Characterization of Shiga-Toxin Producing Escherichia coli Isolated From Cattle and Sheep in Xinjiang Province, China, Using Whole-Genome Sequencing

Yingyu Liu[1], Huoming Li[2], Xuhua Chen[2], Panpan Tong[1], Yan Zhang[1], Mingyue Zhu[1], Zhanqiang Su[1], Gang Yao[1], Ganwu Li[3], and Wentong Cai[2]

[1]Xinjiang Agricultural University
[2]Chinese Academy of Agricultural Sciences Harbin Veterinary Research Institute
[3]Iowa State University

June 8, 2020

Abstract

Shiga toxin-producing Escherichia coli (STEC) is an important foodborne pathogen capable of causing severe gastrointestinal diseases in humans. Cattle and sheep are the natural reservoir hosts of STEC strains. Previously, we isolated 56 STEC strains from anal and carcass swab samples of cattle and sheep in farms and slaughterhouses. In this study, we performed whole-genome sequencing of these isolates and determined their serotypes, virulence profiles, sequence types (STs), and genetic relationships. Our results showed that the 56 isolates belong to 20 different STs, 29 O:H serotypes, and 8 stx subtype combinations. The highly prevalent serotypes were O8:H25 and O87:H16 for bovine and ovine isolates, respectively. Five serotypes of cattle or sheep isolates are novel. The majority (63%) of cattle isolates contain stx1+stx2, subtyped into stx1a, stx2a, and stx2c. In contrast, most of the sheep isolates contain stx1 only, primarily subtyped into stx1a and stx1c. None of the isolates tested eae-positive, but virulence factors such as ehxA and espP were present with variable prevalence rates. The prevalence of saa (19.6%) and espP (12.5%) in cattle isolates is much higher than that in sheep isolates, whereas that of subA (34%), katP (14.3%), and ireA (28.6%) in sheep isolates is considerably higher than that in cattle isolates. Core-genome SNP analysis revealed that the majority of isolates could be clustered based on their serotypes or STs, whereas some clustering is associated with more than one ST or serotype. Seven-gene Multilocus Sequence Typing (MLST) indicated that nine sheep isolates and four cattle isolates were related to a few E. coli isolates associated with human HUS, suggesting their potential in causing severe human infections. Collectively, we described the characteristics of cattle and sheep STEC isolates from Xinjiang, China, which may be utilized in comparative studies of other geographic regions and sources of isolation and for surveillance.

Characterization of Shiga-Toxin Producing *Escherichia coli* Isolated From Cattle and Sheep in Xinjiang Province, China, Using Whole-Genome Sequencing

Running title: Analysis of Shiga-Toxin Producing *Escherichia coli*

Yingyu Liu[1#], Huoming Li[2#], Xuhua Chen[2,3], Panpan Tong[1], Yan Zhang[1], Mingyue Zhu[1], Zhanqiang Su[1], Gang Yao[1], Ganwu Li[2,3*], Wentong Cai[2*]

[1] College of Veterinary Medicine, Xinjiang Agricultural University, Urumqi, China

[2] Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin, China

[3] Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa, USA

# These authors contributed equally to this study.

* Corresponding author:

Wentong Cai, PhD. caiwentong@caas.cn

Ganwu Li, PhD. liganwu@caas.cn

Tel: +86- 451-51997177

Fax: +86- 451-51997166

Summary: Shiga toxin-producing *Escherichia coli* (STEC) is an important foodborne pathogen capable of causing severe gastrointestinal diseases in humans. Cattle and sheep are the natural reservoir hosts of STEC strains. Previously, we isolated 56 STEC strains from anal and carcass swab samples of cattle and sheep in farms and slaughterhouses. In this study, we performed whole-genome sequencing of these isolates and determined their serotypes, virulence profiles, sequence types (STs), and genetic relationships. Our results showed that the 56 isolates belong to 20 different STs, 29 O:H serotypes, and 8*stx* subtype combinations. The highly prevalent serotypes were O8:H25 and O87:H16 for bovine and ovine isolates, respectively. Five serotypes of cattle or sheep isolates are novel. The majority (63%) of cattle isolates contain *stx* 1+*stx* 2, subtyped into*stx* 1a, *stx* 2a, and *stx* 2c. In contrast, most of the sheep isolates contain *stx* 1 only, primarily subtyped into*stx* 1a and *stx* 1c. None of the isolates tested*eae* -positive, but virulence factors such as *ehxA* and*espP* were present with variable prevalence rates. The prevalence of *saa* (19.6%) and *espP* (12.5%) in cattle isolates is much higher than that in sheep isolates, whereas that of *subA*(34%), *katP* (14.3%), and *ireA* (28.6%) in sheep isolates is considerably higher than that in cattle isolates. Core-genome SNP analysis revealed that the majority of isolates could be clustered based on their serotypes or STs, whereas some clustering is associated with more than one ST or serotype. Seven-gene Multilocus Sequence Typing (MLST) indicated that nine sheep isolates and four cattle isolates were related to a few *E. coli* isolates associated with human HUS, suggesting their potential in causing severe human infections. Collectively, we described the characteristics of cattle and sheep STEC isolates from Xinjiang, China, which may be utilized in comparative studies of other geographic regions and sources of isolation and for surveillance.

Keyword: Shiga-toxin producing *Escherichia coli* ; Whole-genome sequencing; Cattle and sheep; Xinjiang province

Introduction

Shiga toxin-producing *Escherichia coli* (STEC) is a potentially deadly foodborne pathogen that causes diarrhea, hemorrhagic colitis (HC), and hemolytic uremic syndrome (HUS) in humans worldwide (Smith, Fratamico, & Gunther, 2014). Enterohemorrhagic *E. coli* (EHEC) is a subgroup of STEC characterized by certain serogroups, which are often associated with outbreaks and severe illnesses. In the US alone, EHEC outbreaks have occurred nearly every year in the last 10 years (CDC, 2017; Fao/Who Stec Expert, 2019). The most common EHEC serogroups are O157, O26, O121, O103, O111, and O145. Although O157 is the predominant serogroup in STEC infections, recent reports have demonstrated that non-O157 STECs are emerging as more impactful pathogens associated with human infections and foodborne illness outbreaks. In several countries including the US, non-O157 associated human infection cases have exceeded that caused by O157 isolates (CDC, 2017; Valilis, Ramsey, Sidiq, & DuPont, 2018). Thus far, over 200 non-O157 serotypes of STEC have been discovered and linked to human diseases across the globe, among which the top six serogroups are: O26, O45, O103, O111, O121, and O145 (Conrad, Stanford, McAllister, Thomas, & Reuter, 2014). Certain serotypes that are usually associated with other pathotypes have recently caused outbreaks such as O104:H4 (Johura et al., 2016; Rasko et al., 2011).

STEC is defined by the production of one or more types of Shiga toxin, which, upon entry into host cells, inhibit the protein synthesis of host cells, eventually leading to cell death. Stx consists of two major types, namely, Stx1 and Stx2, which are closely related and encoded by*stx* 1 and *stx* 2, respectively (Melton-Celsa, 2014). Based on sequence and biological differences, *stx* 1 is classified into three subtypes (*stx* 1a, *stx* 1c, and

stx 1d), and *stx* 2 into seven subtypes (*stx* 2a, *stx* 2b, *stx* 2c,*stx* 2d, *stx* 2e, *stx* 2f, and *stx* 2g) (Scheutz et al., 2012). Certain subtypes have been linked to human infections. For instance, STEC carrying*stx* 2a, *stx* 2c, or *stx* 2d is frequently associated with patients with HUS (Fruth, Prager, Tietze, Rabsch, & Flieger, 2015), whereas *stx* 2e is commonly detected in diseased pigs (Tseng et al., 2014). Aside from Shiga toxins, STEC strains utilize additional virulence factors that allow them to attach, colonize, invade, and cause damage. The locus of enterocyte effacement (LEE) pathogenicity island is responsible for producing attaching and effacing (A/E) lesions on intestinal epithelial cells. Specific contributing factors include intimin encoded by the *eae* gene, secreted effector proteins (Esp), an intimin receptor encoded by the *tir* gene, and others present in the LEE island (Galli, Miliwebsky, Irino, Leotta, & Rivas, 2010). However, LEE-negative STEC strains have been shown to be culprits in some human HUS cases. These could possess other genes responsible for attachment and colonization such as STEC autoagglutinating adhesin (*saa*) (Paton, Srimanote, Woodrow, & Paton, 2001). Key virulence genes also consist of *ehxA* , *espP* ,*etpD* , *toxB* , *katP* , *subA* , *saa* , and*sab* genes. The EHEC hemolysin encoded by *ehxA* is a cytotoxin expressed during infection and is produced by a great majority of EHEC serogroups, mostly frequently linked to HUS (Bielaszewska, Aldick, Bauwens, & Karch, 2014; Schmidt, Kernbach, & Karch, 1996). The presence of *ehxA* along with *stx* 2 and*eae* represents a significant risk factor for severe clinical manifestations, particularly HUS (Boerlin et al., 1999).

As DNA sequencing cost decreases, whole-genome sequencing (WGS) has become increasingly popular in characterizing clinical/environmental bacterial isolates because the serotype, virulence genes, and antimicrobial resistance profile can be well predicted from whole-genome sequences. Additionally, DNA sequence data can provide much better resolution for strain discrimination than any subtyping method used so far for outbreak detection and bacteria tracking, rendering it the most powerful tool in revealing phylogenetic relationships (Bergholz, Moreno Switt, & Wiedmann, 2014; Dallman et al., 2015). Importantly, WGS-based analysis tools have been developed and applied to the characterization of STEC isolates (Ferdous et al., 2016; Gonzalez-Escalona & Kase, 2019; Gonzalez-Escalona et al., 2016). With sequencing coverages between 30 and 40, serotypes, virulence genes, and antibiotic resistance genes/sites can be accurately predicted, thereby providing a faster, cheaper, and better analytical method (Lindsey, Pouseele, Chen, Strockbine, & Carleton, 2016).

Cattle and sheep are ruminants that have been demonstrated to be important reservoirs for non-O157 STEC strains. Foods such as uncooked meat, unpasteurized milk, and vegetables contaminated by STEC strains have been frequently associated with human HC and HUS cases (CDC, 2017) and constitute a constant threat to human health (Djordjevic et al., 2004). Xinjiang Province is one of China's largest provinces with vast pasture land and is thus known for breeding and husbandry of cattle and sheep. Large quantities of foods derived from cattle and sheep are consumed by locals as well as transported to neighboring regions. Cattle and/or sheep isolates from other parts of China were earlier studied using traditional molecular and typing methods (Bai et al., 2016; Fan et al., 2019). Previously, we isolated 56 STEC strains from cattle and sheep in Xinjiang, but the molecular characteristics and phylogeny of these STEC isolates remain untested. Thus, we here characterized this collection of STEC strains in detail based on whole-genome sequencing (WGS) and assessed their pathogenic potential.

Materials and Methods

The collection of STEC isolates

A total of 56 STEC strains were gleaned from anal or carcass swabs of cattle and sheep in Xinjiang Province from April 2016 to September 2018. Among them, 27 bovine isolates were obtained from 5 different farms and slaughterhouses in City of Yili, while 29 ovine isolates from 3 farms and slaughterhouses in Cities of Changji and Korla. These isolates have been shown to carry *stx* 1 and/or *stx* 2 by PCR genotyping (Liu et al., 2020; Zhang et al., 2018). *E. coli*strains were cultured in Luria-Bertani broth (BD, China) or MacConkey Agar medium (AOBOX, Beijing, China).

DNA preparation, genome sequencing, and assembly

E. coli was inoculated into Luria Bertani broth (BD, China) and grown at 37°C with shaking for at least

12 h. Cells were collected by centrifugation, and genomic DNA (gDNA) was extracted using a Takara MiniBest Genomic DNA Extraction kit (Takara, China). DNA quality was measured by a NanoPhotometer (Implen, Germany) and a Qubit fluorometer (dsDNA HS) kit (Life Technologies). Indexed genomic libraries were prepared with the Nextera XT kit (Illumina, CA), and sequenced on Illumina MiSeq using 2×300 bp chemistry. Post-processed reads were assembled using SPAdes Genome Assembler version 3.8.1-Linux (Bankevich et al., 2012; Bolger, Lohse, & Usadel, 2014). Assembly quality was assessed by both custom scripts and abyss-fac command in ABySS (www.bcgsc.ca/platform/bioinfo/software/abyss) to determine N50, total length of contigs, and other parameters. Next-generation sequencing (NGS) data were deposited in the National Center for Biotechnology Information (NCBI) database under the Sequence Read Archive (SRP256501) accession (BioProject: PRJNA625565).

Data analysis

The serotype, virulence genes, and *stx* subtypes were identified with SerotypeFinder 1.1 tool (Joensen, Tetzschner, Iguchi, Aarestrup, & Scheutz, 2015) and VirulenceFinder 1.2 (Joensen et al., 2014) of the Center for Genomic Epidemiology (CGE) website, and BLASTN of NCBI, using the assembled genomes. For the CGE server, 85% was selected as identity threshold, and 60% to be the percentage of minimum overlapping gene length (Ferdous et al., 2016).

Based on the assembled genomes, Parsnp (Treangen, Ondov, Koren, & Phillippy, 2014) in the Harvest package was used to align the core genomes of the STEC strains, followed by the creation of a maximum likelihood (ML) tree. The parameters for the ML tree generation were defaulted with the evolutionary model of General Time Reversible (GTR) and 1,000 resamples for bootstrapping. The tree and the molecular features of each isolate were visualized using iTOL (Letunic & Bork, 2016).

The seven housekeeping genes (i.e., *adk* , *fumC* ,*gyrB* , *icd* , *mdh* , *purA* , and *recA*) for MLST were defined according to the *E. coli* MLST website (*http://mlst.warwick.ac.uk/mlst/dbs/E.coli*) (Larsen et al., 2012). STs for each isolate were assigned based on the allelic profile of the seven housekeeping genes, which was then used to analyze the phylogenetic relationships among bacterial isolates. MLST sequence information about HUSEC isolates was retrieved from the online database (www.ehec.org) (Mellmann et al., 2008). A minimum spanning tree was created based on STs of our isolates and HUSEC strains using BioNumerics (Meng et al., 2014). Detailed information is available in Supplemental Fig. S1.

Ethics statement

Owners of the farms and slaughterhouses were informed of the study and expressed their approval for sampling of their animals. All experimental procedures involving animals were approved by the Animal Welfare and Ethics Committee of Xinjiang Agricultural University, Xinjiang Province, China (Animal protocol number: 2018020).

Results

Determination of serogroups and serotypes by WGS

A collection of 56 STEC isolates were subjected to WGS by Illumina Miseq platform, achieving an average coverage of ˜36×. Among the 56 isolates, 27 were isolated from cattle and 29 from sheep. The genome sequences identified 21 O serogroups, 17 H types, and a total of 29 O:H serotypes (Fig. 1). Sheep isolates belonged to 15 serotypes, whereas cattle isolates belonged to 14 serotypes, suggesting a high degree of diversity in serotypes. Seven isolates were O-serotype unknown (Ounk), likely because some of the O types are not present in the O-serotype database, or the *wzx* and*wzy* genes were not properly assembled. Two isolates were assigned as O153/O178. These two O types cannot be distinguished as these share identical *wzx* and *wzy* gene sequences. The predominant serotype was Ounk:H4, which accounted for 7 isolates from cattle; other prevalent serotypes were O8:H25 (5 isolates from cattle), O87:H16 (4 isolates from sheep), O76:H19 (4 isolates from sheep), O5:H19 (4 isolates from sheep), O108:H21 (4 isolates from sheep), and O8:H21 (3 isolates from cattle and sheep). Three serotypes (i.e., O81:H31, O66:H45, and O21:H25) contained two isolates each, whereas 20 serotypes were each represented by only one strain. Notably, no isolate was

identified to belong to O157 or the top six O serogroups (i.e., O26, O45, O103, O111, O121, and O145) that are associated with human non-O157 STEC infections.

Virulence factors

Among the 56 STEC isolates, 31 (55.4%) were positive for *stx* 1 only, 3 (5.4%) for *stx* 2 only, and 22 (39.3%) for*stx* 1+*stx* 2. *Stx* 1 was subtyped into *stx* 1a (35/53=66%) and *stx* 1c (18/53=34%), and *stx* 2 into*stx* 2a (14/25=56.0%), *stx* 2b (5/25=20.0%), and*stx* 2c (6/25=24.0%). These *stx* genes gave rise to 8 subtype combinations (*stx* 1a, *stx* 1c, *stx* 2a,*stx* 2b, *stx* 2c, *stx* 1a+*stx* 2a,*stx* 1a+*stx* 2c, and *stx* 1c+*stx* 2b), with*stx* 1a being the most prevalent (17/56=30.4%). The prevalence of other subtype combinations was 25.0% (14/56, *stx* 1c), 23.2% (13/56, *stx* 1a+*stx* 2a), 8.9% (5/56,*stx* 1a+*stx* 2c), and 7.1% (4/56, *stx* 1c+*stx* 2b). Two subtypes, *stx* 1c and *stx* 2b, are exclusively associated with sheep isolates, whereas *stx* 2c was exclusively related to cattle isolates. Cattle isolates are more likely to have the*stx* 1+*stx* 2 combination (17/27= 63.0%) than the sheep isolates (5/29= 17.2%). *Stx* 1-only genotype is more prevalent in sheep isolates (23/29=79.3%) than in cattle isolates (8/27=29.6%) (Fig. 1).

None of the isolates was positive for *eae* as well as adhesins*toxB* and *efa1* (data not shown). *Iha* , which encodes an adherence-conferring protein similar to *irgA* of *Vibrio cholerae* , is present in a majority of strains (32/56=57.1%).*Saa* , which encodes an autoagglutinating adhesin that is usually identified in LEE-negative strains, is present in 11 cattle strains (11/27=40.7%), but not in the sheep strains. In contrast, *ireA*is only encoded by sheep isolates (16/29=55.2%), but not cattle isolates. Nineteen strains carried *subA* , and sheep strains apparently had a higher percentage (16/29=55.2%) compared to cattle strains (3/27=11.1%). Hemolysin-encoding gene *ehxA* is present in 26 strains (46.4%), which is concordant to the findings of previous studies that show that *ehxA* is present in about half of non-O157 STEC strains (Hussein & Bollinger, 2005). Only cattle isolates possess the *espP* gene (7/27=25.9%).*Irp2* and *fyuA* carried by high pathogenicity island (HPI) were found in 13 isolates (23.2%). Eight strains (14.3%) harbored*katP* (Fig. 1).

In silico **Multilocus Sequence Typing (MLST) and core-genome SNP analysis**

The 56 STEC isolates were typed into 20 sequence types (STs). STs of four isolates were not matched in the MLST database, which suggests that they may belong to novel STs. ST10 is the most prevalent ST, which is represented by 11 strains. Other frequent STs include ST3249, ST40, ST447, and ST675, each containing over 4 strains. Eight STs were represented only once. ST40, ST447, and ST675 are represented by sheep isolates from Korla; and ST3249 from cattle isolates were obtained in 2016-2018. For the STs that contain more than two isolates, a great majority of these (75%) contain strains of the same O serotype.

Two isolates SG18-D13 and CG17-d36-2 assigned O153/O178 are phylogenetically distantly related, suggesting that these are very likely different serotypes. Two ST43 isolates (CD15-213 and SG17-J10) and 11 ST10 isolates belonging to various serotypes were clustered together and assigned to CC10. SD18-A3 and the clade of SG18-D1, SG18-D27, and SG18-D26 are almost identical in terms of ST, O:H serotype, and virulence profile, except for SD18-A3 that lacks the*stx* 2b gene. Nonetheless, our tree based on core-genome SNP analysis can clearly separate SD18-A3 from the other three, whereas a classical MLST could not (Supplemental Fig. 1).

Some isolates are identical with respect to the characteristics tested, suggesting that these likely originated from the same clone, e.g., the ST675 group. Similarly, SG18-2-2, SG18-2-1, and SG18-2-7 are nearly identical with respect to serotype, ST (the same novel ST type), and virulence factors. Despite the fact that these three strains and SG18-2-15 were isolated from the same location at the same time and clustered together, SG18-2-15 apparently gained a few virulence genes, including HPI-encoded *irp2* and *fyuA* and plasmid-borne*ehxA* and *subA* . SG18-6-2 stands out as it belongs to the unique ST25 and unique O128:H2 serotype. Although CG17-d36-3 and CG16-B21 both belonging to ST58 are in the same clade, these were assigned different serotypes and have distinct virulence gene profiles (Fig. 1).

Phylogenetic relationships of the STEC isolates with human HUS-causing *E. coli* (HUSEC)

A minimum spanning tree was constructed based on MLST STs of our isolates and HUSEC isolates to eva-

luate the potential risk in causing human infection (Figs. 2 and S1). The results suggested that some isolates are related to certain strains from HUSEC collection. ST675 contains four sheep isolates and HUSEC039, all belonging to serotype O76:H19. They all contain *stx* 1c only and *ehxA* . ST25 includes a sheep strain SG18-6-2 and HUSEC028, both typed into O128:H2. These contain both *ehxA* and *iha* , and carry*stx* 1c+*stx* 2b. ST40 contains four sheep isolates (*stx* 1c, *eae* -, *ehxA* -, and *saa* -) and HUSEC023 (*stx* 2d, *eae* -, *ehxA* -, and *saa* -), belonging to O108 and O112, respectively. ST101 contains two cattle isolates (O81,*stx* 1a+*stx* 2a, *eae* -, *saa* +, and *iha* +) and HUSEC025 (O55, *stx* 1, *eae* -, *saa* +, and*iha* +). ST43 contains three strains, one cattle isolate CD15-213, one sheep isolate SG17-J10, and HUSEC001. These are very different in terms of O serogroup and virulence gene contents. Isolate CD15-213 tested positive for all nine virulence genes, implying high virulence potential (Fig. 1).

Discussion

Ruminants, particularly cattle and sheep, have long been considered as important reservoirs for STEC strains. To our knowledge, this is the first study that has characterized cattle and sheep STEC strains isolated from Xinjiang Province, China, a region well known for animal husbandry and grazing. Cattle and sheep can harbor a wide range of STEC serotypes in their gut (Hussein & Bollinger, 2005; Zweifel, Blanco, Blanco, Blanco, & Stephan, 2004). In line with previous findings, our isolates demonstrate a high diversity of serotypes. Serotypes identified in this report were compared to those isolated in other studies from various sources. Isolates in this study belong to 23 O serogroups and 29 O:H serotypes (Fig. 1), of which 12 were isolated from cattle only and 13 from sheep only. Minimal overlap between isolates of two different animal origins was observed. However, many studies indicate that all sheep isolate serogroups found here were isolated from bovine feces and/or carcasses (Arthur, Barkocy-Gallagher, Rivera-Betancourt, & Koohmaraie, 2002; Hussein & Bollinger, 2005). Notably, the O serotypes O3, O15, O81, O116, O129, and O140 of cattle isolates here were rarely reported to be present in sheep isolates (Blanco et al., 2003; Martins et al., 2015; Oporto, Esteban, Aduriz, Juste, & Hurtado, 2008). O129, which was formerly suggested to be EPEC, was not reported in cattle isolates in previous studies, although it was earlier isolated from human feces (Blanco et al., 2006; Schwaiger, Grif, Pierard, Karch, & Allerberger, 1999). Some serogroups share a common ancestor with other pathotypes such as EPEC and EAEC, and integration of a*stx* -containing bacteriophage may convert these into a more virulent variant (Steyert et al., 2012). All serogroups here were previously shown to be associated with human STEC infections, but with different degrees of correlation (CDC, 2017). For example, O5 and O76 isolates caused numerous human infection cases in recent years, but O128 and O113 resulted in a moderate number of cases, whereas O66 and O81 were responsible for a very limited number of infections (CDC, 2017). Among all the serotypes, O8:H21, O76:H19, O104:H7, O113:H4, and O128:H2 have been linked to diarrhea or HUS cases (CDC, 2017; Islam et al., 2007; Monaghan et al., 2012; Mora et al., 2007). Twenty-two serotypes found in this study were reported elsewhere among isolates from various origins such as cattle, goat, and human patients. Serotypes O66:H45 and O74:H39 have not previously been reported in STEC isolates of sheep; O15:H10, O21:H11, and O6:H21 have not previously been reported in STEC isolates of cattle. Compared to a study of 126 beef cattle STEC isolates from a farm in Sichuan Province, China, we share 7 serogroups but only 1 serotype, O81:H31, suggesting high diversity of O serotypes among different regions of isolation.

Aside from the O:H serotypes, the presence of a single *stx* 1 or*stx* 2 or a combination of these and *stx* subtypes has been recognized as potential indicator of STEC pathogenicity. STEC strains that produce *stx* 1 only are generally mild in pathogenicity, whereas strains that produce *stx* 2 alone are more frequently associated with severe diseases such as bloody diarrhea (BD) or HUS (Johura et al., 2016; Melton-Celsa, 2014). The presence of*stx* 1+*stx* 2 may suggest medium pathogenicity between that of*stx* 1 only and *stx* 2 only (Arthur et al., 2002; Brandal et al., 2015). In a study of 361 non-O157 STEC isolates from cattle carcasses, the prevalence of*stx* 2 only, *stx* 1 only, and *stx* 1+*stx* 2 isolates was shown to be 36.6%, 50.2%, and 12.2%, respectively (Arthur et al., 2002). For the Sichuan cattle isolates from China, the corresponding prevalence is 35.7%, 24.6%, and 39.7%, respectively (Fan et al., 2019). For our cattle isolates, the corresponding prevalence is 7.4%, 31.4%, and 61.2%, respectively. The lower percentage of*stx* 2-only isolates and higher percentage of*stx* 1+*stx* 2 isolates suggest a lower possibility of highly virulent strains but a higher possibility of moderate-to-high virulent isolates from cattle. In a study of 379 sheep non-O157 STEC fecal swab isolates, the prevalence

of *stx* 2 only, *stx* 1 only, and*stx* 1+*stx* 2 isolates was shown to be 1.3%, 56.2%, and 42.5%, respectively (Blanco et al., 2003). In a study of 70 ovine isolates, the corresponding numbers are 14.3%, 52.8%, and 32.9%, respectively (Martins et al., 2015). The corresponding prevalence in our sheep isolates is 3.4%, 79.4%, and 17.2%, respectively. Thus, compared to sheep isolates, cattle isolates are more likely to contain *stx* 2 alone or *stx* 1+*stx* 2, implying cattle isolates have a high probability of causing severe disease.

For cattle isolates, all *stx* 1 are *stx* 1a subtype, and a majority of *stx* 2 subtypes is *stx* 2a (62.5%), with others being *stx* 2c. For sheep isolates, *stx* 1 subtypes are*stx* 1a (35.7%) and *stx* 1c (64.3%); *stx* 2 subtypes are mostly *stx* 2b [83.3%, similar to the reported 84.8% (Martins et al., 2015)], with*stx* 2a being 16.7%. High prevalence of *stx* 1c-positive (but not other subtypes) sheep isolates was also found in other studies [57% *stx* 1c among the *stx* 1-positive strains (Zweifel et al., 2004)], supporting the notion that sheep represent the main reservoir of *stx* 1c-carrying strains (Brett et al., 2003).*Stx* 1c-positive strains are not frequently associated with HC or HUS but tend to trigger asymptomatic infection or mild diarrhea (Brandal et al., 2015). STECs carrying*stx* 2b have been associated with sporadic HUS cases, indicating its potential to cause severe infections in humans, although the*stx* 2b subtype in most cases is linked to lower virulence (de Boer et al., 2015) (Buvens et al., 2012). The *stx* 1a,*stx* 2a, *stx* 2c, and *stx* 2d subtypes are most often associated with HC and HUS (Fao/Who Stec Expert, 2019; Persson, Olsen, Ethelberg, & Scheutz, 2007). Stx2a is more potent than Stx2b, Stx2c, and Stx1 in cell cultures and mouse models (Fuller, Pellino, Flagler, Strasser, & Weiss, 2011). Therefore, by judging from the *stx* subtypes, cattle isolates tend to have high virulence potential, whereas sheep isolates may be assumed to be of low virulence.

None of the isolates in this study contain the *eae* gene. A low occurrence of *eae* in STEC isolates from domestic ruminants was also observed elsewhere (Amezquita-Lopez, Quinones, Lee, & Chaidez, 2014; Schilling et al., 2012). In a study of 521 STEC sheep isolates, *eae* was detected only in 0.8% of the isolates (S. Sanchez, 2010). In contrast, other adhesin genes such as *iha* , *saa* , and *subA* were present in many strains but with different prevalence rates. It was reported *iha* and *saa* are carried by more than half of the STEC strains isolated from yak and cattle (Bai et al., 2013; Bosilevac & Koohmaraie, 2011). For our cattle isolates, *iha* and *saa* prevalence was 42.6% and 40.7%, respectively. It has been suggested that *saa* is correlated with the presence of the large STEC virulence plasmid (Toma et al., 2004). Indeed, the isolates positive with *saa* also harbored *ehxA* , which is usually encoded on the virulence plasmid (Blanco et al., 2003; Martins et al., 2015; Oporto et al., 2008). In many isolates,*ehxA* and *espP* are genetically linked (Fan et al., 2019), particularly the O157 serogroup (Islam et al., 2008), but in our study, these do not always coexist. In particular,*ehxA* -positive cattle isolates were all tested negative in*espP* , suggesting that these may not always encode on the same plasmid. We found low prevalence of *katP* (3.7%) and *subA*(11.1%) in cattle isolates, which is comparable to another study of China isolates (Fan et al., 2019). In a study of 60 ovine isolates, 67% of these are *saa* -positive, but our ovine isolates did not harbor *saa* , thereby suggesting a unique regional trait. There were few studies reporting the virulence profile of non-O157 sheep isolates; thus our study provides comprehensive and valuable information on virulence genes of non-O157 sheep isolates.

The genetic relatedness of our isolates to the human HUSEC collection was explored. The O76:H19 serotype is very common in sheep isolates (Zweifel et al., 2004). The four ST675 isolates belonging to O76:H19 are similar to HUSEC039. Although lacking*eae* and containing a mild-virulence subtype *stx* 1c, these contain *ehxA* , *subA* , and *iha* , which could increase their virulence potential. Another human disease-associated serotype O128:H2 frequently isolated from sheep (Zweifel et al., 2004) contains isolate SG18-6-2 and HUSEC028. For HUSEC023 and the 4 sheep isolates of ST40, the former contains a virulent *stx* 2d, but the 4 isolates contain*stx* 1c only. In the case of ST101, two cattle isolates clustered with HUSEC025, but the former contains *stx* 2a, whereas HUSEC025 is*stx* 2-negative. The ST43 cattle strain CD15-213 clustered with HUSEC001 contains two virulent *stx* subtypes, *stx* 1a and*stx* 2c, as well as the key virulence genes *ehxA* ,*espP* , *subA* , *saa* , *iha* , and*fyuA* /*irp2* . Thus, CD15-213 may exhibit the highest pathogenicity. Remarkably, this serotype was shown to be present in Morocco market meat (Badri, Fassouane, Filliol, Hassar, & Cohen, 2011), raising significant food safety concerns. Taken together, the aforementioned cattle isolates may pose the greatest risk. In addition, they should be paid much attention in future surveillance work.

In conclusion, we used WGS analysis to characterize the serotype, virulence gene profile, and genetic diversity, and relationship of a collection of non-O157 STEC isolates from cattle and sheep. This method enables us to reveal much more detail of bacterial isolates, with the highest discriminatory power possible. Several sheep and cattle grouped with HUSEC strains, particularly strain CD15-213, may have the potential to cause serious human infections. Local human HC and HUS cases should be closely monitored, and its relatedness to STEC isolates from various sources should be investigated in order for better food-borne disease prevention.

Acknowledgements

We thank Xinyang Zhang, Dawei Yang and Yixin Peng for their help in bacterial isolation and DNA preparation. This work was supported by The National Key Research and Development Program of China (2018YFD0500500), National Natural Science Foundation of China Young Scholars Project (31902242), and National Key Research and Development Program of China (31702267).

Conflict of interests

The authors declare that they have no conflict of interests.

Data availability statement

The data that support the findings of this study are openly available in National Center for Biotechnology Information (NCBI) database under the Sequence Read Archive (SRP256501) accession (BioProject: PRJNA625565).

References

Amezquita-Lopez, B. A., Quinones, B., Lee, B. G., & Chaidez, C. (2014). Virulence profiling of Shiga toxin-producing *Escherichia coli* recovered from domestic farm animals in Northwestern Mexico. *Front Cell Infect Microbiol, 4* , 7. *doi* :10.3389/fcimb.2014.00007

Arthur, T. M., Barkocy-Gallagher, G. A., Rivera-Betancourt, M., & Koohmaraie, M. (2002). Prevalence and characterization of non-O157 Shiga toxin-producing *Escherichia coli* on carcasses in commercial beef cattle processing plants.*Appl Environ Microbiol, 68* (10), 4847-4852.*doi* :10.1128/aem.68.10.4847-4852.2002

Badri, S., Fassouane, A., Filliol, I., Hassar, M., & Cohen, N. (2011). Detection of Shiga toxin-producing*Escherichia coli* in meat marketed in Casablanca (Morocco).*Cell Mol Biol (Noisy-le-grand), 57 Suppl* , OL1476-1477.

Bai, X., Hu, B., Xu, Y., Sun, H., Zhao, A., Ba, P., . . . Xiong, Y. (2016). Molecular and Phylogenetic Characterization of Non-O157 Shiga Toxin-Producing *Escherichia coli* Strains in China. *Front Cell Infect Microbiol, 6* , 143.*doi* :10.3389/fcimb.2016.00143

Bai, X., Zhao, A., Lan, R., Xin, Y., Xie, H., Meng, Q., . . . Xiong, Y. (2013). Shiga toxin-producing*Escherichia coli* in yaks (Bos grunniens) from the Qinghai-Tibetan Plateau, China. *PLoS One, 8* (6), e65537.*doi* :10.1371/journal.pone.0065537

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., . . . Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology, 19* (5), 455-477. *doi* :10.1089/cmb.2012.0021

Bergholz, T. M., Moreno Switt, A. I., & Wiedmann, M. (2014). Omics approaches in food safety: fulfilling the promise? *Trends in Microbiology, 22* (5), 275-281.*doi* :10.1016/j.tim.2014.01.006

Bielaszewska, M., Aldick, T., Bauwens, A., & Karch, H. (2014). Hemolysin of enterohemorrhagic*Escherichia coli* : structure, transport, biological activity and putative role in virulence. *International Journal of Medical Microbiology, 304* (5-6), 521-529. *doi* :10.1016/j.ijmm.2014.05.005

Blanco, M., Blanco, J. E., Dahbi, G., Mora, A., Alonso, M. P., Varela, G., . . . Blanco, J. (2006). Typing of intimin (eae) genes from enteropathogenic *Escherichia coli* (EPEC) isolated from children with diarrhoea in

Montevideo, Uruguay: identification of two novel intimin variants (muB and xiR/beta2B). *Journal of Medical Microbiology, 55* (Pt 9), 1165-1174. *doi* :10.1099/jmm.0.46518-0

Blanco, M., Blanco, J. E., Mora, A., Rey, J., Alonso, J. M., Hermoso, M., . . . Blanco, J. (2003). Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from healthy sheep in Spain. *Journal of Clinical Microbiology, 41* (4), 1351-1356. *doi* :10.1128/jcm.41.4.1351-1356.2003

Boerlin, P., McEwen, S. A., Boerlin-Petzold, F., Wilson, J. B., Johnson, R. P., & Gyles, C. L. (1999). Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. *Journal of Clinical Microbiology, 37* (3), 497-503.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30* (15), 2114-2120. *doi* :10.1093/bioinformatics/btu170

Bosilevac, J. M., & Koohmaraie, M. (2011). Prevalence and characterization of non-O157 shiga toxin-producing *Escherichia coli* isolates from commercial ground beef in the United States. *Appl Environ Microbiol, 77* (6), 2103-2112. *doi* :10.1128/AEM.02833-10

Brandal, L. T., Wester, A. L., Lange, H., Lobersli, I., Lindstedt, B. A., Vold, L., & Kapperud, G. (2015). Shiga toxin-producing *Escherichia coli* infections in Norway, 1992-2012: characterization of isolates and identification of risk factors for haemolytic uremic syndrome. *Bmc Infectious Diseases, 15* , 324. *doi* :10.1186/s12879-015-1017-6

Brett, K. N., Ramachandran, V., Hornitzky, M. A., Bettelheim, K. A., Walker, M. J., & Djordjevic, S. P. (2003). stx1c Is the most common Shiga toxin 1 subtype among Shiga toxin-producing *Escherichia coli* isolates from sheep but not among isolates from cattle. *Journal of Clinical Microbiology, 41* (3), 926-936. *doi* :10.1128/jcm.41.3.926-936.2003

Buvens, G., De Gheldre, Y., Dediste, A., de Moreau, A. I., Mascart, G., Simon, A., . . . Pierard, D. (2012). Incidence and virulence determinants of verocytotoxin-producing *Escherichia coli* infections in the Brussels-Capital Region, Belgium, in 2008-2010. *Journal of Clinical Microbiology, 50* (4), 1336-1345. *doi* :10.1128/JCM.05317-11

CDC. (2017). National Enteric Disease Surveillance: Shiga toxin-producing *Escherichia coli* (STEC) Annual Report, 2015. *CDC Anual Report, 1* (1), 20-22.

Conrad, C. C., Stanford, K., McAllister, T. A., Thomas, J., & Reuter, T. (2014). Further development of sample preparation and detection methods for O157 and the top 6 non-O157 STEC serogroups in cattle feces. *J Microbiol Methods, 105* , 22-30. *doi* :10.1016/j.mimet.2014.06.020

Dallman, T. J., Byrne, L., Launders, N., Glen, K., Grant, K. A., & Jenkins, C. (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* (8), 1672-1680. *doi* :10.1017/S0950268814002696

de Boer, R. F., Ferdous, M., Ott, A., Scheper, H. R., Wisselink, G. J., Heck, M. E., . . . Kooistra-Smid, A. M. (2015). Assessing the public health risk of Shiga toxin-producing *Escherichia coli* by use of a rapid diagnostic screening algorithm. *Journal of Clinical Microbiology, 53* (5), 1588-1598. *doi* :10.1128/JCM.03590-14

Djordjevic, S. P., Ramachandran, V., Bettelheim, K. A., Vanselow, B. A., Holst, P., Bailey, G., & Hornitzky, M. A. (2004). Serotypes and virulence gene profiles of shiga toxin-producing *Escherichia coli* strains isolated from feces of pasture-fed and lot-fed sheep. *Appl Environ Microbiol, 70* (7), 3910-3917. *doi* :10.1128/AEM.70.7.3910-3917.2004

Fan, R., Shao, K., Yang, X., Bai, X., Fu, S., Sun, H., . . . Xiong, Y. (2019). High prevalence of non-O157 Shiga toxin-producing *Escherichia coli* in beef cattle detected by combining four selective agars. *Bmc Microbiology, 19* (1), 213. *doi* :10.1186/s12866-019-1582-8

Fao/Who Stec Expert, G. (2019). Hazard Identification and Characterization: Criteria for Categorizing Shiga Toxin-Producing *Escherichia coli* on a Risk Basis(dagger).*J Food Prot, 82* (1), 7-21. *doi* :10.4315/0362-028X.JFP-18-291

Ferdous, M., Friedrich, A. W., Grundmann, H., de Boer, R. F., Croughs, P. D., Islam, M. A., . . . Rossen, J. W. (2016). Molecular characterization and phylogeny of Shiga toxin-producing *Escherichia coli* isolates obtained from two Dutch regions using whole genome sequencing. *Clin Microbiol Infect, 22* (7), 642 e641-649. *doi* :10.1016/j.cmi.2016.03.028

Fruth, A., Prager, R., Tietze, E., Rabsch, W., & Flieger, A. (2015). Molecular epidemiological view on Shiga toxin-producing *Escherichia coli* causing human disease in Germany: Diversity, prevalence, and outbreaks. *International Journal of Medical Microbiology, 305* (7), 697-704.*doi* :10.1016/j.ijmm.2015.08.020

Fuller, C. A., Pellino, C. A., Flagler, M. J., Strasser, J. E., & Weiss, A. A. (2011). Shiga toxin subtypes display dramatic differences in potency. *Infect Immun, 79* (3), 1329-1337. *doi* :10.1128/IAI.01182-10

Galli, L., Miliwebsky, E., Irino, K., Leotta, G., & Rivas, M. (2010). Virulence profile comparison between LEE-negative Shiga toxin-producing *Escherichia coli* (STEC) strains isolated from cattle and humans. *Veterinary Microbiology, 143* (2-4), 307-313. *doi* :10.1016/j.vetmic.2009.11.028

Gonzalez-Escalona, N., & Kase, J. A. (2019). Virulence gene profiles and phylogeny of Shiga toxin-positive*Escherichia coli* strains isolated from FDA regulated foods during 2010-2017. *PLoS One, 14* (4), e0214620.*doi* :10.1371/journal.pone.0214620

Gonzalez-Escalona, N., Toro, M., Rump, L. V., Cao, G., Nagaraja, T. G., & Meng, J. (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* (13), 3900-3912.*doi* :10.1128/AEM.00498-16

Hussein, H. S., & Bollinger, L. M. (2005). Prevalence of Shiga toxin-producing *Escherichia coli* in beef cattle. *J Food Prot, 68* (10), 2224-2241.*doi* :10.4315/0362-028x-68.10.2224

Islam, M. A., Heuvelink, A. E., de Boer, E., Sturm, P. D., Beumer, R. R., Zwietering, M. H., . . . Talukder, K. A. (2007). Shiga toxin-producing *Escherichia coli*isolated from patients with diarrhoea in Bangladesh. *Journal of Medical Microbiology, 56* (Pt 3), 380-385.*doi* :10.1099/jmm.0.46916-0

Islam, M. A., Mondol, A. S., de Boer, E., Beumer, R. R., Zwietering, M. H., Talukder, K. A., & Heuvelink, A. E. (2008). Prevalence and genetic characterization of shiga toxin-producing *Escherichia coli* isolates from slaughtered animals in Bangladesh. *Appl Environ Microbiol, 74* (17), 5414-5421.*doi* :10.1128/AEM.00854-08

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., & Aarestrup, F. M. (2014). Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic *Escherichia coli* .*Journal of Clinical Microbiology, 52* (5), 1501-1510.*doi* :10.1128/Jcm.03617-13

Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., & Scheutz, F. (2015). Rapid and Easy In Silico Serotyping of *Escherichia coli* Isolates by Use of Whole-Genome Sequencing Data. *Journal of Clinical Microbiology, 53* (8), 2410-2426. *doi* :10.1128/Jcm.00008-15

Johura, F. T., Parveen, R., Islam, A., Sadique, A., Rahim, M. N., Monira, S., . . . Alam, M. (2016). Occurrence of Hybrid *Escherichia coli* Strains Carrying Shiga Toxin and Heat-Stable Toxin in Livestock of Bangladesh. *Front Public Health, 4* , 287. *doi* :10.3389/fpubh.2016.00287

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., . . . Lund, O. (2012). Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria.*Journal of Clinical Microbiology, 50* (4), 1355-1361.*doi* :10.1128/Jcm.06094-11

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res, 44* (W1), W242-W245. *doi* :10.1093/nar/gkw290

Lindsey, R. L., Pouseele, H., Chen, J. C., Strockbine, N. A., & Carleton, H. A. (2016). Implementation of Whole Genome Sequencing (WGS) for Identification and Characterization of Shiga Toxin-Producing *Escherichia coli* (STEC) in the United States. *Frontiers in Microbiology, 7* , 766.*doi* :10.3389/fmicb.2016.00766

Liu, Y., Zhu, M., Su, X., Xu, L., Liu, J., Zheng, X., . . . Zheng, X. (2020). Analysis of drug resistance and the Shiga toxin-producing *Escherichia coli* in sheep from Korla region of Xinjiang. *Journal of Agricultural Science and Technology, 22* (3), 122-128. *doi* :10.13304/j.nykjdb.2019.0053

Martins, F. H., Guth, B. E., Piazza, R. M., Leao, S. C., Ludovico, A., Ludovico, M. S., . . . Pelayo, J. S. (2015). Diversity of Shiga toxin-producing *Escherichia coli* in sheep flocks of Parana State, southern Brazil. *Veterinary Microbiology, 175* (1), 150-156. *doi* :10.1016/j.vetmic.2014.11.003

Mellmann, A., Bielaszewska, M., Kock, R., Friedrich, A. W., Fruth, A., Middendorf, B., . . . Karch, H. (2008). Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli* . *Emerg Infect Dis, 14* (8), 1287-1290. *doi* :10.3201/eid1408.071082

Melton-Celsa, A. R. (2014). Shiga Toxin (Stx) Classification, Structure, and Function. *Microbiol Spectr, 2* (4), EHEC-0024-2013.*doi* :10.1128/microbiolspec.EHEC-0024-2013

Meng, Q., Bai, X. N., Zhao, A. L., Lan, R. T., Du, H. M., Wang, T., . . . Xiong, Y. W. (2014). Characterization of Shiga toxin-producing *Escherichia coli*isolated from healthy pigs in China. *Bmc Microbiology, 14* .*doi* :Artn 5

10.1186/1471-2180-14-5

Monaghan, A. M., Byrne, B., McDowell, D., Carroll, A. M., McNamara, E. B., & Bolton, D. J. (2012). Characterization of farm, food, and clinical Shiga toxin-producing*Escherichia coli* (STEC) O113. *Foodborne Pathogens and Disease, 9* (12), 1088-1096. *doi* :10.1089/fpd.2012.1257

Mora, A., Blanco, M., Blanco, J. E., Dahbi, G., Lopez, C., Justel, P., . . . Blanco, J. (2007). Serotypes, virulence genes and intimin types of Shiga toxin (verocytotoxin)-producing *Escherichia coli* isolates from minced beef in Lugo (Spain) from 1995 through 2003. *Bmc Microbiology, 7* , 13. *doi* :10.1186/1471-2180-7-13

Oporto, B., Esteban, J. I., Aduriz, G., Juste, R. A., & Hurtado, A. (2008). *Escherichia coli* O157:H7 and non-O157 Shiga toxin-producing E. coli in healthy cattle, sheep and swine herds in Northern Spain. *Zoonoses Public Health, 55* (2), 73-81. *doi* :10.1111/j.1863-2378.2007.01080.x

Paton, A. W., Srimanote, P., Woodrow, M. C., & Paton, J. C. (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* (11), 6999-7009.*doi* :10.1128/IAI.69.11.6999-7009.2001

Persson, S., Olsen, K. E., Ethelberg, S., & Scheutz, F. (2007). Subtyping method for *Escherichia coli*shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *Journal of Clinical Microbiology, 45* (6), 2020-2024. *doi* :10.1128/JCM.02591-06

Rasko, D. A., Webster, D. R., Sahl, J. W., Bashir, A., Boisen, N., Scheutz, F., . . . Waldor, M. K. (2011). Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med, 365* (8), 709-717.*doi* :10.1056/NEJMoa1106920

S. Sanchez, R. M., A. Garcia, J. M. Benitez, J. Blanco, J. E. Blanco, J. M. Alonso, J. Rey. (2010). Variation in the prevalence of non-O157 Shiga toxin-producing in four sheep flocks during a 12-month longitudinal study. *Small Ruminant Research* (93), 144-148.

Scheutz, F., Teel, L. D., Beutin, L., Pierard, D., Buvens, G., Karch, H., . . . O'Brien, A. D. (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *Journal of Clinical Microbiology, 50* (9), 2951-2963. *doi* :10.1128/JCM.00860-12

Schilling, A. K., Hotzel, H., Methner, U., Sprague, L. D., Schmoock, G., El-Adawy, H., . . . Geue, L. (2012). Zoonotic agents in small ruminants kept on city farms in southern Germany. *Appl Environ Microbiol, 78* (11), 3785-3793.*doi* :10.1128/AEM.07802-11

Schmidt, H., Kernbach, C., & Karch, H. (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. *doi* :10.1099/00221287-142-4-907

Schwaiger, M., Grif, K., Pierard, D., Karch, H., & Allerberger, F. (1999). Shiga toxin-producing*Escherichia coli* was the third most frequent bacterial cause of diarrhea in Austria during July and August of 1998. *Clin Microbiol Infect, 5* (10), 645-647. *doi* :10.1111/j.1469-0691.1999.tb00423.x

Smith, J. L., Fratamico, P. M., & Gunther, N. W. t. (2014). Shiga toxin-producing *Escherichia coli* .*Advances in Applied Mircobiology, Vol 67, 86* , 145-197.*doi* :10.1016/B978-0-12-800262-9.00003-2

Steyert, S. R., Sahl, J. W., Fraser, C. M., Teel, L. D., Scheutz, F., & Rasko, D. A. (2012). Comparative genomics and stx phage characterization of LEE-negative Shiga toxin-producing *Escherichia coli* . *Front Cell Infect Microbiol, 2* , 133. *doi* :10.3389/fcimb.2012.00133

Toma, C., Martinez Espinosa, E., Song, T., Miliwebsky, E., Chinen, I., Iyoda, S., . . . Rivas, M. (2004). Distribution of putative adhesins in different seropathotypes of Shiga toxin-producing *Escherichia coli* . *Journal of Clinical Microbiology, 42* (11), 4937-4946.*doi* :10.1128/JCM.42.11.4937-4946.2004

Treangen, T. J., Ondov, B. D., Koren, S., & Phillippy, A. M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biology, 15* (11). *doi* :Artn 524

10.1186/S13059-014-0524-X

Tseng, M., Fratamico, P. M., Bagi, L., Delannoy, S., Fach, P., Manning, S. D., & Funk, J. A. (2014). Diverse virulence gene content of Shiga toxin-producing*Escherichia coli* from finishing swine. *Appl Environ Microbiol, 80* (20), 6395-6402. *doi* :10.1128/AEM.01761-14

Valilis, E., Ramsey, A., Sidiq, S., & DuPont, H. L. (2018). Non-O157 Shiga toxin-producing*Escherichia coli* -A poorly appreciated enteric pathogen: Systematic review. *Int J Infect Dis, 76* , 82-87.*doi* :10.1016/j.ijid.2018.09.002

Zhang, Y., Liu, Y., Zheng, X., Zhang, C., Su, Z., & Zhang, X. (2018). Isolation and identification of the pathogenic non-O157 STEC in Beef slaughterhouse. *China animal husbandry and veterinary medicine, 45* (4), 1035-1040.*doi* :10.16431/j.cnki.1671-7236.2018.04.025

Zweifel, C., Blanco, J. E., Blanco, M., Blanco, J., & Stephan, R. (2004). Serotypes and virulence genes of ovine non-O157 Shiga toxin-producing *Escherichia coli* in Switzerland. *Int J Food Microbiol, 95* (1), 19-27.*doi* :10.1016/j.ijfoodmicro.2004.01.015

Figure legend

Figure 1. Genomic and phylogenetic analysis of the 56 Shiga toxin-producing *Escherichia coli* (STEC) strains sequenced in this study. A maximum-likelihood phylogenetic tree was generated using the core-genome Single Nucleotide Polymorphisms (SNPs) and midpoint rooted. The isolate locations are represented by color bars. The presence of a virulence gene is denoted by a pink pie. "-" in the sequence type (ST) column indicates unrecognized ST in the MLST database; "-" in the*stx* columes indicates the absence of a *stx* gene. The first letter in the name of each isolate indicates its animal origin: S, sheep; C, cattle.

Figure 2. Genetic relatedness of the Shiga toxin-producing *Escherichia coli* (STEC) isolates in this study to the human HUSEC isolates based on MLST. Each pie represents a cluster of isolates assigned to the same ST; the size of a pie is proportional to the amount of isolates in the group. The colors of

and within each pie indicates the source of the isolates: green, sheep; red, cattle; blue, HUSEC. The numbers on the lines between pies indicates the number of allelic difference between two STs.