Interrogation of single nucleotide polymorphisms in *gnd* provides a novel method for molecular serogrouping of clinically important Shiga toxin producing *Escherichia coli* (STEC) targeted by regulation in the United States, including the “big six” non-O157 STEC and STEC O157

J. R. Elder^{1,2}, M. Bugarel¹, H. C. den Bakker^{1,3}, G. H. Loneragan¹ and K. K. Nightingale^{1,2*}

1. Department of Food and Animal Sciences, Texas Tech University, Lubbock, TX 79409
2. Department of Animal Sciences, Colorado State University, Fort Collins, CO 80523
3. Department of Food Science and Technology, Cornell University, Ithaca, NY 14853

*Corresponding Author

201 Animal and Food Sciences Building
Department of Animal and Food Sciences
Texas Tech University
Lubbock, TX 79409
Voice: 806-392-2815
Email: kendra.nightingale@ttu.edu

ABSTRACT

Escherichia coli O157:H7 has frequently been associated with foodborne infections and is considered an adulterant in raw non-intact beef in the U.S. Shiga toxin-producing *E. coli* (STEC) belonging to serogroups O26, O45, O103, O111, O121, and O145 (known as the “big six” non-O157) were estimated to cause >70% of foodborne infections attributed to non-O157 serogroups in the U.S., as a result, these six serogroups have also been targeted by regulation in the U.S. The purpose of this study was to develop a rapid and high-throughput molecular method to group STEC isolates into seven clinically important serogroups (i.e., O157 and the “big six” non-O157 serogroups) targeted by regulation in the U.S. by interrogating single nucleotide polymorphisms (SNPs) in *gnd*. A collection of 195 STEC isolates, including isolates belonging to O157:H7 (n=18), O26 (n=21), O45 (n=19), O103 (n=24), O111 (n=24), O121 (n=23), O145 (n=21), and ten other STEC serogroups (n=45), was assembled and characterized by full *gnd* sequencing to identify informative SNPs for molecular serogrouping. A multiplex SNP typing assay was developed to interrogate twelve informative *gnd* SNPs by single base pair extension chemistry and used to characterize the STEC isolate collection assembled here. SNP types were assigned to each isolate by the assay and polymorphisms were confirmed with *gnd* sequence data. O-serogroup-specific SNP types were identified for each of the seven clinically important STEC serogroups, which allowed the differentiation of these seven STEC serogroups from other non-O157 STEC serogroups. Although serogroups of the “big six” non-O157 STEC and O157:H7 contained multiple SNP types per O-serogroup, there were no overlapping SNP types between serogroups. Our results demonstrate that molecular serogrouping of STEC isolates by interrogation of informative SNPs in *gnd* represents an alternative to traditional serogrouping by agglutination for rapid and high-throughput identification of clinically important STEC serogroups targeted by regulation for surveillance and epidemiological investigations.

KEYWORDS:

Shiga toxin producing *Escherichia coli* (STEC), *E. coli* O157:H7, non-O157 STEC, molecular serotyping, SNP typing, *gnd*

1. INTRODUCTION:

Shiga toxin-producing *Escherichia coli* (STEC) infections can produce a wide range of clinical manifestations of disease from mild diarrhea to hemorrhagic colitis (HC) and hemolytic uremic syndrome (HUS), especially in the elderly and children under the age of 5 (Gould et al., 2009). *E. coli* O157:H7 was recognized as an etiological agent of hemolytic uremic syndrome (HUS) in 1985 (Karmali et al., 1985) and was declared an adulterant in raw non-intact beef and its components in the 1990s (USDA, 2011). From 2000 to 2006 the reported incidence of laboratory confirmed *E. coli* O157:H7 infections reported by FoodNet sites was 1.5 cases per 100,000 (Gould et al., 2009). In a more recent survey of FoodNet sites, the incidence of laboratory confirmed *E. coli* O157:H7 infections decreased to 0.92 cases per 100,000 (CDC, 2014). During that same year the rate of laboratory confirmed non-O157 STEC infections was 1.43 cases per 100,000 (CDC, 2014). Non-O157 STEC were estimated to cause 112,752 infections each year in the U.S. (Scallan et al., 2011). As a result of the emerging importance of non-O157 STEC, the U.S. Department of Agriculture Food Safety Inspection Service (USDA:FSIS) deemed non-O157 STEC to represent a public health threat equivalent to that of *E. coli* O157:H7 (Eblen, 2008) and declared non-O157 STEC belonging to serogroups O26, O45, O103, O111, O121, and O145 (known as the “big six”) as adulterants in raw non-intact beef in 2011 (USDA, 2011). Non-O157 STEC have also been linked to a significant proportion of HUS cases in other countries, including Germany and Austria, Australia, Italy, Denmark, and Argentina (Elliot et al., 2001, Ethelberg et al., 2004, Gerber et al., 2002, Lopez, 1989, Tozzi et al., 2003).

Although a large number of STEC O-serogroups have been linked to human disease, it appears that a subset of serogroups have greater pathogenic potential and are isolated from foodborne disease patients at a higher frequency than others (Bettelheim, 2003, Hughes et al., 2006). In the U.S., serogroups O26, O103, and O111 accounted for 74% of laboratory confirmed infections attributed to non-O157 STEC (CDC, 2014). In Switzerland, serogroups O26, O145, O103, and O121 accounted for 51.5% of non-O157 strains isolated from patients suspected of having STEC infections. Serogroups O26, O103,

O111, O121, and O145 are among the most frequently isolated non-O157 STEC serogroups from patients suffering from diarrhea in Switzerland, Argentina, Peru, UK, Spain, Germany, Belgium, Ireland, Czech Republic, and Finland (Bielaszewska et al., 1996, Blanco et al., 2003, Buvens et al., 2010, Carroll et al., 2005, Contreras et al., 2011, Eklund et al., 2001, Käppeli et al., 2011, Rivero et al., 2010, Werber et al., 2008, Willshaw et al., 2001). Karmali et al. (2003) recognized that a few serotypes account for the majority of outbreaks and severe clinical manifestations of disease attributed to non-O157 STEC. Accordingly, STEC were classified into seropathotypes with the two seropathotypes of greatest clinical significance including seropathotype A (i.e., O157:H7 and O157:NM) and seropathotype B (i.e., serotypes O26:H11, O103:H2, O111:NM, O121:H19, and O145:NM) (Karmali et al., 2003).

Many real-time PCR assays, targeting O-antigen specific genes of the “big six,” have been developed to screen enrichments for the presence or absence of these serogroups in a variety of food matrices (e.g., Bai et al., 2012, Fratamico et al., 2003, Fratamico et al., 2005, Fratamico et al., 2009, Lin et al., 2011a, Lin et al., 2011b, Paton and Paton, 1998, Perelle et al., 2005, Perelle et al., 2007; Wasilenko et al., 2014). Targeting serogroup specific genes limits the number of serogroups, which may be identified via the assay, requiring specific primers and probe for each serogroup that is targeted by the assay. Most of these real-time PCR assays require two to three multiplex reactions to detect the “big six” non-O157 STEC and associated virulence factors (e.g., *eae*, *stxI* and *stxII*) and a separate real-time PCR assay is typically required to detect *E. coli* O157. If a sample screens positive for one of the “big six” non-O157 STEC or *E. coli* O157 conventional microbiological culture is required to obtain and isolates for confirmation and subtyping. Classical *E. coli* serotyping via slide agglutination relies on the use of antisera raised against the O-types of reference *E. coli* serogroups. The classical method of H-typing also relies on antisera, but molecular methods based on PCR restriction fragment length polymorphism (PCR-RFLP) are available (Machado et al., 2000). The large number of antisera required for O-serogrouping severely limits the number of laboratories capable of performing conventional serotyping of *E. coli* isolates. Sequence-based typing of a gene or

genes conserved among STEC to identify O-serogroup is more accessible than conventional serotyping and more informative than PCR-based methods targeting serogroup specific targets.

Gilmour et al. (2007) proposed a sequence-based serogrouping method relying on the sequencing of a 643 bp region of the *gnd* gene. Following development of the method, Gilmour et al., implemented *gnd* sequencing to identify the serogroup of clinical isolates belonging to serotypes O121:H19, O26:H11, and O177:NM. Partial *gnd* sequencing was also used to successfully identify the O-serogroup of the *E. coli* O104:H4 strain linked to the 2011 outbreak in Germany attributed to contaminated sprouts (Mellman et al., 2011). The authors of this study found that *gnd* alleles were conserved within each serogroup investigated and therefore proposed that *gnd* alleles could be used to identify STEC serogroups. Full or partial sequencing of *gnd* is not an ideal method for rapid and high-throughput grouping of STEC isolates into O-serogroups due to the bioinformatics analyses required to identify unique alleles. Also since the majority of *gnd* sequence data is conserved across STEC serogroups and therefore much of the information gained through DNA sequencing is non-informative. The overall aim of this study was to develop a rapid and high-throughput multiplex SNP genotyping assay for serogrouping of presumptive STEC isolates that would confirm the O-serogroup of clinically important STEC targeted by regulation in the U.S. (i.e. O157 and the “big six” non-O157 STEC) with 12 informative SNPs in *gnd*. Although other studies have described molecular methods for confirmation and subtyping of presumptive non-O157 STEC isolates (e.g., Norman et al., 2012, Toro et al., 2013, Timmons et al., 2016), this is the first study describing a molecular serogrouping method capable of identifying *E. coli* isolates belonging to all seven clinically important serogroups targeted by regulation in the U. S. (i.e., O157 and the “big six” non-O157 STEC) based on interrogation of informative SNPs in *gnd*.

2. MATERIALS AND METHODS:

2.1. Assembly of STEC isolate collection for development of *gnd* SNP genotyping

assay. A representative collection of 195 STEC isolates from human clinical cases, animal,

and food sources was assembled (Table 1). The isolate set was assembled from STEC isolates in our existing collection and isolates obtained from (i) the *E. coli* Reference Center (The Pennsylvania State University, University Park, PA), (ii) the STEC Center (Michigan State University, East Lansing, MI) and (iii) the U. S. Meat Animal Research Center (Clay Center, NE). The STEC isolates in the inclusivity panel belonged to serogroups O157 (n=18), O26 (n=21), O45 (n=19), O103 (n=24), O111 (n=24), O121 (n=23), O145 (n=21). STEC isolates belonging to serogroups O2, O4, O5, O8, O14, O30, O73, O113, O118 O126, O128, O147, O153, O156, O171, O172, and O174 made up the exclusivity panel. Serotypes for the STEC isolate collection presented in Table 1 were determined by conventional slide agglutination that was performed at one of the centers listed above.

2.2. PCR Amplification of *gnd*. A segment within the *gnd* gene was amplified from all isolates with primers JE11GNDFW1 (5' TAT GGC AGT GAT GGG GCG 3') and JE12GNDRV1 (5' RTA AGA RAC GAT TTT RCC CAG 3'), which produced a 969 bp amplicon. PCR was performed with GoTaq[®] Flexi polymerase (Promega) on a 2720 Thermal Cycler (Life Technologies). Each 25 µl reaction consisted of 1X GoTaq Flexi PCR buffer, 1 mM MgCl₂, 0.1 mM dNTPs, 0.5 µM of *gnd*-F, 1.0 µM *gnd*-R, 1 U GoTaq[®] Flexi polymerase (Promega), and 1.0 µL of template. Cycling parameters were 94°C for 7 min, followed by 20 cycles of 94°C for 1 min, 62°C for 1 min, 72°C for 1 min where the annealing temperature was reduced by 0.5°C per cycle, and another 20 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min, followed by a final extension period of 7 min at 72°C.

2.3. DNA sequencing. PCR products were purified by column purification with the QIAquick PCR Purification Kit (Qiagen; Valencia, CA) or the GenElute PCR Cleanup Kit (Sigma Aldrich; St. Louis, MO) according to the manufacturers instructions prior to sequence analysis. Sequencing was performed at the Colorado State University's Proteomics and Metabolomics Facility (Fort Collins, CO) using the PCR primers listed above, Big Dye Terminator chemistry, and AmpliTaq-FS DNA Polymerase. Electrophoresis of sequencing reactions was performed on an ABI PRISM 3100 DNA

analyzer. All nucleotide sequences were assembled with Seqman[®] software (DNA Star, Lasergene, Madison, WI). After sequences were assembled, an alignment was created with MegAlign[®] software (version 9.1.0, DNA Star, Lasergene). A trimmed alignment of 670 bp from each isolate was used to identify unique *gnd* allelic types with DnaSP software version 5.10.01.

2.4. Extension primer design for *gnd* SNP serogrouping assay. The partial *gnd* sequences from STEC isolates were used to develop primers for single base extension sequencing of each targeted SNP. Each primer was designed to be complimentary to the 5'-end sequence of the targeted SNP (on either the sense or antisense strand) and that the 3' nucleotide of the primer was one 5' base away from the targeted SNP. By designing the primers in this manner the polymerase extends the primer with a single fluorescently-labeled dideoxynucleotidetriphosphates (ddNTP) base that corresponds to the targeted SNP. Primers were also designed to have annealing temperatures similar to the others included in the same multiplex reaction, between 56-60°C. SNPs were chosen to best capture the diversity within *gnd* by selecting those that had three or more unique nucleotides at each polymorphic site. The aligned *gnd* sequences were used to select the SNPs with a conservative 5' region to minimize mismatching of primers, but when mismatching was unavoidable, degenerate primers were designed. Primers that incorporated degenerate nucleotides to avoid mismatching, were included at higher concentrations within each reaction to account for the reduced annealing efficiency. To allow for separation of fluorescently labeled primers by capillary electrophoresis, non-complementary 5' poly-A tails were incorporated in some of the primers so that within each reaction all primers differed in size by at least eight nucleotides. Primer sequences and corresponding concentrations are shown in Table 2.

2.5. Multiplex *gnd* SNP genotyping for molecular serogrouping of clinically important STEC. Purified PCR products were obtained by the method described above in section 2.3. Further purification of amplicons was necessary to degrade unused primers that resulted in extraneous peaks in electropherograms following capillary electrophoresis to separate single base extension products. Further purification was performed using

Exonuclease I (*ExoI*; Fermentas, Glen Burnie, MD) and Shrimp Alkaline Phosphatase (SAP; Fermentas) or ExoSAP-IT[®] (Affymetrix; Santa Clara, CA). The purified products were combined with ABI PRISM[®] SNaPshot[™] Multiplex Ready Mix (Life Technologies) and primers were extended with fluorescently labeled ddNTPs with PCR products serving as template. Primer concentrations (Table 2) were optimized to produce uniform peaks within each multiplex reaction. After completion of single base extension, the products were treated with alkaline calf intestinal phosphatase to limit interference from unincorporated ddNTPs. The treated products were then combined with Genescan-120 LIZ[™] size standard (Life Technologies) and Hi-Di (deionized) formamide (Life Technologies), before denaturation at 95°C for 5 min. Capillary electrophoresis of samples was performed on the ABI 3100 DNA Analyzer according to the SNaPshot[™] kit instructions. Analyses of electropherograms were performed using Peak Scanner software (version 1.0; Life Technologies). Each STEC isolate was interrogated for 12 *gnd* SNPs in three multiplex single base extension reactions and the combined SNPs read from all three reactions were read to determine the 12 SNP *gnd* profile (Figure 1).

2.6. Cluster analysis. To identify the most informative SNPs and those that could be removed from the analysis to lower cost and improve sensitivity cluster analysis was performed on SNP types that represented the “big six” non-O157 STEC and O157 isolates in the inclusivity panel using the PROC VARCLUS procedure of SAS (version 9.1, SAS Institute Inc; Cary, NC). A SNP from each of the four clusters identified by the procedure was selected based on the $1-R^2$ ratio; the SNPs with the lowest value were those with the highest correlation with the other SNPs included in the same cluster and had the lowest correlation with SNPs in other clusters and were therefore the best SNPs for differentiating “big six” non-O157 STEC serogroups from each other.

2.7. Verification of O-serogroup. The few STEC isolates with *gnd* alleles that differed from majority of other isolates belonging to the same serogroup were analyzed with two commercial real-time PCR kits (i.e., BAX[®] System Real-Time PCR Assay for STEC; DuPont Qualicon, Wilmington, DE and GeneDisc STEC; Pall, Port Washington, NY) and Dryspot kits (Oxoid, Basingstoke, Hampshire, UK) assay as a screen to determine if the

original serogroup information provided was correct. Conventional serotyping by slide agglutination was performed for isolates with discordant serogrouping results at the *E. coli* Reference Center (The Pennsylvania State University, University College, PA).

3. RESULTS AND DISCUSSION:

3.1. Analyses of *gnd* sequences revealed 12 informative SNPs for the identification of STEC belonging to O-serogroups O26, O45, O103, O111, O121, O145, and O157. We identified 42 unique *gnd* allelic types in the collection of 195 STEC isolates included in this study. All isolates belonging to serogroup O26 carried one *gnd* allelic type, while O45, O103, O111, O121, O145 and O157 isolates carried multiple *gnd* allelic types. However, STEC isolates within each serogroup represented by multiple *gnd* allelic types, the majority of isolates represented one predominant allelic type (Table 3). One sequence representing each of the 42 unique *gnd* allelic types was used to create a phylogenetic tree (not shown) to visualize clustering of isolates. The phylogenetic tree illustrated that the majority of isolates belonging to a particular O-serogroup clustered together to form a well-supported clade. There were some isolates belonging to each serogroup (n=17 in total), as determined by conventional serogrouping by slide agglutination, that did not cluster with the majority of isolates belonging to their respective serogroup. The isolates that did not cluster as expected were further characterized by commercial PCR assays for detection and classification of non-O157 STEC (i.e., BAX[®] System Real-Time PCR Assay for STEC and the Pall GeneDisc 9 EHEC identification assay), and Drypsot agglutination test kits as an initial screen to verify the original serogroup information provided by the repository. All but three isolates had discordant or negative serogroup results were sent the *E. coli* reference center (The Pennsylvania State University) to be re-serotyped by slide agglutination (Table 4). Results from re-serotyping support that 13 isolates were originally misclassified by conventional serotyping by slide agglutination and updated serogroup information was used for further analyses. With the exception of TTU E1-133 (serogroup O20), which shared the same *gnd* allelic type as the majority of isolates belonging to O111 (allelic type 1), and a single O4 and a single O126 isolates, which also shared the same *gnd*

allele (allelic type 10), all other isolates carried O-serogroup specific *gnd* allelic types, supporting that SNPs in *gnd* could identify and classify the seven clinically important STEC O-serogroups targeted here.

The most informative *gnd* SNPs were selected using an alignment of sequences representing each of the “big six” non-O157 STEC serogroups and O157. By selecting the most polymorphic SNPs, we limited the total number of SNPs required to classify isolates belonging to the “big six” non-O157 STEC and O157 into their respective serogroups and differentiate them from other non-O157 STEC isolates. In the development of this assay, we sought to identify serogroup specific SNP allelic types (i.e., serogroup specific SNP combinations) instead of attempting to locate serotype specific SNPs. For example, in order to differentiate O26, O45, and O103 isolates from each other, two SNPs are necessary instead of a single serogroup specific SNP for each. In this particular case O45 and O103 strains both have a G present at SNP1 (Table 5), which differentiates these strains from O26. Isolates belonging to serogroups O45 and O103 can be differentiated by SNP5, where most O45 isolates have a G while the majority of O103 isolate have a C. By using this approach the assay has improved flexibility over assays that utilize serogroup specific SNPs. While including all 12 SNPs has increased discriminatory power, our cluster analyses showed that the “big six” non-O157 STEC and O157 serogroups can be differentiated from each other with as few as four SNPs (i.e., SNP5, SNP6, SNP7, and SNP8). Therefore this assay can be adapted to have high specificity by interrogating all 12 SNPs or reduced to four SNPs in order to lower the cost of performing the assay. The distribution of SNP types determined by these four SNPs are shown in Table 1 and 4 SNP *gnd* SNP types observed for the “big six” non-O157 STEC and O157 isolates are summarized in Table 3.

3.2. SNP Based Serogrouping is a Reliable Alternative to Conventional and Other Methods for Identification of O-Serogroup. The SNP genotyping assay developed here was able to classify 195 STEC isolates into 49 *gnd* SNP types using twelve SNPs (Table 1). More SNP types were identified than allelic types due to discrepancies between the *gnd* sequencing and the SNP typing assay. Using *gnd* SNP types, the STEC isolates belonging

to the seven serogroups targeted for regulatory and surveillance purposes (i.e. O26, O45, O103, O111, O121, O145, and O157) were classified into their respective serogroup (Table 3). Serogroup O26 had two unique SNP types, O45 had three, O103 had three, O111 had five, O121 had two, O145 had two, and O157 had two (Table 3). Although multiple SNP types were observed for each serogroup except O121, a single SNP and allelic type represented the majority of strains within each O-serogroup. Although serogroup O26 had two SNP types, 20 of 21 isolates belonged to the same *gnd* 12 SNP type (SNP type 7). This was also true of serogroups O45, O103, O111, O145, and O157, where 89.5%, 91.7%, 70.8%, 95.2%, and 94.4% of isolates, respectively, were represented by a single SNP type (Table 3).

Unique SNP types were observed for almost all *gnd* allelic types, indicating that full sequencing is not necessary to identify *gnd* allele, in most cases the 12 informative SNPs identified here can capture the diversity from the full *gnd* sequence. The unique *gnd* alleles 15, 26, and 27 found in serogroup O121 isolates were not differentiated by SNP typing, and were all identified as SNP type 19. The O103 strain TTU A1-014 and the O145 strain TTU E1-190 also had unique *gnd* allele types that differed slightly from the majority of isolates within their respective O-serogroups. SNP typing of these strains did not identify their unique *gnd* alleles and classified them as the same SNP type as the majority of isolates within their O-serogroup. Since the objective of this study was to develop an assay that could confirm or identify O-serogroup it was not necessary to identify the SNPs responsible for these unique *gnd* allelic types within the O-serogroups.

A total of 38 discrepancies were observed out of 2,340 total SNPs interrogated. The disagreements found between the *gnd* sequence and the SNP-typing results were found in 21 of 195 isolates used in the study, 10 of which belonged to “big six” serogroups and three of which were found in isolates originally misclassified by conventional serotyping. These were observed in isolates belonging to the serogroups O2, O8, O45, O111, O145, and O153. Although the SNP types observed for these strains did not match the sequence, the observed SNP types were still able to differentiate these STEC into their respective serogroups. Additionally 13 SNPs out of the 2,340 total were not successfully interrogated

in six isolates, including isolates belonging to serogroups O2, O111 and O128. The O45 and O145 isolates where discrepancies were found between the sequence and SNP typing results belonged to *gnd* allelic types different from the majority of isolates with the same O-serogroup. These isolates may have had other regions similar enough to the targeted sequences that extension primers annealed at the wrong site and the identification of the targeted SNPs was inhibited. Multiple degenerate nucleotides in the sequence of these primers make it difficult to identify the competing sequences that may be present. As mentioned above, an isolate belonging to serogroup O20 shared allelic types with the O111 serogroup. These isolates could not be differentiated by the SNP typing assay either and this was also the case for the O4 and O126 isolates that shared the same allelic type. However, all other O-serogroups had allelic types and SNP types that differentiated them from all other serogroups, and 98.3% of the 2,340 SNPs were successfully interrogated.

A significant advantage of using the SNaPshot™ Multiplex kit for characterizing SNPs is the ease in incorporating more SNPs in the assay if there are multiple serogroups that share the same SNP type. Additionally, this assay does not rely on serogroup specific genes so that as other serogroups arise as clinically significant they may be identified with little to no modification of the current assay. Identification of new serogroups may be as simple as determining the *gnd* sequence from a subset of isolates and using the sequence to determine the possible SNP types of that serogroup. This allows for expansion of the assay to identify the serogroup of potentially all STEC serogroups including serogroups O55, O91, O104, and O113, which have been associated with hemolytic uremic syndrome (Mellmann et al., 2008).

3.3. Conclusions. Although other molecular serotyping methods for non-O157 STEC exist (e.g., Norman et al., 2012, Toro et al., 2013, Timmons et al., 2016), the method presented here possesses several advantages over conventional serotyping and existing molecular methods. The SNP typing assay is more informative than current PCR based methods, providing the allelic type in most cases, which corresponds to the O-serogroup, rather than simply positive or negative for amplification of O-serogroup specific genes. By targeting the most informative SNPs within *gnd* one can identify the allelic type without the need to

conduct sequence analysis and tree building, making the interpretation of the assay less complicated than full sequencing of *gnd*.

Since this assay targets the *gnd* gene, it could potentially work on all STEC serotypes. Because commercial assays typically target serogroup specific genes, multiple *E. coli* strains and enterobacteriaceae do not interfere with the interpretation of results. However, with the exception of Neogen's NeoSeekTM assay, it is impossible to identify whether *stx*, intimin, and "big six" O-serogroup specific genes detected in initial screens are present in the same strain with existing PCR methodologies and commercial kits. This assay is not likely to replace existing commercial PCR assays, however it is not limited in the number of serogroups it can identify and could be used as a reliable method for confirmation of STEC serogroup after strains are isolated.

Identifying the *gnd* allele confers the ability to serogroup isolates other than the "big six" and O157 STEC, which would prove to be valuable if other O-serogroups, such as O104:H4, continue to emerge as significant sources of human illness. Existing PCR assays and commercial kits selectively target "big six" serogroups and would fail to identify the serogroup of seropathotype C strains, such as O5, O91, O113, and O104 which also cause HC and HUS (Paton et al., 1999, Pradel et al., 2008, Vidal et al., 2008). Diagnostic tools used in public health laboratories would ideally not select for a small subset of STEC serogroups. Presently, diagnostic methods have been well established for the detection of STEC O157:H7 but the same is not true of non-O157 STEC, which has led to a knowledge-gap regarding the public health burden of non-O157 STEC.

In this study, we developed and validated a molecular serogrouping assay based on interrogation of informative SNPs in *gnd*. This assay provides an alternative to conventional serogrouping and will facilitate rapid and confirmation of clinically important STEC serogroups currently targeted in the U.S. Confirmation of presumptive STEC colonies from screen positive enrichments currently represents a significant and costly problem. The assay can also be used to facilitate routine surveillance of the public health burden of disease caused by STEC and microbial source tracking investigations.

ACKNOWLEDGEMENTS:

This research was supported by the National Integrated Food Safety Initiative of the USDA Cooperative State Research, Education, and Extension Service, grant number no. 51110-04333. The funding source did not influence the design or outcome of this study. We would like to recognize the contributions of Dr. Chitrita DebRoy of Pennsylvania State University, Dr. Shannon Manning of Michigan State University, Dr. Robert Elder, and Dr. Joseph Bosilevac of USDA Meat Animal Research Center for providing the STEC strains used to develop the molecular serogrouping method d in this study.

FIGURE LEGENDS:

Figure 1. Electropherogram of single base extension reaction for E1-210. Blue peaks indicate the incorporation of ddGTP, green indicates ddATP, red indicates ddTTP, and black indicates ddCTP. Panels A, B, and C indicate multiplex reactions 1, 2 and 3, respectively, as described in Table 2. The allelic type of the isolate shown above was assigned as allelic type 8 is read as CGGT (panel A) TACT (panel B) GGAT (panel C).

REFERENCES:

1. **Bai J, Paddock Z, Shi X, Li S, An B, Nagaraja T.** 2012. Applicability of a multiplex PCR to detect the seven major Shiga toxin-producing *Escherichia coli* based on genes that code for serogroup-specific O-antigens and major virulence factors in cattle feces. *Foodborne Pathog. and Dis.* **6**:541–548.
2. **Bettelheim KA.** 2003. Non-O157 verotoxin-producing *Escherichia coli*: a problem, paradox, and paradigm. *Exp. Biol. Med.* **228**:333–344.
3. **Bielaszewska M, Janda J, Blahová K, Feber J, Potuzník V, Soucková A.** 1996. Verocytotoxin-producing *Escherichia coli* in children with hemolytic uremic syndrome in the Czech Republic. *Clin. Nephrol.* **46**:42–44.
4. **Blanco J, et al.** 2003. Verotoxin-producing *Escherichia coli* in Spain: prevalence, serotypes, and virulence genes of O157:H7 and non-O157 VTEC in ruminants, raw beef products, and humans. *Exp. Biol. Med.* **228**:345–351.
5. **Boerlin P, McEwen SA, Boerlin-Petzold F, Wilson JB, Johnson RP, Gyles CL.** 1999. Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. *J. Clin. Microbiol.* **37**:497–503.
6. **Buven G, Lauwers S, Piérard D.** 2010. Prevalence of subtilase cytotoxin in verocytotoxin-producing *Escherichia coli* isolated from humans and raw meats in Belgium. *Eur. J. Clin. Microbiol. Infect. Dis.* **29**:1395–1399.
7. **Carroll AM, Gibson A, McNamara EB.** 2005. Laboratory-based surveillance of human verocytotoxigenic *Escherichia coli* infection in the Republic of Ireland, 2002–2004. *J. Med. Microbiol.* **54**:1163–1169.
8. **CDC.** 2011. Vital signs: incidence and trends of infection with pathogens transmitted commonly through food—foodborne diseases active surveillance network, 10 U.S. sites, 1996–2010. *MMWR Morb. Mortal. Wkly. Rep.* **60**:749–755.
9. **Contreras CA, et al.** 2011. Phylogenetic relationships of Shiga toxin-producing *Escherichia coli* isolated from Peruvian children. *J. Med. Microbiol.* **60**:639–646.
10. **Eblen DR.** 2008. Public health importance of non-O157 Shiga toxin-producing *Escherichia coli* (non-O157 STEC) in the US Food Supply. U.S. Food Inspection Service.

11. **Eklund M, Scheutz F, Siitonen A.** 2001. Clinical isolates of non-O157 Shiga toxin-producing *Escherichia coli*: serotypes, virulence characteristics, and molecular profiles of strains of the same serotype. *J. Clin. Microbiol.* **39**:2829–2834.
12. **Elliot EJ, Robins-Browne R, Hogg G.** 2001. Characteristics of Shiga toxin-producing *E. coli* causing HUS in Australia. *J. Gastroenterol. and Hepatol.* **16**:16.
13. **Ethelberg S, et al.** 2004. Virulence factors for hemolytic uremic syndrome, Denmark. *Emerg. Infect. Dis* **10**:842–847.
14. **Fratamico PM, Briggs CE, Needle D, Chen C-Y, DebRoy C.** 2003. Sequence of the *Escherichia coli* O121 O-antigen gene cluster and detection of enterohemorrhagic *E. coli* O121 by PCR amplification of the *wzx* and *wzy* genes. *J. Clin. Microbiol.* **41**:3379–3383.
15. **Fratamico PM, DebRoy C, Miyamoto T, Liu Y.** 2009. PCR detection of enterohemorrhagic *Escherichia coli* O145 in food by targeting genes in the *E. coli* O145 O-antigen gene cluster and the Shiga toxin 1 and Shiga toxin 2 genes. *Foodborne Pathog. and Dis.* **6**:605–611.
16. **Fratamico PM, DebRoy C, Strobaugh Jr. TP, Chen C-Y.** 2005. DNA sequence of the *Escherichia coli* O103 O-antigen gene cluster and detection of enterohemorrhagic *E. coli* O103 by PCR amplification of the *wzx* and *wzy* genes. *Can. J. Microbiol.* **51**:515–522.
17. **Gerber A, Karch H, Allerberger F, Verweyen HM, Zimmerhackl LB.** 2002. Clinical course and the role of Shiga toxin–producing *Escherichia coli* infection in the hemolytic-uremic syndrome in pediatric patients, 1997–2000, in Germany and Austria: a prospective study. *J. Infect. Dis.* **186**:493–500.
18. **Gilmour MW, Olson AB, Andrysiak AK, Ng L-K, Chui L.** 2007. Sequence-based typing of genetic targets encoded outside of the O-antigen gene cluster is indicative of Shiga toxin-producing *Escherichia coli* serogroup lineages. *J. Med. Microbiol.* **56**:620–628.
19. **Gould LH, Demma L, Jones TF, Hurd S, Vugia DJ, Smith K, Shiferaw B, Segler S, Palmer A, Zansky S, Griffin PM.** 2009. Hemolytic uremic syndrome and death in persons with *Escherichia coli* O157:H7 infection, Foodborne Diseases Active Surveillance Network sites, 2000–2006. *Clin. Infect. Dis.* **49**:1480–1485.
20. **Hu Y, Zhang Q., Meitzler JC.** 1999. Rapid and sensitive detection of *Escherichia coli* O157:H7 in bovine faeces by a multiplex PCR. *J. Appl. Microbiol.* **87**:867–876.

21. **Hughes JM, Wilson ME, Johnson KE, Thorpe CM, Sears CL.** 2006. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin. Infect. Dis. **43**:1587–1595.
22. **Käppeli U, Hächler H, Giezendanner N, Beutin L, Stephan R.** 2011. Human infections with non-O157 Shiga toxin-producing *Escherichia coli*, Switzerland, 2000–2009. Emerg. Infect. Dis. **17**:180–185.
23. **Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, Reid-Smith R, Isaac-Renton J, Clark C, Rahn K, Kaper JB.** 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.
24. **Karmali MA, Petric M, Lim C, Fleming PC, Arbus GS, Lior H.** 1985. The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing *Escherichia coli*. J. Infect. Dis. **151**:775–782.
25. **Lacher DW, Walk ST, Wick LM, Whittam TS.** 2004. Multiplex PCR detection of intimin and Shiga toxin genes. The STEC Center.
26. **Lin A, Nguyen L, Lee T, Clotilde LM, Kase JA, Son I, Carter JM, Lauzon CR.** 2011. Rapid O-serogroup identification of the ten most clinically relevant STECs by Luminex microbead-based suspension array. J. Microbiol. Methods **87**:105–110.
27. **Lin A, Sultan O, Lau HK, Wong E, Hartman G, Lauzon CR.** 2011. O-serogroup specific real time PCR assays for the detection and identification of nine clinically relevant non-O157 STECs. Food Microbiol. **28**:478–483.
28. **Lopez EL, Diaz M, Grinstein S, Devoto S, Mendilaharsu F, Murray BE, Ashkenazi S, Ruboglio E, Woloj M, Vasquez M, Turco M, Pickering LK, Cleary TG.** 1989. Hemolytic uremic syndrome and diarrhea in Argentine Children: the role of Shiga-like toxins. J. Infect. Dis. **160**:469–475.
29. **Machado J, Grimont F, Grimont PAD.** 2000. Identification of *Escherichia coli* flagellar types by restriction of the amplified *fliC* gene. Res. Microbiol. **151**:535–546.
30. **Manuel C.** 2011. Master's Thesis. Department of Animal Sciences, Colorado State University, Fort Collins, CO.

31. **Mellmann A, Bielaszewska M, Köck R, Friedrich AW, Fruth A, Middendorf B, Harmsen D, Schmidt MA, Karch H.** 2008. Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli*. *Emerg. Infect. Dis.* **14**:1287–1290.
32. **Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H.** 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS ONE* **6**:e22751.
33. **Meng J, Zhao S, Doyle MP, Mitchell SE, Kresovich S.** 1997. A multiplex PCR for identifying Shiga-like toxin-producing *Escherichia coli* O157 : H7. *Lett. Appl. Microbiol.* **24**:172–176.
34. **Norman, KN, Strockbine NA, and Bono JL.** 2012. Association of nucleotide polymorphisms within the O-antigen gene cluster of *Escherichia coli* O26, O45, O103, O111, and O145 with serogroups and genetic subtypes. *Appl. Environ. Microbiol.* **78**:6689-6703.
35. **Paton AW, Paton JC.** 1998. Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for *stx1*, *stx2*, *eaeA*, enterohemorrhagic *E. coli* *hlyA*, *rfbO111*, and *rfbO157*. *J. Clin. Microbiol.* **36**:598–602.
36. **Paton AW, Woodrow MC, Doyle RM, Lanser JA, Paton JC.** 1999. Molecular characterization of a Shiga toxigenic *Escherichia coli* O113:H21 strain lacking *eae* responsible for a cluster of cases of hemolytic-uremic syndrome. *J. Clin. Microbiol.* **37**:3357–3361.
37. **Perelle S, Dilasser F, Grout J, Fach P.** 2005. Detection of *Escherichia coli* serogroup O103 by real-time polymerase chain reaction. *J. Appl. Microbiol.* **98**:1162–1168.
38. **Perelle S, Dilasser F, Grout J, Fach P.** 2007. Screening food raw materials for the presence of the world's most frequent clinical cases of Shiga toxin-encoding *Escherichia coli* O26, O103, O111, O145 and O157. *Int. J. Food Microbiol.* **113**:284–288.

39. **Pradel N, Bertin Y, Martin C, Livrelli V.** 2008. Molecular analysis of Shiga toxin-producing *Escherichia coli* strains isolated from hemolytic-uremic syndrome patients and dairy samples in France. *Appl. Environ. Microbiol.* **74**:2118–2128.
40. **Rivero MA, Passucci JA, Rodriguez EM, Parma AE.** 2010. Role and clinical course of verotoxigenic *Escherichia coli* infections in childhood acute diarrhoea in Argentina. *J. Med. Microbiol.* **59**:345–352.
41. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. *Emerg. Infect. Dis.* **17**:7–15.
42. **Timmons C, Trees E, Ribot EM, Gerner-Smidt P, LaFon P, Im S, and Ma LM.** 2016. Multiple-locus variable number tandem repeat analysis for strain discrimination of non-O157 Shiga toxin producing *Escherichia coli*. *J. Microbiol. Methods.* 125:70-80.
43. **Toro M, Najjar MB, Ju W, Brown E, Zhao S, and Meng J.** 2013. Molecular serotyping of Shiga toxin-producing *Escherichia coli* using suspension array. *Foodborne Pathog. Dis.* 10:478-480.
44. **Tozzi AE, Caprioli A, Minelli F, Gianviti A, De Petris L, Edefonti A, Montini G, Ferretti A, De Palo T, Gaido M, Rizzoni G.** 2003. Shiga toxin-producing *Escherichia coli* infections associated with hemolytic uremic syndrome, Italy, 1988–2000. *Emerg. Infect. Dis.* **9**:106–108.
45. **USDA.** 2011. Shiga Toxin-Producing *Escherichia coli* in certain raw beef products. *Federal Register* **76**:58157– 58165.
46. **Vidal M, Prado V, Whitlock GC, Solari A, Torres AG, Vidal RM.** 2008. Subtractive hybridization and identification of putative adhesins in a Shiga toxin-producing *eae*-negative *Escherichia coli*. *Microbiology* **154**:3639–3648.
47. **Werber D, Beutin L, Pichner R, Stark K, Fruth A.** 2008. Shiga toxin-producing *Escherichia coli* serogroups in food and patients, Germany. *Emerg. Infect. Dis.* **14**:1803–1806.
48. **Willshaw GA, Cheasty T, Smith HR, O'brien SJ, Adak GK.** 2001. Verocytotoxin-producing *Escherichia coli* (VTEC) O157 and other VTEC from human infections in England and Wales. *J. Med. Microbiol.* **50**:135–142.

Table 1. Description of shiga toxin producing *Escherichia coli* isolates used to develop the *gnd* SNP genotyping assay and description of results from subtyping by *gnd* sequencing, the 12 SNP *gnd* genotyping assay developed here and *gnd* SNP genotyping using 4 informative SNPs identified through cluster analyses.

Isolate ID	Isolate Source ^a	<i>stx1</i> ^b	<i>stx2</i>	<i>eae</i>	Confirmed O-serogroup	<i>gnd</i> allelic type	<i>gnd</i> 12 SNP profile ^c	<i>gnd</i> 12 SNP type	<i>gnd</i> 4 SNP profile	<i>gnd</i> 4 SNP type
A1- 001	Human	+	+	+	111	1	CGTCGTCGAGCT	11	GTTC	19
A1- 002	Cow	+	+	+	111	1	GCTCGTCGAGCT	31	CTTC	11
A1- 003	Cow	+	+	+	111	1	CGTCGTCGAGCT	11	GTTC	19
A1- 004	Pig	+	+	+	111	1	GCTCGTCGAGCT	31	CTTC	11
A1- 005	Human	+	+	+	111	1	CCTCGTCGAGCT	9	CTTC	11
A1- 006	antelope	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
A1- 007	Pig	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
A1- 008	Human	+	-	-	NT	2	GAGGAGTTTCCT	25	AGGT	8
A1- 009	Cow	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
A1- 010	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
A1- 011	Horse	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
A1- 012	Mouse	+	-	-	103	5	GAGGAGTCAACA	20	AGGT	8
A1- 013	Goat	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
A1- 014	Cow	+	+	+	103	6	GCGGTGCTTACT	30	CGGC	10
A1- 015	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
A1- 016	Cow	-	+	-	147	7	GAGGAACCAACT	16	AAGC	4
A1- 017	Pig	-	+	-	147	7	GAGGAACCAACT	16	AAGC	4
A1- 018	Pig	-	+	-	147	7	GAGGAACCAACT	16	AAGC	4
A1- 019	Pig	-	+	-	147	7	GAGGAACCAACT	16	AAGC	4
A1- 020	Pig	-	+	-	147	29	GAGGAACCAACT	16	AAGC	4
A1- 021	Cow	+	+	-	2	30	GGGCAATCGACT	36	GAGT	14
A1- 022	Cow	+	+	-	2	8	GAG-AACCAACT	29	AAGC	4
A1- 023	Cow	+	+	-	2	8	GAG-AACCAACT	29	AAGC	4
A1- 024	Cow	+	+	-	2	8	GAG-AACCAACT	29	AAGC	4
A1- 025	Cat	+	+	-	2	9	GAGGCGGTGCCG	26	AGGG	7
A1- 026	Antelope	+	-	-	5	9	GGGGAGTTGCCG	41	GAGT	14
A1- 027	Goat	+	-	-	5	9	GGGGAGTTGCG	41*	GGGT	17
A1- 028	Food	+	+	+	5	9	GGGGAGTTGCCG	41	GAGT	14
A1- 029	Human	+	-	-	5	9	AGGGAGTTGCCG	8	GGGT	17
A1- 030	Cow	+	-	+	5	9	GGGGAGTTGCCG	41	GAGT	14
A1- 031	Rabbit	+	-	+	153	31	GGGGTATTGCCT	42	GAGT	14
A1- 032	Rabbit	+	-	+	153	31	GGGGTATTGCCT	42	GAGT	14
A1- 033	Cow	+	+	-	153	31	GGGGAATTGCCT	40	GAGT	14
A1- 034	Rabbit	+	-	+	153	31	GGGGAATTGCCT	40	GAGT	14
A1- 035	Rabbit	+	-	+	153	31	GGGGTATTGCCT	42	GAGT	14
A1- 036	Okapi	+	-	-	128	32	GGGGTGTTCCT	44	GGGT	17
A1- 037	Okapi	+	-	-	128		GG-TGG----	47	GG--	15
A1- 038	Sheep	+	+	-	128	33	GGGTAGTCGACA	45	GGGT	17
A1- 039	Sheep	+	+	-	128	33	GAGGAGTCGACA	21	AGGT	8
A1- 040	Sheep	+	+	-	128	33	GAGGAGTCGACA	21	AGGT	8
A1- 041	Pig	+	-	-	8	34	GAGGAGTTGACT	22	AGGT	8
A1- 042	Cow	+	+	-	8	35	GAGTAACTGACT	28	AAGC	4
A1- 043	Food	+	-	-	8	36	GAGGAACCGACT	18	AAGC	4
A1- 044	Pig	-	+	-	8	37	AAAGAGTTGACT	3	AGAT	6
A1- 045	Pig	-	+	-	8	38	AAAGAATTGACA	1	AAAT	3
E1- 004	NA	+	-	-	126	10	GGGGTGCCAACT	43	GGGC	16
E1- 018	NA	+	-	-	4	10	GGGGTGCCAACT	43	GGGC	16
E1- 025	NA	+	-	+	26	3	AGGGAGTCAACC	6	GGGT	17
E1- 028	NA	-	+	-	14	11	GGAGCATTGACC	33	GAAT	12

E1- 038	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 040	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 042	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 043	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 044	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 045	NA	+	-	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 046	NA	-	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 047	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 049	NA	+	-	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 053	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 054	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 055	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 056	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 058	NA	+	+	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 061	NA	+	-	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 066	NA	+	-	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 075	NA	+	-	+	157	39	GAAGAGTTGACT	15	AGAT	6
E1- 091	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 092	Human	+	+	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 093	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 094	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 095	Human	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 096	Human	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 097	Human	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 098	Human	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 099	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 100	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 101	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 102	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 103	Human	-	-	-	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 104	Human	+	+	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 105	Human	+	-	+	111	1	GAGGGTCGAGCT	27	ATGC	9
E1- 106	Human	+	+	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 107	Human	-	+	-	113	14	GAAAAGTTGACT	14	AGAT	6
E1- 108	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 109	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 110	Human	+	-	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 112	Human	-	-	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 113	Human	+	-	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 114	Human	-	+	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 115	Cow	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 116	Cow	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 117	Cow	-	-	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 118	Calf	-	+	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 119	Cow	-	+	+	172	17	GGGGAATCAACA	39	GAGT	14
E1- 120	Cow	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 121	Calf	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 122	Cow	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 123	Cow	+	+	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 124	NA	-	-	+	157	12	AAAGAGCTGACT	2	AGAC	5
E1- 125	NA	-	-	+	45	18	GTTTCGCTAAAG	49	TGTC	20
E1- 126	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 127	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 128	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 129	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 130	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 132	Human	-	-	-	45	13	GGGCAACCGACT	35	GAGC	13

E1- 133	Human	+	-	-	30	1	CGTCGTCGAGCT	11	GTTC	19
E1- 134	Human	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 135	Human	+	-	-	171	19	GAGGAGTTGCTT	24	AGGT	8
E1- 136	Human	+	+	-	73	20	GGGGAACCTGACC	38	GAGC	13
E1- 137	Human	+	-	-	45	18	GGTTCGCTAAAG	46	GGTC	18
E1- 138	Human	-	-	-	NT	40	GTGGAACCGACG	48	TAGC	19
E1- 139	Human	+	-	-	45	13	GGGCAACCGACT	35	GAGC	13
E1- 140	Human	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 141	Human	+	-	-	103	21	GAGGAACCGACA	17	AAGC	4
E1- 142	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 143	Human	+	+	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 144	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 145	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 146	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 147	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 148	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 149	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 150	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 151	Human	+	-	-	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 152	Human	+	-	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 153	Human	+	-	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 154	Human	+	-	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 155	Human	+	-	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 156	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 157	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 158	Human	+	-	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 159	Human	+	-	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 160	Human	+	-	-	121	27	GAGGAACCGCCT	19	AAGC	4
E1- 161	Human	+	-	-	121	27	GAGGAACCGCCT	19	AAGC	4
E1- 162	Human	+	-	-	121	27	GAGGAACCGCCT	19	AAGC	4
E1- 163	Human	+	-	-	121	26	GAGGAACCGCCT	19	AAGC	4
E1- 164	Human	+	-	-	121	27	GAGGAACCGCCT	19	AAGC	4
E1- 165	Human	+	-	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 166	Human	+	+	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 167	Human	-	+	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 168	Human	-	-	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 169	Cow	+	-	-	NT	22	CGGTAACCTGACT	10	GAGC	13
E1- 170	Cow	+	-	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 171	Cow	-	+	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 172	Cow	+	+	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 173	Cow	+	-	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 174	Cow	-	-	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 175	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 176	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 177	Human	+	+	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 178	Cow	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 179	Cow	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 180	Human	-	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 181	Human	-	-	-	45	13	GGGCAACCGACT	35	GAGC	13
E1- 182	Human	+	+	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 183	Human	+	+	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 184	Human	+	+	+	111	1	G-GCCTCGAGCT	34	-TGC	1
E1- 185	Human	+	+	+	111	1	C-TCGTCG-GCT	13	-TTC	2
E1- 186	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 187	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 188	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 189	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4

E1- 190	Cow	?	?	+	145	23	AAGCTGTTGACT	5	AGGT	8
E1- 191	Human	?	?	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 192	Human	+	-	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 193	Human	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 194	Beef	+	-	+	26	3	AGGGAGTCAACG	7	GGGT	17
E1- 195	Beef	?	+	-	174	24	GGAGAATCGACT	32	GAAT	12
E1- 196	Human	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 197	Human	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 198	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 199	Beef	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 200	Beef	-	+	-	171	25	GAGGAGTTGCCT	23	AGGT	8
E1- 201	Human	+	+	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 202	Human	+	-	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 203	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 204	Beef	+	-	-	121	26	GAGGAACCGCCT	19	AAGC	4
E1- 205	Beef	+	-	-	121	27	GAGGAACCGCCT	19	AAGC	4
E1- 206	Human	+	+	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 207	Beef	+	?	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 208	Human	+	-	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 209	Human	+	-	+	103	4	GCGGTGCTTACT	30	CGGC	10
E1- 210	Cow	+	-	+	118	22	CGGTAAC TGACT	10	GAGC	13
E1- 211	Human	+	+	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 213	Human	+	-	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 214	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 215	Human	?	?	+	145	16	AAGCTGTTGACT	5	AGGT	8
E1- 216	Cow	+	+	+	111	1	CGTCGTCGAGCT	11	GTTC	19
E1- 217	Human	+	-	+	156	28	GGGCCGCTGACT	37	GGGC	16
E1- 218	Rabbit	-	-	+	145	41	AAGCTGTTGACA	4	AGGT	8
E1- 219	Human	+	-	+	111	42	CGTCGTCGATCT	12	GTTC	18
E1- 220	Human	+	-	+	45	13	GGGCAACCGACT	35	GAGC	13
E1- 221	Human	-	+	+	121	15	GAGGAACCGCCT	19	AAGC	4
E1- 222	Human			+	121	15	GAGGAACCGCCT	19	AAGC	4

^a "NA" indicates information on isolate source was not available.

^b "+" indicates gene was present, "-" indicates gene was absent and "?" indicates information was not available or results were indeterminate.

^c "-" indicates SNP was not determined by single base extension

Table 2. Oligonucleotides used for single base extension sequencing of informative single nucleotide polymorphisms in *gnd*

Primer name	Primer sequence ^{a,b}	Concentration	Multiplex reaction no.	Allele ^c	Product size (bp) ^d
1FA	GTG GTG ARG ARG GYG C	0.275μM	1	A (green)	34
				C (black)	32
				G (blue)	33
5RA	(A) ₁₆ RT ARG TWA YRC AYG GYT C	1.15μM	1	A (green)	36
				C (black)	36
				G (blue)	35
				T (red)	38
7FA	(A) ₂₅ GGA CYA GCC AGA GCK C	0.075μM	1	A (green)	40
				G (blue)	40
				T (red)	42
3FB	(A) ₂₈ GTC CTT CYA TYA TGC CWG GTG G	0.5μM	1	A (green)	48
				C (black)	47
				G (blue)	49
				T (red)	47
2RA	(A) ₅ CCW GGC ATR ATR GAW GG	0.5μM	2	A (green)	35
				C (black)	31
				G (blue)	34
				T (red)	36
6FA	(A) ₂₁ GTG ATY YTG GAY GAA GC	0.75μM	2	A (green)	40
				G (blue)	39
				T (red)	39
8FA	(A) ₂₈ TTC TCT GAA AGA BCA GCG	0.25μM	2	C (black)	46
				T (red)	47
9FA	(A) ₃₄ YTT YAT YGG TAC CGG YGT	0.75μM	2	C (black)	51
				G (blue)	50
				T (red)	53
12FA	RAT MAG CGA CAG YGG TTC	0.4μM	3	A (green)	33
				G (blue)	31
10RA	(A) ₁₃ GCG GCA ACR CGC TG	0.95μM	3	A (green)	37
				C (black)	38
13RA	(A) ₁₅ CAT CCA GGA TCA CAT CAA C	0.25μM	3	C (black)	36
				T (red)	38
14FA	(A) ₂₀ TGC TGA AGC CTA TTC TCT GCT	0.4μM	3	A (green)	44
				T (red)	46

^aPoly-A tails used to separate primers in length are indicated by (A) followed by the number of adenine residues included in the tail.

^bDegenerate nucleotides included in some primers to improve annealing efficiency, where B=[C, G, or T], K=[G or T], M=[A or C], R=[A or G],

W=[A or T], Y=[C or T].

^cColors in electropherograms that represent dideoxynucleotide incorporated.

^dProduct size will differ from expected size due to the influence of the incorporated fluorescent dye on the mobility of the primer.

Table 3. Distribution of isolates within the “big six” non-O157 STEC and O157 O-serogroups within each subtype as determined using sequence data from *gnd* sequencing, the 12 SNP multiplex assay or four *gnd* SNPs identified through cluster analyses.

Serogroup (No. Isolates)	<i>gnd</i> allelic type (No. Isolates) ^a	<i>gnd</i> 12 SNP type ^b	<i>gnd</i> 4 SNP type ^c
O26 (21)	3 (21)	6 (1), 7 (20)	17 (21)
O45 (19)	13 (17) , 18 (2)	35 (17) , 16 (1), 49 (1)	13 (17) , 18 (1), 20 (1)
O103 (24)	4 (21) , 5 (1), 6 (1), 21 (1)	17 (1), 20 (1), 30 (22)	10 (24)
O111 (24)	1 (23) , 42 (1)	9 (1), 11 (17) , 12 (1), 13 (1), 27 (1)	1 (1), 2 (1), 9 (1), 11 (3), 19 (18)
O121 (23)	15 (16) , 26 (2), 27 (5)	19 (23)	4 (23)
O145 (21)	16 (19) , 23 (1), 41 (1)	4 (1), 5 (20)	8 (21)
O157 (18)	12 (17) , 39 (1)	2 (17) , 15 (1)	5 (17) , 6 (1)

^aPredominant SNP type observed for each O-serogroup is indicated in bold type.

^b*gnd* SNP based on all 12 SNPs used to develop the assay.

^c*gnd* SNP type based on 4 SNPs (i.e., SNP5, SNP6, SNP7 and SNP8) identified through cluster analyses to be sufficient to differentiate the “big six” non-O157 STEC and O157 O-serogroups.

Table 4. Characterization of isolates with unique *gnd* alleles that did not cluster within well-supported clades for respective serogroup.

Isolate	Original O-serogroup	Confirmed O-serogroup	Latex Agglutination ^a	BAX ^b	GeneDisc
TTU A1-008	26	Not typeable	-	-	-
TTU A1-012	103	103	103	103	103
TTU E1-075	157	157	157	N/A	157
TTU E1-107	121	113	N/A	-	113
TTU E1-119	157	172	-	N/A	-
TTU E1-125	45	45	N/A	-	-
TTU E1-133	20	30	-	-	-
TTU E1-135	45	171	N/A	-	-
TTU E1-136	45	73	N/A	-	-
TTU E1-137	45	45	N/A	-	-
TTU E1-138	45	Not typeable	N/A	-	-
TTU E1-141	103	103	103	103	103
TTU E1-169	145	Not typeable	145	-	-
TTU E1-195	26	174	-	-	-
TTU E1-200	103	171	-	-	-
TTU E1-210	145	118	-	-	-
TTU E1-217	145	156	-	-	-

^aLatex agglutination kits were not available for O-types 45 and 121^bBAX[®] System Real-Time PCR Assay for STEC does not identify O-type 157

Table 5. Nucleotide identified by single base extension sequencing of single nucleotide polymorphisms in *gnd* for isolates belonging to O-serogroups

Primer	Allele ^a	O-serogroup ^b
1FA	A (green)	5, 8, 26, 145, 157
	C (black)	2, 20, 111, 145
	G (blue)	2, 4, 5, 8, 14, 26, 45, 103, 111, 121 , 126, 128, 145 , 147, 153, 157
5RA	A (green)	2, 8, 26, 45, 103, 121 , 128, 145 , 147, 157
	C (black)	103, 111
	G (blue)	2, 4, 5, 20, 26, 45, 111 , 126, 145 , 153, 157
	T (red)	45
7FA	A (green)	8, 14, 26, 121, 157
	G (blue)	2, 5, 8, 26, 45, 103, 111, 121 , 126, 128, 145 , 147, 153, 157
	T (red)	20, 45, 111
3FB	A (green)	121
	C (black)	2, 20, 45, 111, 145
	G (blue)	4, 5, 8, 14, 26, 45, 103, 121 , 126, 128, 147, 153, 157
	T (red)	2, 8, 45, 128, 145
2RA	A (green)	2, 5, 8, 26, 45, 103, 121 , 128, 145 , 147, 153, 157
	C (black)	2, 14, 45, 111, 145
	G (blue)	20, 111 , 128
	T (red)	4, 103, 145 , 153
6FA	A (green)	2, 8, 14, 26, 45, 103, 121, 145 , 147, 153
	G (blue)	2, 4, 5, 8, 26, 45, 103, 121 , 126, 128, 145, 157
	T (red)	20, 111
8FA	C (black)	2, 4, 8, 20, 45, 103, 111, 121 , 126, 145 , 147, 157
	G (blue)	103
	T (red)	2, 5, 8, 14, 26, 45, 103, 121 , 128, 145 , 153, 157
9FA	C (black)	2, 4, 8, 26, 45, 103, 121 , 126, 128, 146, 157
	G (blue)	20, 111
	T (red)	5, 8, 14, 26, 45, 103, 121 , 128, 145 , 153, 157
12FA	A (green)	26, 103, 111
	G (blue)	14, 121 , 128, 145
10RC	A (green)	103 , 128, 145 , 147
	C (black)	121
	C (red)	45
1		
3RA	T (black)	103
14FA	A (green)	103 , 128
	C (black)	14, 26
	T (red)	8, 111, 121, 145

^a Colors indicate resultant nucleotide sequenced by single base extension^b The “bix six” non-O157 STEC serogroups and O157 sequences are indicated bold type.