2016

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

Chris Timmons
Oklahoma State University

Eija Trees
Centers for Disease Control and Prevention

Efrain M. Ribot
Centers for Disease Control and Prevention

Peter Gerner-Smidt
Centers for Disease Control and Prevention

Patti LaFon
Centers for Disease Control and Prevention

See next page for additional authors

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.

Table 2
Sixty non-O157 STEC isolates from CDC.

Serogroup	Isolate ID	State	Serotype	Epidemiological information	*Xba*I pattern[a]	*Bln*I pattern	MLVA pattern
O145	2010C-3517	MI	O145:NM	Cluster 1004MIENM-1	ENMX01.0025	ENMA26.0018	001
	2010C-3515	MI	O145:NM	Cluster 1004MIENM-1	ENMX01.0016	ENMA26.0017	001
	2010C-3507	OH	O145:NM	Cluster 1004MIENM-1	ENMX01.0016	ENMA26.0017	001
	2010C-3508	OH	O145:NM	Cluster 1004MIENM-1	ENMX01.0016	ENMA26.0017	002
	2010C-3513	MI	O145:NM	Cluster 1004MIENM-1	ENMX01.0016	ENMA26.0017	001
	2010C-3526	MI	O145:NM	Cluster 1004MIENM-1	ENMX01.0043	ENMA26.0018	001
	K6208	ND	O145:NM	Sporadic isolate	ENMX01.		003
	2011EL-1210	FL	O145:NM	Sporadic isolate	ENMX01.0112	ENMA26.0085	004
	3060-04	UT	O145:NM	Sporadic isolate	ENMX01.0082		005
	K2387	MD	O145:NM	Sporadic isolate	ENMX01.0040		006
O111	K6807	OK	O111:NM	Cluster 0808OKEXD-1	EXDX01.0005	EXDA26.0029	007
	K6808	OK	O111:NM	Cluster 0808OKEXD-1	EXDX01.0005	EXDA26.0029	008
	K6809	OK	O111:NM	Cluster 0808OKEXD-1	EXDX01.0005	EXDA26.0029	007
	K7091	OK	O111:H8	Cluster 0808OKEXD-1	EXDX01.0005	EXDA26.0029	007
	K5652	IN	O111:NM	Sporadic isolate	EXDX01.		009
	2009EL1340	FL	O111:NM	Sporadic isolate	EXDX01.		010
	2010EL-1239	CO	O111:NM	Cluster 1005COEXD-1	EXDX01.0123	EXDA26.0077	011
	2010EL-1240	CO	O111:NM	Cluster 1005COEXD-1	EXDX01.0130	EXDA26.0077	011
	2010EL-2219	FL	O111:H8	Sporadic isolate	EXDX01.		012
	2010EL-2231	FL	O111:H8	Sporadic isolate	EXDX01.		013
O26	2009EL-1049	OK	O26:H11	Sporadic isolate	EVCX01.0260		014
	2011EL-1012	IN	O26:H9	Sporadic isolate	EVCX01.0103		015
	2011EL-1138	AK	O26:H11	Sporadic isolate	EVCX01.0383		016
	2010EL-1372	WA	O26:NM	Daycare outbreak	EVCX01.0264		017
	2009EL-1480	FL	O26:H11	Sporadic isolate	EVCX01.		018
	2011EL-1233	NV	O26:H11	Sporadic isolate	EVCX01.0071	EVCA26.0236	019
	2010EL-2220	FL	O26:H11	Sporadic isolate	EVCX01.0930		020
	K3621	CO	O26:H11	Sporadic isolate	EVCX01.		021
	K3651	NC	O26:H11	Sporadic isolate	EVCX01.		022
	K5537	MO	O26:H11	Sporadic isolate	EVCX01.		023
O45	05-3031	UT	O45:H2	Sporadic isolate	EH2X01.0003		024
	03-3300	MO	O45:H2	Sporadic isolate	EH2X01.		024
	K3472	NC	O45:H2	Sporadic isolate	EH2X01.0031	EH2A26.0023	025
	K3523	FL	O45:H2	Sporadic isolate	EH2X01.		026
	3506-04	MI	O45:H2	Sporadic isolate	EH2X01.0031	EH2A26.0023	027
	3001-04	MO	O45:H2	Sporadic isolate	EH2X01.0008		027
	3065-04	WI	O45:H2	Sporadic isolate	EH2X01.		028
	3093-04	MA	O45:H2	Sporadic isolate	EH2X01.0021		027
	3095-04	MA	O45:H2	Sporadic isolate	EH2X01.		025
	3105-04	MI	O45:H2	Sporadic isolate	EH2X01.0066		026
O103	2009EL1342	FL	O103:NM	Sporadic isolate	EXWX01.0537		029
	2009EL1295	IN	O103:H2	Sporadic isolate	EXWX01.0540		030
	3546-05	VA	O103:H25	Sporadic isolate	EXWX01.0146		031
	3409-05	VA	O103:H25	Sporadic isolate	EXWX01.0145		032
	K3530	NE	O103:H2	Goat associated	EXWX01.		033
	K3529	NE	O103:H2	Goat associated	EXWX01.		034
	K3435	MO	O103:H2	Sporadic isolate	EXWX01.		035
	2010C-3251	IA	O103:H2	Sporadic isolate	EXWX01.0128	EXWA26.0034	036
	2010C-3219	IA	O103:H2	Sporadic isolate	EXWX01.0128	EXWA26.0034	036
	2009EL-1899	FL	O103:H2	Sporadic isolate	EXWX01.0073	EXWA26.0048	037
O121	K5363	CT	O121:H19	Sporadic isolate	EXKX01.		038
	K5316	CO	O121:H19	Sporadic isolate	EXKX01.0001	EXKA26.0001	039
	K5313	CO	O121:H19	Cluster 0707COEXK-1	EXKX01.0001	EXKA26.0001	039
	K5223	CO	O121:H19	Cluster 0707COEXK-1	EXKX01.0011	EXKA26.0001	039
	K3673	FL	O121:H19	Sporadic isolate	EXKX01.		040
	K3663	CO	O121:H19	Sporadic isolate	EXKX01.0074		041
	K2126	VT	O121:H19	Sporadic isolate	EXKX01.0041		042
	3294-06	WY	O121:H19	Sporadic isolate	EXKX01.0011	EXKA26.0001	043
	3326-06	NY	O121:H19	Sporadic isolate	EXKX01.		044
	K2225	FL	O121:H19	Sporadic isolate	EXKX01.0044		045

[a] New unique patterns were not named in the PFGE database which explains the incomplete pattern names.

obtained from the STEC Center at Michigan State University (MSU) as part of a non-O157 STEC reference set. This set includes four individual strains of each of the six major non-O157 STEC serogroups (O26, O103, O111, O121, O145, and O45) isolated from humans in Australia, Canada, France, Germany, Uruguay, and the United States over a span of 20 years (Table 1). Preliminary validation was carried out with 60 non-O157 STEC isolates obtained from the Enteric Disease Laboratory Branch at the CDC (ten strains from each of the six non-O157 serogroups; Fig. 2). Fifty-eight out of 60 strains were clinical isolates associated with either outbreaks or sporadic cases (Table 2); two were of animal origin. Epidemiological information and PFGE data for all 60 isolates was provided by the CDC (Fig. 2). In addition to the 84 non-O157 STEC isolates, five isolates of STEC O157:H7, two isolates of enteropathogenic *E. coli*, and one strain of *E. coli* K-12 were also analyzed for comparison (Table 3).

2.2. VNTR locus selection

To identify potentially useful VNTR loci for inter- and intra-serogroup discrimination of non-O157 STEC, the published genomes of *E. coli* O26:H11 strain 11368 (NC_013361.1), *E. coli* O103:H2 strain 12009 (NC_013353.1), *E. coli* O111:H-strain 11128 (NC_013364.1),

Table 3
E. coli isolates used for comparison.

E. coli group	Strain	Outbreak source	Isolation year/location
STEC O157:H7	K3995	Spinach outbreak isolate	2006/California
	C7927	Apple cider outbreak isolate	1991/Massachusetts
	F4546	Alfalfa sprout outbreak isolate	1997/Michigan
	EO144	Meat isolate	
	SEA-13B88	Apple juice outbreak isolate	
EPEC	O119:H6		
	O55:H6		
Non-pathogenic	K-12		

E. coli O145:H28 strain RM12581 (CP007136.1), *E. coli* O145:H28 strain RM13514 (CP006027.1), *E. coli* O145:H28 strain RM13516 (CP006262.1), and *E. coli* O145:H28 strain RM12761 (CP007133.1) were scanned for tandem repeats using the Tandem Repeats Finder software (Benson, 1999). Custom parameters were chosen for Tandem Repeats Finder to narrow the number of reported tandem repeat arrays to those comprised of between 4 and 20 bp repeats, with larger tandem repeat copy numbers, and minimal mismatching and indels within the tandem repeat array (Nadon et al., 2013). Once candidate VNTR loci were identified, the flanking sequences of the repeat arrays were searched against NCBI's whole genome shotgun contigs (wgs) database with BLAST since several other non-O157 STEC genomes (in addition to the seven listed above) have been sequenced but not fully assembled.

In accordance with Nadon et al. (2013), selection of a VNTR locus was based on several criteria: a locus had to be present in at least two of the three assembled genomes, had to have a high number of tandem repeat percent matches (>80%), and had to have a low percentage of indels (<3%). These criteria ensured selection of conserved but diverse VNTR loci with common tandem repeat consensus sequences. Following initial selection of possible loci, the flanking sequences of each of the VNTR loci were aligned with ClustalW (Larkin et al., 2007). Only VNTR loci having highly similar flanking sequences were selected to allow optimal primer design and minimize the need for degenerate primers. Additionally, VNTR loci exhibiting differences in tandem repeat copy numbers among the three strains were preferentially selected. The more diverse but often less conserved loci (larger difference in copy number) were selected to help discriminate closely related strains within individual serogroups while the less diverse and more conserved loci (smaller difference in copy number) were selected to help discriminate among different serogroups (Keys et al., 2005). The final selection included ten VNTR loci, seven of which have been previously described but were renamed for the sake of uniformity and due to new PCR primer design (Table 4).

The presence and diversity of the selected loci in STEC O157:H7 strains were evaluated also by comparing each selected VNTR locus with the Tandem Repeats Finder results of the published genomes of

Table 4
Characteristics of the ten VNTR loci used in this study.

Locus name	Alternative[a] name	Array location (5′ end)	Repeat length (nt)	Consensus sequence	Primers (5′-3′)	Primer Tm (°C)	Offset size (nt)	Primer conc. (μM)	Function
SVL-1	O157-2, EHC-2, CVN016	250070 in O111	6	CTCTGA	F: **6FAM**-ACTGTTTC AGCGGTCTCTTCC	60.81	97	0.05	Putative ATP-dependent Clp proteinase ATP-binding chain
					R: ACG CAG ATA CCG TGG AG	61.65			
SVL-2	O157-9, Vhec4, TR1, CVN017	2913106 in O111	6	AGAAAT	F: **PET**-ATCGCCTTCT TCCTCCGTAA	61.08	244	0.05	Hypothetical protein
					R: TCAGGAATGTGG TGGTCTGT	58.94			
SVL-3	O157-11, EHC-1, CVN014	4662685 in O111	6	GGTGCA	**F: VIC**-TGGCAAACAG CACTACCATC	59.72	248	0.04	Predicted protoheme IX synthesis protein HemY
					R: GGACCAGTTAAG CCAGCAAA	60.25			
SVL-4	CVN004	810131 in O103	15	GCAGCAAA AGCCGCA	F: **PET**-GGAAGAAGCA GCGAAGAAAG	59.34	270	0.06	Membrane anchored protein TolA in TolA-TolQ-TolR complex
					R: CATCGGGTGCCAGT TTTATG	61.27			
SVL-5		3051096 in O103	6	GCGCTG	F: **VIC**-GTCGTCTGTG GGATGCTCAA	62.27	159	0.05	Hydrogenase 4, Membrane subunit HyfF
					R: CAGCAATAACAG CAGGACGA	60.01			
SVL-6		2922513 in O103	9	CAGTGC AGC	F: **6FAM**-AATTAGGA AAAGCATCAGCCG	60.57	242	0.07	Putative adenine methylase, Putative integrase, *stx*2 converting phage
					R: CCTCCCATCGTTTC TGTTTCC	62.98			
SVL-10	EH111-14	3346927 in O111	7	TCAAAGA	F: **VIC**-TTTGATGCAA TGGTGGAGTG	60.52	166	0.05	Putative integrase
					R: CACAAAGTGAGA GTCCGAAAA	57.99			
SVL-11	O157-37	35162 in O111 plasmid 3	6	CTGCTA	F: **NED**-ATTCTGCTGT GGGCTTCTGT	59.87	90	0.05	Plasmid located, no known function
					R: AATCAGAGCGGC AGGAAAA	60.87			
SVL-12	EHC-6	52289 in O26 plasmid 2	9	AACAGC CGC	F: **NED**-CCGCAAGGGA AGCAGAAG	62.02	197	0.04	Plasmid located, no known function
					R: TGCTGTTCCATCTC TTCTTCC	59.42			
SVL-23		63087 in O121:H19 str. MT#2 EC1660_contig_31	6	TCTCCC	F: **PET**-AAATCGGGCG GGAAGAAG	62.38	361	0.05	Dihydrodipicolinate synthase DapA
					R: GGGCGTAAAAAG CAATAAAGG	59.98			

[a] O157-x loci are from Keys et al. (2005); EHC-x and EH111-x loci are from Izumiya et al. (2010); CVN0xx loci are from Løbersli et al. (2012); Vhecx loci are from Lindstedt et al. (2003); TR loci are from Noller et al. (2003).

four STEC O157:H7 strains (EDL933 (NC_002655.2), Sakai (NC_002695.1), EC4115 (NC_011353.1), and TW14359 (NC_013008.1)). All loci except SVL-10 and SVL-12 were present also in STEC O157:H7 but with less flanking sequence similarity.

2.3. DNA preparation

Bacterial strains were grown overnight at 37 °C on trypticase soy agar (TSA). Two to three colonies were suspended in 100 μL of sterile distilled water and boiled for 10 min at 100 °C. The suspension was cooled briefly and centrifuged at 10,000 rpm (8165 × *g*) for 10 min. The undiluted supernatant was used as template DNA for PCR amplification and stored at −20 °C.

2.4. Primer design and PCR amplification

PCR primers for amplification of selected VNTR loci were designed from highly similar VNTR flanking sequences identified by multiple sequence alignment with ClustalW using Primer3 software (Untergrasser et al., 2012), followed by an evaluation of primer thermodynamics using the Mfold web server (Zuker, 2003), then by a BLAST search against the NCBI nucleotide (nr/nt) database for primer specificity analysis. PCR primers were designed to minimize multiplex reactions and to allow all multiplex PCRs to occur at the same thermal cycling conditions. Therefore, all primers were designed with minimal 3′ self-complementary sequences and with similar lengths, GC contents, and melting temperatures. Primers amplifying previously identified loci were redesigned to have characteristics similar to those of all other primers in this study. Additionally, MultiPLX 2.1 (Kaplinski et al., 2005) was used to evaluate the potential for primer dimer formation among all ten primer sets. Since the specific size range of the amplified fragments for each VNTR locus was unknown, all primers were designed to allow multiplexing of any combination of primer sets (i.e. minimal potential for primer dimer formation).

Initial screening of the amplification effectiveness of the ten primer sets was carried out with the 24-isolate non-O157 STEC reference set from the STEC Center at MSU and visualized by agarose gel electrophoresis. Based on the amplicon sizes, the primer sets were combined into three multiplex PCR reactions. Reaction 1 contained primer sets SVL-1, SVL-3, and SVL-4, reaction 2 contained primer sets SVL-2, SVL-6, SVL-10, and SVL-12, and reaction 3 contained primer sets SVL-5, SVL-11, and SVL-23.

Forward PCR primers were fluorescently labeled to allow accurate sizing by multicolor capillary electrophoresis (Table 4). Unlabeled reverse primers were synthesized by Integrated DNA Technologies (Coralville, IA) and fluorescently labeled forward primers were synthesized by Life Technologies (Foster City, CA). The PCR amplification conditions were designed to mimic, as closely as possible, the PCR reaction conditions and reagent concentrations currently used for MLVA by PulseNet (Hyytiä-Trees et al., 2010). PCR amplification was performed in final volumes of 10 μL consisting of 1.5 μL of 5× Colorless GoTaq Reaction Buffer (Promega, Madison, WI), 0.4 μL of 50 mM $MgCl_2$ (bringing final $MgCl_2$ concentration to 2.0 mM), 1.0 U of GoTaq DNA Polymerase (Promega), 0.2 mM of PCR Nucleotide Mix (Promega), and 1.0 μL of DNA template. Primer concentrations were adjusted to allow optimal peak heights for confident fragment size calling. The amplification conditions consisted of an initial denaturation step at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 30 s, with a final extension step at 72 °C for 15 min with an Eppendorf MasterCycler (ThermoFisher, Waltham, MA).

2.5. Fragment analysis

Amplified PCR products were diluted 1:60 in sterile distilled water. A 1.0 μL aliquot of the diluted PCR product was added to 8.6 μL of Hi-Di Formamide (Life Technologies) and 0.4 μL of GeneScan 600LIZ size standard (Life Technologies). PCR products were sized using an Applied Biosystems 3730 Genetic Analyzer (Life Technologies).

2.6. Pulsed-field gel electrophoresis

PFGE was performed for all 84 non-O157 STEC isolates according to the standardized PulseNet protocol (Ribot et al., 2006). All isolates were analyzed using *Xba*I restriction enzyme (Roche Applied Science, Indianapolis, IN). Twenty-three isolates from the CDC were also analyzed using *Bln*I restriction enzyme (Roche Applied Science) (Table 2). PFGE patterns were analyzed with BioNumerics software version 5.01 (Applied Maths, Kortrijk, Belgium), uploaded to the PulseNet PFGE pattern database, and named according to the standard nomenclature system (Swaminathan et al., 2001).

2.7. Analysis of VNTR data

Fragment data were evaluated with GeneMapper software (Life Technologies) and fragment peak tables from GeneMapper were imported into BioNumerics (Applied Maths) for analysis. A custom VNTR allele assignment script in BioNumerics was used to translate fragment size data to copy numbers. Partial repeats were rounded up or down to the closest complete tandem repeat number in accordance with the scheme developed by Hyytiä-Trees et al. (2010). For each locus, alleles were named according to the number of tandem repeats, whereas null alleles, defined as no PCR amplification at a given locus, were designated as −2.0 to differentiate between null alleles and VNTR loci with no tandem repeats (i.e. a copy number of "0"). Null alleles were confirmed by singleplex PCR visualized by agarose gel electrophoresis to rule out the lack of amplification due to multiplex PCR complications. The diversity index (D_I) for each locus was calculated in BioNumerics based on Simpson's diversity index according to the formula $D_I = 1 - \Sigma\,(\text{allelic frequency})^2$ (Hunter and Gaston, 1988; Weir, 1990). Dendrograms were constructed with BioNumerics using a categorical multi-state coefficient and UPGMA (unweighted pair group method with arithmetic mean) clustering. Minimum spanning trees were constructed with BioNumerics using the Manhattan coefficient. Outbreak related isolates with indistinguishable PFGE patterns using two restriction enzymes were used to evaluate the epidemiological concordance of the MLVA scheme in comparison to PFGE.

3. Results

3.1. Selection of VNTR loci

A comparison of reported short tandem repeat structures for four STEC O157:H7 strains, two non-pathogenic *E. coli* strains, and seven non-O157 STEC strains revealed more VNTR diversity among STEC O157 than among non-O157 STEC and generic *E. coli*. While the total number of reported tandem repeats were similar between STEC O157:H7 strains and non-O157 STEC strains, about twice as many tandem repeat arrays with high copy numbers were identified in STEC O157:H7 strains than in non-O157 STEC strains (Table 5). The number of tandem repeats having higher copy numbers among the non-O157 STEC strains was more similar to those found in two strains of generic *E. coli* K-12, which have an approximately 800 Kb smaller genome. The ten selected VNTR loci exhibited differing levels of diversity among the genomic sequences of the seven fully assembled non-O157 STEC genomes in GenBank, as well as among the NCBI *E. coli* whole genome shotgun contigs (wgs) database.

Since the majority of bacterial genomes code for proteins, it was expected that most VNTR arrays would be located within genes. Of the ten selected VNTR loci evaluated, eight are located on the bacterial chromosome and two on plasmids. According to BLAST searches against the NCBI nucleotide database, all chromosomal VNTR loci are located within

Table 5
Comparison of genome size, number of reported tandem repeat arrays, and number of tandem repeat arrays with copy numbers greater than 5.0, according to Tandem Repeats Finder software, for four STEC O157:H7, three non-O157 STEC, and two *E. coli* K-12 strains.

E. coli strain	GenBank accession number	Genome size (bp)	Total number of tandem repeats[a]	Number of tandem repeats with copy number ≥ 5.0
O157:H7 EC4115	NC_011353.1	5572075	177	19
O157:H7 TW14359	NC_013008.1	5528136	174	20
O157:H7 EDL933	NZ_CP008957.1	5528445	167	17
O157:H7 Sakai	NC_002695.1	5498450	159	15
O111:H-11128	NC_013364.1	5371077	126	7
O26:H11 11368	NC_013361.1	5697240	129	9
O103:H2 12009	NC_013353.1	5449314	123	6
O145:H28 RM12581	NZ_CP007136.1	5585611	148	6
O145:H28 RM13514	NZ_CP006027.1	5585613	148	6
O145:H28 RM13516	NZ_CP006262.1	5402276	135	9
O145:H28 RM12761	NZ_CP007133.1	5402281	135	9
K-12 DH10B	NC_010473.1	4686137	87	9
K-12 W3110	NC_007779.1	4646332	89	9

[a] As reported by Tandem Repeats Finder software with default parameter settings.

sequences coding for known or putative proteins but the plasmid located VNTR loci had no known functions (Table 4).

3.2. Evaluation of selected VNTR loci

All VNTR loci were polymorphic, ranging from 4 to 22 alleles per locus (Table 6) and no isolates of different serogroups shared an indistinguishable MLVA type. A high number of null alleles were observed for several serogroups, especially among serogroups O45 and O121. Although not ideal, null alleles were still useful for discrimination with several loci (Tables 6 and 7). A low to moderate diversity index was observed for the ten selected loci and was similar for each of the loci when comparing the two sets of isolates from CDC and MSU (Table 6). Only SVL-3 had a relatively high overall diversity index of 0.895. Loci SVL-11 and SVL-23 had very low diversity indices but were retained since they aided in discrimination between serogroups O111 and O121, respectively. Locus SVL-11 was highly polymorphic only within serogroup O111, as was expected since this locus is located on a plasmid and may be fairly specific for serogroup O111. SVL-1 was the most polymorphic locus with 22 different alleles, but had only a moderate overall diversity index (0.791) due to a lack of diversity in serogroups O103, O45, and O145 (Table 6). Loci SVL-2, SVL-3, SVL-6 and SVL-11 also exhibited high levels of polymorphism with 9, 15, 12, and 10 alleles, respectively (Table 6).

Five STEC O157:H7 strains (C7927, EO144, F4546, K3995, and SEA-13B88), two EPEC strains (O119 and O55), and one strain of generic *E. coli* K-12 were MLVA typed with the selected loci and compared to the MLVA types of the 84 non-O157 STEC isolates. Although all eight strains had a unique MLVA type, PCR amplification was not possible at most loci. Dendrograms generated by BioNumerics separated the STEC O157:H7 isolates from all others when compared with both sets of non-O157 STEC isolates from CDC and MSU (data not shown).

3.3. MLVA typing of 84 non-O157 STEC isolates

A total of 65 unique MLVA types were identified among the 84 non-O157 STEC isolates tested: 45 MLVA types among the 60 isolates from CDC and 22 MLVA types among the 24 isolates from MSU (3 O45 isolates from MSU were indistinguishable by MLVA from 2 separate groups of O45 isolates from CDC). Serogroups generally clustered together in minimum spanning trees (Fig. 1). All serogroups differed from each other by one or more tandem repeats at three or more loci (Fig. 1).

3.3.1. O26

The highest level of discriminatory capability was achieved in serogroup O26. All 14 isolates tested exhibited a unique MLVA type that differed from all other O26 isolates by at least one locus. Thirteen different alleles were observed in locus SVL-1 alone. The high level of serogroup O26 discrimination was achieved with just four loci (Table 7). Omitting all loci except SVL-1, SVL-2, SVL-3, and SVL-6 had no effect on the discriminatory capability. Therefore, a STEC O26-specific MLVA assay may be possible when loci SVL-1, SVL-2, SVL-3, and SVL-6 are targeted. Following further evaluation of the congruence

Table 6
VNTR loci characteristics among 60 clinical non-O157 STEC isolates from the CDC and 24 isolates of a non-O157 STEC reference set from the STEC Center at Michigan State University.

		VNTR locus									
		SVL-1	SVL-2	SVL-3	SVL-4	SVL-5	SVL-6	SVL-10	SVL-11	SVL-12	SVL-23
60 isolates from CDC	Fragment range (nt)	112–254	253–295	276–365	387–452	135–210	255–394	175–278	105–172	210–289	405–422
	No. of alleles	17	8	12	4	6	9	4	9	4	6
	Null alleles (%)	0	47	0	0	0	35	28	75	55	6
	Allelic range	3–26	2–8	5–20	8–12	0–7	1–17	1–16	3–14	1–10	7–10
	Diversity index	0.75	0.729	0.898	0.561	0.532	0.798	0.543	0.433	0.532	0.275
24 isolates from MSU	Fragment range (nt)	106–211	253–301	270–372	405–452	169–186	255–342	175–181	147–213	210–307	416–428
	No. of alleles	13	7	9	4	5	8	3	4	5	4
	Null alleles (%)	13	58	0	25	4	50	38	83	46	4
	Allelic range	2–19	2–9	4–21	9–12	2–5	1–11	1–2	10–20	1–12	9–11
	Diversity index	0.88	0.696	0.873	0.609	0.543	0.746	0.54	0.308	0.638	0.239
84 isolates from CDC and MSU	Fragment range (nt)	106–254	253–295	270–372	387–452	135–210	255–394	175–278	105–213	210–307	405–422
	No. of alleles	22	9	15	5	7	12	4	10	7	7
	Null alleles (%)	4	49	0	6	1	39	31	76	52	6
	Allelic range	2–26	2–9	4–21	8–12	0–7	1–17	1–16	3–20	1–12	7–11
	Diversity index	0.791	0.717	0.895	0.585	0.529	0.791	0.538	0.396	0.558	0.262

Table 7
Comparison of VNTR loci characteristics for six non-O157 STEC serogroups.

Serogroup		VNTR locus									
		SVL-1	SVL-2	SVL-3	SVL-4	SVL-5	SVL-6	SVL-10	SVL-11	SVL-12	SVL-23
O26	Fragment range (nt)	131–254	259–295	288–318	451–452	169–170	256–336	175–176	132–133	288–289	416–418
	No. of alleles	13	6	6	1	1	7	1	2	2	1
	Null alleles (%)	0	0	0	0	0	7	0	93	93	0
	Allelic range	6–26	2–8	7–12	12	2	1–10	1	7	10	9
	Diversity index	0.989	0.769	0.868	0	0	0.802	0	0.143	0.143	0
O111	Fragment range (nt)	131–211	265–301	299–372	405–452	169–170	255–343	175–182	135–160	210–211	416–418
	No. of alleles	7	6	8	3	1	6	3	6	3	1
	Null alleles (%)	7	14	0	14	0	29	7	14	7	0
	Allelic range	6–19	3–9	6–21	9–12	2	1–11	1–2	8–14	1–5	9
	Diversity index	0.846	0.736	0.912	0.385	0	0.835	0.473	0.747	0.275	0
O103	Fragment range (nt)	112–118	264–266	282–318	405–452	135–210	341–394	175–176	105–106	210–211	416–418
	No. of alleles	2	4	6	3	4	3	2	2	2	1
	Null alleles (%)	0	71	0	7	7	79	21	93	36	0
	Allelic range	3–4	2–4	6–12	9–12	0–5	11–17	1	3	1	9
	Diversity index	0.143	0.495	0.868	0.473	0.396	0.385	0.363	0.143	0.495	0
O121	Fragment range (nt)	106–150	0	293–353	418–419	169–170	279–343	0	132–133	254–255	405–422
	No. of alleles	5	1	6	2	1	4	1	2	2	6
	Null alleles (%)	0	100	0	7	0	0	100	93	93	36
	Allelic range	2–9	n/a	5–17	10	2	4–11	n/a	7	6	7–11
	Diversity index	0.725	0	0.813	0.275	0	0.495	0	0.143	0.143	0.813
O45	Fragment range (nt)	112–113	0	288–305	387–452	180–197	0	175–176	0	210–211	416–418
	No. of alleles	1	1	4	2	3	1	1	1	1	1
	Null alleles (%)	0	100	0	0	0	100	0	100	0	0
	Allelic range	3	n/a	6–10	8–12	4–7	n/a	1	n/a	1	9
	Diversity index	0	0	0.648	0.143	0.582	0	0	0	0	0
O145	Fragment range (nt)	112–180	253–277	270–317	451–452	169–170	291–347	175–278	144–213	281–307	416–418
	No. of alleles	4	4	5	2	1	4	3	3	3	1
	Null alleles (%)	14	14	0	14	0	21	57	71	86	0
	Allelic range	3–14	2–5	4–12	12	2	5–12	1–16	9–20	9–12	9
	Diversity index	0.495	0.495	0.505	0.264	0	0.736	0.615	0.473	0.275	0

with epidemiological data and PFGE, these four loci could potentially be combined in a single multiplex PCR reaction for rapid screening of isolates in a STEC O26 outbreak investigation.

3.3.2. O111

Serogroup O111 had a low percentage of null alleles and the highest loci diversity indices, even though little or no diversity was observed in

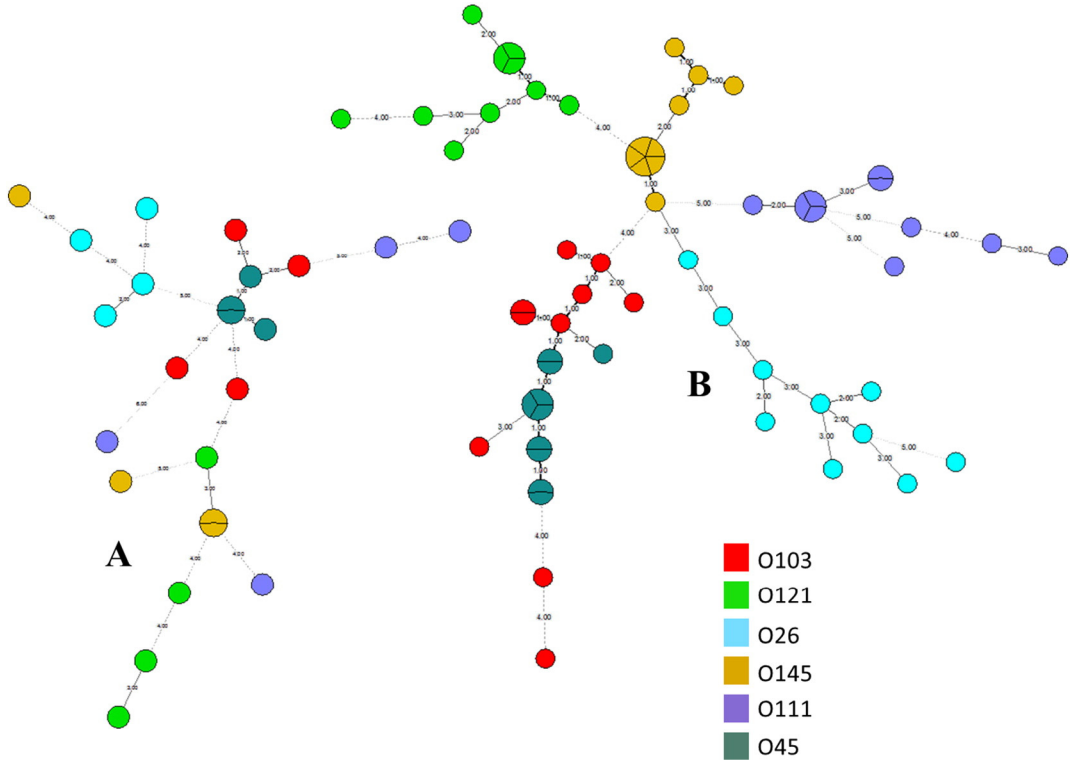

Fig. 1. Minimum spanning trees of (A) 24 non-O157 STEC isolates from the STEC Center at MSU and (B) 60 non-O157 STEC isolates from CDC constructed by BioNumerics using the Manhattan coefficient. Each circle represents a single MLVA type with the size proportional to the number of isolates with that MLVA type. Numbers on branches indicate the number of loci that vary between each MLVA type.

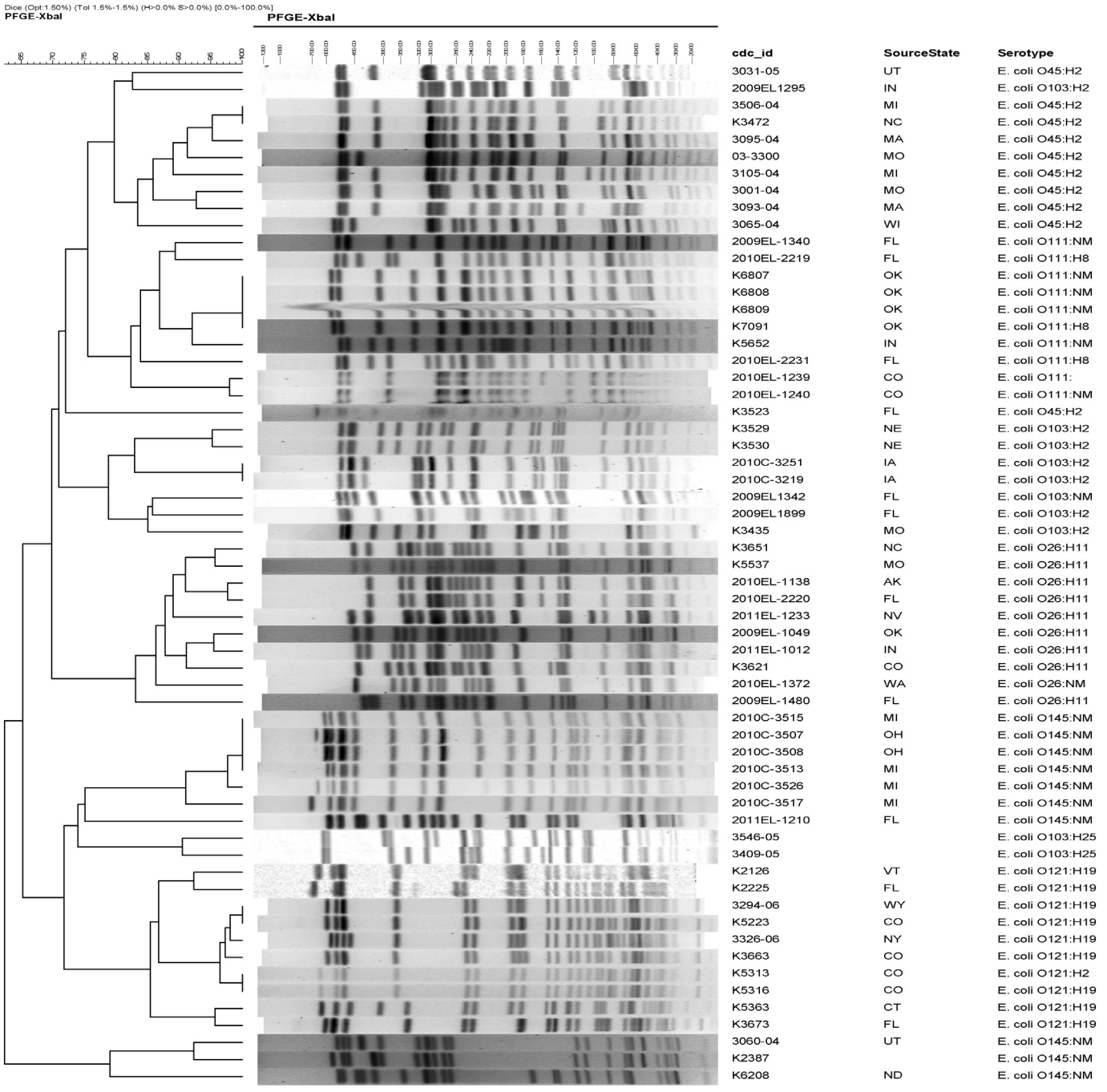

Fig. 2. PFGE dendrogram of 60 clinical non-O157 STEC isolates from the CDC, generated by BioNumerics using categorical coefficient and UPGMA clustering.

five loci (SVL-4, SVL-5, SVL-10, SVL-12, and SVL-23). The remaining five loci had a moderate to high level of diversity, ranging from 0.736 to 0.912 (Table 7). A total of 11 unique O111 MLVA types were observed, with two groups of indistinguishable MLVA types. The three isolates composing one of the groups were also indistinguishable by PFGE using *Bln*I and *Xba*I, while the two isolates composing the other group were distinguishable by PFGE.

3.3.3. O103

Serogroup O103 exhibited low to moderate diversity at most loci. Only locus SVL-3 had a high diversity index of 0.868 and only loci SVL-3, SVL-4, SVL-5, and SVL-12 were required to provide the observed level of discrimination (Table 7). One pair of indistinguishable MLVA types were observed among 13 unique MLVA types for the 14 isolates tested. The two O103 isolates indistinguishable by MLVA, 2010C-3251 and 2010C-3219, were also indistinguishable by PFGE.

3.3.4. O121

Four loci exhibited moderate diversity in serogroup O121. Only SVL-1, SVL-3, SVL-6, and SVL-23 were needed to provide the observed level of discrimination (Table 7). Twelve unique MLVA types were observed among the 14 O121 isolates. One group of 3 indistinguishable isolates by MLVA was observed. Two of the three isolates (K5313 and K5316) were also indistinguishable by PFGE with *Bln*I and *Xba*I.

3.3.5. O45

The lowest level of diversity was observed among serogroup O45. No PCR amplification was possible for loci SVL-2, SVL-6, and SVL-11 and no

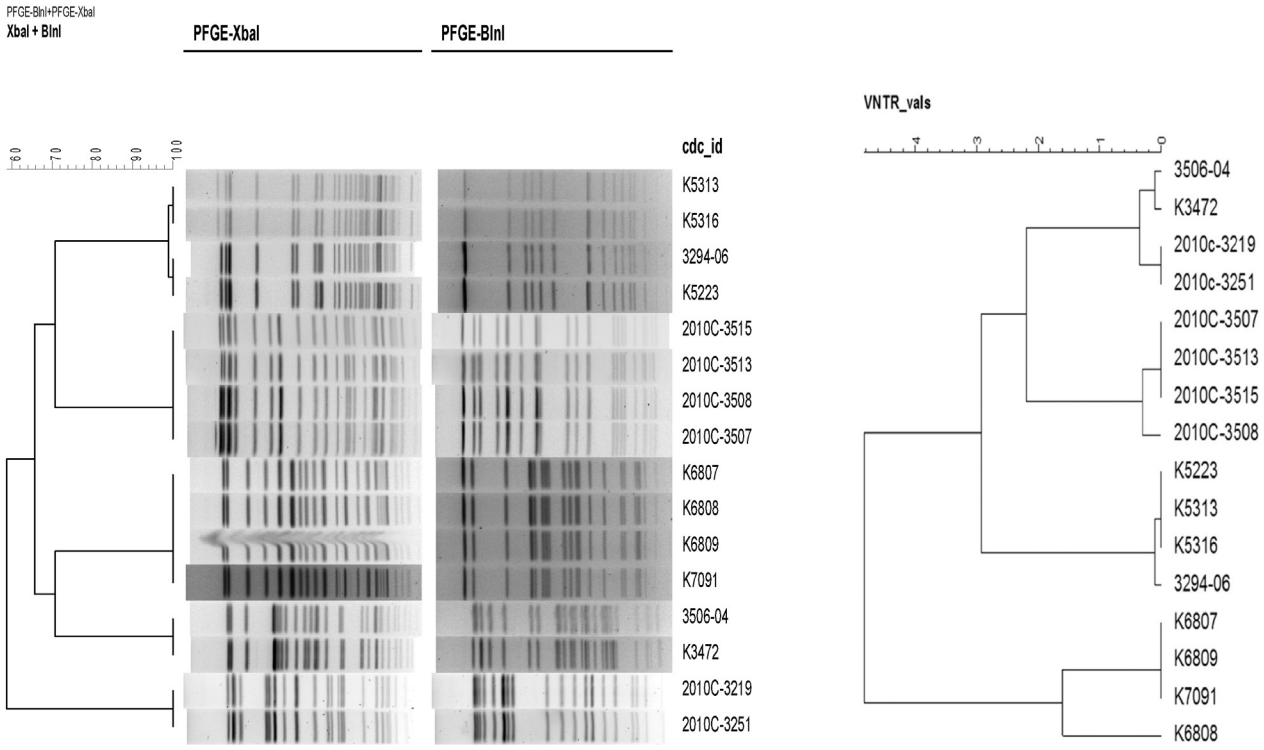

Fig. 3. Comparison of PFGE (left) and MLVA (right) for ten outbreak related and six sporadic non-O157 STEC isolates, comprising 6 groups of indistinguishable PFGE patterns by both *Xba*I and *Bln*I. Only 2 of the 6 groups (isolates 3506-04 and K3472; isolates 2010C-3219 and 2010C-3251) were also indistinguishable by MLVA or clustered similarly.

diversity was observed for loci SVL-1, SVL-10, SVL-12, and SVL-23 (Table 7). Among the 14 isolates tested, only six unique MLVA types were observed with three groups of indistinguishable MLVA types. The isolates constituting these groups were not epidemiologically related. However, two isolates indistinguishable by PFGE (K3472 and 3506-04) were distinguishable by MLVA, differing by one tandem repeat at a single locus (SVL-5).

3.3.6. O145

Low diversity indices were observed also for all ten loci among serogroup O145. Although PCR amplification was possible at all loci, no diversity was observed for 6 loci: SVL-1, SVL-2, SVL-4, SVL-5, SVL-12, and SVL-23 (Table 7). Of 14 isolates tested, nine MLVA types were observed with two groups of indistinguishable MLVA types. The first group, isolates 4865/96 and GS-G5578620, isolated in Germany and Nebraska, respectively, were of different serotypes and had no logical epidemiological connection. The second group of indistinguishable O145 MLVA types consisted of isolates 2010C-3513, 2010C-3515, 2010C-3507, 2010C-3526-1, and 2010C-3517. Three of these five isolates (2010C-3513, 2010C-3515, and 2010C-3507) were also indistinguishable by PFGE.

3.4. Correlation of MLVA data with PFGE and epidemiological data.

3.4.1. 60 CDC isolates

Compared to PFGE, a similar level of discrimination was possible with MLVA. While the total number of PFGE patterns (50) was slightly higher than the number of MLVA types (45), the number of unique bacterial subtypes stayed the same for all serogroups except for O45 and O145. Fifteen of the 58 clinical isolates from the CDC were outbreak related, and three of the four outbreaks with multiple isolates included displayed multiple PFGE patterns (Table 2). MLVA correctly grouped together isolates from two out of four outbreaks, although a sporadic isolate matched the outbreak pattern by both PFGE and MLVA in one of the outbreaks (O121:H19). In the O145:NM outbreak, isolate 2010C-3508 differed from the other four isolates at MLVA loci SVL-11 and SVL-12 (Fig. 3), even though it was a PFGE match to the outbreak. In one of the two O111:NM outbreaks, one isolate was different from the main outbreak MLVA profile even though it was a PFGE match. In the second O111:NM outbreak the two isolates included differed by PFGE but not by MLVA. Only six MLVA profiles were detected among the ten sporadic O45:H2 strains even though there were nine different PFGE patterns. The two isolates matching by PFGE had different MLVA profiles.

3.4.2. 24 MSU isolates

All 24 non-O157 STEC isolates from the STEC Center at MSU had unique PFGE patterns when *Xba*I was used, while 22 different MLVA types were observed. Two O145 isolates (4865/96 and GS G5578620) and two O45 isolates (MI01-88 and DA-21) were indistinguishable by MLVA but had no known epidemiological connection and were of different serotypes.

4. Discussion

Successful identification and traceback of foodborne illness outbreaks caused by bacterial pathogens requires bacterial subtyping techniques that are highly discriminatory, reproducible, portable, objective, versatile, and allow high throughput (Nadon et al., 2013). While MLVA performs very well when assessed by these criteria, the method often has a major weakness: versatility (Nadon et al., 2013). Most published and well validated MLVA protocols are only useful for typing a subset or a group of bacterial pathogens, such as a single serogroup or serotype within a species. The discriminatory power and therefore the epidemiological value of MLVA is usually decreased when a broad and highly diverse collection of strains from a bacterial species are targeted. Thus, the versatility of MLVA is limited by its specificity. The value of MLVA is that it allows evaluation of multiple relatively rapidly changing regions of a bacterial genome, identifying minor differences among highly genetically similar strains. As a result, strains that are more distantly related

are not as efficiently typed and accurate evaluation of epidemiological congruence might not be possible.

A single MLVA assay for multiple serogroups of pathogenic *E. coli* that is comprised of a small enough number of VNTR loci to allow rapid and routine strain typing of clinical and environmental/food isolates while allowing better discrimination than the current gold standard subtyping technique, PFGE, has been attempted previously (Lindstedt et al., 2007; Izumiya et al., 2010; Løbersli et al., 2012). The major limiting factor for further development of such assays may be the lack of availability of closed genomes of clinically relevant *E. coli* strains. Draft and partially assembled bacterial genome sequences do not assemble accurately in repeat regions due to the short read length produced by the predominant DNA sequencing technologies commonly used and therefore do not allow optimal identification of candidate VNTR loci for MLVA assay development. Additionally, MLVA may eventually go by the wayside as whole genome sequencing technologies are becoming less expensive, potentially allowing whole genome comparisons of epidemiologically related isolates. However, MLVA is currently still a valuable and highly discriminatory method that is commonly used to augment PFGE data in foodborne illness outbreak investigations.

Non-O157 STEC serogroups have been isolated with increasing frequency in recent years but no MLVA scheme for any non-O157 STEC serogroups has yet been adopted for use by PulseNet. The purpose of this study was to investigate the possibility of developing a single, highly discriminatory MLVA protocol for the six most commonly isolated non-O157 STEC serogroups in the United States. Using all of the currently available assembled non-O157 STEC genomes and whole genome shotgun sequence contigs for non-O157 STEC strains deposited in NCBI's GenBank database, ten VNTR loci were identified, allowing for inter- and intra-serogroup strain discriminatory capability similar to PFGE. While the number of non-O157 STEC isolates used in this study was small, the relatively high congruence of MLVA, PFGE, and epidemiological data for five of the six serogroups tested illustrates the potential usefulness of the developed scheme, following further optimization.

Strain discrimination by the developed MLVA scheme was relatively high among serogroups O26, O111, O103, and O121, with similar discrimination to PFGE. Less strain discrimination than PFGE was observed for serogroups O45 and O145 even though the epidemiological concordance was better than PFGE for O145:NM. Even in the available closed genome sequences used for VNTR identification, O26, O111, and O103 exhibited more tandem repeat diversity than all four strains of O145. While the whole genome shotgun contigs (wgs) database of NCBI was searched with candidate VNTR flanking sequences identified among the closed genomes, this only aided in optimal PCR primer design and did not aid in identification of diverse VNTR loci (except for SVL-23, which was only diverse in O121).

Among outbreak related isolates, the developed MLVA scheme differentiated among few isolates with indistinguishable PFGE profiles, which will complicate data interpretation. Conversely, several isolates indistinguishable by MLVA were distinguishable by PFGE (Table 2). This observation confirms that the maximum possible strain discrimination often requires the use of more than one bacterial subtyping method. However, for surveillance epidemiological concordance is more desirable instead of maximum strain discrimination. Much like other MLVA protocols currently used by PulseNet, the developed MLVA scheme could potentially be used to augment PFGE data for non-O157 STEC isolates associated with foodborne illness outbreaks.

Since multiple serogroups were targeted in this study, potentially highly diverse VNTR loci were chosen to aid in intra-serogroup discrimination and potentially less diverse VNTR loci were chosen to aid in inter-serogroup discrimination. Several loci exhibited little or no intra-serogroup diversity but had distinct inter-serogroup diversity, helping discriminate between serogroups (Table 7). For example, locus SVL-4 contained 12 tandem repeats in all 14 O26 isolates, nine tandem repeats in 11 of 14 O111 isolates, and ten tandem repeats in 12 of 14 O121 isolates. VNTR loci located on plasmids may also serve as useful serogroup identifiers. Locus SVL-11 was located on an O111 plasmid and was highly diverse among this serogroup. All chromosomally located VNTR loci were contained within DNA sequences coding for known or putative proteins (Table 4). It has been speculated that tandem repeat arrays that are located within genes and having repeat lengths in multiples of three, therefore not altering the open reading frame, are likely to be more diverse than those located outside of gene sequences (Keys et al., 2005). One of the selected VNTR loci (SVL-10) contained a tandem repeat that was not a multiple of three. As expected, this locus exhibited low overall diversity and only aided in the discrimination of one serogroup (O145).

The genomic location of locus SVL-6 was of special interest. Based on a BLAST search against the NCBI database, SVL-6 was located within a gene sharing high similarity to a *stx*2 converting phage (Smith et al., 2012). Stx2 is one of the major virulence factors of STEC and is frequently associated with the development of HUS (Nataro and Kaper, 1998). It is believed that the *stx*2 gene can be acquired by *E. coli* following contact with *stx*2 converting phages and subsequent incorporation of the sequence into previously non-pathogenic or less pathogenic *E. coli* genomes (Scheutz et al., 2011; Grande et al., 2014). As expected, SVL-6-specific PCR primers allowed amplification among serogroups O157, O26, O111, O103, O121, and O145—the serogroups most commonly associated with Shiga toxin production—but not in 2 EPEC strains or in *E. coli* K-12. However, the lack of amplification among serogroup O45 could not be explained but could be due to nucleotide polymorphisms in the primer annealing location.

The developed prototype non-O157 STEC MLVA scheme is simple and rapid with easy-to-interpret and portable results. Among the six non-O157 STEC serogroups tested, the characteristics of the ten selected VNTR loci varied considerably and it may be possible to tailor the developed MLVA scheme for each serogroup by retaining the most diverse loci and discarding the least diverse. However, when typing all six serogroups simultaneously, discarding any of the ten loci decreased the inter-serogroup discriminatory capability. Unless more closed genome sequences are available for comparison, a higher overall level of discrimination might not be possible. Before the developed prototype MLVA scheme could be used to evaluate epidemiologically related isolates, further extensive validation of the proposed method with a large panel of outbreak related and sporadic isolates is necessary. The resultant data should be compared to PFGE for all isolates to gain a more complete understanding of the usefulness of this method for intra- and inter-serogroup discrimination of epidemiologically related and non-related non-O157 STEC isolates. Additionally, in order to deploy the assay in multiple laboratories with different capillary electrophoresis platforms, a set of isolates with all ten VNTRs sequenced will need to be defined so that the fragment sizing data can be normalized to the actual sequenced copy number.

Acknowledgments

Funding for this project was provided by the United States Department of Agriculture National Needs Fellowship grant 2010-38420-20423 and the European Union Seventh Framework Program: “Plant and Food Biosecurity,” grant agreement No. 261752. We would like to thank the STEC Center of Michigan State University and the Centers for Disease Control and Prevention (CDC) for providing non-O157 STEC isolates with accompanying information. Additionally, we would like to thank the National Institute of Microbial Forensics and Food and Agricultural Biosecurity (NIMFFAB) and the Biochemistry and Molecular Biology Core Facility at Oklahoma State University.

References

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27 (2), 573–580.

Besser, R.E., Griffin, P.M., Slutsker, L., 1999. *Escherichia coli* O157:H7 gastroenteritis and the hemolytic uremic syndrome: an emerging infectious disease. Annu. Rev. Med. 50, 355–367.

Bettelheim, K.A., 2007. The non-o157 shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. Crit. Rev. Microbiol. 33 (1), 67–87.

Boxrud, D., Pederson-Gulrud, K., Wotton, J., Medus, C., Lyszkowicz, E., Besser, J., Bartkus, J.M., 2007. Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-field gel electrophoresis, and phage typing for subtype analysis of *Salmonella enterica* serotype Enteritidis. J. Clin. Microbiol. 45 (2), 536–543.

Brooks, J.T., Sowers, E.G., Wells, J.G., Greene, K.D., Griffin, P.M., Hoekstra, R.M., Strockbine, N.A., 2005. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. J. Infect. Dis. 192 (8), 1422–1429.

Centers for Disease Control and Prevention (CDC), 2012. National Shiga Toxin-producing *Escherichia coli* (STEC) Surveillance Overview. US Department of Health and Human Services, Atlanta, Georgia.

Cho, S., Boxrud, D.J., Bartkus, J.M., Whittam, T.S., Saeed, M., 2007. Multiple-locus variable-number tandem repeat analysis of *Salmonella* Enteritidis isolates from human and non-human sources using a single multiplex PCR. FEMS Microbiol. Lett. 275 (1), 16–23.

Ethelberg, S., Olsen, K., Flemming, S., Jensen, C., Schiellerup, P., Engberg, J., Munk Petersen, A., Olesen, B., Gerner-Smidt, P., Molbak, K., 2004. Virulence Factors for Hemolytic Uremic Syndrome. Denmark, Emerging Infectious Diseases Available at: http://wwwnc.cdc.gov/eid/article/10/5/03-0576.htm.

Grande, L., Michelacci, V., Tozzoli, R., Ranieri, P., Maugliani, A., Caprioli, A., Morabito, S., 2014. Whole genome sequence comparison of vtx2-converting phages from Enteroaggregative Haemorrhagic *Escherichia coli* strains. BMC Genomics 15 (1), 574.

Gyles, C.L., 2007. Shiga toxin-producing *Escherichia coli*: an overview. J. Anim. Sci. 85 (Suppl. 1), E45–E62.

Hunter, P.R., Gaston, M.A., 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. 26, 2465–2466.

Hyytiä-Trees, E., Smole, S.C., Fields, P.A., Swaminathan, B., Ribot, E.M., 2006. Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 (STEC O157). Foodborne Pathog. Dis. 3, 118–131.

Hyytiä-Trees, E., Lafon, P., Vauterin, P., Ribot, E., 2010. Multi-laboratory validation study of standardized multiple-locus VNTR analysis (MLVA) protocol for Shiga toxin-producing *Escherichia coli* O157 (STEC O157): a novel approach to normalize fragment size data between capillary electrophoresis platforms. Foodborne Pathog. Dis. 7, 129–136.

Izumiya, H., Pei, Y., Terajima, J., Ohnishi, M., Hayashi, T., Iyoda, S., Watanabe, H., 2010. New system for multilocus variable-number tandem-repeat analysis of the enterohemorrhagic *Escherichia coli* strains belonging to three major serogroups: O157, O26, and O111. Microbiol. Immunol. 54 (10), 569–577.

Johnson, R.P., Clarke, R.C., Wilson, J.B., 1996. Growing concerns and recent outbreaks involving non-O157:H7 serotypes of verotoxigenic *Escherichia coli*. J. Food Prot. 59, 1112–1122.

Johnson, K.E., Thorpe, C.M., Sears, C.L., 2006. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin. Infect. Dis. 43 (12), 1587–1595.

Kaplinski, L., Andreson, R., Puurand, T., Remm, M., 2005. MultiPLX: automatic grouping and evaluation of PCR primers. *Bioinformatics* 21 (8), 1701–1702 Available at: http://bioinfo.ut.ee/multiplx/.

Karama, M., Gyles, C.L., 2010. Methods for genotyping verotoxin-producing *Escherichia coli*. Zoonoses Public Health 57 (7–8), 447–462.

Karmali, M.A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., Kaper, J.B., 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 (11), 4930–4940.

Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., Hugh-Jones, M.E., 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. 182 (10), 2928–2936.

Keys, C., Kemper, S., Keim, P., 2005. Highly diverse variable number tandem repeat loci in the *E. coli* O157:H7 and O55:H7 genomes for high-resolution molecular typing. J. Appl. Microbiol. 98, 928–940.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. ClustalW and ClustalX version 2. Bioinformatics 23 (21), 2947–2948.

Lindstedt, B.A., Heir, E., Gjernes, E., Kapperud, G., 2003. DNA fingerprinting of *Salmonella enterica* subsp. *enterica* serovar Typhimurium with emphasis on phage type DT104 based on the variable number tandem repeat loci. J. Clin. Microbiol. 41, 1469–1479.

Lindstedt, B.A., Vardund, T., Aas, L., Kapperud, G., 2004. Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. J. Microbiol. Methods 59, 163–172.

Lindstedt, B.-A., Brandal, L.T., Aas, L., Vardund, T., Kapperud, G., 2007. Study of polymorphic variable-number of tandem repeats loci in the ECOR collection and in a set of pathogenic *Escherichia coli* and *Shigella* isolates for use in a genotyping assay. J. Microbiol. Methods 69, 197–205.

Løbersli, I., Haugum, K., Lindstedt, B.-A., 2012. Rapid and high resolution genotyping of all *Escherichia coli* serotypes using 10 genomic repeat-containing loci. J. Microbiol. Methods 88 (1), 134–139.

Nadon, C.A., Trees, E., Ng, L.K., Møller Nielsen, E., Reimer, A., Maxwell, N., Kubota, K.A., Gerner-Smidt, P., MLVA Harmonization Working Group, 2013. Development and application of MLVA methods as a tool for inter-laboratory surveillance. Eur. Surg. 18 (35) (pii=20565).

Nataro, J.P., Kaper, J.B., 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. 11, 142–201.

Noller, A.C., McEllistrem, M.C., Pacheco, A.G.F., Boxrud, D.J., Harrison, L.H., 2003. Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic *Escherichia coli* O157:H7 isolates. J. Clin. Microbiol. 41 (12), 5389–5397.

O'Brien, A.D., Tesh, V.L., Donohue-Rolfe, A., Jackson, M.P., Olsnes, S., Sandvig, K., Lindberg, A.A., Keusch, G.T., 1992. Shiga toxin: biochemistry, genetics, mode of action, and role in pathogenesis. Curr. Top. Microbiol. Immunol. 180, 65–94.

Ribot, E.M., Fair, M.A., Gautom, R., Cameron, D.N., Hunter, S.B., Swaminathan, B., Barret, T.J., 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. Foodborne Pathog. Dis. 3, 59–67.

Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.-A., Roy, S.L., Jones, J.L., Griffin, P.M., 2011. Foodborne illness acquired in the United States—major pathogens. Emerg. Infect. Dis. 17, 7–15.

Scheutz, F., Nielsen, E.M., Frimodt-Moller, J., Boisen, N., Morabito, S., Tozzoli, R., Caprioli, A., 2011. Characteristics of the enteroaggregative Shiga toxin/verotoxin-producing *Escherichia coli* O104:H4 strain causing the outbreak of haemolytic uraemic syndrome in Germany, May to June 2011. Eur. Surg. 16 (24), 19889.

Smith, D., Rooks, D., Fogg, P., Darby, A., Thomson, N., McCarthy, A., Allison, H., 2012. Comparative genomics of Shiga toxin encoding bacteriophages. BMC Genomics 13 (1), 311.

Swaminathan, B., Barrett, T.J., Hunter, S.B., Tauxe, R.V., 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. 7, 382–389.

Tarr, P.I., Gordon, C.A., Chandler, W.L., 2005. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. Lancet 365, 1073–1086.

Untergrasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3 — new capabilities and interfaces. Nucleic Acids Res. 40 (15), e115 Available at: http://bioinfo.ut.ee/primer3/.

Weir, B.S., 1990. Genetic Data Analysis. Sinauer Associates, Sunderland, MA.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31 (13), 3406–3415 Available at: http://unafold.rna.albany.edu/?q=mfold/mfold-references.