Epidemiol. Infect. (2015), **143**, 104–107. © Cambridge University Press 2014
doi:10.1017/S0950268814000739

SHORT REPORT
Inter-laboratory comparison of multi-locus variable-number tandem repeat analysis (MLVA) for verocytotoxin-producing *Escherichia coli* O157 to facilitate data sharing

A. HOLMES[1]*, N. PERRY[2], G. WILLSHAW[2], M. HANSON[1] AND L. ALLISON[1]

[1] *Scottish Escherichia coli O157/VTEC Reference Laboratory, Royal Infirmary of Edinburgh, Edinburgh, UK*
[2] *Gastrointestinal Bacterial Reference Unit, Public Health England, London, UK*

Received 17 January 2013; Final revision 10 March 2014; Accepted 10 March 2014; first published online 4 April 2014

SUMMARY

Multi-locus variable number tandem repeat analysis (MLVA) is used in clinical and reference laboratories for subtyping verocytotoxin-producing *Escherichia coli* O157 (VTEC O157). However, as yet there is no common allelic or profile nomenclature to enable laboratories to easily compare data. In this study, we carried out an inter-laboratory comparison of an eight-loci MLVA scheme using a set of 67 isolates of VTEC O157. We found all but two isolates were identical in profile in the two laboratories, and repeat units were homogeneous in size but some were incomplete. A subset of the isolates ($n = 17$) were sequenced to determine the actual copy number of representative alleles, thereby enabling alleles to be named according to international consensus guidelines. This work has enabled us to realize the potential of MLVA as a portable, highly discriminatory and convenient subtyping method.

Key words: Data sharing, MLVA, verocytotoxin-producing *Escherichia coli* O157, typing.

Effective typing of verocytotoxin-producing *Escherichia coli* O157 (VTEC O157) is essential for outbreak investigations and epidemiological surveillance. Multi-locus variable number tandem repeat analysis (MLVA) is a highly discriminatory and reproducible molecular typing method for a variety of pathogens, including VTEC O157 [1]. Although published MLVA methods exist for VTEC/STEC O157 [2–5], there is a paucity of information on inter-laboratory harmonization, and as yet there is no internationally standardized nomenclature. A multi-laboratory validation of a VTEC O157 scheme targeting eight tandem repeat loci in two multiplex PCR reactions [5, 6] has been reported; however, the strategy used for data comparison was only possible if a standardized protocol was used. Recently, the results of an international consensus for the development and use of MLVA for inter-laboratory surveillance were reported proposing the sequencing of a representative set of 'calibration' strains to normalize raw data to actual fragment sizes, thereby enabling laboratories to share data despite using different platforms, equipment and reagents [7].

The Gastrointestinal Bacterial Reference Unit (GBRU) at Colindale, Public Health England, provides reference services for VTEC O157 isolates from England and Wales, and has utilized MLVA as a front-line typing tool for epidemiology and surveillance since 2008. As the majority of European *E. coli* O157 confirmed cases are reported by the UK and Ireland (~80% in 2011) [8], the ability to easily compare MLVA results within and between these countries is imperative to facilitate national

* Author for correspondence: Dr A. Holmes, Scottish *Escherichia coli* O157/VTEC Reference Laboratory, Royal Infirmary of Edinburgh, 51 Little France Crescent, Edinburgh, EH16 4SA. (Email: anne.holmes@luht.scot.nhs.uk)

surveillance and investigation of cross-border outbreaks. Therefore, the aim of this study was to use a panel of VTEC O157, previously subtyped by GBRU, to determine the reproducibility and robustness of MLVA carried out in the Scottish *E. coli* O157/VTEC Reference Laboratory (SERL) and to identify and address any issues hindering the comparison of data between the laboratories.

The isolates of VTEC O157 were received and typed at GBRU in 2009 and 2010. A wide range of isolates was chosen to represent frequently occurring phage types (21/28, 8, 2) together with representatives of less common types (14, 32, 34), and reflected, as far as possible, all of the alleles for each of the eight loci in the MLVA scheme. DNA was extracted using 10% Chelex resin (Bio-Rad, UK; GBRU) or Instagene Matrix (Bio-Rad; SERL). Amplification of eight MLVA loci was carried out in two quadruplex PCR reactions [6] with some modifications. The following dye labels were used in the forward primers in the amplifications: NED in VNTR3 and VNTR17; 6-FAM in VNTR34, VNTR19, VNTR9 and VNTR36 and VIC in VNTR25 and VNTR37. PCR reactions contained Invitrogen Platinum *Taq* (Invitrogen, UK; GBRU) or Qiagen Multiplex PCR kit reagents (Qiagen, UK; SERL). Sizing of the amplified products was on ABI 3730 (GBRU) or ABI 3130 (SERL) Genetic Analysers with 600 LIZ (Applied Biosystems, UK) as size standard, using 50 cm (GBRU) or 36 cm (SERL) capillaries and POP-7™ polymer. Data were analysed with Peak Scanner v. 1.0 (Applied Biosystems; GBRU) or GeneMapper v. 4.0 (Applied Biosystems; SERL) software, and fragment sizes were imported and analysed in BioNumerics v. 6.6 (Applied Maths, Belgium). EDL933 was included in each run as a positive control.

Loci were named and reported as proposed recently [7]. Alleles were assigned based on their actual sequenced copy number, which was determined by sequencing a subset of the 67 isolates ($n = 17$). For sequencing, the MLVA primers were used in singleplex PCR but the forward primers were unlabelled. The PCR products were cleaned with Illustra ExoProStar 1-Step™ (GE Healthcare Life Sciences, UK) and both forward and reverse strands were sequenced using the ABI Cycle-sequencing kit v. 3.1 (Applied Biosystems). The products were cleaned and run on the ABI 3130xl automated sequencer. Sequences were analysed in BioNumerics and Tandem Repeat Finder [9] was used to confirm copy numbers.

Identical MLVA profiles were obtained for 65/67 (97%) isolates in both laboratories (data not shown). The two discrepancies in MLVA profile were due to single locus difference between strains. In one case, no peak was detected at ECS5426. When this locus was re-amplified using singleplex PCR and agarose gel electrophoresis, no PCR product was detected, suggesting either a failure of primer binding due to a primer binding site mutation or the loss of the locus in the strain. In the other case, a one repeat difference of 7 bp was detected at pO15754 (216 bp *vs.* 223 bp). The most likely explanation for these differences was the strains evolved during subculture.

As shown in Table 1, the inter-laboratory fragment sizing was very reproducible (0–2 bp) despite differences in the Genetic Analyser models, capillary column sizes, PCR reagents and run conditions. Others have previously reported inter-laboratory fragment sizing within 1·5 bp when the same platform was used, while differences of up to 6 bp have been observed when different machines were used [6].

The sizing discrepancies were most notable between the laboratories in locus ECS3490, probably because the difference in the capillary array lengths causes resolution differences for large fragments. Moreover, the likely reason for the overall good reproducibility between the laboratories was that the same polymer type was used for fragment analysis. Differences in the composition of polymers cause different levels of secondary and tertiary structure in the fragments which in turn cause the fragments to migrate differently through the polymer resulting in notable fragment size discrepancies [10].

To determine allele nomenclature we sequenced a subset of 17 genetically diverse strains covering a representative set of alleles for each of the eight loci, which are highlighted in Table 1. As proposed by Nadon *et al.* [7], the smallest, largest and at least every third allele of each locus were sequenced. We found the repeat units were homogeneous in size; however, incomplete repeats were observed. Most notably, ECS3490 loci always ended with an incomplete repeat of 0·8. According to the proposed criteria, incomplete repeats should be rounded down to the nearest complete copy number, therefore the ECS3490 allele for EDL933 was assigned 10, rather than 11 as previously reported [5, 11]. The full MLVA profile for EDL933 was 9-5-6-10-10-6-7-8. To help identify the start and end of the repeat regions, the 5′ and 3′ sequences immediately flanking the tandem repeats were identified (see Table 1). Sequence variation was

https://doi.org/10.1017/S0950268814000739 Published online by Cambridge University Press

Table 1. *Details of the VTEC O157 VNTR loci**

Locus name (old name)	ECS271 (VNTR3)		ECS1520 (VNTR25)		ECS2862 (VNTR19)		ECS3490 (VNTR9)		ECS5331† (VNTR34)		ECS5426 (VNTR17)		pO15746 (VNTR37)		pO15754 (VNTR36)	
5′ flanking seq.	TCTGCTAGAA		TAATTTTGC**T**		ACATCATCAA		AAATCACAAC		AATCATCAGA		TGATGACTTA		TTGTTGTTCA		AACTCAGGCT	
3′ flanking seq.	AACCAAAATC		TGAGAAATAT		GAACATCATC		AAATATTAAA		ACTTAAAA AA		TAACAGTTAA		TGTTCTTCTT		AACAGCGGGT	
Repeat unit	AAGGTG		**T**GCAAA		GACCAC		AAATAG		AGTTAAATAA**TTCG**CAG**G**		TATCTT		TGCTAC		ACCTCAC	
Allele no.	SERL	GBRU	SERL	GBRU	SERL	GBRU	SERL	GBRU	SERL	GBRU	SERL	GBRU	SERL	GBRU	SERL	GBRU
2			119	119			478	479								
3			125	125			484	485			135	135				
4	345	344	130–131	130–131	292	293	490	491			141–143	141	167	167	126–127	127
5	350–351	350	136	135–137	298	298–299	495–496	496–497			148–149	147–148	173	173	133	133–134
6	356–357	356	142	142	304	304–305	501	502	202	203	154–155	154	179	179	140	140
7	363	362	148	148	310	310	507	508	220–221	221	160	160	185	185	147	147
8	368–369	368			316	316	513	514	238	238–239	166–167	166	191	191	154	154
9	374–375	374					519	520	256	256–257	173	172	197	197	161	161
10	380–381	380			328	328	525	526	274	274–275			203	203	168	168
11	386–387	386			333	334	530–531	532			185	185	209	209	174	175
12	393	392			339	340	536	538			191	191	215	215	181	181
13	398–399	398			345	346	542	543–544			197	197	221	221	188	188
14	405	404			351	352	548	549					227	227	195	195
15	410–411	410			357	358	553–554	555					233	233	202	202
16	416–417	416			363	362	559	561								
17	422–423	422					565	567							216	216
18	428	428					571	572								223
19	434	434					577	578								
20							582–583	584								
21	446	446														
22	453	452														
23	459	458														
24	464–465	464														

SERL, Scottish *E. coli* O157/VTEC Reference Laboratory; GBRU, Gastrointestinal Bacteria Reference Unit.

* Consensus 5′ and 3′ flanking sequences and repeat unit sequences are shown; sequence variation is highlighted in bold; and representative alleles that were sequenced are underlined.

† Five different repeat unit sequences were identified for ECS5331: AGTTAAATAATCTACAGA, AGTTAAATAATATACAGA, AGTTAAATAATATACAGG, AGTTAAATAATTCGCAGG, and AGTTAAATAACTCGCGGG.

observed in the 5′ flanking sequences of ECS1520 in two of the 17 strains tested (TAATTTTGC**T**→ TAATTTTGC**C**). Sequence variation was also identified in one of the ECS1520 repeat units in one strain (**TGCAAA**→**AGCAAA**), and in all sequenced strains five different repeat unit sequences were identified for ECS5331 (see Table 1).

Problems identified in comparing MLVA data produced in different laboratories have been reported previously [6]. In this study, we used one of the MLVA protocols available for VTEC O157 and followed recent guidelines based on international consensus to name alleles based on actual copy numbers enabling the successful comparison of data generated in our laboratories, despite the use of different equipment and reagents. Similar work has been reported for *Salmonella* Typhimurium MLVA [12, 13]. Since the completion of the present study, we have exchanged real-time MLVA data to aid cross-border outbreak investigations, and are implementing a quality assurance programme to ensure the continual provision of high-quality typing data, which could be extended to other countries. The panel of 17 reference strains can be provided to other laboratories interested in calibrating their VNTR data to actual sequenced copy numbers and is available for inclusion in an extended European or international calibration set. The international harmonization of MLVA for VTEC would facilitate the development of a central database and a consensus nomenclature for allelic profiles to enable the rapid identification of known or novel strains for international surveillance and outbreak detection of VTEC.

ACKNOWLEDGEMENTS

The authors thank Leila Shaw and Tracy Dransfield for technical assistance.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Lindstedt BA, *et al.*** Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. *Eurosurveillance* 2013; **18**: pii=20385
2. **Lindstedt BA, Vardund T, Kapperud G.** Multiple-locus variable number tandem-repeats analysis of *Escherichia coli* O157 using PCR multiplexing and multi-colored capillary electrophoresis. *Journal of Microbiology Methods* 2004; **58**: 213–22.
3. **Noller AC, *et al.*** Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic *Escherichia coli* O157:H7 isolates. *Journal of Clinical Microbiology* 2003; **41**: 5389–5397.
4. **Keys C, Kemper S, Keim P.** Highly diverse variable number tandem repeat loci in the *E. coli* O157:H7 and O55:H7 genomes for high-resolution molecular typing. *Journal of Applied Microbiology* 2005; **98**: 928–940.
5. **Hyytiä-Trees E, *et al.*** 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 Pathogens Disease*. 2006; **3**: 118–131.
6. **Hyytia-Trees E, *et al.*** Multilaboratory validation study of standardized multiple-locus variable-number tandem repeat analysis protocol for shiga toxin-producing *Escherichia coli* O157: a novel approach to normalize fragment size data between capillary electrophoresis platforms. *Foodborne Pathogens and Disease* 2010; **7**: 129–136.
7. **Nadon CA, *et al.*** Development and application of MLVA methods as a tool for inter-laboratory surveillance. *Eurosurveillance* 2013; **18**: 20 565–20 575.
8. **European Centre for Disease Prevention and Control (ECDC).** Annual Epidemiological Report. Reporting on 2009 surveillance data and 2010 epidemic intelligence data. Stockholm: ECDC, 2011.
9. **Benson G.** Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 1999: **27**: 573–580.
10. **Rosenblum BB, *et al.*** Improved single-strand DNA sizing accuracy in capillary electrophoresis. *Nucleic Acids Research* 1997; **25**: 3925–3929.
11. **Lindstedt BA, *et al.*** DNA fingerprinting of Shiga-toxin producing *Escherichia coli* O157 based on multiple-locus variable-number tandem-repeats analysis (MLVA). *Annals of Clinical Microbiology and Antimicrobials* 2003; **2**: 12.
12. **Larsson JT, *et al.*** Proof-of-concept study for successful inter-laboratory comparison of MLVA results. *Eurosurveillance* 2013; **18**: 20 566–20 573.
13. **Larsson JT, *et al.*** Development of a new nomenclature for *Salmonella typhimurium* multilocus variable number of tandem repeats analysis (MLVA). *Eurosurveillance* 2009; **14**: pii=19174.