Standardisation of multilocus variable-number tandem-repeat analysis (MLVA) for subtyping of *Salmonella enterica* serovar Enteritidis

K L Hopkins (katie.hopkins@hpa.org.uk)¹, T M Peters¹, E de Pinna¹, J Wain¹

1. Health Protection Agency Microbiology Services Colindale, London, United Kingdom

Citation style for this article:

Citation style for this article: Hopkins KL, Peters TM, de Pinna E, Wain J. Standardisation of multilocus variable-number tandem-repeat analysis (MLVA) for subtyping of *Salmonella enterica* serovar Enteritidis. Euro Surveill. 2011;16(32):pii=19942. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19942>

Article published on 11 August 2011

Salmonella enterica serovar (S.) Enteritidis is an important cause of food-borne infection in Europe and the United States. Further subtyping of isolates is necessary to support epidemiological data for the detection of outbreaks and identification of the vehicle of infection. Multilocus variable-number tandem-repeat analysis (MLVA) is reportedly more discriminatory and produces data that are easier to share via databases than other molecular subtyping methods. However, lack of standardisation of the methodology and interpretive criteria for data analysis has meant that comparison of data between laboratories can be problematic. On the basis of MLVA profiles of 298 *S. Enteritidis* isolates received at the Health Protection Agency's *Salmonella* Reference Unit and sequence analysis of selected isolates, we propose a MLVA scheme for *S. Enteritidis* based on five loci (SENTR₄, SENTR₅, SENTR₆, SENTR₇ and SE-3) that have been selected from previously published *S. Enteritidis* MLVA schemes. A panel of reference strains has been developed that can be used by laboratories to normalise their raw fragment data to actual fragment sizes. We also provide recommendations for analysing and interpreting MLVA data. We urge laboratories to consider implementing these guidelines, thereby allowing direct comparison of data between laboratories irrespective of the platform used for fragment analysis, to facilitate international surveillance and investigation of international outbreaks.

Introduction

Non-typhoidal *Salmonella enterica* cause a considerable disease burden, with an estimated 93.8 million cases of infection worldwide every year, resulting in 155,000 deaths [1]. In the European Union, this pathogen is ranked second among the causes of bacterial gastrointestinal disease; *S. enterica* serovar (S.) Enteritidis is responsible for approximately 60% of salmonella infections in humans, making it the leading cause of salmonellosis [2]. Further subtyping of isolates is needed to support classical epidemiological data for the detection of outbreaks and identification of the vehicle of infection.

Phage typing is a phenotypic method traditionally used for surveillance and subtyping of salmonellae but is performed in only a few laboratories due to the requirement for standardised phage panels [3]. Although useful in detecting outbreaks caused by isolates exhibiting less common phage reaction patterns, the technique may otherwise lack discriminatory capacity – 20% of salmonellosis cases reported in Europe in humans in 2006 were caused by *S. Enteritidis* phage type (PT) 4 [2]. Subsequently, DNA fingerprinting techniques such as pulsed-field gel electrophoresis (PFGE), which is currently the gold standard for subtyping of salmonellae, are used in outbreak investigations to supplement phage typing data where further strain discrimination is required. However, the lack of genetic variation within the serovar Enteritidis population can make discrimination beyond phage type difficult. In a multicentre European study conducted between 2001 and 2004, over half of the serovar Enteritidis strains produced PFGE profile SENTXB.0001 regardless of phage type [4]. Strong associations between phage type and a particular PFGE pattern also make strain differentiation within certain phage types challenging [4].

Multilocus variable-number tandem-repeat analysis (MLVA) targets rapidly evolving genomic elements known as tandem repeats (TRs). The application of MLVA to a variety of bacterial species including several salmonellae serovars has led to the conclusion that the technique is more discriminatory than other molecular methods, and is reproducible, quicker, easier to perform and produces data that are easier to analyse and share via databases. To date, several schemes for MLVA subtyping of *S. Enteritidis* have been published [5-10]. However, the use of different loci in each protocol (or different primers for the same loci), different sequencer platforms, dye chemistries and size standards used for fragment analysis, differences in interpretation of loci where incomplete TRs or TRs of heterogeneous sequence occur, and different ways of assigning allele numbers means comparison of data between laboratories can be problematic. In addition, with few data available on the stability

of the loci, it is uncertain whether TRs may evolve so rapidly that variation leading to multiple types could emerge during an outbreak caused by a single ancestral isolate. Such concerns threaten to diminish the utility of MLVA for *S. Enteritidis* outbreak detection unless specific guidance is developed for performing MLVA and a consensus reached for the interpretation of MLVA data, as has been proposed for *S. Typhimurium* [11].

In this study we analysed the DNA sequence of the TR regions at the nine loci used in the MLVA method proposed by Malorny et al. to determine the actual number, stability and heterogeneity of the TRs [8]. This scheme was chosen due to its emphasis on discrimination within PTs 4 and 8, which rank in the top three most prevalent PTs in several European countries [2]. We also examined other published MLVA schemes for *S. Enteritidis* to determine whether other loci could be used to supplement the scheme.

Here we propose a standardised MLVA typing scheme for *S. Enteritidis* targeting five loci, where profiles are assigned based on the number of TRs at each locus. In addition, we developed a panel of reference strains that can be used by laboratories to normalise their raw fragment data to actual fragment sizes, thereby allowing

direct comparison of data between laboratories irrespective of the platform used for fragment analysis.

Methods

S. Enteritidis strains

To evaluate the MLVA method, 298 *S. Enteritidis* strains were selected from the Health Protection Agency (HPA) *Salmonella* Reference Unit culture collection, which consists of strains originating from human clinical specimens, animals, food and the environment. The strain panel comprised 91 strains from the phage typing 'type strain' panel (the first recorded isolations of each phage type), 88 strains isolated in England and Wales (30 strains of PT14b, 21 strains of PT4, 15 strains of PT8, nine strains of PT42 and one to three strains of PTs 1, 1b, 2, 3, 5c, 6, 21, 22, 59 and 14c) and 15 PT14b strains from University Hospital Galway, Ireland. In addition, strains isolated during well-characterised outbreaks in 2009 and 2010 associated with *S. Enteritidis* PTs 4 (n=23), 8 (n=26), 14 (n=11) and 14b (n=40, representing 12 geographically distinct outbreaks, A-L), were analysed to determine TR stability during an outbreak and therefore the utility of MLVA for outbreak detection (Table 1). A total of 16 strains that cover the range of alleles seen at each locus were selected as the reference strain panel.

TABLE 1

Outbreaks due to infection with *Salmonella enterica* serovar Enteritidis characterised in this study

Outbreak	Phage type	Date of isolate receipt in the laboratory	Number of isolates tested (total number of isolates)	Source (number of isolates, when more than one source)	MLVA profile ^a (number of isolates, when more than one profile)
A	14b	7–20 Oct 2009	2 (9)	Faeces	3-10-9-5-4-4-1-8-8
B	14b	29 Sep–9 Oct 2009	3 (3)	Faeces (2), unknown (1)	3-10-9-5-4-4-1-8-8
C	14b	22–28 Sep 2009	1 (2)	Faeces	3-10-9-5-4-4-1-8-8
D	14b	20 Aug 2009	1 (6)	Faeces	3-10-9-5-4-4-1-8-8
E	14b	11–24 Aug 2009	3 (4)	Faeces	3-10-9-5-4-4-1-8-8
F	14b	9–30 Sep 2009	1 (8)	Faeces	3-10-9-5-4-4-1-8-8
G	14b	23–24 Sep 2009	2 (3)	Faeces	3-10-9-5-4-4-1-8-8
H	14b	9 Sep 2009	2 (2)	Faeces (1), egg mayonnaise (1)	3-10-9-5-4-4-1-8-8
I	14b	15–22 Sep 2009	1 (9)	Faeces	3-10-9-5-4-4-1-8-8
J	14b	18–30 Sep 2009	10 (34)	Faeces (8), refrigerator door (1), poultry meat (1)	3-10-9-5-4-4-1-8-8 (9) 3-10-9-5-4-4-1-8-8 (2)
K	14b	24–28 Sep 2009	2 (2)	Faeces	3-11-9-5-4-4-1-8-8
L	14b	7 Aug–29 Sep 2009	12 (20)	Faeces (10), blood (2)	2-12-9-12-4-3-2-8-9 (9) 3-11-9-5-4-4-1-8-8 (2) 3-10-9-5-4-4-1-8-8 (1)
–	4	15–25 Sep 2009	23 (34)	Faeces (13), unknown (10)	3-10-9-5-4-4-1-8-8
–	8	20 Aug–30 Sep 2009	26 (67)	Faeces	1-9-9-7-4-3-2-8-9
–	14	29 Mar–15 Apr 2010	11 (11)	Faeces	2-8-7*-4-5-5-NA-8-6 (10) 2-7-7*-5-5-5-NA-8-6 (1)

MLVA: multilocus variable-number tandem-repeat analysis; NA: no amplification at this locus.

^a Numbers in bold indicate locus variants. 7* refers to an allele with seven tandem repeats (TRs), but TR2 lacks a 6 base pair (bp) insert and TR6 is missing 21 bp.

Strain NCTC 13349 from the National Collection of Type Cultures was included in our fragment analysis protocol as a positive control.

Comparison of published MLVA schemes for *S. Enteritidis* strains

Published MLVA schemes for *S. Enteritidis* other than that of Malorny et al. [8] were examined to determine whether other loci [5-7,9,10] could be used to supplement the scheme using the Tandem Repeats Database [12]. We used the TR region targeted and length and sequence similarity of the TR units to compare the loci.

MLVA typing

MLVA was performed using previously described primers [8]. Forward primers for loci SENTER1, SENTER4, SE-3 and SE-7 were labelled with the fluorescent dye VIC, SENTER2, SENTER5 and SENTER7 with 6-FAM, and SENTER3 and SENTER6 with NED. The nine loci were amplified in one multiplex PCR (12.5 µl volume) using a Multiplex PCR Kit (Qiagen, United Kingdom), 0.5 pmol of the primers amplifying loci SENTER4 and SENTER7, 1 pmol of primers targeting SENTER5, SENTER6 and SE-3, 5 pmol of primers targeting SENTER1, 7.5 pmol of primers targeting SENTER3 and SE-7, 10 pmol primers targeting SENTER2 and 1 µl cell lysate prepared by emulsifying one colony in 100

µl of sterile distilled water and boiling for 10 minutes. PCR cycling conditions were as previously described [8]. Amplification products were diluted 1:40 in sterile distilled water and 1 µl aliquots of this dilution were mixed with 10 µl Hi-Di formamide (Applied Biosystems, United Kingdom) and 0.5 µl GeneScan1200 LIZ Size Standard (Applied Biosystems) before being subjected to capillary electrophoresis using POP7 polymer on an ABI 3730 DNA Analyzer (Applied Biosystems) spectrally calibrated to run filter set G5.

Data were imported into Peak Scanner software (Applied Biosystems) where each fragment was identified according to colour and size. Naming of profiles was based on a string of allele numbers (in order of SENTER7-SENER5-SENER2-SENER6-SENER3-SENER4-SE3-SENER1-SE7) showing the actual number of repeats at each locus. We assigned 'NA' (no amplification) to loci that failed to amplify, in accordance with guidance issued by Larsson et al., to distinguish between strains where the locus is absent and those where there are no TRs but the flanking regions are present [11].

Validation of MLVA results

A combination of agarose gel electrophoresis (for loci SENTER1-3 and SE-7) and DNA sequencing (for loci SENTER4 to SENTER7 and SE-3) was used to determine

TABLE 2

Characteristics of tandem-repeat loci in published *Salmonella enterica* serovar Enteritidis MLVA schemes

Locus ^a	Alternative name(s) [reference]	Repeat size in base pairs	Genome position, in nucleotides ^b	Number of repeats ^b	Percentage conservation of tandem repeats ^b
SENER1	SE-10 [5], STTR1 [14]	45	774,282 – 774,633	8	92
SENER2	STTR7 [14]	39 ^c	954,671–955,038	9	92
SENER3	–	93	1,697,207–1,697,618	4	93
SENER4	SE-1 [5–7,10], ENTR13 [9]	7	2,504,795–2,504,828	4	100
SENER5	SE-5 [5–7], STTR5 [9,10,14]	6	3,073,233–3,073,292	10	100
SENER6	SE-2 [5–7,10], ENTR20 [9]	7	4,617,691–4,617,726	5	100
SENER7	SE-9 [5–7]	9	533,269–533,296	3	100
SE-3	SE-3 [5–7]	12 ^d	2,073,359–2,073,396	3	68
SE-7	SE-7 [5–7]	61 ^e	2,961,431–2,961,886	8	100: 29 bp 5'-sequence ^e 24: 32 bp 3'-sequence ^e
–	SE-4 [5–7]	117	2,530,891–2,531,224	2	75
–	SE-6 [5–7], STTR3 [10,14]	33	3,511,025–3,511,367	10	90
–	SE-8 [5–7]	87	2,812,925–2,813,011	1	–
–	ENTR6 [9]	175 ^f	1,363,028–1,363,570	3	97
–	STTR9 [10,14]	Unknown ^g	3,134,588–3,134,731	Unknown ^g	Unknown ^g

bp: base pair; MLVA: multilocus variable-number tandem-repeat analysis; NCTC: National Collection of Type Cultures; TR: tandem repeat.

^a Nomenclature according to Malorny et al. [8].

^b In NCTC 13349 (GenBank accession number AM933172); only complete repeats are included.

^c Second TR is 45 bp.

^d Third TR is 14 bp.

^e 29 bp conserved 5'-sequence, together with a 32 bp variable 3'-sequence.

^f Second TR is 174 bp.

^g The Tandem Repeats Database was unable to identify a TR at this locus, presumably as there is only one copy in NCTC 13349. The TR sequence reportedly located at this locus in serovar Typhimurium [11] is not found in serovar Enteritidis.

the number of TRs in the first 100 strains tested. Thereafter DNA sequencing was used only to confirm the number of TRs in novel amplified fragments identified by capillary electrophoresis. Amplification was performed on cell lysates in a monoplex PCR with the same primer sequences used for MLVA but with unlabelled forward primers. Sequencing was performed in one direction only using the forward primer for all loci except SENTRY7, which was sequenced with the reverse primer. Loci of the 16 strains chosen as the reference strain panel were sequenced in both directions to determine the sequence of the TRs and flanking regions.

Sequencing data were imported into Bionumerics version 6.1 (Applied Maths, Belgium) as categorical data and the numbers of TRs at each locus calculated; only complete TRs were included in the analysis. Standard minimum spanning trees generated in Bionumerics using the single and double locus variance priority rules were used to visualise the relationships between strains. Alignments of TR sequences were performed and degree of sequence identity between copies of each TR calculated using BioEdit version 7.0.9.0 [13].

Results

Comparison of published MLVA schemes for *S. Enteritidis*

The majority of loci targeted in published MLVA schemes for *S. Enteritidis* do not contain true TRs, as the repeat sequence varied in length and sequence within NCTC 13349; only loci SENTRY4 to SENTRY7 contained TRs that were 100% conserved (Tables 2 and 3). There was considerable overlap in the loci targeted between the scheme of Malorny et al. [8] and other published schemes, though with the exception of loci SE-3 and SE-7, different primers were used. No additional loci were therefore identified that could enhance the discriminatory capacity of the Malorny et al. scheme.

MLVA validation using DNA sequencing

Agarose gel electrophoresis and DNA sequencing were used to determine the number of TRs at each locus in the first 100 isolates analysed. This confirmed that fragment size always correlated with multiples of TRs. Subsequent strains were analysed by fragment analysis and sequencing was used to determine the number of TRs in any novel-sized fragments. The 16 reference strains were selected from the 298 strains examined to

TABLE 3

Characteristics of *Salmonella enterica* serovar Enteritidis loci targeted by the MLVA scheme described by Malorny et al.^a

Locus	Length of 5' flanking region	Sequence of 5' flanking region ^b	Tandem repeat sequence ^c	Sequence of 3' flanking region ^b	Length of 3' flanking region	Allele number ^d	Number of alleles identified in this study
ENTRY1	28	GGCCGAAAGAA	GCGGYRAAAGCRGCRGC GGASGCKAAGAARAAAG CGGMDGCCGMR	AAAGCGGCCG	52	(X-80)/45	2
ENTRY2	110	TCAGCAACCT	GTAGCRCKCARCC<TCAGC C>RCAGTAYCAGCAGCCGC ARCARCCG	ACAGCGCCGC	51	(X-167)/39 ^e	4
ENTRY3	46	TATCCGGCGG	CGAYCCGCGYAAAGCCGC RGTGGARGCGCYATCGC YCGCGCCAAAGCCCGYAA RCAGGAGCAGCAGGCCGG ARGCGAACCTGCCGAACC GGY	CGACCCGCGTA	75	(X-121)/93	2
ENTRY4	50	TAAATGATTT	GTTGGTA	GTTGGTGATC	41	(X-91)/7	7 ^f
ENTRY5	127	CCACCATCAC	GACCAT	CATGGTCACA	84	(X-211)/6	11
ENTRY6	77	ATAGCCAGAA	GATCCG	GCTGCGCCTT	68	(X-145)/6	9
ENTRY7	27	TTCTGGCGCA	GCGAATATG	GTCTGGCAGC	81	(X-108)/9	3
SE-3	124	GATGGTATTG	TTTTCCATATTG ^g	GTTTCCTTAA	172	(X-296)/12	3 ^e
SE-7	127	GCAACCCAAC	CGGTTTATCCCCGTGGCG CGGGGAACACRNNNNVNN VNNHNNNNNNNNNNNN NNHNNNNBN	TTTACTACA	102	(X-229)/61	8

MLVA: multilocus variable-number tandem-repeat analysis; NCTC: National Collection of Type Cultures; TR: tandem repeat.

^a Source: [8].

^b Only the 10 bases adjacent to the start and finish of the TR region in NCTC 13349; incomplete TRs are excluded from the TR region.

^c As determined in NCTC 13349 where M=A or C; R=A or G; S=C or G; Y=C or T; K=G or T; B=C, G or T; D=A, G or T; H=A, C or T; V=A, C or G and N=A, C, G or T. '<' indicates nucleotides inserted into the TR.

^d X is the amplicon length as determined by sequencing, which may differ from the size determined by capillary electrophoresis.

^e The number of TRs may need to be rounded up if TR2 is lacking the 6 bp insert.

^f Includes a null variant where no fragment is amplified by PCR.

^g Differs from the TR sequence in the published scheme based on observations made during this study.

TABLE 4

Reference strains for MLVA of *Salmonella enterica* serovar Enteritidis

Reference strain	SENTR7		SENTR5		SENTR2		SENTR6		SENTR3		SENTR4		SE-3		SENTR1		SE-7	
	Length in bp ^a	Number of TRs	Length in bp ^a	Number of TRs	Length in bp ^a	Number of TRs	Length in bp ^a	Number of TRs	Length in bp ^a	Number of TRs	Length in bp ^a	Number of TRs	Length in bp ^a	Number of TRs	Length in bp ^a	Number of TRs	Length in bp ^a	Number of TRs
HPA001	135	3	265	9	518	9	173	4	493	4	119	4	308	1	440	8	717	8 ^b
HPA002	135	3	301	15	479	8	180	5	493	4	119	4	320	2	440	8	839	10
HPA003	126	2	277	11	518	9	180	5	493	4	112	3	320	2	440	8	778	9
HPA004	135	3	289	13	518	9	180	5	493	4	119	4	309	1	440	8	656	7
HPA005	135	3	271	10	518	9	187	6	493	4	119	4	309	1	440	8	717	8 ^b
HPA006	117	1	265	9	518	9	194	7	493	4	112	3	320	2	440	8	778	9
HPA007	126	2	295	14	518	9	208	9	493	4	112	3	320	2	440	8	778	9
HPA008	126	2	277	11	518	9	215	10	493	4	112	3	320	2	440	8	778	9
HPA009	126	2	283	12	518	9	229	12	493	4	112	3	320	2	440	8	778	9
HPA010	126	2	235	4	518	9	208	9	493	4	126	5	0	NA	440	8	717	8 ^c
HPA011	126	2	247	6	518	9	187	6	493	4	126	5	308	1	440	8	717	8 ^c
HPA012	126	2	253	7	518	9	229	12	493	4	133	6	308	1	440	8	595	6
HPA013	126	2	259	8	518	9	201	8	493	4	126	5	320	2	440	8	778	9
HPA014	126	2	271	10	518	9	236	13	493	4	126	5	308	1	440	8	778	9
HPA015	126	2	301	15	518	9	201	8	493	4	140	7	0	NA	440	8	656	7
HPA016	126	2	253	7 ^d	395 ^e	6 ^e	194	7	586	5	147	8	0 ^f	NA	440	8	778	9

bp: base pair; MLVA: multilocus variable-number tandem-repeat analysis; NA: no amplification at this locus; NCTC: National Collection of Type Cultures; TR: tandem repeat.

^a Length of fragment as determined by sequencing, which may differ from the size determined by capillary electrophoresis.^b Same order and sequence as NCTC 13349.^c Different order and sequence from NCTC 13349 (see Figure 1, panel D).^d Sequence of first three TRs is GACCAC-GACCAC-GGCCAT.^e The second tandem-repeat lacks a 6 bp insertion compared with the expected 401 bp amplicon.^f ABI 3730 DNA Analyzer fragment trace exhibits a 346 bp VIC-labelled fragment, which results from amplification with primers SENTR1-F and SENTR2-R.

FIGURE 1

Sequence alignments showing variation in sequence and order of tandem repeat (TR) sequences at *Salmonella enterica* serovar Enteritidis loci (A) SENTR1, (B) SENTR2, (C) SENTR3 and (D) SE-7 in NCTC 13349 and the 16 reference strains

A SENTR1

```

      10      20      30      40
      . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . |
TR1 – NCTC 13349 g c g g c g a a a g c g g c g g c g g a c g c g a a g a a a a a g c g g a g g c c g a a
TR2 – NCTC 13349 g c g g c g a a a g c g g c g g c g g a c g c g a a g a a g a a g c g g a g g c c g a g
TR2 – HPA016     G C G G C G A A A G C G G C G G C G G A G G C G A A G A A G A A G C G G A G G C C G A G
TR3 – NCTC 13349 g c g g c g a a a g c g g c g g c g g a c g c g a a g a a g a a g c g g a a g c c g a g
TR4 – NCTC 13349 g c g g t a a a a g c g g c g g c g g a c g c g a a g a a g a a g c g g a a g c c g a a
TR5 – NCTC 13349 g c g g c g a a a g c g g c g g c g g a g g c g a a g a a g a a g c g g a a g c c g a a
TR6 – NCTC 13349 g c g g c g a a a g c g g c g g c g g a g g c g a a g a a g a a g c g g a t g c c g a g
TR6 – HPA015     G C G G C G A A A G C G G C G G C G G A G G C G A A G A T G A A A G C G G A T G C C G A G
TR7 – NCTC 13349 g c g g c g a a a g c g g c g g c g g a g g c g a a g a a g a a g c g g a t g c c g c g
TR7 – HPA005     G C G G C G A A A G C G G C G G C G G A G G C G A A G A A G A A G C G G A T G C C G C A
TR8 – NCTC 13349 g c g g c g a a a g c a g c g g c g g a c g c t a a g a a g a a g c g g c t g c c g a a

```

B SENTR2

```

      10      20      30      40
      . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . |
TR1 – NCTC 13349 g t a g c g c c g c a g c c ----- g c a g t a t c a g c a g c c g c a g c a a c c g
TR1 – HPA016     G T A G C G C C G C A G C C ----- G C A G T A T C A G C A G C C G C A G C A G C C G
TR2 – NCTC 13349 g t a g c g c c g c a g c c t c a g c c a c a g t a c c a g c a g c c g c a a c a g c c g
TR2 – HPA016     G T A G C G C C G C A G C C ----- A C A G T A T C A G C A G C C G C A G C A G C C G
TR3 – NCTC 13349 g t a g c g c c g c a a c c ----- g c a g t a t c a g c a g c c g c a g c a g c c g
TR3 – HPA016     G T A G C G T C G C A G C C ----- A C A G T A T C A G C A G C C G C A A C A G C C G
TR4 – NCTC 13349 g t a g c g c c t c a g c c ----- g c a g t a t c a g c a g c c g c a a c a g c c g
TR4 – HPA016     G T A G C G C C T C A G C C ----- A C A G T A C C A G C A G C C G C A A C A G C C G
TR5 – NCTC 13349 g t a g c g c c g c a g c c ----- g c a g t a t c a g c a g c c g c a g c a g c c g
TR5 – HPA016     G T A G C G C C G C A A C C ----- G C A G T A T C A G C A G C C G C A G C A G C C G
TR6 – NCTC 13349 g t a g c g c c g c a g c c ----- g c a g t a t c a g c a g c c g c a a c a g c c g
TR7 – NCTC 13349 g t a g c a c c g c a g c c ----- g c a g t a t c a g c a g c c g c a g c a g c c g
TR8 – NCTC 13349 g t a g c g c c g c a g c c ----- a c a g t a t c a g c a g c c g c a g c a g c c g
TR8 – HPA002     G T A G C G C C G C A G C C ----- A C A G T A T C A G C A G C C G C A A C A G C C G
TR9 – NCTC 13349 g t a g c g c c g c a g c c ----- a c a g t a t c a g c a g c c g c a a c a g c c g

```

C SENTR3

```

      10      20      30      40      50      60      70
      . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . |
TR1 – NCTC 13349 c g a c c c g c g t a a a g c c g c a g t g g a g g c g g c t a t c g c t c g c g c c a a a g c c c g c a a a c a g g a g c a g c a g g c c
TR2 – NCTC 13349 c g a t c c g c g a a a g c c g c g g t g g a g g c g g c c a t c g c c c g c g c c a a a g c c c g c a a a c a g g a g c a g c a g g c c
TR2 – HPA016     C G A T C C G C G C A A A G C C G C G G T G G A G G C G G C C A T C G C C C G C G C C A A A G C C C G C A A A C A G G A G C A G C A G G C C
TR3 – NCTC 13349 c g a c c c g c g a a a g c c g c g g t g g a a g c g g c t a t c g c c c g c g c c a a a g c c c g c a a g c a g g a g c a g c a g g c c
TR3 – HPA016     C G A C C C G C G C A A A G C C G C G G T G G A A G C G G C T A T C G C C C G C G C C A A A G C C C G C A A G C A G G A G C A G C A G A C C
TR4 – NCTC 13349 c g a t c c g c g t a a a g c c g c g g t g g a a g c g g c c a t c g c c c g c g c c a a a g c c c g t a a g c a g g a g c a g c a g g c c
TR4 – HPA016     C G A T C C G C G T A A A G C C G C G G T G G A A G C G G C T A T C G C C C G C G C C A A A G C C C G C A A G C A G G A G C A G C A G G C C

      80      90
      . . . . | . . . . | . . . . | . . . . | . . . . |
TR1 – NCTC 13349 g g a a g c g a a c c t g c c g a a c c g g t
TR2 – NCTC 13349 g g a a g c g a a c c t g c c g a a c c g g t
TR2 – HPA016     G G A A G C G A A C C T G C C G A A G C G G T
TR3 – NCTC 13349 g g a g g c g a a c c t g c c g a a c c g g t
TR3 – HPA016     G G A A G C G A A C C T G C C G A A C C G A T
TR4 – NCTC 13349 g g a a g c g a a c c t g c c g a a c c g g c
TR4 – HPA016     G G A A G C G A A C C T G T C G A A C C G G C

```


D SE-7

bp: base pair; NCTC: National Collection of Type Cultures; TR: tandem repeat.

TR sequences of the 16 reference strains are shown only where they differed by one or more bases from the NCTC 13349 sequence. Locus SE-7 consists of a 29 bp 100% conserved 5'-sequence, together with a 32 bp variable 3'-sequence that shares only 24% similarity between the TRs in NCTC 13349. Letters A–N in panel D correspond to TRs with different sequences. TRs A and A', and D and D' differ by only one single nucleotide polymorphism. These are not true TRs and therefore we propose that loci SENTER1, SENTER2, SENTER3 and SE-7 are omitted from the multilocus variable-number tandem-repeat analysis (MLVA) scheme, as such variation could not be detected using fragment analysis.

represent most of the alleles observed for each locus and the amplicons from each locus were sequenced (Table 4). Sequencing confirmed the number of TRs and alignments revealed that loci SENTER1, SENTER2, SENTER3 and SE-7 exhibit variation in the sequence of the TR unit within a strain and between strains (Table 3, Figure 1). Single nucleotide polymorphisms (SNPs) were identified in the TRs of reference strains HPA002, HPA005, HPA015 and HPA016 compared with NCTC 13349 (Figure 1, panels A–C), while reference strains HPA004 and HPA015, and HPA003, HPA014 and HPA016 shared the same number of TRs at locus SE-7 (seven and nine TRs, respectively) but the order of the TRs was different in each strain (Figure 1, panel D). The TR units at locus SENTER5 were identical in all strains examined except for reference strain HPA016 and strain H101440321; in these two strains the first three repeats were GACCAC-GACCAC-GGCCAT instead of the expected GACCAT. On the basis of sequencing data for locus SE-3, we propose that the TR unit sequence at locus SE-3 should be changed from <T>WATTG<G>BTTTCCW (where W=A or T, B=C, G or T and '<'>' indicates nucleotides inserted into the TR) to TTTTCCATATTG. This TR unit is consistently 12 bp long and 100% conserved in strains carrying multiple copies, unlike the TR unit previously proposed (Table 2). The TRs at loci SENTER4, SENTER5, SENTER6, SENTER7 and SE-3 were identical in all strains and all copies were the same within a strain. The flanking regions were generally conserved, with SNPs being identified only in the 5' flanking region at locus SENTER3 in HPA016 and the 5' and 3' flanking regions of locus SENTER5 in HPA016. Fragment sizes obtained

by capillary electrophoresis were consistently smaller than actual fragment sizes as determined by sequencing, ranging from a one base pair (bp) to a 19 bp difference. This difference was most pronounced in fragments amplified from loci harbouring the longer TR units (SENTER1, SENTER2, SENTER3 and SE-7).

Diversity of MLVA patterns among *S. Enteritidis* strains

MLVA typing targeting the nine loci described by Malorny et al. [8] identified 71 profiles consisting of different allele number combinations among the panel of 298 *S. Enteritidis* strains consisting of 93 different phage types (Figure 2, panel A). The most common profile (3-10-9-5-4-4-1-8-8) was observed in 121 isolates belonging to PT14b (n=52, including 26 of the 40 outbreak isolates), PT4 (n=31, including 23 belonging to the outbreak) and a further 38 sporadic isolates belonging to 37 different phage types. This profile was also shared by control strain NCTC 13349. We noted some associations between PT and MLVA profile among sporadic isolates. Sporadic PT4 isolates with similar MLVA profiles (3-[9/10/11/12/13]-9-[4/5]-4-4-1-8-[7/8]) clustered together except the 'type strain', which produced profile 2-13-9-4-4-4-1-8-8. The most common profile among sporadic PT4 isolates was 3-10-9-5-4-4-1-8-8 (n=8). Sporadic PT8 isolates also had similar profiles ([1/2]-[9/10/11]-9-[7/8/10]-4-[3/5]-2-8-9), with the exception of one isolate that produced profile 3-15-8-5-4-4-2-8-10. The most common profile among sporadic PT8 isolates was 1-9-9-7-4-3-2-8-9 (n=5). PT14b isolates were broadly divided into two profiles

depending on whether they were sensitive to all antimicrobials tested (2-[12/13/14]-9-[9/12]-4-3-2-8-[8/9]) or resistant to quinolones (3-[10/11]-9-[5/6]-[4/5]-4-1-8-8); the most common profiles among sporadic isolates were 2-13-9-9-4-4-3-2-8-9 (n=5) and 3-10-9-5-4-4-1-8-8 (n=24), respectively. Null variants (where no amplicon is produced from a locus) were produced for locus SENTER4 in PT26 and locus SE-3 in PTs 9b, 9c, 14, 14b, 14c, 40, 42 and 52.

Outbreak investigations to determine the stability of tandem repeats

All isolates from the outbreaks caused by PT4 and PT8 were indistinguishable by MLVA typing; the MLVA profile assigned to the PT4 outbreak was 3-10-9-5-4-4-1-8-8 and to the PT8 outbreak was 1-9-9-7-4-3-2-8-9 (Table 1). Of the 11 PT14 outbreak isolates, 10 had profile 2-8-7*-4-5-5-NA-8-6 and one isolate was a double locus variant (DLV) assigned profile 2-7-7*-5-5-5-NA-8-6 (where 7* refers to an allele with seven TRs, but TR2 lacks a 6 bp insert and TR6 is missing 21 bp and numbers in bold indicate locus variants). Of the 40 PT14b isolates that were representative of nine of 12 outbreaks (outbreaks A to I) occurring across England and Wales in 2009 associated with a common food source, 16 shared the MLVA profile 3-10-9-5-4-4-1-8-8 [15]. In addition to this profile, a single locus variant (SLV) was seen in one of 10 isolates from outbreak J and was assigned profile 3-10-9-5-4-5-1-8-8. Both isolates selected to represent outbreak K were SLVs assigned profile 3-11-9-5-4-4-1-8-8. Nine of 12 isolates from outbreak L were assigned profile 2-12-9-12-4-3-2-8-9, with the remainder assigned profiles 3-11-9-5-4-4-1-8-8 (two isolates) and 3-10-9-5-4-4-1-8-8 (one isolate).

Discussion

The ability to identify isolates belonging to an outbreak and differentiate them from concurrent sporadic isolates is essential for the investigation of communicable diseases. Without a discriminatory typing method coupled with good classical epidemiological data, it would be extremely difficult to identify the source and route of transmission of infection, thereby making it almost impossible to implement appropriate intervention strategies. This is particularly important for highly clonal bacterial groupings such as *S. Enteritidis*, where the heterogeneity between isolates is limited. However, lack of standardisation of the methodology and interpretive criteria for data analysis has meant that comparison of MLVA data between laboratories can be problematic, thereby hindering attempts at international surveillance and investigation of outbreaks involving more than one country. Given the multinational distribution of some food products, collaboration between countries can be crucial in identifying cases and in tracing the source of infection.

Previously published *S. Enteritidis* MLVA schemes have named profiles based on allele numbers that may or may not reflect the number of TRs at each locus; in

addition, the schemes vary in their treatment of partial TRs. This lack of congruence means not only that data cannot be easily compared between laboratories using different naming schemes, but also that it is difficult to assess the true relationship between isolates exhibiting variation in their MLVA profiles where allele numbers do not accurately reflect the number of TRs. We chose to follow the same principle adopted by MLVA schemes for *S. Typhimurium* and other bacterial pathogens by naming alleles based on the number of TRs at each locus [11,16,17]. We also recommend that only whole TRs are included and that partial TRs are rounded down to the nearest integer to simplify reporting of MLVA data. The length of partial TRs as determined by sequencing was constant; therefore including them in data analysis did not improve the discriminatory capacity of the technique.

Comparison of the published MLVA schemes for *S. Enteritidis* did not identify any further loci that could be added to the scheme of Malorny et al. [8] to increase the discriminatory capacity. In addition, we recommend that loci SENTER1, SENTER2, SENTER3 and SE-7, which harbour the longest TR units in the scheme, are excluded due to the observed variation within TR unit sequence, which could not be detected reliably by capillary electrophoresis. Unless TR regions are sequenced, this variation would be overlooked and strains would be clustered incorrectly. Removal of primers targeting SENTER1 and SENTER2 would also prevent amplification of a 346 bp VIC-labelled fragment, which results from amplification with primers SENTER1-F and SENTER2-R, in a small number of strains (Table 4). Loci SENTER1, SENTER2 and SENTER3 previously showed low Nei's diversity indices (0.07), while SE-7 exhibited a higher diversity index of 0.63 [8]. Removing these four loci before cluster analysis revealed that this had little effect on the number of profiles detected (65 profiles compared with 71 with nine loci) and had no major effect on clustering in the minimal spanning tree (Figure 2). Boxrud et al. also observed that removing low-diversity loci from their analyses had little effect on the discriminatory capacity of MLVA [5]. Longer TR units have been suggested to serve as a molecular clock, leading to the possibility that sequencing of these loci could be used to determine phylogenetic relationships within *S. Enteritidis*, as has been proposed for *Mycobacterium tuberculosis* [5,18]. We also propose that the TR unit sequence at locus SE-3 should be changed to TTTCCATATTG; this TR unit is consistently 12 bp long and 100% conserved in strains carrying multiple copies, unlike the TR unit previously proposed (Table 2) [5,7,8]. The MLVA scheme would therefore consist of five loci (allele string reported as SENTER7-SE-5-SE-6-SE-4-SE-3), which are of consistent length and 100% conserved.

MLVA fragment sizes obtained by capillary electrophoresis frequently differ from the sequenced length due to variations in the sequencer model, size marker and primer fluorophores used. DNA composition of

FIGURE 2

Minimum spanning trees of MLVA of 298 *Salmonella enterica* serovar Enteritidis isolates based on data from (A) nine loci and (B) five loci

A

B

PT14 outbreak
PT8 outbreak
PT4 outbreak
PT14b outbreaks A–J and L (MLVA profile 3-10-9-5-4-4-1-8-8)
PT14b outbreaks K and L (MLVA profile 3-11-9-5-4-4-1-8-8)
PT14b outbreak L (MLVA profile 2-12-9-12-4-3-2-8-9)

MLVA: MLVA: multilocus variable-number tandem-repeat analysis; PT: phage type.

Each circle represents a different MLVA type, with its size proportional to the number of strains with that MLVA profile. Wedges in circles represent the proportion of isolates with a particular MLVA profile that belong to a specific outbreak. Numbers on branches indicate the number of loci that vary between each MLVA profile. Grey shading indicates clonal complexes created based on maximum neighbour distance of changes at one locus and a minimum of two MLVA profiles per complex.

the fragment also plays a part, as demonstrated in this study by locus SE-7 of reference strains HPAo3, HPAo14 and HPAo16, which differed in size by up to 3 bp despite each having nine TRs (data not shown). A panel of 31 *S. Typhimurium* strains with fragment sizes verified by sequencing was compiled by the Statens Serum Institut, Denmark, and made available to allow laboratories with different set-ups to normalise their *S. Typhimurium* MLVA data to the actual fragment sizes [11]. We therefore recommend that laboratories use the set of 16 reference strains described herein (Table 4) to ensure compatibility of *S. Enteritidis* MLVA data between laboratories.

In this study, 71 different MLVA profiles were identified among the 298 strains, indicating that MLVA shows promise as a subtyping method for *S. Enteritidis*. MLVA was capable of subdividing isolates within a phage type, and in most instances multiple isolations of a phage type tended to cluster together by MLVA, as has previously been shown [8,19]. However, isolates of different phage types may also share the same MLVA profile, as was shown here by 41% of isolates sharing profile 3-10-9-5-4-4-1-8-8 despite belonging to 39 different phage types. This is perhaps not surprising considering the two subtyping methods are determining strain diversity using two very different approaches, but highlights the importance of not relying on a single subtyping method and of combining laboratory data with accurate and meaningful epidemiological data when defining relationships between strains. We suggest that a combination of phage typing (where available) and MLVA may be useful for characterisation of *S. Enteritidis* isolates, as has previously been suggested [19].

We were concerned that TRs may evolve so rapidly that multiple types could emerge during the course of an outbreak. Isolates from 15 different outbreaks belonging to four different phage types were subtyped by MLVA to determine stability of the TRs. Previous studies have found that MLVA profiles remain stable during the course of an *S. Enteritidis* outbreak [5,8]. The data presented here suggests that, as with *S. Typhimurium* [20], SLVs may occur sporadically during an outbreak (Figure 2). A DLV was identified among isolates from an outbreak caused by an unusual phage type, PT14, with strong epidemiological evidence to link this isolate to the outbreak. Only 16 isolates of PT14 have been identified since 1981, with the last report of two cases in 1997 (HPA *Salmonella* dataset). This indicated that DLVs may also be detected during an outbreak. Outbreak L was unusual in that two distinct MLVA profiles differing at six of the nine loci were identified, suggesting involvement of two different PT14b strains. This observation was confirmed by the two MLVA profiles belonging to strains with distinct PFGE profiles and exhibiting different antimicrobial resistance phenotypes (data not shown). On the basis of these data, the cut-off to allow classification of *S. Enteritidis* isolates as part of an outbreak could be defined as a difference of one TR

at no more than two loci, with the analysis of more outbreaks needed to confirm this.

In conclusion, we propose an MLVA scheme for *S. Enteritidis* based on five loci (SENTR7, SENTR5, SENTR6, SENTR4 and SE-3) that show little or no variation in sequence length and diversity. A panel of reference strains has been developed that can be used by laboratories to normalise their raw fragment data to actual fragment sizes. Since this study was completed, two novel alleles have been identified at loci SENTR4 and SENTR6. These loci will be sequenced, the strains added to the reference panel and made available to laboratories on request. In addition, we encourage laboratories that have identified novel alleles to send us the strains, to add to the reference panel. We also provide here recommendations for analysing and interpreting data. We urge laboratories to consider implementing these guidelines, thereby allowing direct comparison of data between laboratories irrespective of the platform used for fragment analysis. MLVA profiles identified during outbreaks of *S. Enteritidis* may then be reported via the Epidemic Intelligence Information System (EPIS) of the European Centre for Disease Prevention and Control (ECDC) to public health laboratories.

References

1. Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, et al. The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin Infect Dis*. 2010;50(6):882-9.
2. European Food Safety Authority (EFSA). The community summary report on trends and sources of zoonoses, zoonotic agents, antimicrobial resistance and foodborne outbreaks in the European Union in 2006. *EFSA Journal*. 2007;130. Available from <http://www.efsa.europa.eu/en/scdocs/scdoc/130r.htm>
3. Ward LR, de Sa JD, Rowe B. A phage-typing scheme for *Salmonella enteritidis*. *Epidemiol Infect*. 1987;99(2):291-4.
4. Peters TM, Berghold C, Brown D, Coia J, Dionisi AM, Echeita A, et al. Relationship of pulsed-field profiles with key phage types of *Salmonella enterica* serotype Enteritidis in Europe: results of an international multi-centre study. *Epidemiol Infect*. 2007;135(8):1274-81.
5. Boxrud D, Pederson-Gulrud K, Wotton J, Medus C, Lyszkowicz E, Besser J, et al. 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*. 2007;45(2):536-43.
6. Cho S, Whittam TS, Boxrud DJ, Bartkus JM, Saeed AM. Allele distribution and genetic diversity of VNTR loci in *Salmonella enterica* serotype Enteritidis isolates from different sources. *BMC Microbiol*. 2008;8:146.
7. Cho S, Boxrud DJ, Bartkus JM, Whittam TS, Saeed M. 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*. 2007;275(1):16-23.
8. Malorny B, Junker E, Helmuth R. Multi-locus variable-number tandem repeat analysis for outbreak studies of *Salmonella enterica* serotype Enteritidis. *BMC Microbiol*. 2008;8:84.
9. Beranek A, Mikula C, Rabold P, Arnhold D, Berghold C, Lederer I, et al. Multiple-locus variable-number tandem repeat analysis for subtyping of *Salmonella enterica* subsp. *enterica* serovar Enteritidis. *Int J Med Microbiol*. 2009;299(1):43-51.
10. Ross IL, Heuzenroeder MW. A comparison of two PCR-based typing methods with pulsed-field gel electrophoresis in *Salmonella enterica* serovar Enteritidis. *Int J Med Microbiol*. 2009;299(6):410-20.
11. Larsson JT, Torpdahl M, Petersen RF, Sorensen G, Lindstedt BA, Nielsen EM. Development of a new nomenclature for *Salmonella Typhimurium* multilocus variable number of tandem repeats analysis (MLVA). *Euro Surveill*. 2009;14(15):pii=19174.

Available from: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19174>

12. Gelfand Y, Rodriguez A, Benson G. TRDB--the Tandem Repeats Database. *Nucleic Acids Res.* 2007;35(Database issue):D80-7.
13. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser.* 1999;41:95-8.
14. Lindstedt BA, Vardund T, Aas L, Kapperud G. Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. *J Microbiol Methods.* 2004;59(2):163-72.
15. Health Protection Agency (HPA). S. Enteritidis infections in England in 2009: national case control study report. Health Protection Report. 2010;4:2-3. London: HPA. [Accessed 14 Apr 2011]. Available from: <http://www.hpa.org.uk/hpr/archives/2010/news0610.htm#pt14b>
16. Noller AC, McEllistrem MC, Pacheco AG, Boxrud DJ, Harrison LH. Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic *Escherichia coli* O157:H7 isolates. *J Clin Microbiol.* 2003;41(12):5389-97.
17. Liang SY, Watanabe H, Terajima J, Li CC, Liao JC, Tung SK, et al. Multilocus variable-number tandem-repeat analysis for molecular typing of *Shigella sonnei*. *J Clin Microbiol.* 2007;45(11):3574-80.
18. Spurgiesz RS, Quitugua TN, Smith KL, Schupp J, Palmer EG, Cox RA, et al. Molecular typing of *Mycobacterium tuberculosis* by using nine novel variable-number tandem repeats across the Beijing family and low-copy-number IS6110 isolates. *J Clin Microbiol.* 2003;41(9):4224-30.
19. Cho S, Whittam TS, Boxrud DJ, Bartkus JM, Rankin SC, Wilkins MJ, et al. Use of multiple-locus variable number tandem repeat analysis and phage typing for subtyping of *Salmonella* Enteritidis from sporadic human cases in the United States. *J Appl Microbiol.* 2010;108(3):859-67.
20. Hopkins KL, Maguire C, Best E, Liebana E, Threlfall EJ. Stability of multiple-locus variable-number tandem repeats in *Salmonella enterica* serovar Typhimurium. *J Clin Microbiol.* 2007;45(9):3058-61.