**:4757-4763.
696 52. **Mellmann, A., M. Bielaszewska, R. Kock, A. W. Friedrich, A. Fruth, B.**
697 **Middendorf, D. Harmsen, M. A. Schmidt, and H. Karch**. 2008. Analysis of
698 collection of hemolytic uremic syndrome-associated enterohemorrhagic
699 *Escherichia coli*. Emerg. Infect. Dis. **14**:1287-1290.
700 53. **Bielaszewska, M., R. Prager, R. Kock, A. Mellmann, W. Zhang, H. Tschape,**
701 **P. I. Tarr, and H. Karch**. 2007. Shiga toxin gene loss and transfer in vitro and in
702 vivo during enterohemorrhagic *Escherichia coli* O26 infection in humans. Appl.
703 Environ. Microbiol. **73**:3144-3150.
704 54. **McAllister, L. J., S. J. Bent, N. K. Petty, E. Skippington, S. A. Beatson, J. C.**
705 **Paton, and A. W. Paton**. 2016. Genomic Comparison of Two O111:H-
706 Enterohemorrhagic Escherichia coli Isolates from a Historic Hemolytic-Uremic
707 Syndrome Outbreak in Australia. Infect. Immun. **84**:775-781.
708 55. **Bielaszewska, M., A. W. Friedrich, T. Aldick, R. Schurk-Bulgrin, and H.**
709 **Karch**. 2006. Shiga toxin activatable by intestinal mucus in Escherichia coli
710 isolated from humans: predictor for a severe clinical outcome. Clin Infect. Dis.
711 **43**:1160-1167.
712 56. **Scheutz, F., L. D. Teel, L. Beutin, D. Pierard, G. Buvens, H. Karch, A.**
713 **Mellmann, A. Caprioli, R. Tozzoli, S. Morabito, N. A. Strockbine, A. R.**

Melton-Celsa, M. Sanchez, S. Persson, and A. D. O'Brien. 2012. Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. J Clin Microbiol. **50**:2951-2963.

57. **Beutin, L., J. A. Hammerl, J. Reetz, and E. Strauch**. 2014. Shiga toxin 2A-encoding bacteriophages in enteroaggregative *Escherichia coli* O104:H4 strains. Emerg. Infect Dis. **20**:1567-1568.

58. **Melton-Celsa, A. R. and A. D. O'Brien**. 1998. Structure, biology, and relative toxicity of Shiga toxin family members for cells and animals, p. 121-128. *In*: J. B. Kaper and A. D. O'Brien (eds.), *Escherichia coli* O157:H7 and other Shiga toxin-producing *E. coli* strains. ASM Press, Washington, DC.

59. **Shin, S. W., M. K. Shin, M. Jung, K. M. Belaynehe, and H. S. Yoo**. 2015. Prevalence of Antimicrobial Resistance and Transfer of Tetracycline Resistance Genes in Escherichia coli Isolates from Beef Cattle. Appl. Environ. Microbiol. **81**:5560-5566.

60. **Liu, Y. Y., Y. Wang, T. R. Walsh, L. X. Yi, R. Zhang, J. Spencer, Y. Doi, G. Tian, B. Dong, X. Huang, L. F. Yu, D. Gu, H. Ren, X. Chen, L. Lv, D. He, H. Zhou, Z. Liang, J. H. Liu, and J. Shen**. 2016. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect. Dis. **16**:161-168.

61. **Karami, N., A. Martner, V. I. Enne, S. Swerkersson, I. Adlerberth, and A. E. Wold**. 2007. Transfer of an ampicillin resistance gene between two Escherichia coli strains in the bowel microbiota of an infant treated with antibiotics. J Antimicrob Chemother **60**:1142-1145.

62. **Ruppitsch, W., A. Pietzka, K. Prior, S. Bletz, H. L. Fernandez, F. Allerberger, D. Harmsen, and A. Mellmann**. 2015. Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of Listeria monocytogenes. J. Clin. Microbiol. **53**:2869-2876.

63. **Chen, Y., N. Gonzalez-Escalona, T. S. Hammack, M. W. Allard, E. A. Strain, and E. W. Brown**. 2016. Core Genome Multilocus Sequence Typing for Identification of Globally Distributed Clonal Groups and Differentiation of Outbreak Strains of Listeria monocytogenes. Appl. Environ. Microbiol. **82**:6258-6272.

64. **Gonzalez-Escalona, N., R. G. Gavilan, E. W. Brown, and J. Martinez-Urtaza**. 2015. Transoceanic Spreading of Pathogenic Strains of *Vibrio parahaemolyticus* with Distinctive Genetic Signatures in the *rec*A Gene. PLoS. One. **10**:e0117485.

65. **Strauss, L., M. Stegger, P. E. Akpaka, A. Alabi, S. Breurec, G. Coombs, B. Egyir, A. R. Larsen, F. Laurent, S. Monecke, G. Peters, R. Skov, B. Strommenger, F. Vandenesch, F. Schaumburg, and A. Mellmann**. 2017. Origin, evolution, and global transmission of community-acquired Staphylococcus aureus ST8. Proc. Natl. Acad. Sci. U. S A **114**:E10596-E10604.

66. **Gonzalez-Escalona, N., K. A. Jolley, E. Reed, and J. Martinez-Urtaza**. 2017. Defining a Core Genome Multilocus Sequence Typing Scheme for the Global Epidemiology of Vibrio parahaemolyticus. J Clin Microbiol **55**:1682-1697.

67. **Ghanem, M., L. Wang, Y. Zhang, S. Edwards, A. Lu, D. Ley, and M. El-Gazzar**. 2018. Core Genome Multilocus Sequence Typing: a Standardized Approach for Molecular Typing of Mycoplasma gallisepticum. J Clin Microbiol **56**.

760 68. **Martinez-Urtaza, J., A. R. van, M. Abanto, J. Haendiges, R. A. Myers, J.**
761 **Trinanes, C. Baker-Austin, and N. Gonzalez-Escalona**. 2017. Genomic
762 Variation and Evolution of Vibrio parahaemolyticus ST36 over the Course of a
763 Transcontinental Epidemic Expansion. MBio **8**.
764 69. **Karch, H., T. Meyer, H. Russmann, and J. Heesemann**. 1992. Frequent loss of
765 Shiga-like toxin genes in clinical isolates of Escherichia coli upon subcultivation.
766 Infect. Immun. **60**:3464-3467.
767 70. **Okello, E., K. Moonens, J. Erume, and G. H. De**. 2015. Enterotoxigenic
768 Escherichia coli strains are highly prevalent in Ugandan piggeries but disease
769 outbreaks are masked by antibiotic prophylaxis. Trop. Anim Health Prod. **47**:117-
770 122.
771 71. **Ochoa, T. J. and C. A. Contreras**. 2011. Enteropathogenic escherichia coli
772 infection in children. Curr. Opin. Infect. Dis. **24**:478-483.
773 72. **Trabulsi, L. R., R. Keller, and T. A. T. Gomes**. 2002. Typical and Atypical
774 Enteropathogenic Escherichia coli. Emerg. Infect. Dis **8**:508-513.
775 73. **Elliott, S. J., V. Sperandio, J. A. Giron, S. Shin, J. L. Mellies, L. Wainwright,**
776 **S. W. Hutcheson, T. K. McDaniel, and J. B. Kaper**. 2000. The locus of
777 enterocyte effacement (LEE)-encoded regulator controls expression of both LEE-
778 and non-LEE-encoded virulence factors in enteropathogenic and
779 enterohemorrhagic Escherichia coli. Infect Immun. **68**:6115-6126.
780 74. **McWilliams, B. D. and A. G. Torres**. 2014. Enterohemorrhagic *Escherichia coli*
781 adhesins. Microbiol Spectrum **2**:EHEC-0003-2013.

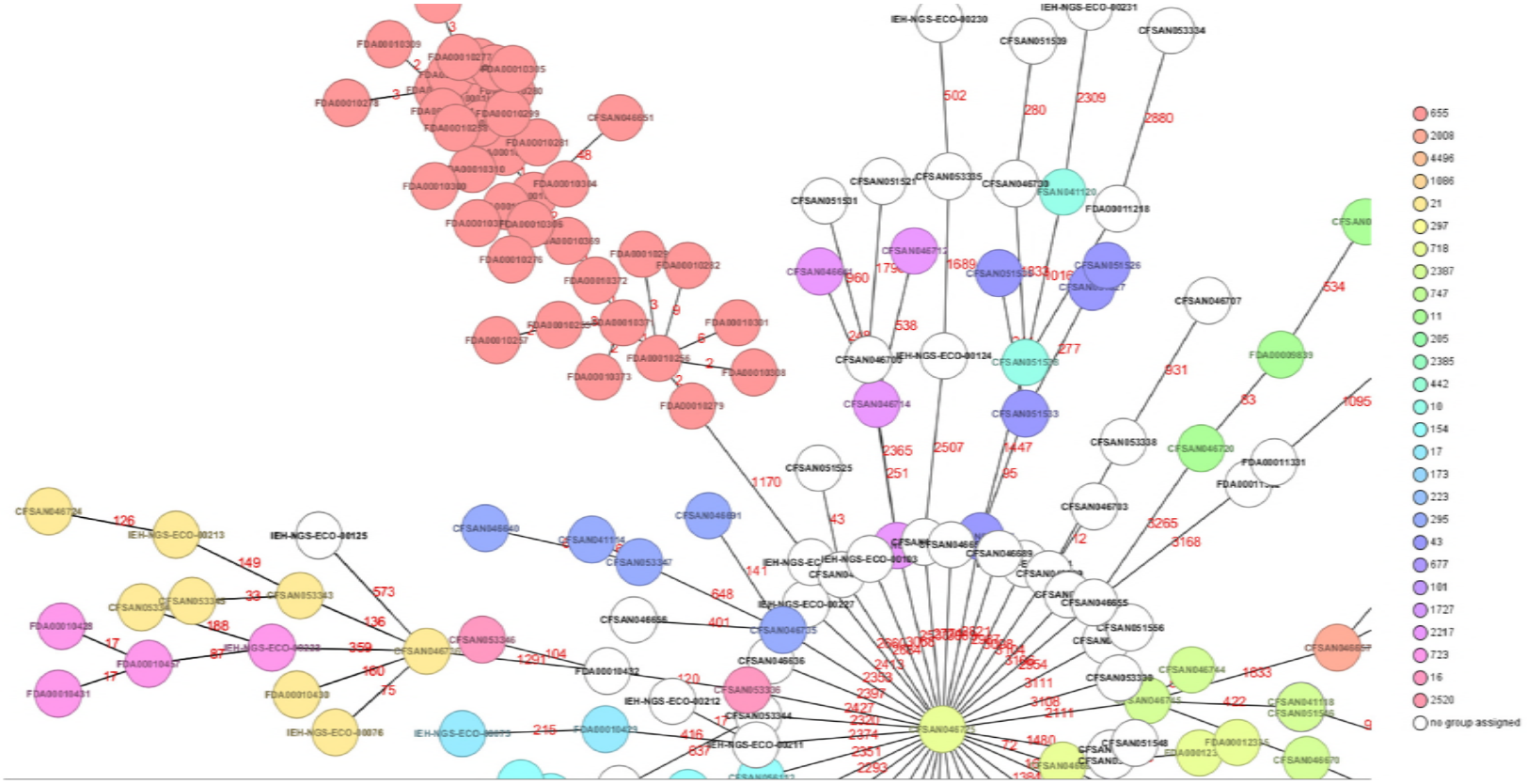
75. **Rohde, H., J. Qin, Y. Cui, D. Li, N. J. Loman, M. Hentschke, W. Chen, F. Pu, Y. Peng, J. Li, F. Xi, S. Li, Y. Li, Z. Zhang, X. Yang, M. Zhao, P. Wang, Y. Guan, Z. Cen, X. Zhao, M. Christner, R. Kobbe, S. Loos, J. Oh, L. Yang, A. Danchin, G. F. Gao, Y. Song, Y. Li, H. Yang, J. Wang, J. Xu, M. J. Pallen, J. Wang, M. Aepfelbacher, and R. Yang**. 2011. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. N Engl J Med **365**:718-724.

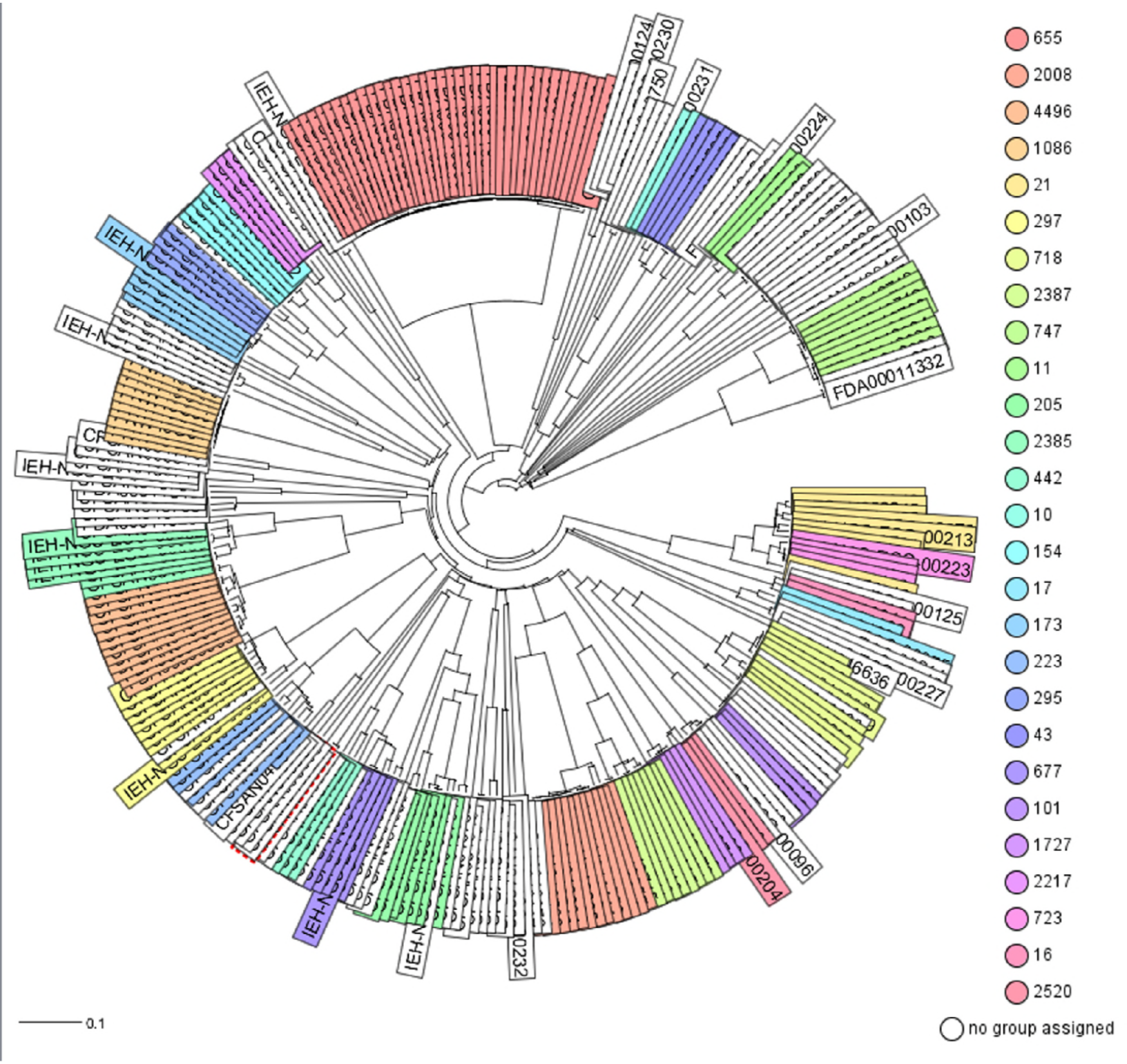
76. **Galia, W., P. Mariani-Kurkdjian, E. Loukiadis, S. Blanquet-Diot, F. Leriche, H. Brugere, A. Shima, E. Oswald, B. Cournoyer, and D. Thevenot-Sergentet**. 2015. Genome Sequence and Annotation of a Human Infection Isolate of *Escherichia coli* O26:H11 Involved in a Raw Milk Cheese Outbreak. Genome Announc. **3**.

77. **Bugarel, M., L. Beutin, F. Scheutz, E. Loukiadis, and P. Fach**. 2011. Identification of genetic markers for differentiation of Shiga toxin-producing, enteropathogenic, and avirulent strains of *Escherichia coli* O26. Appl. Environ. Microbiol. **77**:2275-2281.

78. **Allard, M. W., E. Strain, D. Melka, K. Bunning, S. M. Musser, E. W. Brown, and R. Timme**. 2016. Practical Value of Food Pathogen Traceability through Building a Whole-Genome Sequencing Network and Database. J Clin Microbiol. **54**:1975-1983.

79. **Oanh, T. K. N., V. K. Nguyen, H. De Greve, and B. M. Goddeeris**. 2012. Protection of Piglets against Edema Disease by Maternal Immunization with Stx2e Toxoid. Infect. Immun. **80**:469-473.





TABLES

Table 1. Frequency of STEC isolated by food commodity.

Commodities	No. strains	%
Spinach	88	31.88
Flour	59	21.38
Lettuce	35	12.68
Cilantro (coriander)	32	11.59
Cheese	9	3.26
Leafy greens	8	2.90
Kale	7	2.54
Basil	6	2.17
Pepper	6	2.17
Alfalfa sprouts	3	1.09
Cantaloupe	3	1.09
Parsley	3	1.09
Tomatoes	3	1.09
Creamy soy Nut butter	2	0.72
pizza dough dry mix	2	0.72
sprouts	2	0.72
Almond	1	0.36
oats animal feed	1	0.36
Celery	1	0.36
Clover sprouts	1	0.36
cucumbers	1	0.36
animal feed	1	0.36
enviromental	2	0.72
Total	276	16.30

The *stx*-negative strains were eliminated from this analysis.

Table 2. Frequency of STECs isolated from food commodities per year.

Year[a]	Frequency of Isolation	%	STs	Commodities	State
2003	3	1.09	3017, 641, 446	lettuce, celery, tomato	TX
2004	4	1.45	655, 205, 642	lettuce, cantaloupe, cilantro	TX, CA, MD
2005	1	0.36	677	lettuce	CA
2006	4	1.45	33,211,764,496	Alfalfa sprouts, lettuce	MI, CA, MN
2007	2	0.72	297, 295	Cantaloupe, lettuce	NY, OH
2008	10	3.62	21, 329, 6475, 642, 2217, 4496, 2008, 1385	Spinach, lettuce	CA, WA, MI
2009	11	3.99	223, 58, 11, 6509, 718, 4496, 2389, 6638, 2008, 5530	Spinach, lettuce, flour	FL, CA, OH, MI, TX, MD, WI, NY
2010	34	12.32	6640, 2520, 154, 661, 443, 205, 718, 173, 2387, 10, 706, 692, 4173, 1431, 4496, 5299, 3, 295, 5435, 2008, 1727	Spinach, lettuce, cilantro, sprouts, hot pepper, tomato	TX, CA, FL, WI, NY, CO, WA,
2011	38	13.77	6642, 942, 297, 6641, 955, 11, 679, 6639, 2161, 21, 692, 2217, 4496, 88, 306, 5435, 205, 2008, 5973, 691, 724	Spinach, cantaloupe, lettuce, Almond, Alfalfa sprouts, cilantro, hot pepper	MI, MN, CA, TX, OH, FL, CO
2012	40	14.49	993, 16, 223, 5975, 2520, 443, 1611, 119, 718, 677, 297, 173, 2387, 101, 21, 5602, 747, 906, 5395, 2385, 4496, 101, 442, 295	Cilantro, Flour, Spinach, Lettuce, Cherry tomatoes	TX, FL, NC, NY, CO, OH, MI, CA, WA
2013	13	4.71	223, 297, 325, 2388, 1611, 394, 677, 156, 937, 2385, 35, 515	Cilantro, basil, sprouts, parsley, flour	GA, CA, TX, TN, AZ
2014	36	13.04	993, 17, 16, 2520, 297, 329, 442, 679, 677, 297, 173, 657, 2387, 21, 342, 43, 906, 675, 6632, 2385, 88, 3759, 2217, 306, 29, 5960, 10	Lettuce, kale, cheese, clover sprouts, animal feed, basil, Leafy Greens, spinach	OH, CA, WI, AZ, WA, KY, OR, PA
2015	12	4.35	723, 1967, 1817, 25, 398, 32, 442, 205, 446, 21, 342, 40	Environmental, flour, lettuce, Leafy Greens, spinach, pepper, kale	WA, NE, AZ, CA, OR
2016	53	19.20	655, 747, 723, 4496, 154, 17, 1792, 297, 21, 33	Flour, kale, spinach	MO, CO, MI, OK, CA
2017	15	5.43	1112, 5082, 662, 1086, 162	cucumbers, soy nut butter, pepper, flour	KY
Total	276	100.00			

[a]- STECs isolated during years 2003-2009 are included as historical STECs and their prevalence was not used for determining the STEC frequency in foods regulated by the FDA per year.

Table 3. STs observed and number of strains included in each ST. Additionally information is provided for strains belonging to those STs such as: link to human cases, link to EHEC cases, and known serotypes. These additional reports are based on what it is reported in the *E. coli* section of the Enterobase database (http://enterobase.warwick.ac.uk).

ST[a]	No. Strains	%	Human cases	Reported as EHEC[b]	Known serotypes
655	38	13.67	+	+	O121:H19
4496	12	4.32	+	NR	O8:H28
2008	12	4.32	+	NR	Ounk:H2/40
21	8	2.88	+	+	O26:H11/-
297	9	3.24	+	- (UPEC, APEC)	diverse serotypes
205	7	2.52	+	- (UPEC)	NR
43	5	1.80	+	- (ETEC, EAEC)	O6:H10
154	5	1.80	+	- (EPEC, APEC)	diverse serotypes
173	5	1.80	+	NR	O181:H49
295	5	1.80	+	- (EAEC, UPEC, ExPEC)	diverse serotypes
677	5	1.80	+	+	diverse serotypes
747	8	2.88	+	- (ETEC)	diverse serotypes
11	3	1.08	+	+	O157:H7/-
16	2	0.72	+	+	O111:H8/2/-
17	2	0.72	+	+	O103:H2/-
25	1	0.36	+	+	O128:H2
29	1	0.36	+	+ (also EPEC)	O26:H11
32	1	0.36	+	+	O145:H-
33	2	0.72	+	+	O91:H14
119	1	0.36	+	+	O165:H25/28
223	4	1.44	+	+ (also UPEC, EAEC)	diverse serotypes
306	2	0.72	+	+	O84:H2/K+
329	2	0.72	+	+ (also EAEC)	diverse serotypes
657	1	0.36	+	+	diverse serotypes
675	1	0.36	+	+	O76:H19
679	3	1.08	+	+	O163:H19
724	1	0.36	+	+	O154/Ounk:H20
5299	2	0.72	+	NR	O8:H49
325	1	0.36	+	NR	O15:H16/K+
6639	1	0.36	+	NR	O174:H21/36
661	2	0.72	+	NR	O174:H2
662	2	0.72	+	NR	diverse serotypes
691	2	0.72	+	NR	diverse serotypes

692	3	1.08	+	NR	O74:H42
718	3	1.08	+	NR	O168:H8
723	4	1.44	+	NR	O103:H11
942	1	0.36	+	NR	O116:H28
955	1	0.36	+	NR	O139:H1/6
993	2	0.72	+	NR	O100:H30
1792	1	0.36	+	NR	O111:H8
1817	1	0.36	+	NR	O104:H7
1967	2	0.72	+	NR	O103:H2
2388	1	0.36	+	NR	O15
2520	3	1.08	+	NR	O116:H49
3759	1	0.36	+	NR	NR
5973	2	0.72	+	NR	Ounk:H2
6475	2	0.72	+	NR	O17/077:H45
10	2	0.72	+	- (mainly EAEC, UPEC, ETEC)	diverse serotypes
35	1	0.36	+	- (EPEC or UPEC)	O154:H4, O145:H34/31
40	1	0.36	+	- (EAEC)	O111ac:H21
58	1	0.36	+	- (EAEC, UPEC, ExPEC, APEC)	diverse serotypes
88	2	0.72	+	- (EAEC, UPEC, ExPEC, APEC)	diverse serotypes
101	3	1.08	+	- (EAEC, UPEC, ExPEC, APEC)	diverse serotypes
156	1	0.36	+	- (UPEC, EXPEC)	diverse serotypes
162	1	0.36	+	- (UPEC, APEC)	O8:H19
342	2	0.72	+	- (EPEC)	O177:NM
394	1	0.36	+	- (EAEC, UPEC)	diverse serotypes
398	1	0.36	+	- (ExPEC)	diverse serotypes
442	3	1.08	+	- (EPEC)	O146:H21
443	2	0.72	+	- (UPEC)	NR
446	2	0.72	+	- (APEC)	diverse serotypes
515	1	0.36	+	- (EAEC)	O2:H9
641	1	0.36	+	- (ExPEC)	diverse serotypes
642	3	1.08	+	- (EPEC)	diverse serotypes
706	1	0.36	+	- (UPEC)	diverse serotypes
906	3	1.08	+	- (UPEC)	diverse serotypes
1431	1	0.36	+	- (ExPEC)	O8:H19/30
1727	4	1.44	+	- (mostly nonpathogen)	diverse serotypes
937	1	0.36	+	- (ExPEC, non pathogen)	O43:H2
1385	1	0.36	-	- (APEC)	Ounk:H4
1611	2	0.72	-	- (APEC)	diverse serotypes
5082	1	0.36	-	NR	NR

332	1	0.36	-	NR	O171:H2
5395	1	0.36	-	NR	O74:H8
5435	2	0.72	-	NR	Ounk:H16
5530	1	0.36	-	NR	Ounk:H21
5602	2	0.72	-	NR	36:H28
5960	1	0.36	-	NR	NR
5975	1	0.36	-	NR	O113:H21
6509	1	0.36	-	NR	O168:H8
6632	1	0.36	-	NR	O8:H16
6638	1	0.36	-	NR	Ounk:H19
3017	1	0.36	-	NR	O116:H21
6640	1	0.36	-	NR	O113:H21
6641	1	0.36	-	NR	O130:H11
6642	1	0.36	-	NR	O113:H21
1112	1	0.36	-	- (nonpathogen)	diverse serotypes
1176	1	0.36	-	- (nonpathogen)	O36:H14
2161	1	0.36	-	- (nonpathogen)	O180:H14
2217	4	1.44	-	- (nonpathogen)	diverse serotypes
2389	1	0.36	-	- (nonpathogen)	NR
1086	10	3.60	-	- (nonpathogen)	diverse serotypes
2385	7	2.52	-	- (nonpathogen)	O8:H19
2387	8	2.88	-	- (nonpathogen)	O185:H7
4173	2	0.72	-	- (nonpathogen)	O79:H2

NR- not reported, UPEC (uropathogenic *E. coli*), EPEC (enteropathogenic *E. coli*), ETEC (Enterotoxigenic *E. coli*), APEC (Avian pathogenic *E. coli*), EAEC (Enteroaggregative *E. coli*) and ExPEC (Extraintestinal pathogenic *E. coli*).

[a]-Determined by *in silico* analysis of the WGS assemblies.

[b]-when negative, the reported *E. coli* type is stated.

Table 4. *in silico* characterization of STECs from this study for presence of virulence genes and their serotype.

strains	ST	serotype	*stx*1 type	*stx*2 type	eae type	*ehxA*	*espP*	*etpD*	*toxB*	*katP*	*subA*	*saa*	*sab*
CFSAN041120	10	O2:H27	-	a	-	+	-	-	-	-	-	-	-
CFSAN051538	10	Ounk:H32	a	-	-	-	+	-	-	-	-	-	-
CFSAN046715	11	O157:H7	-	a	gamma-1	+	+	+	+	+	-	-	-
CFSAN046720	11	O157:H7	-	a	gamma-1	+	+	+	+	+	-	-	-
FDA00009839	11	O157:H7	-	a	gamma-1	+	+	+	-	+	-	-	-
CFSAN053336	16	O111:H8	a	-	theta-2	+	-	-	-	-	-	-	-
CFSAN053346	16	O111:H8	a	a	theta-2	+	-	-	-	-	-	-	-
FDA00010429	17	O103:H2	a	-	epsilon	+	-	+	-	-	-	-	-
IEH-NGS-ECO-00075	17	O103:H2	a	-	epsilon	+	+	-	+	-	-	-	-
CFSAN046724	21	O26:H11	a	-	beta-1	+	+	-	-	+	-	-	-
CFSAN046724	21	O26:H11	a	-	beta-1	+	-	-	-	+	-	-	-
CFSAN053342	21	O103:H11	a	-	beta-1	+	+	-	+	+	-	-	-
CFSAN053343	21	O26:H11	a	-	beta-1	+	+	-	+	+	-	-	-
CFSAN053345	21	O26:H11	a	-	beta-1	+	+	-	+	+	-	-	-
FDA00010430	21	O26:H11	a	-	beta-1	-	+	-	-	+	-	-	-
IEH-NGS-ECO-00076	21	O26:H11	a	-	beta-1	+	+	-	+	+	-	-	-
IEH-NGS-ECO-00213	21	O26:H11	a	-	beta-1	+	+	-	+	-	-	-	-
IEH-NGS-ECO-00227	25	O128ac:H2	c	-	-	-	-	-	-	-	+	-	+
IEH-NGS-ECO-00125	29	Ounk:H11	-	a	beta-1	+	+	-	-	-	-	-	-
IEH-NGS-ECO-00224	32	O145:H28	a	d	gamma-1	+	+	-	+	+	-	-	-
CFSAN051773	33	O91:H14	a	d	-	+	-	-	-	-	+	+	+
CFSAN053344	33	O91:H14	a	d	-	+	-	-	-	-	+	-	+

IEH-NGS-ECO-00232	40	Ounk:H21	c	-	-	-	-	-	-	+	-	-	-
CFSAN051526	43	O6:H10	c	-	-	-	-	-	-	+	-	-	-
CFSAN051527	43	O6:H10	c	-	-	-	-	-	-	+	-	-	-
CFSAN051533	43	O6:H10	c	-	-	-	-	-	-	+	-	-	-
CFSAN051535	43	O6:H10	c	-	-	-	-	-	-	-	-	-	-
CFSAN051537	43	O6:H10	c	-	-	-	-	-	-	+	-	-	-
CFSAN046659	58	O116:H21	-	a	-	+	+	-	-	-	+	+	+
CFSAN046700	88	O8:H9	-	e	-	-	-	-	-	-	-	-	-
CFSAN051531	88	O8:H30	-	e	-	-	-	-	-	-	-	-	-
CFSAN046737	101	O82:H8	-	a	-	+	+	-	-	-	-	+	+
CFSAN046738	101	O82:H8	-	a	-	+	+	-	-	-	-	+	+
CFSAN046749	101	O21:H21	-	a	-	+	-	-	-	-	+	+	+
CFSAN046750	119	O165:H28	a	a	epsilon-2	+	+	-	-	+	-	-	-
CFSAN046672	154	O88:H25	a	a	-	+	-	-	-	-	-	+	+
CFSAN046673	154	O88:H25	a	a	-	+	-	-	-	-	-	+	+
CFSAN046682	154	O134:H38	a	d	-	+	-	-	-	-	-	+	+
CFSAN056112	154	O88:H25	a	a	-	+	-	-	-	-	-	+	+
CFSAN056113	154	O88:H25	a	a	-	+	-	-	-	-	-	+	+
CFSAN051557	156	O174:H28	-	d	-	-	-	-	-	-	-	-	-
FDA00011520	162	O8:H19	-	d	-	-	-	-	-	-	-	-	-
CFSAN046678	173	O181:H49	a	a	-	+	+	-	-	-	+	+	+
CFSAN046740	173	O181:H49	-	d	-	+	+	-	-	-	+	+	+
CFSAN046747	173	O181:H49	-	d	-	+	+	-	-	-	+	+	+
IEH-NGS-ECO-00108	173	O181:H49	-	a	-	+	-	-	-	-	+	+	+
CFSAN046751	173	O181:H49	-	d	-	+	-	-	-	-	+	+	+
CFSAN041109	205	Ounk:H19	-	a	-	+	-	-	-	-	+	+	+
CFSAN046650	205	O153/O178:H19	-	a	-	+	+	-	-	-	+	+	+

CFSAN046690	205	O153/O178:H19	a	a	-	+	+	-	-	-	+	+	+
CFSAN046704	205	Ounk:H19	-	a	-	+	+	-	-	-	+	+	+
CFSAN051550	205	Ounk:H19	-	a	-	+	+	-	-	-	+	+	+
CFSAN051551	205	Ounk:H19	-	a	-	+	+	-	-	-	+	+	+
FDA00009731	205	O153/O178:H19	a	d	-	+	-	-	-	-	-	+	+
CFSAN046660	223	O113:H21	-	a	-	+	+	-	-	-	+	+	+
CFSAN046665	223	O113:H21	-	a	-	+	+	-	-	-	+	+	+
CFSAN053329	223	O113:H21	-	a	-	+	+	-	-	-	+	+	+
CFSAN046734	223	O113:H21	-	a	-	+	+	-	-	-	+	+	+
CFSAN041114	295	Ounk:H16	-	a	-	+	-	-	-	-	+	+	+
CFSAN046640	295	Ounk:H16	c	b	-	-	-	-	-	-	-	-	+
CFSAN046691	295	Ounk:H11	-	a	-	+	+	-	-	-	+	+	+
CFSAN053347	295	Ounk:H16	-	a	-	+	+	-	-	-	+	+	+
CFSAN046735	295	Ounk:H11	-	a	-	+	-	-	-	-	+	+	+
CFSAN041108	297	O130:H11	a	d	-	+	-	-	-	-	+	+	+
CFSAN046639	297	O130:H11	-	a	-	+	-	-	-	-	+	+	+
CFSAN046719	297	O130:H11	-	d	-	+	-	-	-	-	+	+	+
CFSAN046753	297	O179:H8	-	a	-	+	+	-	-	-	+	+	+
CFSAN051516	297	O130:H11	-	a	-	+	-	-	-	-	+	+	+
CFSAN051524	297	O130:H11	-	a	-	+	-	-	-	-	+	+	+
CFSAN051549	297	O130:H11	a	d	-	+	-	-	-	-	+	+	+
CFSAN056111	297	O130:H11	-	a	-	+	-	-	-	-	+	+	+
IEH-NGS-ECO-00087	297	O179:H8	-	a	-	+	-	-	-	-	+	+	+
CFSAN046717	306	O98:H21	a	-	zeta	+	+	-	-	-	-	-	-
CFSAN051525	306	O98:H21	a	-	zeta	+	+	-	-	-	-	-	-
CFSAN051544	325	O15:H16	-	g	-	-	-	-	-	-	-	-	-
CFSAN046643	329	O136:H16	a	-	-	+	-	-	-	-	-	-	-

CFSAN053337	329	O136:H16	a	-	-	+	-	-	-	-	-	-	-
CFSAN046636	332	O171:H2	-	c	-	-	-	-	-	-	-	-	-
CFSAN053335	342	O5:Hunk	a	-	beta-1	+	+	-	-	-	-	-	-
IEH-NGS-ECO-00230	342	O5:Hunk	a	-	beta-1	+	-	-	-	-	-	-	-
CFSAN053330	394	O17/O77:H18	-	d	-	+	+	-	-	-	+	+	+
IEH-NGS-ECO-00231	398	O136:H20	c	-	-	-	-	-	-	+	-	-	-
CFSAN046746	442	O91:H21	-	a	-	+	-	-	-	-	-	+	+
CFSAN051529	442	O146:H21	c	b	-	+	-	-	-	-	+	-	+
CFSAN051540	442	O146:H21	c	-	-	+	-	-	-	-	+	-	+
CFSAN046688	443	O153/O178:H19	a	d	-	+	-	-	-	-	-	+	+
CFSAN046752	443	O153/O178:H19	a	d	-	-	-	-	-	-	-	+	+
CFSAN046631	446	O22:H8	-	c	-	-	-	-	-	-	-	-	+
FDA00009425	446	O22:H8	-	c	-	-	-	-	-	-	-	-	-
CFSAN053334	515	Ounk:H29	-	d	-	-	-	-	-	-	-	-	-
CFSAN046632	641	O117:H10	a	-	-	-	+	-	-	-	-	+	-
CFSAN046646	642	O187:H52	c	-	-	-	-	-	-	-	-	-	-
CFSAN046648	642	O187:H52	c	-	-	-	-	-	-	-	-	-	-
CFSAN046649	642	O187:H52	c	-	-	-	-	-	-	-	-	-	-
CFSAN046651	655	O121:H19	-	a	epsilon-2	+	-	-	-	-	-	-	-
FDA00010253	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010254	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010255	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010256	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010257	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010258	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010259	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010276	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-

FDA00010277	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010278	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010279	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010280	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010281	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010282	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010283	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010284	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010285	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010296	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010297	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010298	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010299	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010300	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010301	655	O121:H19	-	a	epsilon-2	-	+	-	-	-	-	-	-
FDA00010302	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010303	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010304	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010305	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010306	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010307	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010308	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010309	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
FDA00010310	655	O121:H19	-	a	epsilon-2	-	+	-	-	-	-	-	-
FDA00010369	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010370	655	O121:H19	-	a	epsilon-2	-	+	-	-	-	-	-	-
FDA00010371	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-

FDA00010372	655	O121:H19	-	a	epsilon-2	+	+	-	-	-	-	-	-
FDA00010373	655	O121:H19	-	a	epsilon-2	-	-	-	-	-	-	-	-
IEH-NGS-ECO-00103	**657**	O183:H18	a	d	-	+	+	-	-	-	+	+	+
CFSAN041119	661	O174:H2	-	c	-	+	+	-	-	-	-	+	+
CFSAN051547	661	O174:H2	-	c	-	+	-	-	-	-	-	+	+
FDA00011331	662	O73 or O17/O77:H45	a	d	-	+	+	-	-	-	-	+	+
FDA00011332	662	O73 or O17/O77:H45	a	d	-	+	+	-	-	-	-	+	+
CFSAN051530	**675**	O76:H19	c	-	-	+	-	-	-	-	+	-	+
CFSAN046652	677	Ounk:H21	-	d	-	-	-	-	-	-	-	-	-
CFSAN046748	677	O174:H21	a	d	-	+	+	-	-	-	+	+	+
CFSAN051519	677	O174:H21unk	-	a	-	-	-	-	-	-	-	-	+
CFSAN051520	**677**	O174:H21	-	a	-	-	-	-	-	-	-	-	+
IEH-NGS-ECO-00080	677	O174?	-	d	-	-	-	-	-	-	-	-	-
CFSAN046699	679	O163:H19	-	d	-	+	+	-	-	-	+	+	+
CFSAN046705	679	O163:H19	-	a	-	+	+	-	-	-	+	+	+
CFSAN051541	679	O163:H19	a	d	-	+	+	-	-	-	+	+	+
CFSAN053338	691	Ounk:H20	a	d	-	+	+	-	-	-	-	+	+
CFSAN046703	691	Ounk:H20	a	-	-	+	+	-	-	-	-	+	+
CFSAN046709	691	Ounk:H20	a	-	-	+	+	-	-	-	-	+	+
CFSAN041113	692	O74:H42	a	d	-	+	+	-	-	-	-	+	+
CFSAN051548	692	O74:H42	a	d	-	+	+	-	-	-	-	+	+
CFSAN051553	692	O74:H42	a	d	-	+	+	-	-	-	-	+	+
CFSAN046689	706	O32:H1	-	d	-	-	-	-	-	-	-	-	+
CFSAN046666	718	O168:H8	-	a	-	+	-	-	-	-	-	-	-
CFSAN046693	718	O168:H8	-	d	-	-	-	-	-	-	-	-	-
CFSAN046725	718	O168:H8	-	d	-	-	-	-	-	-	-	-	-
FDA00010428	723	O103:H11	a	-	beta-1	-	-	-	-	+	-	-	-

FDA00010431	723	O103:H11	a	-	beta-1	+	-	-	-	+	-	-	-
FDA00010457	723	O103:H11	a	-	beta-1	+	+	-	-	+	-	-	-
IEH-NGS-ECO-00223	723	O103:H11	a	-	beta-1	+	+	-	+	+	-	-	-
CFSAN046707	724	Ounk:H20	a	-	-	+	+	-	-	-	+	+	+
CFSAN046733	747	O73 or O17/O77:H45	-	a	-	+	+	-	-	-	-	+	+
CFSAN046741	747	O73 or O17/O77:H45	-	a	-	+	+	-	-	-	-	+	+
CFSAN046742	747	O73 or O17/O77:H45	-	a	-	+	+	-	-	-	-	+	+
CFSAN046743	747	O73 or O17/O77:H45	-	a	-	+	+	-	-	-	-	+	+
FDA00010882	747	O73 or O17/O77:H45	-	a	-	+	+	-	-	-	-	+	+
FDA00010883	747	O73 or O17/O77:H45	-	a	-	+	+	-	-	-	-	+	+
FDA00010884	747	O73 or O17/O77:H45	-	a	-	+	+	-	-	-	-	+	+
FDA00010885	747	O73 or O17/O77:H45	-	a	-	+	+	-	-	-	-	+	+
CFSAN041116	906	O74:H8	-	d	-	+	+	-	-	-	+	+	+
CFSAN051555	906	O74:H8	-	d	-	+	+	-	-	-	+	+	+
IEH-NGS-ECO-00096	906	O74:H8	-	a	-	+	+	-	-	-	+	+	+
CFSAN051545	937	O43:H2	a	a	-	+	-	-	-	-	-	+	+
CFSAN046721	942	O116:H28	a	-	-	+	+	-	-	-	-	+	+
CFSAN046713	955	O139:H1	-	e	-	-	-	-	-	-	-	-	-
CFSAN051539	993	O100:H30	-	e	-	-	-	-	-	-	-	-	-
CFSAN046730	993	O100:H30	-	e	-	-	-	-	-	-	-	-	-
FDA00011815	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-
FDA00011816	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-
FDA00011817	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-
FDA00011818	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-
FDA00011819	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-
FDA00011820	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-
FDA00011821	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-

FDA00011822	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-
FDA00011823	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-
FDA00011824	1086	O8:H14	-	d/e	-	-	-	-	-	-	-	-	-
FDA00011218	1112	O142:H27?	-	e	-	-	-	-	-	-	-	-	-
CFSAN046653	1176	O36:H14	-	g	-	+	-	-	-	-	-	-	-
CFSAN046655	1385	Ounk:H4	-	d	-	-	-	-	-	-	-	-	-
CFSAN046683	1431	O8:H19	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN046728	1611	O159:H19	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN051515	1611	O159:H19	-	d/e	-	-	-	-	-	-	-	-	-
CFSAN046668	1727	Ounk:H7	-	c	-	-	-	-	-	-	-	-	-
CFSAN046669	1727	Ounk:H7	-	c	-	-	-	-	-	-	-	-	-
CFSAN046671	1727	Ounk:H7	-	c	-	-	-	-	-	-	-	-	-
CFSAN046692	1727	Ounk:H7	-	c	-	-	-	-	-	-	-	-	-
FDA00010432	1792	O111:H8	a	a	theta-2	+	-	-	-	-	-	-	-
IEH-NGS-ECO-00221	1817	O104:H7	c	-	-	-	-	-	-	+	-	-	-
IEH-NGS-ECO-00211	1967	O103:H2	a	-	epsilon	+	+	-	-	+	-	-	-
IEH-NGS-ECO-00212	1967	O103:H2	a	-	epsilon	+	+	-	-	-	-	-	-
CFSAN041110	2008	Ounk:H2	-	d	-	-	-	-	-	-	-	-	-
CFSAN041111	2008	Ounk:H2	-	d	-	-	-	-	-	-	-	-	-
CFSAN041112	2008	Ounk:H2	-	d	-	-	-	-	-	-	-	-	-
CFSAN046642	2008	Ounk:H2	-	a	-	-	-	-	-	-	-	-	-
CFSAN046657	2008	Ounk:H2	-	d/e	-	-	-	-	-	-	-	-	-
CFSAN046684	2008	Ounk:H2	-	d/c	-	-	-	-	-	-	-	-	-
CFSAN046685	2008	Ounk:H2	-	d/c	-	-	-	-	-	-	-	-	-
CFSAN046686	2008	Ounk:H2	-	d/c	-	-	-	-	-	-	-	-	-
CFSAN046687	2008	Ounk:H2	-	d/c	-	-	-	-	-	-	-	-	-
CFSAN046708	2008	Ounk:H2	-	d/c	-	-	-	-	-	-	-	-	-

CFSAN046710	2008	Ounk:H2	-	d/c	-	-	-	-	-	-	-	-	-
CFSAN051552	2008	Ounk:H2	-	d/c	-	-	-	-	-	-	-	-	-
CFSAN046701	2161	O180:H14	-	d/e	-	-	-	-	-	-	-	-	-
CFSAN046641	2217	O45:H16	a	-	-	-	+	-	-	-	-	+	-
CFSAN046712	2217	O76:H21	a	-	-	-	-	-	-	-	-	+	-
CFSAN046714	2217	O8:H16	a	-	-	-	+	-	-	-	-	+	-
CFSAN051522	2217	O84:H38	a	-	-	-	+	-	-	-	-	+	-
CFSAN046723	2385	O8:H19	a	a	-	+	+	-	-	-	+	+	+
CFSAN051517	2385	O8:H19	a	a	-	+	+	-	-	-	+	+	+
CFSAN051518	2385	O8:H19	a	a	-	+	+	-	-	-	+	+	+
IEH-NGS-ECO-00082	2385	O8:H19	a	a	-	+	+	-	-	-	+	+	+
IEH-NGS-ECO-00088	2385	O8:H19	-	a	-	+	+	-	-	-	+	+	+
IEH-NGS-ECO-00089	2385	O8:H19	-	a	-	+	+	-	-	-	+	+	+
IEH-NGS-ECO-00090	2385	O8:H19	a	c	-	+	+	-	-	-	+	+	+
CFSAN041115	2387	O185:H7	-	c	-	-	-	-	-	-	-	-	-
CFSAN041118	2387	O185:H7	-	c	-	-	-	-	-	-	-	-	-
CFSAN046670	2387	O185:H7	-	c	-	-	-	-	-	-	-	-	-
CFSAN051528	2387	O185:H7	-	c	-	+	+	-	-	-	-	-	-
CFSAN051546	2387	O185:H7	-	c	-	+	+	-	-	-	-	-	-
CFSAN051554	2387	O185:H7	-	c	-	-	-	-	-	-	-	-	-
CFSAN046744	2387	O185:H7	-	c	-	-	-	-	-	-	+	+	+
CFSAN046745	2387	O185:H7	-	c	-	-	+	-	-	-	+	+	+
CFSAN053339	2388	O15:H27	-	d	-	-	-	-	-	-	-	-	-
CFSAN046656	2389	Ounk:H11	-	d	-	+	+	-	-	-	+	+	+
CFSAN046679	2520	O116:H49	-	a	-	+	+	-	-	-	+	+	+
CFSAN046731	2520	O116:H49	-	a	-	+	+	-	-	-	+	+	+
IEH-NGS-ECO-00204	2520	O116:H49	a	a	-	+	+	-	-	-	+	+	+

CFSAN046633	3017	O116:H21	a	a	-	+	-	-	-	-	+	+	+
IEH-NGS-ECO-00070	3759	O8:H49	-	a	-	+	-	-	-	-	+	+	+
CFSAN046667	4173	O79:H2	a	-	-	-	-	-	-	-	-	-	-
CFSAN046676	4173	O79:H2	a	-	-	-	-	-	-	-	-	-	-
CFSAN046637	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN046644	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN046654	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN046661	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN046680	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN046695	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN046718	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN046726	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN046727	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
FDA00009866	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
FDA00009867	4496	O8:H28	-	d/e	-	-	-	-	-	-	-	-	-
CFSAN046729	4496	O8:H28	-	d/e	-	+	-	-	-	-	-	-	-
FDA00011519	5082	O180:H14	-	d/e	-	-	-	-	-	-	-	-	-
CFSAN046674	5299	O8:H49	a	-	-	-	-	-	-	-	-	+	+
CFSAN046675	5299	O8:H49	a	-	-	+	+	-	-	-	-	+	+
CFSAN046722	5395	O74:H8	a	-	-	+	+	-	-	-	+	+	+
CFSAN046694	5435	Ounk:H16	-	a	-	+	-	-	-	-	+	+	+
CFSAN046696	5435	Ounk:H16	-	a	-	+	+	-	-	-	+	+	+
CFSAN046664	5530	Ounk:H21	-	d/e	-	-	-	-	-	-	-	-	-
CFSAN041117	5602	O36:H28	-	g	-	+	-	-	-	-	-	-	-
CFSAN051556	5602	O36:H28	-	g	-	+	-	-	-	-	-	-	-
IEH-NGS-ECO-00105	5960	Ounk:H19	-	a	-	+	+	-	-	-	+	+	+
CFSAN046702	5973	Ounk:H2	-	d	-	+	+	-	-	-	+	+	+

CFSAN046716	5973	Ounk:H2	-	c	-	+	+	-	-	-	+	+	+
CFSAN046739	5975	O113:H21	-	a	-	+	+	-	-	-	+	+	+
CFSAN046645	6475	O17/O77:H45	a	a	-	+	+	-	-	-	-	+	+
CFSAN046647	6475	O17/O77:H45	a	a	-	+	+	-	-	-	-	-	-
CFSAN046663	6509	O168:H8	-	d/e	-	+	-	-	-	-	-	-	-
CFSAN051521	6632	O8:H16	-	d	-	-	+	-	-	-	-	+	-
CFSAN046662	6638	Ounk:H19	-	a	-	+	-	-	-	-	+	+	+
CFSAN046711	6639	O174:H21	-	c	-	-	+	-	-	-	-	-	-
CFSAN046677	6640	O113:H21	-	d	-	+	+	-	-	-	+	+	+
CFSAN046706	6641	O130:H11	a	-	-	+	-	-	-	-	+	+	+
CFSAN046697	6642	O113:H21	-	a	-	+	+	-	-	-	+	+	+

Table 5. Presence of antimicrobial resistance genes identified by *in silico* analysis in the 276 STEC genomes analyzed in this study.

Strains	*aadA*[a]	*aph3*[a]	*strA*[a]	*strB*[a]	*bla TEM*[b]	*floR*[d]	*QnrB*[e]	*sul1*[f]	*sul2*[f]	*sul3*[f]	*tetA*[g]	*tetB*[g]	*tetC*[g]	*dfrA*[ha]
CFSAN053338	-	-	+	+	-	-	-	-	-	-	-	-	-	-
CFSAN046714[g]	-	-	-	-	-	-	-	+	-	+	-	-	+	-
CFSAN051552	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN046710	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN046642	-	-	+	+	-	-	-	-	+	-	-	+	-	-
IEH-NGS-ECO-00231	-	-	-	-	-	-	-	-	-	-	+	-	-	-
CFSAN053336	-	+	+	+	-	-	-	-	+	-	+	-	-	-
FDA00009425	-	-	+	+	-	+	-	-	+	-	+	-	-	-
FDA00011218	-	-	-	-	-	-	-	-	-	-	+	-	-	-
IEH-NGS-ECO-00230[i]	+	-	+	+	-	+	-	+	+	+	-	-	-	-
CFSAN051526	+	-	-	-	-	-	-	+	-	+	+	-	-	-
CFSAN046636	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN046669	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN046668	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN053334	-	-	+	+	-	+	-	-	+	-	+	-	-	-
CFSAN046687	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN051521	-	-	-	-	-	-	-	-	-	-	+	-	-	-
CFSAN046730	-	+	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN041112	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN041111	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN046671	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN046708	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN046685	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN046684	-	-	+	+	-	-	-	-	+	-	-	+	-	-
CFSAN051527	+	-	-	-	-	-	-	+	-	+	+	-	-	-
CFSAN046686	-	-	+	+	-	-	-	-	+	-	-	+	-	-

CFSAN051535	-	-	+	+	+	-	-	-	+	-	+	-	-	-
CFSAN046693	-	-	+	+	-	-	-	-	+	-	-	-	-	-
CFSAN046713	-	-	-	+	+	-	+	-	+	-	+	-	-	+
CFSAN046725	-	-	-	+	-	-	-	-	+	-	-	-	-	-
CFSAN051531	-	-	-	+	-	-	-	-	+	-	+	-	-	+
CFSAN051539[j]	+	+	-	-	+	-	-	-	-	+	-	+	-	-
CFSAN041110	-	-	+	+	-	-	-	-	+	-	-	+	-	-

[a]Aminoglycoside, [b]Beta-lactamase, [c]Macrolide, [d]Phenicol, [e]Quinolone, [f]Sulphonamide, [g]Tetracycline, and [h] Trimethoprim.
[g] strain carrying *blaOXA*[b] gene. [I] strain carrying *blaCMY-2*[b] gene. [j] strain carrying *mef(B)*[c], *cml*[d] and *cmlA1*[d] genes.

Table 6. ST, serotype and virulence profile of intimin positive non-STEC isolated from FDA regulated foods (2003-2017).

strains	ST	Serotype	eae type	*astA*	*bfpA*	*efa1*	*espC*	*espF*	*espJ*	*gad*	*nleA*	*nleB*	*nleC*	*espB*	*espI*	*espK*	*celb*	*cif*	*ehxA*	*iss*	*lp*
205	10	Ounk:H40	theta-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	
107	17	O103:H2	beta-1	-	-	-	-	-	+	-	+	-	+	+	-	+	-	+	-	-	
101	20	O15:H2	beta-1	-	-	+	-	+	+	-	+	+	-	+	-	-	-	+	-	+	
110	20	O15:H2	beta-1	-	-	+	-	+	+	+	+	-	-	+	-	-	-	+	-	+	
123	20	Ounk:H2	beta-1	-	-	-	-	+	+	-	+	+	-	+	-	-	-	+	-	+	
226	20	O128ac:H2	beta-1	+	-	-	-	-	-	-	+	+	-	+	-	-	-	+	-	+	
233	20	O51:H2	beta-1	-	-	-	-	+	+	-	+	+	-	+	-	-	-	+	-	+	
208	20	O51:H2	beta-1	+	-	-	-	+	+	-	+	+	-	+	-	-	-	+	-	+	
94	28	O167:H6	beta-2	-	+	-	-	-	-	+	+	+	-	-	-	-	-	+	-	-	
109	28	Ounk:Hunk	beta-2	-	-	-	-	-	-	-	+	+	-	-	-	-	-	+	-	-	
100	28	O167:H6	beta-2	-	+	-	-	-	-	+	+	-	-	-	-	-	-	+	-	-	
83	122	O63:H6	alpha-2	-	-	-	+	+	-	+	-	+	-	-	-	-	-	+	-	-	
86	122	O63:H6	alpha-2	-	-	-	+	+	-	-	-	+	-	-	-	-	-	+	-	-	
91	122	O63:H6	alpha-2	-	-	-	+	+	-	+	-	+	-	-	-	-	-	+	-	-	
95	327	O156:H8	theta-2	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	+	
229	327	O156:H8	theta-2	-	-	-	-	-	+	-	-	+	-	-	-	-	-	+	-	+	
93	328	Ounk:H7	beta-1	-	-	+	-	+	+	-	+	+	-	+	-	-	-	+	-	-	
98	328	Ounk:H7	beta-1	-	-	+	-	+	+	-	+	+	-	+	-	-	-	+	-	-	
209	337	O108:H21	theta-2	-	-	-	-	-	+	-	-	+	-	-	-	-	-	+	-	+	
210	337	O108:H21	theta-2	-	-	-	-	-	+	-	-	+	-	-	-	-	-	+	-	+	
214	337	O108:H21	theta-2	-	-	-	-	-	+	-	-	+	-	-	-	-	-	+	-	-	
124	342	O145:H28	beta-1	+	-	+	-	-	+	-	+	+	-	+	+	-	+	-	+	+	

225	442	Ounk:H21	theta-2	-	-	-	-	-	+	+	-	-	-	-	-	-	-	+	-	-
228	442	Ounk:H21	theta-2	+	-	-	-	-	+	+	-	+	-	-	-	-	+	+	-	-
234	442	Ounk:H21	theta-2	-	-	-	-	+	+	-	-	+	-	-	-	-	+	+	-	-
235	442	O146:H21	theta-2	+	-	-	-	-	+	+	-	+	-	-	-	-	-	+	-	-
99	517	Ounk:H19	epsilon-2	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-
84	582	O132:H34	alpha-2	-	-	-	+	-	-	+	-	+	+	-	-	-	-	-	-	-
85	582	O132:H34	alpha-2	-	-	-	+	+	-	+	-	+	+	-	-	-	-	-	-	-
97	582	O132:H34	alpha-2	-	-	-	+	+	-	-	-	+	+	-	-	-	-	-	-	-
215	800	O121:H19	beta-1	+	-	+	-	+	+	-	+	+	+	+	-	-	-	-	-	+
206	1092	O179:H31	zeta-3	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-
111	2166	Ounk:H19	theta-2	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	+
222	4268	Ounk:H45	delta	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-
236	5965	O107:H45	delta	-	-	-	-	+	+	-	-	+	-	-	-	-	-	+	-	-
E2348/69[b]	15	O127:H6	alpha-1	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-

[a]All strains designation start with IEH-NGS-ECO, except the [b]prototypic EPEC strain.

All strains were positive for *espA, tir, pssA*, and *air* genes. Only E2348/69b was positive for the perA gene.