Journal of Microbiological Methods 129 (2016) 85–93

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

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 [a,b], M. Bugarel [a], H.C. den Bakker [a,c], G.H. Loneragan [a], K.K. Nightingale [a,b,*]

[a] Department of Food and Animal Sciences, Texas Tech University, Lubbock, TX 79409, United States
[b] Department of Animal Sciences, Colorado State University, Fort Collins, CO 80523, United States
[c] Department of Food Science and Technology, Cornell University, Ithaca, NY 14853, United States

ARTICLE INFO

Article history:
Received 9 June 2016
Received in revised form 6 July 2016
Accepted 7 July 2016
Available online 16 July 2016

Keywords:
Shiga toxin producing *Escherichia coli* (STEC)
E. coli O157:H7
Non-O157 STEC
Molecular serotyping
SNP typing
gnd

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.

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 based on data from 2006 through 2013, the incidence of laboratory confirmed *E. coli* O157:H7 infections decreased to 1.15 cases per 100,000 (CDC, 2014). During that same time frame (2006–2013) the rate of laboratory confirmed non-O157 STEC infections was 1.17 cases per 100,000 (CDC, 2011). 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,

* Corresponding author at: 201 Animal and Food Sciences Building, Department of Animal and Food Sciences, Texas Tech University, Lubbock, TX 79409, United States.
E-mail address: kendra.nightingale@ttu.edu (K.K. Nightingale).

http://dx.doi.org/10.1016/j.mimet.2016.07.005
0167-7012/

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 et al., 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, 2011). 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, b, Paton and Paton, 1998, Perelle et al., 2005, 2007). 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 (Mellmann 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 1× 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.