Journal of Microbiological Methods 125 (2016) 70–80

Contents lists available at ScienceDirect

Journal of Microbiological Methods

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

Multiple-locus variable-number tandem repeat analysis for strain discrimination of non-O157 Shiga toxin-producing *Escherichia coli*

Chris Timmons [a], Eija Trees [b], Efrain M. Ribot [b], Peter Gerner-Smidt [b], Patti LaFon [b], Sung Im [b], Li Maria Ma [a,*]

[a] National Institute for Microbial Forensics & Food and Agricultural Biosecurity, Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater OK 74078, United States
[b] Centers for Disease Control and Prevention, Atlanta, GA 30329, United States

ARTICLE INFO

Article history:
Received 24 February 2016
Received in revised form 4 April 2016
Accepted 6 April 2016
Available online 9 April 2016

Keywords:
Escherichia coli O157
Non-O157 STEC
MLVA
Strain discrimination
Outbreak investigation

ABSTRACT

Non-O157 Shiga toxin-producing *Escherichia coli* (STEC) are foodborne pathogens of growing concern worldwide that have been associated with several recent multistate and multinational outbreaks of foodborne illness. Rapid and sensitive molecular-based bacterial strain discrimination methods are critical for timely outbreak identification and contaminated food source traceback. One such method, multiple-locus variable-number tandem repeat analysis (MLVA), is being used with increasing frequency in foodborne illness outbreak investigations to augment the current gold standard bacterial subtyping technique, pulsed-field gel electrophoresis (PFGE). The objective of this study was to develop a MLVA assay for intra- and inter-serogroup discrimination of six major non-O157 STEC serogroups—O26, O111, O103, O121, O45, and O145—and perform a preliminary internal validation of the method on a limited number of clinical isolates. The resultant MLVA scheme consists of ten variable number tandem repeat (VNTR) loci amplified in three multiplex PCR reactions. Sixty-five unique MLVA types were obtained among 84 clinical non-O157 STEC strains comprised of geographically diverse sporadic and outbreak related isolates. Compared to PFGE, the developed MLVA scheme allowed similar discrimination among serogroups O26, O111, O103, and O121 but not among O145 and O45. To more fully compare the discriminatory power of this preliminary MLVA method to PFGE and to determine its epidemiological congruence, a thorough internal and external validation needs to be performed on a carefully selected large panel of strains, including multiple isolates from single outbreaks.

1. Introduction

Escherichia coli is a genetically diverse enteric bacterial species that is an essential constituent of the natural gut micro flora of many warm-blooded organisms. Most *E. coli* strains are commensal, but some are pathogenic to humans. The most severe and life-threatening human illness caused by *E. coli*, hemolytic uremic syndrome (HUS), is associated with the production of one or more Shiga toxins and expression of a few other virulence determinants (O'Brien et al., 1992; Ethelberg et al., 2004; Gyles, 2007; Besser et al., 1999; Tarr et al., 2005). Of over 100 Shiga toxin-producing *E. coli* (STEC) serogroups identified by the World Health Organization, O157 is the most commonly isolated serogroup in the United States and causes the highest percentage of illnesses (Scallan et al., 2011; Johnson et al., 1996; CDC, 2012). However, non-O157 STEC serogroups have been increasingly associated with human illness in recent years and have caused several major outbreaks (Brooks et al., 2005; Johnson et al., 2006; Bettelheim, 2007). Non-O157 STEC serogroups O26, O111, O103, O121, O45, and O145 are the most frequently isolated in the United States and are often referred to as the 'big 6' non-O157 STEC serogroups (Karmali et al., 2003).

Molecular bacterial subtyping methods are essential tools in outbreak investigations involving STEC, from the initial identification of clusters of foodborne illness, the outbreak investigation process, and while monitoring the effectiveness of product recalls. The PulseNet network coordinated by the United States Centers for Disease Control and Prevention (CDC) and the Association of Public Health Laboratories (APHL) is the national molecular subtyping network that functions as a foodborne illness cluster detection tool. The primary bacterial subtyping method used by PulseNet is pulsed-field gel electrophoresis (PFGE), the current gold standard bacterial subtyping method for foodborne pathogens (Swaminathan et al., 2001). Although the good epidemiological congruence and high bacterial strain discriminatory capability of PFGE are well documented by the success of the PulseNet network, the technique has several drawbacks. PFGE is a time-consuming and laborious method requiring a high level of technical skill and rigorous standardization to allow inter-laboratory data sharing. Additionally, in some cases PFGE does not allow optimal discrimination among closely related bacterial isolates (Hyytiä-Trees et al., 2006). To overcome these

* Corresponding author.
E-mail address: li.ma@okstate.edu (L.M. Ma).

http://dx.doi.org/10.1016/j.mimet.2016.04.005
0167-7012/

limitations, PulseNet has begun to augment PFGE data of outbreak-related bacterial isolates with DNA sequence- and PCR-based methods.

Multiple-locus variable-number tandem repeat analysis (MLVA) is a molecular subtyping method based on detection of differing numbers of tandem repeats within several distinct variable-number tandem repeat (VNTR) loci throughout a bacterial genome (Keim et al., 2000). Following PCR amplification of VNTR loci, the amplified DNA fragments are sized or sequenced and compared among different strains. The tandem repeat copy number of each VNTR locus can be designated as a discrete allele type denoted by an integer corresponding to the number of tandem repeats at a given locus, with the string of allele types for several VNTR loci constituting a MLVA type, allowing data comparison among multiple laboratories over extended periods of time (Hyytiä-Trees et al., 2006). MLVA is currently used by PulseNet to help discriminate among highly clonal isolates of *Salmonella* Typhimurium DT104 (Lindstedt et al., 2003; Lindstedt et al., 2004), *Salmonella* Enteritidis (Cho et al., 2007; Boxrud et al., 2007), and O157 STEC (Hyytiä-Trees et al., 2010).

The current O157 STEC MLVA protocol used by PulseNet (Hyytiä-Trees et al., 2010), an optimized and modified 8-locus version of the MLVA method developed by Keys et al. (2005), has proven to be useful in outbreak investigations, allowing a high level of discrimination in conjunction with PFGE. However, this protocol was developed specifically for O157 STEC and PCR amplification of many of the VNTR loci is not possible in non-O157 STEC serogroups (Izumiya et al., 2010; Lindstedt et al., 2007). Given the increasing isolation rates of non-O157 STEC, a MLVA method optimized for these pathogens is needed. However, most MLVA methods target a single serogroup or serotype and development of a MLVA method targeting multiple serogroups poses notable challenges (Karama and Gyles, 2010). The discriminatory power at the serotype level is likely to be decreased if multiple serogroups are targeted in a single protocol since loci conserved enough to be present in multiple serotypes might not provide the necessary level of discrimination. In addition, the most diverse loci and slight differences in VNTR locus flanking sequences among several serogroups can make optimal PCR primer design difficult. As a result, maximum strain discrimination may necessitate individual MLVA protocols for each serogroup. However, a single MLVA protocol for multiple serogroups would be more practical in public health laboratories and the difficulties associated with developing such a protocol can be overcome.

Two notable MLVA schemes for multiple *E. coli* serogroups have been recently developed and used to subtype non-O157 STEC (Løbersli et al., 2012; Izumiya et al., 2010). The MLVA scheme by Løbersli et al. (2012) was originally designed to discriminate among all *E. coli* serogroups (not just STEC), validated by typing the *E. coli* reference (ECOR) collection (Lindstedt et al., 2007), and subsequently optimized by discarding the least informative loci and adding two VNTR loci and one CRISPR (clustered regularly interspaced short palindromic repeat) locus (Løbersli et al., 2012). The MLVA scheme by Izumiya et al. (2010) was designed to target STEC serogroups O157, O111, and O26, essentially by adding nine VNTR loci to the O157-specific MLVA protocol developed by Hyytiä-Trees et al. (2006). Although both of these MLVA schemes have been found to be useful in outbreak investigations, when targeting the 'big 6' non-O157 STEC serogroups, the scheme by Izumiya et al. (2010) may be too narrow while the scheme developed by Løbersli et al. (2012) may be too broad. By searching for diverse VNTR loci present in the seven currently available and fully-assembled 'big-6' non-O157 STEC genomes in GeneBank, it may be possible to develop a novel MLVA scheme that allows increased discrimination for the 'big 6' non-O157 STEC. Of the above mentioned *E. coli* MLVA schemes, only Izumiya et al. (2010) used assembled non-O157 STEC genomes (O26 and O111) in addition to four O157:H7 STEC genomes for identifying potentially discriminatory VNTR loci. Thus, the objective of this study was to develop a robust and highly discriminatory MLVA scheme primarily for the six major non-O157 STEC serogroups—O26, O111, O103, O121, O45, and O145—by independently identifying diverse and informative VNTR loci from seven assembled non-O157 STEC genomes (O26(1), O111(1), O103(1), and O145(4)). The concordance of the MLVA data with PFGE data is presented and the MLVA assay was also used to type O157 STEC, generic *E. coli*, and enteropathogenic *E. coli* for comparison.

2. Materials and methods

2.1. Bacterial strains

A total of 92 *E. coli* strains were used in this study. Initial assay development and optimization was done with 24 non-O157 STEC strains

Table 1
Twenty-four human isolates of the non-O157 STEC reference set.[a]

O	H	Isolate ID	Isolation location	Isolation date	Clinical manifestation	MLVA pattern[b]
26	11	DEC10B	Australia	1986	Diarrhea (bloody)	046
26	11	97-3250	USA (Idaho)	1997	HUS (expired)	047
26		MT#10	USA (Mont.)	1999–2000		048
26	N	TB352A	USA (Wash.)	1991	Diarrhea (chronic)	049
45	2	M103-19	USA (Mich.)	2003		050
45	2	MI01-88	USA (Mich.)	2001		027
45	2	MI05-14	USA (Mich.)	2006		025
45	NM	DA-21	USA (Fla.)	1999	Diarrhea (bloody)	027
103	2	MT#80	USA (Mont.)	1999–2000		051
103	6	TB154A	USA (Wash.)	1991	Diarrhea	052
103	25	8419	USA (Idaho)			053
103	N	PT91-24	USA (Wash.)	1990		054
111	2	RD8	France	1992	HUS (outbreak)	055
111	8	3215-99	USA (TX)	1999	HC (outbreak)	056
111	11	0201 9611	USA (Conn.)	2003		057
111	NM	3007-85	USA (Neb.)	1985		058
121	19	MDCH-4	USA (Mich.)	2000		059
121	19	MT#2	USA (Mont.)	1998		060
121		MT#18	USA (Mont.)	1999–2000		061
121	[19]	DA-5	USA (Mass.)	1998	Diarrhea (bloody)	062
145	16	DEC10I	Canada	1987	HC (HUS)	063
145	[28]	4865/96	Germany	1996	HUS	064
145	NM	GS G5578620	USA (Neb.)	1998	Diarrhea	064
145	NT	IH 16	Uruguay			065

[a] MLVA pattern designations were determined in this study.
[b] Information provided by the STEC Center of Michigan State University.