Analysis of *Salmonella enterica* Serovar Typhimurium Variable-Number Tandem-Repeat Data for Public Health Investigation Based on Measured Mutation Rates and Whole-Genome Sequence Comparisons

Karolina Dimovski, a Hanwei Cao, a Odilia L. C. Wijburg, a Richard A. Strugnell, a Radha K. Mantena, a Margaret Whipp, a Geoff Hogg, a Kathryn E. Holt a,c

Microbiological Diagnostic Unit Public Health Laboratory, University of Melbourne, Parkville, Victoria, Australia; Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria, Australia; Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria, Australia

Variable-number tandem repeats (VNTRs) mutate rapidly and can be useful markers for genotyping. While multilocus VNTR analysis (MLVA) is increasingly used in the detection and investigation of food-borne outbreaks caused by *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) and other bacterial pathogens, MLVA data analysis usually relies on simple clustering approaches that may lead to incorrect interpretations. Here, we estimated the rates of copy number change at each of the five loci commonly used for *S. Typhimurium* MLVA, during *in vitro* and *in vivo* passage. We found that loci STTR5, STTR6, and STTR10 changed during passage but STTR3 and STTR9 did not. Relative rates of change were consistent across *in vitro* and *in vivo* growth and could be accurately estimated from diversity measures of natural variation observed during large outbreaks. Using a set of 203 isolates from a series of linked outbreaks and whole-genome sequencing of 12 representative isolates, we assessed the accuracy and utility of several alternative methods for analyzing and interpreting *S. Typhimurium* MLVA data. We show that eBURST analysis was accurate and informative. For construction of MLVA-based trees, a novel distance metric, based on the geometric model of VNTR evolution coupled with locus-specific weights, performed better than the commonly used simple or categorical distance metrics. The data suggest that, for the purpose of identifying potential transmission clusters for further investigation, isolates whose profiles differ at one of the rapidly changing STTR5, STTR6, and STTR10 loci should be collapsed into the same cluster.

Genetic subtyping of bacterial isolates is critical in the investigation of food-borne infections. Multilocus variable-number tandem-repeat (VNTR) analysis (MLVA) is high-resolution typing method that has become widespread among public health laboratories for the investigation of *Salmonella* and other food-borne outbreaks (1–3) (see Fig. S1 in the supplemental material). A range of methods are available for subtyping of *Salmonella* (4), but MLVA is particularly well suited to outbreak investigation as it provides high resolution, is efficient and robust, and generates repeatable results suitable for sharing between laboratories using the same standardized techniques (1, 5–7). An MLVA profile is usually expressed as a string of numbers of length *N*, which represents the number of copies of repeated sequences at each of a set of *N* loci under analysis. For example, the profile 3-4-3-10:12 indicates 3 repeat copies at locus 1, 4 repeat copies at locus 2, and so on. Where repeat copies differ in length, allele codes can be assigned to distinguish specific types of repeat combinations (8).

*Salmonella enterica*, in particular, serovar Typhimurium (*S. Typhimurium*), is among the most common causes of food-borne outbreaks globally. MLVA was first proposed for the analysis of *S. Typhimurium* in 2003 by Lindstedt and colleagues (9), who then developed a multiplex assay targeting the five most variable loci (STTR3, STTR5, STTR6, STTR9, and STTR10) (10). The 5-locus multiplex scheme has since been widely adopted in Europe, Australia, and elsewhere (3, 8, 11). Since repeat lengths in STTR3 can differ, this locus is sometimes expressed using allele codes as presented in reference 8 or as the total size in base pairs. Similar schemes are also in use or development for other *S. enterica* serovars, including Enteritidis, Gallinarum, Heidelberg, Dublin, and Typhi (12–18), and almost a hundred papers reported the use of MLVA for *Salmonella* analysis between 2004 and 2013 (see the supplemental material).

Analysis of MLVA profiles typically relies on distance-based methods for constructing trees or dendrograms to infer the relationships between isolates with different profiles (see, e.g., the PulseNet MLVA protocols [http://www.pulsenetinternational.org/protocols/mlva/] and the supplemental material) or for identifying clusters of closely related isolates (see, e.g., references 11 and 19). Dendrogram or tree construction typically uses either “simple” distance (the number of loci that differ) or “categorical” distance (the number of repeat differences summed across loci) (see, e.g., Table S1 in the supplemental material, which summarizes the MLVA results reported in a cross section of papers from 2012). Cluster identification typically starts by identifying groups

Received 6 May 2014 Accepted 27 May 2014
Published ahead of print 23 June 2014

Address correspondence to Kathryn E. Holt, khol@unimelb.edu.au.

Supplemental material for this article is available at http://dx.doi.org/10.1128/JB.01820-14.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.01820-14
of isolates in a given spatial and temporal range that share identical MLVA profiles (11,19) and sometimes expands to include isolates that differ slightly in their profiles. However, these simplistic approaches share two key problems. First, they treat all loci equally; e.g., a single-copy change at locus 1 is considered equivalent to a single-copy change at locus 2. Yet it is clear from the variations in locus-specific diversity metrics for many organisms, including *Salmonella* (20), and from *in vitro* passage experiments in other organisms (21, 22) that different VNTR loci mutate at different rates. Second, these approaches do not account for the fact that multiple copies of a tandem-repeat sequence can be gained or lost in multiple or single events; e.g., a decrease of 2 copies could be the result of a single deletion event in which 2 copies were lost or the result of 2 separate deletion events, with each affecting 1 copy. Yet *in vitro* passage studies in *Yersinia pestis* and *Escherichia coli* have shown that multicopy changes can occur and that the probability of a change of a given size occurring can be approximated using a geometric model for VNTR evolution (GMVE) (22). The eBURST approach to multi locus analysis (23), developed for multilocus sequence typing (MLST), has also been applied to MLVA data. This method links profiles that share loci, creating chains or complexes of related profiles, but does not attempt to infer relationships beyond this.

It is generally assumed that MLVA profiles (i.e., repeat copy numbers) change rapidly enough to distinguish bacterial clones from one another—e.g., to distinguish outbreak strains from other circulating strains—but not so rapidly that clonal relationships are masked by changes arising during the course of an outbreak. However, the rates at which MLVA profiles change have not been directly investigated for *Salmonella*; hence, it is unclear how *S. Typhimurium* MLVA profiles should be interpreted in the context of outbreak detection and investigation. Parallel serial-passage experiments (PSPE) have been used to directly estimate the rate of VNTR mutations, and the relative rates of single-copy changes, among several pathogenic bacteria, including *Y. pestis, E. coli* O157:H7, and *Burkholderia pseudomallei* (21, 22, 24). This allows GMVE model-based analysis and interpretation of MLVA data for these organisms (22, 24). Some public health laboratories give more weight to selected loci when interpreting *S. Typhimurium* MLVA profiles (personal communications, Microbiological Diagnostic Unit [MDU] staff members); however, this is not based on direct measurement of mutation rates and has not been formalized. The persistence and VNTR stability of *S. Typhimurium* DT41 were recently studied through a series of experiments in chicken flocks in Denmark (25). The study detected a few changes, at loci STTR5, STTR6, and STTR10 but not STTR3 or STTR9, suggesting that the VNTRs used for *S. Typhimurium* MLVA are quite stable but that the levels of stability differ among loci. Similar patterns were also reported during *in vitro* passage of a range of *S. Typhimurium* phage types and for MLVA profiles from Belgian isolates (26). However, mutation rates were not estimated directly in either of these studies and they did not provide enough data to construct a GMVE model for *S. Typhimurium*.

In the present study, we aimed to estimate the different rates of change in copy number at the five loci targeted by the *S. Typhimurium* MLVA scheme, during *in vitro* and *in vivo* passage of a clinical isolate. We then used these rates to assess alternative methods for analyzing *S. Typhimurium* MLVA profiles, including tree construction and cluster identification, using whole-genome sequencing of a small number of representative strains to determine the true underlying relationships.

### MATERIALS AND METHODS

**Bacterial isolates and sequencing.** All *S. Typhimurium* isolates in this study shared a phage type that is a variant of definitive type 135 (DT135). This variant does not yet have an official phage type designation but is commonly referred to as 135a or 135b and is among the most common phage types detected in Australian food-borne outbreaks (11, 27-29). All isolates were sourced from the Microbiological Diagnostic Unit Public Health Laboratory, Victoria. *S. Typhimurium* used in the passage experiments was isolated from a human gastroenteritis patient during a 2005 outbreak in Tasmania, and its MLVA profiles are 2-11-10-10-212 using the European scheme (8) and 3-13-11-11-523 using the Australian nomenclature for the same data. MLVA was performed on a total of 201 outbreak-linked *S. Typhimurium* 135a isolates and on 2 isolates from sporadic cases in Tasmania, as summarized in Table 1 and Fig. 1. *S. Typhimurium* ST5 and 11 other Tasmanian isolates (see Table 1 and Fig. 2) were subjected to whole-genome sequencing on an Illumina HiSeq platform; details of sequencing and analysis are reported elsewhere (30).

**In vitro passage.** Parallel serial-passage experiments (PSPE), modeled on those described in references 21, 22, and 31, were performed using isolate ST5a. Brie 21, the PSPE procedure involved 100 independent clonal lineages subcultured on Luria-Bertani (LB) media (Difco) plates at 24-h intervals for a total of 10 passages. At each stage of each passage, one colony was picked at random and split into two; one half-colony was subcultured to continue the passage, and the other half-colony was suspended in LB-15% glycerol and stored at −80°C for later MLVA. DNA was extracted from all 100 lineages at the final time point (i.e., after 10 passages) and subjected to MLVA testing. Where mutations were identified at the final time point, the half-colonies stored from earlier time points from the same clonal lineage were tested to determine the precise stage of passage during which each mutation arose and to determine whether the observed change in copy number was the result of a single mutational event or multiple mutational events.

The number of generations per 24-h subculture was estimated by picking at random three replicate colonies following 24 h of growth of ST5 on LB media plates, suspending each colony in 1 ml phosphate-buffered saline (Media Preparation Unit, Department of Microbiology and Immunology, University of Melbourne), plating the bacteria using serial dilution, and counting CFU, as described in reference 21. The total number of generations during 24-h *in vitro* passage was therefore equal to the average number of generations per colony × 100 lineages × 10 passages; specifically, 28.6 × 100 × 10 = 28,600 generations.

<p>| TABLE 1 Summary of 203 <em>S. Typhimurium</em> isolates analyzed in this study |
|-----------------------------|-------|--------|-------------------------------|</p>
<table>
<thead>
<tr>
<th>State</th>
<th>Yr</th>
<th>n</th>
<th>Outbreak*</th>
<th>Sequenced isolate(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tasmania</td>
<td>2005</td>
<td>8</td>
<td>2005 (OB1)</td>
<td>STm8</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>79</td>
<td>2005 (OB2)</td>
<td>STm1a, STm2, STm3, STm5</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>42</td>
<td>2005 (OB3)</td>
<td>STm11</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>10</td>
<td>2007 (OB6)</td>
<td>STm9</td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>45</td>
<td>2008 (OB7)</td>
<td>STm6, STm7, STm12</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>1</td>
<td>Sporadic</td>
<td>STm10</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>1</td>
<td>Sporadic</td>
<td>STm1</td>
</tr>
<tr>
<td>Victoria</td>
<td>2002</td>
<td>2</td>
<td>2002 (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>3</td>
<td>2002 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>3</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>5</td>
<td>2008 (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>4</td>
<td>2008 (2)</td>
<td></td>
</tr>
</tbody>
</table>

* OB, outbreak.

* Passaged isolate.
Spleens were harvested from euthanized mice, homogenized, and cultured on xylene lysozyme deoxycholate (XLD) media (Oxoid) for selective isolation of Salmonella. Following overnight culture, three half-colonies were randomly sampled from each homogenate and subjected to DNA extraction and MLVA profiling. One of the three remaining half-colonies was randomly selected and subcultured overnight and was used to inoculate a further five mice in the next round of passage.

Multilocus VNTR analysis. Genomic DNA was extracted from colonies or half-colonies using a QIAxtractor instrument with a DX reagent kit (Qiagen). DNA was subjected to MLVA profiling using a multiplex assay targeting five VNTR loci (10). MLVA profiles were expressed as repeat copy numbers for loci STTR9, STTR8, STTR6, and STTR10, followed by an allele code for STTR3, using the methods and nomenclature described in reference 8. For example, profile 02-11-10-10-10-212 indicates 2 repeat copies for locus STTR9, 11 repeat copies for STTR8, 10 repeat copies for STTR6, 10 repeat copies for STTR10, and allele 212 at locus STTR3. For cluster analysis, which relies on analysis of repeat copy numbers, STTR3 was expressed as the number of repeat copies estimated from the total amplicon size, according to reference 8.

Statistical analysis. All statistical analyses were performed in R (33). Distance-based analysis of MLVA profiles involved the calculation of a pairwise distance metric representing the distance between pairs of strains according to the specified distance metric, followed by tree inference by clustering of the distance matrix using the upgma function of the phangorn R package. All distances were calculated with and without locus weights, with \( w_k \) representing weight for locus \( k \) defined as \( 1/\tau_k \), where \( \tau_k \) represents the mutation rate for locus \( k \) (determined by in vitro passage in this study). Simple pairwise distance determinations were based on the total number of loci at which two MLVA profiles differ, defined as dist \((i,j) = \sum_k (n_{ik} \neq n_{jk}) \times w_k \) where \( k \) represents the locus index and \( n_{ik} \) represents the copy number at locus \( k \) in strain \( i \). Categorical distance incorporates the number of VNTR copies at which two MLVA profiles differ, defined as dist \((i,j) = \sum_k (f(n_{ik} - n_{jk})) \times w_k \) where \( f(n_{ik}, n_{jk}) = 0 \) if \( n_{ik} = n_{jk} \) and \( f(n_{ik}, n_{jk}) = 1 - P \times \frac{P(n_{ik} \neq n_{jk})}{P(n_{ik} \neq n_{jk}) - 1} \) and \( P \) is the rate of single-copy mutational changes as a proportion of all changes (as observed during in vitro passage).

An alternative method, using a combination of maximum parsimony (MP) to infer possible phylogenies and calculation of likelihoods of each possible phylogeny under a GMVE model, was performed as previously described (22) but with a modification to include locus weights. Briefly, maximum-parsimony trees were inferred using the Pauv program and the Wagner parsimony measure for discrete characters. The likelihood for each tree was calculated by inferring all the copy number changes that are required to explain the tree (using the clfigest command in Pauv) and calculating the product of the probabilities of each change, weighted by locus weights, such that \( \text{prob}(tree) = \sum_k (x_k \times \prod (1 - P)^{x_k}) \times w_k \) where \( x_k \) is the copy number in change \( k \) and \( w_k \) is the weight of the locus affected by change \( k \), as defined above.

Results. Estimation of VNTR mutation rates at S. Typhimurium MLVA loci during experimental passage. In order to determine the rates of copy number change at the five VNTR loci targeted by the European S. Typhimurium scheme (8, 10), we performed a series of 100 parallel serial-passage experiments (PSPE) using a representative clinical isolate of S. Typhimurium phage type 1356, STm5. Each clonal lineage was passaged 10 times (see Materials and Methods), resulting in a total of 1,000 24-h subcultures. We calculated the mean number of generations in a colony during 24...
h of passage to be 28.6 generations (95% confidence interval [CI], 28.2 to 29.0), similar to those estimated previously for *Salmonella* (25 generations) (34), *E. coli* O157:H7 (28 generations) (21), and *Y. pestis* (25 generations) (22). Hence, a total of 28,600 (95% CI, 28,200 to 29,000) generations of STm5 growth were observed during 10 rounds of *in vitro* passage in 100 parallel lineages. MLVA profiling of the 100 cultures obtained following 10 rounds of passage identified a total of 10 copy number changes across three of the five loci analyzed, as summarized in Table 2. For each of the passed lineages in which a change was observed, MLVA of subcolonies stored from passages 1 to 9 confirmed that each change was the result of a single mutational event.

We also investigated MLVA mutation rates during *in vivo* passage of STm5 in a murine infection model (see Materials and Methods). We observed a total of four changes in the VNTR copy number at three loci, with patterns similar to *in vitro* change patterns (see Table 2). As we were unable to estimate the number of bacterial generations involved during the course of each murine infection, we compared the *in vitro* and *in vivo* mutation rates on a temporal scale (per-day mutation rates in Table 2). The per-day mutation rates were similar during *in vivo* passage and *in vitro* passage (see Table 2), suggesting that the VNTR loci may be under similar mutational pressures in the *in vivo* and *in vitro* environments.

The mutation rates calculated from *in vitro* passage were inversely correlated with published measures of Simpson’s diversity for the five VNTR loci (10, 20). The evidence for correlation was strongest using Simpson’s diversity estimates obtained from clonal sets of isolates linked to individual outbreaks, i.e., for estimates of diversity that reflect short-term evolution ($R^2 = -0.97$ and $P = 0.005$ versus $R^2 = -0.88$ and $P = 0.05$). This suggests that locus-specific mutation rates per generation could be accurately estimated from the corresponding locus-specific Simpson’s diversity ($D$) measures, using the following formula: rate $= 1 \times 10^{-3}/(1 - D)$. Hence, locus-specific relative mutation rates, or weights ($w$), could be estimated simply as $w = 1/(1 - D)$. We hypothesize that this relationship is likely to apply to other bacteria and thus provides a rational basis for estimating relative locus-specific weights in the absence of experimental data on VNTR mutation rates.

Case study: understanding MLVA profiles during a real outbreak series. *S. Typhimurium* 135@ was responsible for a series of outbreaks in Tasmania, Australia, between 2005 and 2008 (27, 28). A total of 184 *S. Typhimurium* 135@ isolates were collected during these outbreaks and profiled using MLVA. The results of an eBURST analysis of these profiles, plus results of analyses of a further 2 sporadic cases and 17 isolates from outbreaks in the nearby state of Victoria (total $n = 203$; Table 1), are shown in
<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene</th>
<th>Gene function</th>
<th>Repeat length(s) (bp)</th>
<th>No. of changes observed*</th>
<th>Mutations rate</th>
<th>Diversity$^6$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>In vitro</td>
<td>In vivo</td>
<td>Per generation,</td>
<td>Per day,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>in vitro</td>
<td>in vitro,</td>
</tr>
<tr>
<td>STTR9</td>
<td>Intergenic</td>
<td>Upstream of mannitol dehydrogenase</td>
<td>9</td>
<td>0</td>
<td>&lt;3.5E-3</td>
<td>&lt;1.0E-3</td>
</tr>
<tr>
<td>STTR5</td>
<td>yohM</td>
<td>Nickel-cobalt efflux protein</td>
<td>6</td>
<td>2 (+1, +1)</td>
<td>7.00E-03</td>
<td>2.00E-03</td>
</tr>
<tr>
<td>STTR6</td>
<td>Intergenic</td>
<td>Phage related</td>
<td>6</td>
<td>4 (+1, +1, +1, -1, -2)</td>
<td>1 (-3)</td>
<td>1.40E-04</td>
</tr>
<tr>
<td>STTR10</td>
<td>Intergenic</td>
<td>pSLT plasmid</td>
<td>6</td>
<td>4 (+1, +1, +1, +1, -1)</td>
<td>1 (-2)</td>
<td>1.40E-04</td>
</tr>
<tr>
<td>STTR3</td>
<td>bigA</td>
<td>Surface-exposed virulence protein</td>
<td>27/33</td>
<td>0</td>
<td>&lt;3.5E-5</td>
<td>&lt;1.0E-3</td>
</tr>
</tbody>
</table>

* Total numbers of observed changes are given; sizes and directions of changes are given in parentheses (e.g., "+1" indicates a 1-copy increase).

$^6$ Diversity measures are Simpson's diversity values; "Outbreak" refers to diversity measures within sets of isolates epidemiologically defined as belonging to the same outbreak; "General" refers to large isolate collections not limited to outbreaks.
these VNTR loci (STTR5, STTR6, and STTR10) are indicative of a substantially longer evolutionary time than a single locus change would be and that the changes are equivalent to 50 to 100 chromosomal SNPs (i.e., the number of SNPs occurring in 10 to 30 years).

**Locus-specific mutation rates and the GMVE model for tree construction from MLVA data.** We compared the abilities of several alternative analysis methods to recover the underlying phylogenetic relationships depicted in Fig. 2 directly from MLVA profiles of the full set of 203 S. Typhimurium 135 isolates from Tasmania and Victoria (Table 1). The eBURST analysis of the data has already been shown (Fig. 1), and the additional methods considered were as follows: (i) maximum-parsimony reconstruction coupled with GMVE likelihood estimation, as proposed in references 22, 24, and 35, and (ii) unweighted-pair group method using average linkages (UPGMA) clustering using (a) a novel distance metric based on the GMVE model, (b) simple clustering, and (c) categorical clustering. Each method was considered using both unweighted analysis and locus-specific weights based on the locus-specific mutation rates estimated from the passage experiments (Table 2) (see Materials and Methods). The resulting tree topologies were assessed with respect to their ability to (i) differentiate the Tasmanian outbreak clone from other isolates (including sporadic Tasmanian cases and Victorian outbreaks) and (ii) differentiate the Tasmanian outbreak clone isolates from different outbreaks into separate clusters.

Maximum-parsimony (MP) analysis of the MLVA profiles identified 156 MP tree topologies, of which 18 alternative tree topologies shared maximum likelihood under the GMVE model. The majority-rules consensus of these 18 trees did not differentiate the outbreak clone from the Victorian isolates; however, incorporating locus-specific mutation rates into the GMVE likelihood calculations produced better results, resulting in six maximum-likelihood topologies that separated the outbreak clone from the Victorian outbreaks (Fig. 3A).

Clustering using GMVE distances and locus-specific weights correctly separated the outbreak clone from other isolates and also clustered profiles from the same outbreaks (Fig. 3B). This method performed better than unweighted simple or categorical clustering (Fig. 3C and D), which broadly grouped the outbreak clone...
members but did not cluster profiles from the same outbreak. The inclusion of locus-specific weights into simple or categorical clustering did not improve the performance of these metrics. Hence, the data suggest that improvements in the construction of trees from MLVA profiles can be most easily obtained using the GMVE distance coupled with locus-specific weights.

**Identifying clusters from MLVA profiles.** Some public health and surveillance laboratories use regular reviews of S. Typhimurium MLVA profiles to detect potential transmission clusters that may warrant further investigation. Cluster definitions in use include (i) identical profiles, (ii) profiles with ≤1 or a specified number of locus differences, and (iii) more-complex rules about which VNTR loci may differ and the scale of differences. The WGS tree (Fig. 2) shows that a cluster definition based on identical profiles would fail to cluster isolates that are otherwise genetically identical and clearly part of the same transmission chain. For example, S1a3 and S1m5, food and case isolates from the second 2005 outbreak that differ by a single SNP and were collected 6 days apart, differed by one repeat copy at locus STTR6. A more relaxed approach to cluster identification which allows copy number change at one of the three most rapidly changing loci (STTR5, STTR6, and STTR10; see Table 2) would identify clusters consistent with the WGS data. In contrast, in our WGS data set, copy number changes at >1 of the rapidly mutating loci were associated with the much more distant phylogenetic relationships (>75 SNPs, ~15 years of independent evolution).

To further explore the effects of different cluster identification rules, we compared retrospective cluster definitions based on MLVA data for the 203 1350 isolates from Australian outbreaks to outbreak definitions based on epidemiological investigations conducted at the time (Table 1). Using a cluster definition based on identical profiles, only 68% of isolates would have been correctly included in their epidemiologically determined outbreak clusters. Allowing one (or two) differences at one of the three rapidly mutating loci would result in 91% (or 98%) of outbreak isolates being correctly assigned to epidemiologically determined clusters.

**DISCUSSION**

The stability of the MLVA loci through 1,000 rounds of *in vitro* passage suggests that the likelihood of mutations being introduced during regular laboratory manipulation (usually only 1 to 2 subcultures) is very low. That is, MLVA assays as currently performed can be expected to give an accurate indication of the true VNTR status of the bacteria at the point of isolation, with very low risk of observed differences being the result of mutations arising in the laboratory. Furthermore, the similarity in relative rates of change observed between loci during *in vitro* passage and *in vivo* passage, as well as real-world variation during outbreaks, is consistent with the hypothesis that VNTR loci evolve at similar rates under a variety of growth conditions.

These results are also consistent with recent studies of VNTR stability that examined a range of S. Typhimurium strains, albeit through fewer passages. In S. Typhimurium DT41, occasional changes were identified during passage experiments performed *in vivo* (in chickens) and *in vitro* (8 passage chains with daily subculture on blood agar for 100 to 200 days) (25). The changes occurred mostly in STTR6, followed by STTR10 and STTR5. A second study performed 50 12-h passages in LB broth, using 31 S. Typhimurium isolates with different MLVA profiles and a range of phage types (DT104, DT193, DT120, DT195, U302, and DT12) (26).

Changes were observed in STTR6 (*n* = 6 changes), STTR10 (*n* = 4), and STTR5 (*n* = 2), and a single change was also detected in STTR9. It was not possible to calculate rates per generation in the reported studies, which did not directly track clonal lineages. However, there is clear agreement across all these studies regarding the order of instability among the S. Typhimurium MLVA loci, namely, STTR6 > STTR10 > STTR5 > STTR9 and STTR3. This suggests that our results and the mutation rates are likely to be broadly applicable to different S. Typhimurium phage types and strains. However, neither the current study nor the others reported to date were sufficiently powered to detect and quantify subtle differences in per-unit-of-time mutation rates under different growth conditions, and a larger study incorporating much longer total passage times (and ideally incorporating a wider range of *in vitro* and *in vivo* growth conditions) would be required to prove the absolute equivalence of mutation rates under different conditions or in different genetic backgrounds.

The present report reveals clear trends that have immediate utility for deciding how best to handle MLVA data analysis of S. Typhimurium and bacterial pathogens more generally. Using WGS data and detailed epidemiological trace-back methods to determine the true underlying genetic relationships behind a series of outbreaks, our analysis shows that accurate conclusions about this case study could be drawn from MLVA data alone. Similarly, a recent study comparing MLVA and WGS for *analysis of Clostridium difficile* transmission showed that MLVA was as informative as WGS, which is considered the gold standard for transmission tracking (36).

With respect to cluster detection for identifying outbreaks and ruling out unrelated strains, our data indicate that the best cluster definition is one that includes isolates with variation in none or one of the rapidly changing loci STTR5, STTR6, and STTR10 but that excludes strains with more than one difference or with any difference in locus STTR3 or STTR9. This definition is supported by the only other available study of VNTR stability analysis in S. Typhimurium, which also did not observe changes at STTR3 or STTR9 during environmental or *in vivo* passage (25), and by Simpson’s diversity measures for the five VNTR loci calculated among the members of our outbreak set and an independent collection of S. Typhimurium outbreak isolates, as summarized in Table 2 (20).

Finally, our data indicate that, for the purpose of tree construction based on MLVA profiles, a weighted GMVE model is superior to the simple or categorical distances that have been most frequently used in published studies of MLVA data. This is consistent with the expectation that a distance based on the GMVE that attempts to account for the mutational processes which generate variation in MLVA profiles will lead to more-accurate inferences of genetic relationships between isolates. Standard eBURST analysis, which has rarely been used in published studies reporting MLVA data, provided an accurate, if incomplete, representation of the relationships between profiles and does not require prior knowledge of any mutational parameters.

**ACKNOWLEDGMENTS**

This work was supported by an Early Career Researcher grant from the University of Melbourne, a grant from the government of Victoria, Australia, and a Victorian Life Sciences Computation Initiative (VLSCI) (no. VR0082). K.E.H. and O.L.C.W. were supported by fellowships from the National Health and Medical Research Council (NHMRC) of Australia.
REFERENCES


(No. 628930 and no. 454774) and I.C.C. by a NIMRC program grant (no. 606788). The Microbiological Diagnostic Unit (MDU) Public Health Laboratory is funded by the Department of Health Victoria.

We thank the MDU staff members and the staff members of all laboratories who sent isolates for typing.


Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Dimovski, K; Cao, H; Wijburg, OLC; Strugnell, RA; Mantena, RK; Whipp, M; Hogg, G; Holt, KE

Title:
Analysis of Salmonella enterica Serovar Typhimurium Variable-Number Tandem-Repeat Data for Public Health Investigation Based on Measured Mutation Rates and Whole-Genome Sequence Comparisons

Date:
2014-08-01

Citation:
Dimovski, K; Cao, H; Wijburg, OLC; Strugnell, RA; Mantena, RK; Whipp, M; Hogg, G; Holt, KE, Analysis of Salmonella enterica Serovar Typhimurium Variable-Number Tandem-Repeat Data for Public Health Investigation Based on Measured Mutation Rates and Whole-Genome Sequence Comparisons, JOURNAL OF BACTERIOLOGY, 2014, 196 (16), pp. 3036 - 3044

Persistent Link:
http://hdl.handle.net/11343/57537

File Description:
Accepted version