

Development and Evaluation of a Commercial Sequence-Based Strain Typing Service for *Listeria monocytogenes*

Tom Edlind¹ and Yanhong Liu^{2*}

¹MicrobiType LLC, 5110 Campus Drive, Plymouth Meeting, Pennsylvania, USA

²Molecular Characterization of Foodborne Pathogens Research Unit, Eastern Regional Research Center, Agricultural Research Service, U.S. Dept. of Agriculture, Pennsylvania, USA

Abstract

Listeria monocytogenes is an important foodborne pathogen and, relatedly, a persistent contaminant in many food processing facilities. Strain typing is critical to detecting and investigating outbreaks, and can be used to track down sources of contamination. The typing systems in current use have limitations ranging from low resolution and data portability to high cost and technical complexity. The aim of this study was to develop an outsourcing option for *L. monocytogenes* strain typing that addresses these limitations. The NCBI Genomes database representing 109 strains was screened for highly informative, tandem repeat-containing loci. The most promising, termed LmMT1 (0.8-1 kbp) and LmMT2 (0.7-0.8 kbp), exhibited complex patterns of polymorphism (insertions/deletions and single nucleotide polymorphisms), diversity indices of 0.99 (LmMT1) and 0.98 (LmMT2), and were present in all *L. monocytogenes* and from one (LmMT1) to four (LmMT2) additional *Listeria* species represented in NCBI databases. Using a distinct set of strains, Miya et al. (J. Microbiol. Methods, 2012, 90:285-291) previously reported diversity indices of 0.95 and 0.91 for more limited regions (0.3-0.5 kbp) of these same two loci. Phylogenetic analysis of LmMT1 and LmMT2 sequences revealed distinct clusters corresponding to serotype (4b, 1/2a, and 4a complexes) and evolutionary lineage (I, II, and III/IV). Comparisons to PFGE, the current gold standard, suggest that LmMT1 typing is comparably discriminatory. Importantly, strains from four outbreaks formed corresponding clusters, although those from a 2002 outbreak were resolved into related environmental and food/human isolates that challenges their epidemiological connection. In the laboratory, LmMT1 and LmMT2 typing proved to be robust, generating high quality sequence from colonies submitted as non-hazardous heat-inactivated suspensions. Analysis of LmMT1 sequences from 62 diverse strains demonstrated overall agreement with single nucleotide polymorphism-based typing.

Keywords: *Listeria*; Foodborne outbreak; Genotype; Phylogenetic tree; PFGE; MLVA

Abbreviations: PFGE: Pulsed-Field Gel Electrophoresis; MLVA: Multilocus Variable Number of Tandem Repeats Analysis; MLST: Multilocus Sequence Typing; WGS: Whole Genome Sequencing; MLGT: Multilocus Genotyping; Indels: Insertions And Deletions; Snp: Single Nucleotide Polymorphisms

Introduction

Listeria monocytogenes is an environmentally ubiquitous bacterium capable of contaminating various foods, particularly dairy products and processed meats, but also smoked fish and raw fruits and vegetables [1,2]. Contamination is aided by its ability to grow at low temperatures and its tolerance to freezing, drying, and heat; long term persistence in food processing facilities has been well documented [3-6]. Upon ingestion, *L. monocytogenes* can cause severe infection, particularly in immunocompromised individuals, the elderly, pregnant women, and neonates [7]. Its virulence derives from its abilities to survive and grow within host phagocytes and aggressively spread through tissues, mediated by internalin proteins on its surface and induction of host actin polymerization [8,9]. Although most cases of food-borne listeriosis are sporadic, there have been multiple outbreaks in recent years, including one in 2011 responsible for 147 infections and 33 deaths across 28 states that was traced to contaminated cantaloupe (www.cdc.gov/listeria/outbreaks).

Tracking down the source of food-borne listeriosis is a challenging task that begins with laboratory culture (enrichment followed by plating) and species identification (to distinguish *L. monocytogenes* from its non-pathogenic sister species). This may be followed by conventional serotyping, although its limitations including low

resolution (serotypes 4b, 1/2a, and 1/2b predominate) and high cost have largely led to its replacement with molecular typing systems. Among the latter, the gold standard is pulsed-field gel electrophoresis (PFGE) comparing fragment lengths following restriction enzyme digestion of chromosomal DNA [10]. Its primary advantage is high resolution, but its limitations include lengthy turnaround time, technical complexity, high cost, and the inherently low portability of length data, although the latter has been addressed through strict standardization and pattern analysis algorithms as implemented in the PulseNet system [11]. The limitations of serotyping and PFGE have encouraged the development of numerous alternative *L. monocytogenes* typing methods [12-14]. The most promising and practical of these is multilocus variable number of tandem repeats analysis (MLVA), first adapted to *L. monocytogenes* by Murphy et al. [15] and subsequently by other laboratories [16-21]. MLVA typically employs multiplex PCR and capillary electrophoresis to assess length variation due to insertions and deletions (indels) in 9 or so loci that contain tandem repeats, which are typically polymorphic due to slippage between repeat units during DNA replication [22].

***Corresponding authors:** Yanhong Liu, Molecular Characterization of Foodborne Pathogens Research Unit, Eastern Regional Research Center, Agricultural Research Service, U.S. Dept. of Agriculture, Pennsylvania, USA, Tel: 215-233-6587; Fax: 215-233-6581; E-mail: yanhong.liu@usda.ars.gov

Received September 22, 2015; **Accepted** October 05, 2015; **Published** October 12, 2015

Citation: Edlind T, Liu Y (2015) Development and Evaluation of a Commercial Sequence-Based Strain Typing Service for *Listeria monocytogenes*. J Microb Biochem Technol 7: 351-362. doi:10.4172/1948-5948.1000238

Copyright: © 2015 Edlind T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Additional length-based methods include ribotyping [23], rep-PCR [24,25], and amplified fragment length polymorphism [26,27].

The inherently higher portability of DNA sequence-based methods, facilitating day-to-day and lab-to-lab comparisons, is exploited by methods including multilocus sequence typing (MLST) analyzing 7 loci [28], and the related multi-virulence-locus sequence typing analyzing 6 loci [29]. These methods analyze single nucleotide polymorphisms (SNPs) within relatively conserved loci, and hence are most useful for analyzing evolutionary trends rather than epidemiology. Furthermore, their multilocus requirement adds to their complexity and cost. Two additional methods representing variations on sequence analysis include multilocus genotyping (MLGT) which employs >100 probes and flow cytometry to detect the presence or absence of previously documented SNPs [30], and a microarray which detects the presence or absence of >100 genes previously shown to vary among strains [31]. Development of both of these methods hinged on genomics; i.e., the availability in recent years of whole genome sequences (WGS) for large numbers of diverse *L. monocytogenes* strains. Indeed, WGS analysis itself is being explored as a direct approach to *L. monocytogenes* outbreak detection and investigation [32-35]. The WGS sequencing technologies commonly employed generate large numbers of short reads which are subsequently analyzed for SNPs, analogous to MLST and MVLST but providing considerably higher resolution. Unfortunately, short reads preclude the unambiguous sequence assembly of some tandem repeat regions. A general concern involves the substantial investments in equipment, reagents, and trained personnel required to implement MLGT, microarrays, and WGS as routine typing tools.

A more practical alternative to the multilocus or WGS approaches described above would be sequence-based typing of one (or two) highly polymorphic loci. Current representatives of this approach include *Staphylococcus aureus spa* typing [36], *Streptococcus pyogenes emm* typing [37], *Neisseria meningitidis porA* and *fetA* typing [38], and *Campylobacter jejuni flaA* typing [39]. These loci were selected largely based on their previously characterized roles as immunodominant antigens or virulence factors in these four pathogens. Identifying highly informative loci for sequence-based typing of other bacterial pathogens might benefit from a less biased approach. For example, Miya et al. [40] identified several promising loci for sequence-based typing of *L. monocytogenes* based on MLVA data generated in their lab.

The approach used here to identify promising loci for sequence-based typing exploits the current availability of numerous genome sequences, and analogous to MLVA focuses on tandem repeats as major contributors to polymorphism. In addition to indels, it was anticipated that these tandem repeat loci, consistent with their length polymorphism, would also exhibit higher rates of SNPs within flanking regions and within the repeats themselves.

The goal was to develop and evaluate a cost-effective but uncompromising outsourcing option for *L. monocytogenes* strain typing that could be used by food industries to monitor and track down sources of contamination, and by public health and clinical labs for surveillance and outbreak detection and investigation. In addition to identifying promising loci, it was critical to develop a simple and safe procedure for sample submission, robust approaches to sequencing these samples, and a format for typing results that emphasizes interpretability and utility.

Materials and Methods

Bioinformatics

Tandem repeats in complete genome sequences were identified

using the Tandem Repeats Database (TRDB; <https://tandem.bu.edu>). BLASTN searches were conducted on the NCBI website (www.ncbi.nlm.nih.gov) against the NCBI Genomes database, and as needed against the WGS and Nucleotide collection databases. Downloaded sequences were trimmed to common termini, and aligned with clustalw2 (www.ebi.ac.uk/Tools/msa/clustalw2) in PHYLIP format using default parameters. Alignments were analyzed using dnapsars (DNA parsimony), and dendrograms constructed using drawgram, both from the PHYLIP package (<http://evolution.genetics.washington.edu/phylip.html>). Simpson's index of diversity was calculated using the formula $D=1-(\sum n(n-1)/N(N-1))$, where n=number of strains with a given allele and N=number of (epidemiologically unrelated) strains. PCR-based serogroups were determined by BLASTN searches for serogroup-specific genes [41]. For strains lacking experimentally determined PFGE profiles but having complete genome sequences, profiles were modelled in silico (<http://insilico.ehu.es/digest/index.php?mo=Listeria>).

Strains and culture conditions

A diverse set of 60 *L. monocytogenes* strains from the NRRL collection, including representatives of each lineage and serogroup, were obtained from T. Ward (USDA-ARS, Peoria, IL). Strains 1875 and 1877 are laboratory-generated deletion derivatives of F2365 [42], and ATCC strain 19115 was obtained from J. Brewster (USDA-ARS, Wyndmoor, PA). Strains were cultured on brain heart infusion agar at 37°C. As per website guidelines (www.microbitype.com/submission-guidelines), isolated colonies were suspended in provided buffer-containing tubes, bacteria were heat inactivated by incubating at 100°C for 15 min, and tubes were transported to MicrobiType (Plymouth Meeting, PA) by overnight courier at ambient temperature.

PCR and sequence analysis

Proprietary primers for PCR and sequencing were designed based on conserved regions identified by clustalw2 analyses of LmMT1 and LmMT2 loci, and synthesized by IDT (Coralville, Iowa). DNA was purified from heat-inactivated lysates, amplified with Taq polymerase, and subjected to Sanger dideoxynucleotide sequencing using proprietary modifications of standard methods. Chromatograms were visually scanned, and sequences edited as needed. All new LmMT1 sequences generated in this study have been submitted to GenBank with accession numbers KT626013-KT626044.

Results

Identification of candidate typing loci LmMT1 and LmMT2

Tandem repeats were identified within the complete *L. monocytogenes* genome sequences for serotype 4b strain J1776 (GenBank accession CP006598) and serotype 1/2a strain EGD (accession HG421741) using TRDB (<https://tandem.bu.edu>). The repeats plus 500 nucleotides upstream and downstream flanking sequence were used as queries in BLASTN searches of the NCBI Genomes (Organism: *Listeria monocytogenes*) database which, as of March 2015, included 109 strains. Each repeat-containing region was evaluated for degree of polymorphism, presence in all or nearly all strains, and total length <1000 nucleotides to permit coverage by a single sequencing reaction. Two promising candidates were identified that satisfied these criteria.

The first candidate, LmMT1, involves a 0.8-1 kbp region encoding a putative internalin (gene *lmo1136* in strain EGD) with Pro-Val-Asp repeats (25 to 107 residues total). Surveys of the MLVA literature

and SNPs. For LmMT1, a total of 53 alleles (i.e., unique sequences) were resolved from analysis of the 109 strains in the NCBI Genomes database, as shown in the dendrogram (Figure 2). Note, however, that 37 of these 109 strains are annotated as “retail deli” isolates (e.g.,

R8-7914), collected from specific sites at multiple times over a 1 year period, as recently described [6]. These 37 R8 isolates define only 6 total LmMT1 alleles, and 26 of the 37 define only 2 alleles, consistent with their characterization as “persistent” and the predominance of

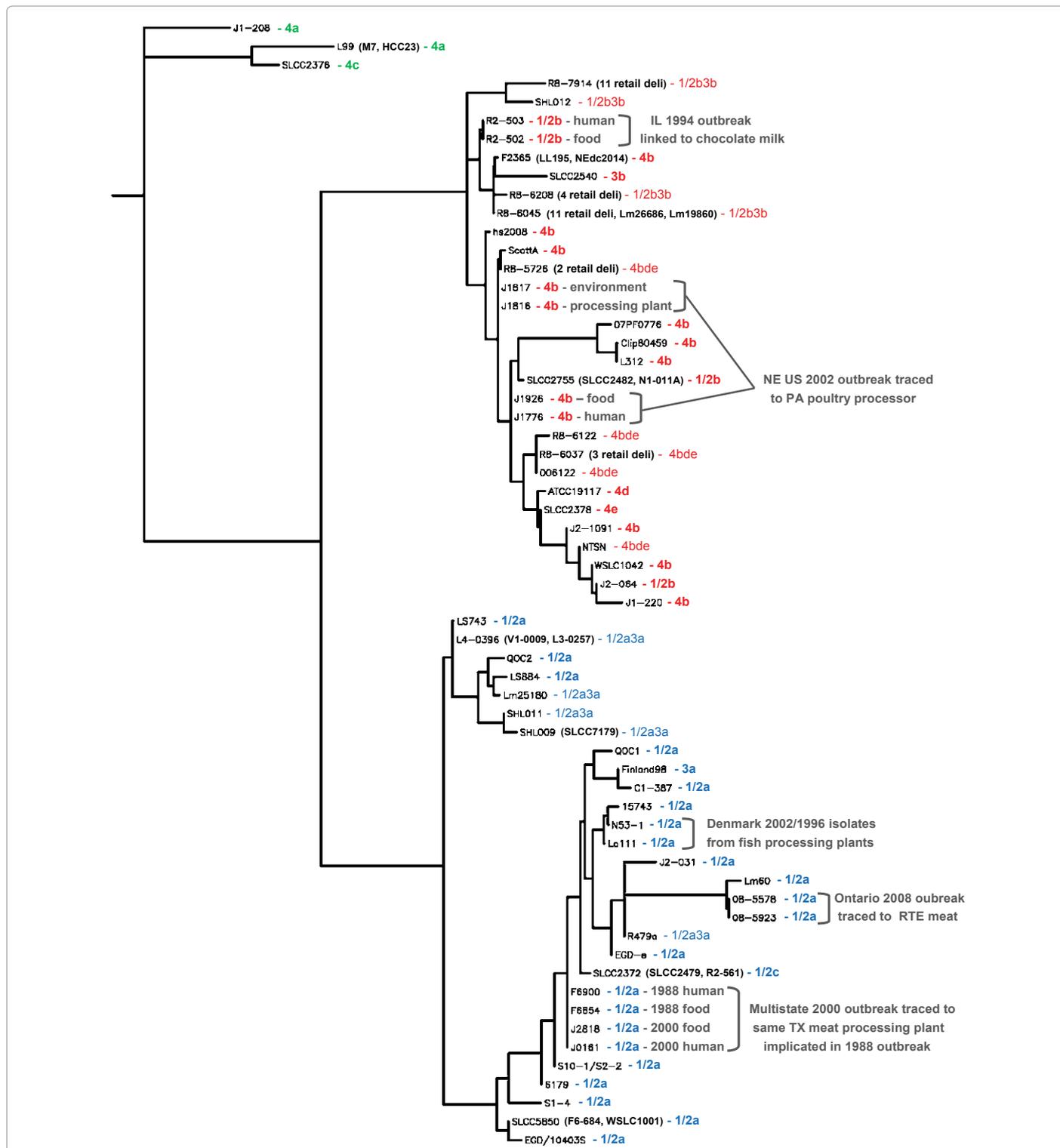
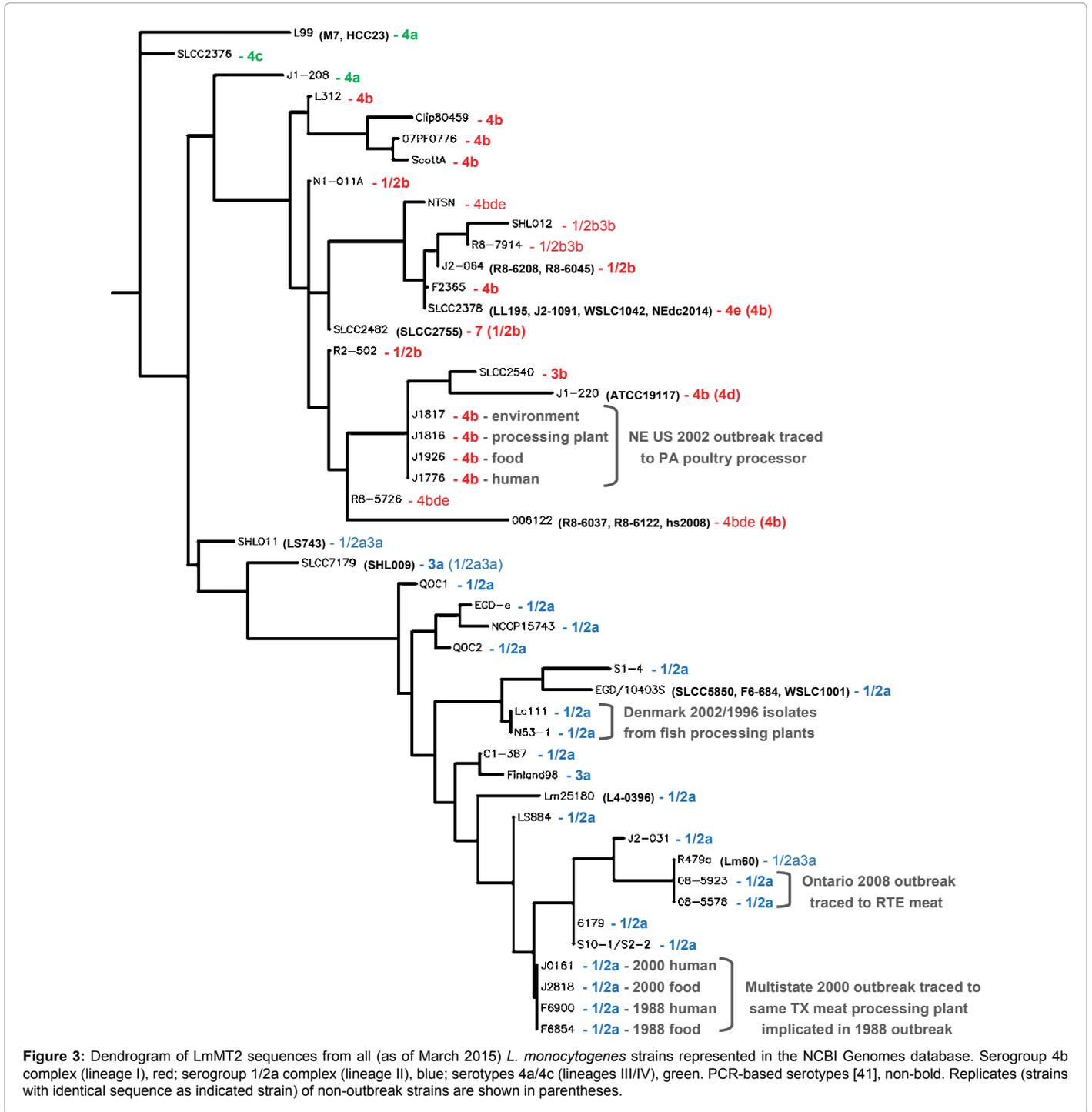


Figure 2: Dendrogram of LmMT1 sequences from all (as of March 2015) *L. monocytogenes* strains represented in the NCBI Genomes database. Serogroup 4b complex (lineage I), red; serogroup 1/2a complex (lineage II), blue; serogroup 4a complex (lineage III), green. PCR-based serotypes [41], non-bold. Replicates (strains with identical sequence as indicated strain, or their numbers in the case of retail deli isolates) of non-outbreak strains are shown in parentheses.

certain PFGE profiles [6]. In the dendrogram, strains that share the same LmMT1 allele are represented by one strain (except for outbreak strains; see below), and the replicate strains (or their number in the case of R8 strains) are shown in parentheses. Based on their NCBI annotations and literature searches, it was estimated that 70 of the 109 *L. monocytogenes* strains in the NCBI Genomes database have no known epidemiological connection (for each retail deli isolate allele, one representative was included). The distribution of these 70 strains into the 53 alleles yielded a Simpson's index of diversity of 0.99.

For LmMT2, phylogenetic analysis (Figure 3) resolved 38 alleles from the full-length sequences for 69 strains represented in the NCBI Genomes database. (The reduced number – 69 strains compared to 109 for LmMT1 – is largely due to reduced representation of R8 strains). Again, based on their annotations and literature searches it was estimated that 59 of these strains are epidemiologically unrelated. This value combined with 38 alleles yielded a Simpson's index of diversity of 0.98. Consistent with this slightly lower value, several sets of strains that were resolved with LmMT1 (Figure 2) were clustered with LmMT2



(Figure 3); notably, the two pairs of “NE US 2002 outbreak” strains and the two “Denmark 2002/1996” strains. The “Ontario 2008” and “Multistate 2000” outbreak strains clustered with both LmMT1 and LmMT2. (complete LmMT2 sequence was available for only one of the “IL 1994” outbreak strains).

Although their intended use is for epidemiology rather than evolutionary analysis, it is noteworthy that both LmMT1 and LmMT2 dendrograms divide the strains into three distinct groups representing lineages I (serotype 4b complex, including 4b, 4d, and 4e), II (serotype 1/2a complex, including 1/2a, 1/2c, 3a, 3c), and III/IV (serotypes 4a, 4c), as previously described [30,43].

LmMT1 typing of epidemiologically related strains with comparison to PFGE

In light of its higher diversity index and ability to resolve closely related strains as noted above, LmMT1 was selected for further analysis. Below it is compared to PFGE with respect to four well characterized outbreaks:

(1) Identical LmMT1 sequences, and unique relative to all other strains in NCBI Genomes and WGS databases, were obtained for four serotype 1/2a strains: two from a 1988 outbreak (F6900, F6854) and two from a 2000 outbreak (J0161, J2818) traced to the same Texas meat processor [5]. These strains are also indistinguishable by PFGE [5], and nearly identical based on whole genome SNP analysis [34]. In addition to confirming their epidemiological connection, this lack of variation over 12 years is remarkable, and demonstrates the stability of LmMT1 typing.

(2) The two serotype 1/2b strains (R2-502 and R2-503) from the 1994 Illinois outbreak traced to chocolate milk [44] have identical, and unique, LmMT1 sequences. All strains associated with this outbreak analyzed by PFGE also had PFGE profiles that were indistinguishable or nearly so [44].

(3) The two serotype 1/2a strains (08-5578, 08-5923) from the 2008 Ontario outbreak traced to ready-to-eat meat have identical, unique LmMT1 sequences. By PFGE, these two strains are nearly indistinguishable, differing by a single AscI band which was shown by genome sequencing to be due to prophage insertion, a recognized source of PFGE profile variation in otherwise related strains [32].

(4) Four serotype 4b strains associated with the 2002 outbreak in northeastern U.S. traced to a Pennsylvania poultry processor [45] are indistinguishable by PFGE and MVLST [46], and also have identical LmMT2 sequences (Figure 3). They are, however, resolved by LmMT1 (Figure 2) into two closely related pairs - human/food strains J1776/J1926 and processing plant/environment strains J1816/J1817 - differing by a 9 bp indel within the repeat region (Figure 4). This split is supported by analysis of additional indel-based polymorphic loci (Figure 4). Furthermore, although both pairs have unique LmMT1 sequences relative to all other NCBI Genomes and WGS strains, the J1816/J1817 pair differs only by two SNPs from recent retail deli isolates such as R8-5726 (Figure 2). These data cast doubt on the presumed connection of environmental strains J1816 and J1817 to the 2002 outbreak [47].

Comparisons between LmMT1 and PFGE for epidemiologically unrelated strains

Comparisons between LmMT1 (Figure 2) and available (experimental or in silico) PFGE typing results placed epidemiologically unrelated strains into one of three groups:

(1) Strains that are indistinguishable by both LmMT1 and PFGE include serotype 4b strains F2365 and LL195; serotype 1/2c or 3c strains SLCC2372, SLCC2479, and R2-561; and serotype 4a strains L99, M7, and HCC23.

(2) Strains that are indistinguishable by LmMT1 but resolved by PFGE include serotype 1/2a strains EGD and 10403S; and serotype 1/2b strains SLCC2755, SLCC2482, and N1-011A.

```

LmMT1
J1776 661-TCTACCGGATCTACCGGGTAATCGGATTAACAGGCTTCTCCAACAGTGGCTGCGTAATT
J1926 TCTACCGGATCTACCGGGTAATCGGATTAACAGGCTTCTCCAACAGTGGCTGCGTAATT
J1816 TCTACCGG-----GGTAATCGGATTAACAGGCTTCTCCAACAGTGGCTGCGTAATT
J1817 TCTACCGG-----GGTAATCGGATTAACAGGCTTCTCCAACAGTGGCTGCGTAATT
*****

J1776 879611-GTAAACCACCAGCTACGGTTGAGTGGGA---GTAAACAAAGCCCTAAAACCTCAGTAATG
J1926 GTAAACCACCAGCTACGGTTGAGTGGGA---GTAAACAAAGCCCTAAAACCTCAGTAATG
J1816 GTAAACCACCAGCTACGGTTGAGTGGGAATAGTAAACAAAGCCCTAAAACCTCAGTAATG
J1817 GTAAACCACCAGCTACGGTTGAGTGGGAATAGTAAACAAAGCCCTAAAACCTCAGTAATG
*****

J1776 1613369-AGCGTCGGCATCCGCATCAGCGTC-----GGCATCCGCATCAGCGTCGGCGTC
J1926 AGCGTCGGCATCCGCATCAGCGTC-----GGCATCCGCATCAGCGTCGGCGTC
J1816 AGCGTCGGCATCCGCATCAGCGTCGGCATCCGCATCAGCGTCGGCGTC
J1817 AGCGTCGGCATCCGCATCAGCGTCGGCATCCGCATCAGCGTCGGCGTC
*****

J1776 2137732-ATCAAACGTAATAGTGTATGGAGTAATAGTGTATGGAGTATTCGTATTTGATGCTGAACC
J1926 ATCAAACGTAATAGTGTATGGAGTAATAGTGTATGGAGTATTCGTATTTGATGCTGAACC
J1816 ATCAAACGTAATAGTGTATGGAGTA-----TTCGTATTTGATGCTGAACC
J1817 ATCAAACGTAATAGTGTATGGAGTA-----TTCGTATTTGATGCTGAACC
*****

```

Figure 4: LmMT1 and three additional polymorphic loci resolve the 2002 outbreak isolate pair J1776 (human) and J1926 (food) from the J1816 (food processing plant) and J1817 (environment) isolate pair. Nucleotide locations in the J1776 genome (GenBank accession NC_021839) are indicated. Polymorphisms were identified by BLASTN searches of the NCBI Genomes database with the J1776 genome (in 100 kbp increments), and screening for indels.

(3) Strains that are resolved by LmMT1 but indistinguishable by PFGE include serotype 4b strains L312 and Clip80459, and serotype 1/2a strains La111 and N53-1.

Typing of non-monocytogenes *Listeria* species

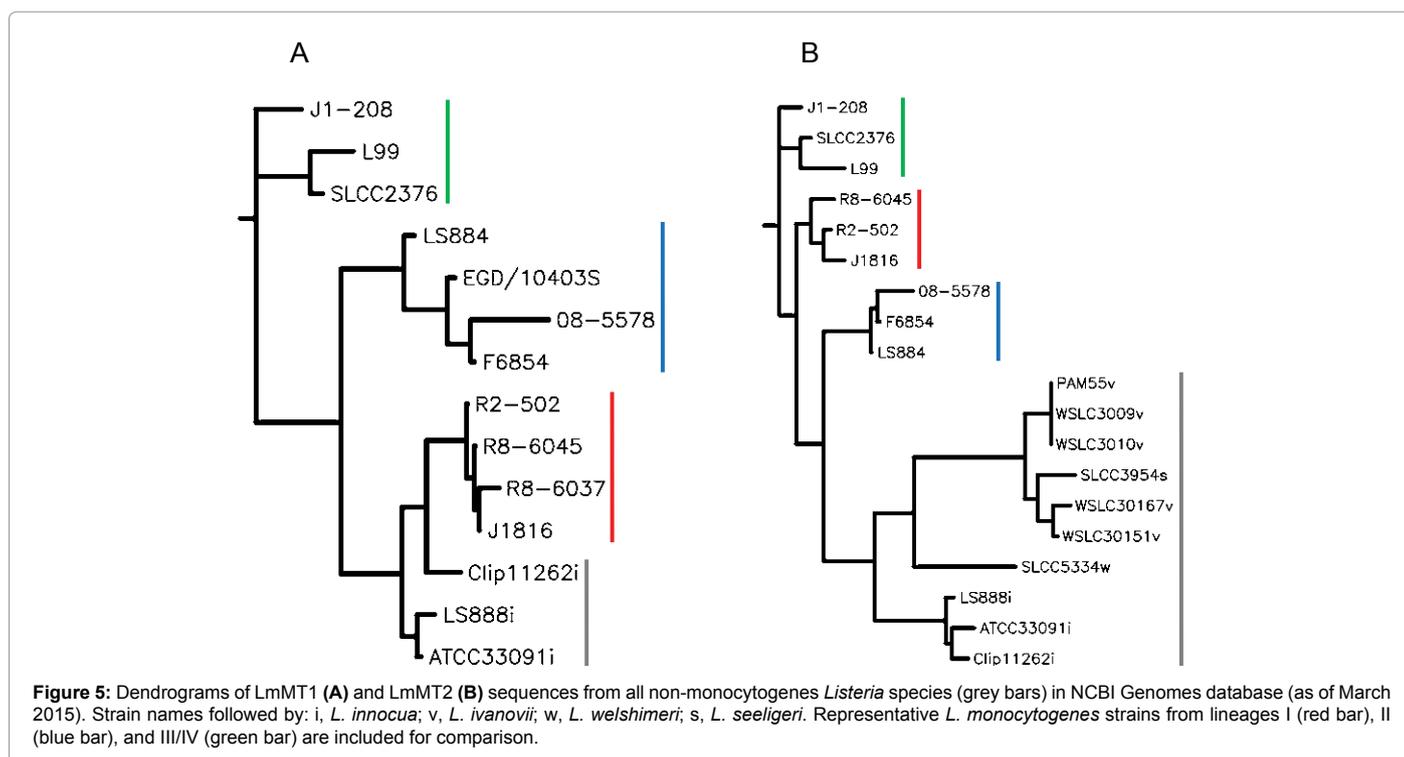
While human pathogenesis is limited to *L. monocytogenes*, *Listeria ivanovii* is a pathogen of ruminants; additional species are non-pathogenic but often co-isolate with, and must be differentiated from, *L. monocytogenes*. Thus, it would be advantageous if LmMT1 or LmMT2 typing clearly distinguished between these species. Eleven strains of four non-monocytogenes *Listeria* species are represented in the NCBI Genomes database (as of March 2015). Among these, the LmMT1 locus is present in only *Listeria innocua*, where it resolves all three strains (Figure 5A). The LmMT2 locus, on the other hand, is present in all 11 non-monocytogenes strains, 10 of which have complete sequences and are resolved into 8 alleles (Figure 5B). For both LmMT1 and LmMT2, the non-monocytogenes strains are clearly resolved from *L. monocytogenes* lineages I, II, and III/IV strains.

Laboratory evaluation

In addition to identifying promising loci, the development of a practical outsourcing option for strain typing relies on a simple and safe procedure for sample submission and robust approaches to sequencing these samples. Non-thermophilic and non-spore-forming bacteria are effectively inactivated by heating to 100°C, and in this form do not require expensive and cumbersome biohazard packaging and shipping. To test the compatibility of heat inactivation with LmMT1 and LmMT2 typing, isolated colonies from *L. monocytogenes* strains representing lineages I (serotypes 1/2b and 4b), II (1/2a and 1/2c), and III (4a) were prepared, submitted, and analyzed as described in Materials and Methods. The resulting chromatograms and sequences (representative results in Figure 6) were of uniformly high quality for both LmMT1 and LmMT2.

Using this protocol, LmMT1 analysis was extended to an additional 56 *L. monocytogenes* strains and 1 *L. innocua* strain from the USDA-ARS collection, and the resulting dendrogram is shown in Figure 7. Of the 62 total strains, 34 defined 32 new alleles (i.e., unique LmMT1 sequence). The remaining 28 had LmMT1 sequences identical to 12 NCBI database strains (underlined in Figure 7). These include: strain F2365 and its laboratory-generated derivatives 1875 and 1877 which matched database F2365, strain ScottA which matched database Scott A, strain 33419 (derived from J0161) which matched database F6854 (epidemiologically related to J0161 as discussed above), and strain 33233 (derived from H7858) which matched database H7858. All of these matches were expected, and demonstrate that routine passaging does not alter LmMT1 sequence. The 59 epidemiologically unrelated strains (excluding *L. innocua*, 1875, and 1877) defined 44 LmMT1 alleles, and yielded a Simpson's diversity index=0.98.

As observed above with NCBI strains, the USDA-ARS strains formed lineage-specific clusters that correspond fully to serogroup (with the exception of two atypical 4b strains in lineage III/IV). PFGE data were not available for these strains, but comparisons of LmMT1 typing to MLGT, which analyzes >100 SNPs using a flow cytometry-based approach [30], demonstrated overall agreement. Specifically, all MLGT-defined lineage I strains (i.e., 1.1 to 1.9) are also LmMT1 lineage I (Figure 4). Similarly, all MLGT-defined lineage II strains (2.1 to 2.9) are also LmMT1 lineage II, and all MLGT-defined lineages III/IV strains (3.15 to 4.4) are also LmMT1 lineages III/IV (including the two atypical serotype 4b strains). Considering strain clusters with identical LmMT1 sequence, 5 of 6 similarly include strains with identical MLGT type (e.g., the three serotype 1/2b strains with MLGT type 1.29). MLGT does, however, resolve a number of strains within several LmMT1 clusters (e.g., three serotype 1/2b strains with MLGT types 1.26, 1.28, and 1.49). Conversely, two MLGT 1.18 strains and 5 of 7 MLGT 2.10 strains were resolved by LmMT1 (Figure 7) due to a combination of indels and SNPs (Figure 8).



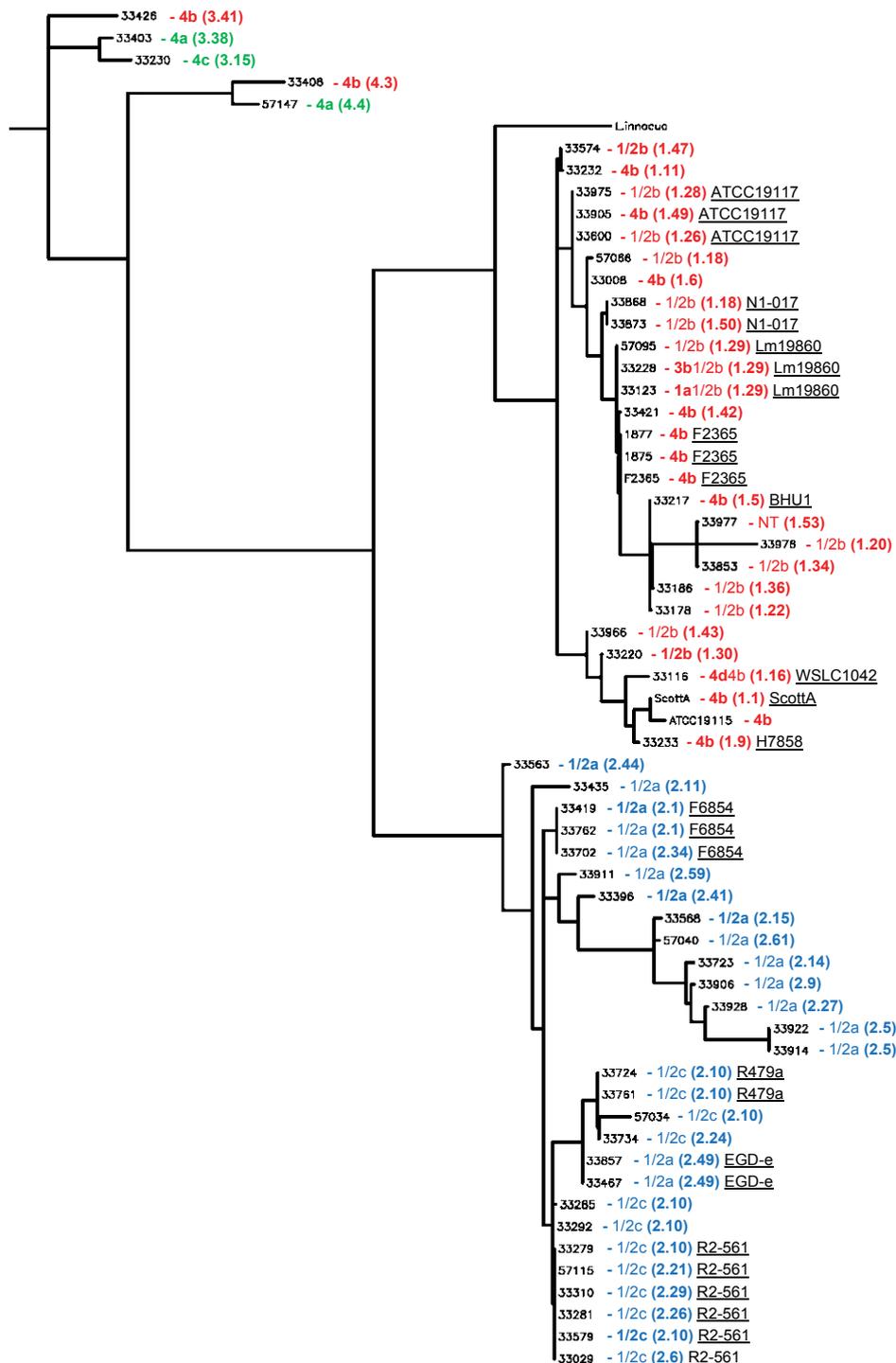


Figure 7: Dendrogram of LmMT1 sequences generated from heat-inactivated colonies of 62 strains in the USDA-ARS collection. Serogroup 4b complex (lineage I), red; serogroup 1/2a complex (lineage II), blue; serotypes 4a/4c/atypical 4b (lineages III/IV), green. PCR-based serotypes [41], non-bold; MLGT type [30], parentheses; identical match to NCBI Genomes or WGS database strain, underlined.

Discussion

The goal of this study was to develop and validate a commercial sequence-based strain typing service for the foodborne pathogen *L. monocytogenes* that meets the following requirements: (1) resolution comparable to PFGE, the current gold standard; (2) clustering

consistent with epidemiological relatedness, evolutionary lineage, and serotype; (3) data portability to facilitate day-to-day and lab-to-lab comparisons; (4) readily interpretable results that reference public domain databases; (5) simple and safe sample submission; (6) turn-around time of 2 to 3 days; and (6) affordable cost. To meet this goal, a

A

```
57066-1/2b (1.18) TCTACCGGATCTACCGGATCTACCGGGGTAATCGGATTAACAGGCTTCTCCAACAGTGGC 660
33868-1/2b (1.18) TCTACCGGATCTACCGG-----GGTAATCGGATTAACAGGCTTCTCCAACAGTGGC 651
*****
```

B

```
57034-1/2c (2.10) TCTACCGGATCTACCGGATCTACCGGATCTACCGGATCTACTGGATCTACCGGATCTACC 660
33579-1/2c (2.10) TCTACCGGATCTACCGGATCTACCGGATCTACCGGAT----- 637
33724-1/2c (2.10) TCTACCGGATCTACCGGATCTACCGGATCTACCGGAT-----CTACC 642
33292-1/2c (2.10) TCTACCGGATCTACCGGATCTACCGGATCTACCGGAT----- 637
33285-1/2c (2.10) TCTACTGGATCTACCGGATCTACCGGATCTACCGGAT----- 637
*****

57034-1/2c (2.10) GGATCTACCGGATCTACCGGATCTACTGGATCTACCGGATCTACCGGATCTACCGGATCT 720
33579-1/2c (2.10) -----CTACTGGATCTACCGGATCTACCGGATCTACCGGATCT 675
33724-1/2c (2.10) GGATCTACCGGATCTACCGGATCTACTGGATCTACCGGATCTACCGGATCTACCGGATCT 702
33292-1/2c (2.10) -----CTACTGGATCTACCGGATCTACCGGATCTACCGGATCT 675
33285-1/2c (2.10) -----CTACTGGATCTACCGGATCTACCGGATCTACCGGATCT 675
*****

57034-1/2c (2.10) ACTGGAGTGACCGAATCAACAGGCTTCTCTAACAAAGGCTGCGTAATTTTACCAGAATAC 780
33579-1/2c (2.10) ACTGGAGTGACCGAATCAACAGGCTTCTCTAACAAAGGCTGCGTAATTTTACCAGAATAC 735
33724-1/2c (2.10) ACTGGAGTGACCGGATCAACAGGCTTCTCTAACAAAGGCTGCGTAATTTTACCAGAATAC 762
33292-1/2c (2.10) ACTGGAGTGACCGGATCAACAGGCTTCTCTAACAAAGGCTGCGTAATTTTACCAGAATAC 735
33285-1/2c (2.10) ACTGGAGTGACCGGATCAACAGGCTTCTCTAACAAAGGCTGCGTAATTTTACCAGAATAC 735
*****
```

Figure 8: ClustalW2 alignments of LmMT1 polymorphic regions from (A) two serotype 1/2b strains with equivalent MLGT type 1.18, and (B) five serotype 1/2c strains with equivalent MLGT type 2.10.

genomics-based approach was employed which led to the identification of LmMT1 and LmMT2. These tandem repeat-containing loci were present in all *L. monocytogenes* genomes, and could be amplified and sequenced by robust, inexpensive methods that are compatible with samples submitted as heat-inactivated colonies. Both demonstrate complex patterns of strain-dependent indels and SNPs that can be readily compared, using standard bioinformatics tools, to publicly available NCBI sequence databases currently representing >300 total strains. Importantly, LmMT1 sequence analysis was demonstrated here to provide strain resolution comparable to PFGE, and hence sufficient for outbreak detection and investigation in the public health sector and tracking down contamination in the food processing sector. Furthermore, LmMT1 and LmMT2 include the more limited TR1 and TR2 regions, respectively, previously shown by Miya et al. [40] using a distinct set of strains to exceed the resolution obtained by virulence gene-based MLST. Finally, the PCR-based technology behind LmMT1 and LmMT2 typing provides the potential to be used with unpurified and mixed samples, while technologies such as PFGE and whole genome sequencing rely on pure cultures. This is highly relevant in light of the trend toward culture-independent diagnostic methods [48].

Two approaches are currently used to summarize and share typing data. Dendrograms (e.g., Figure 2) are the most informative, as they can reveal epidemiological connections between isolates while also providing evolutionary perspective. A second, space-spacing approach is to apply a type designation in a nomenclature format that, unfortunately, is often cryptic and uninformative (i.e., a new type is arbitrarily assigned the next highest number). We propose a combination of these two approaches for sequence-based single locus

typing systems such as LmMT1, where identical matches to NCBI database strains can be readily determined by BLASTN search. For example, USDA-ARS strains 33868 and 33873 are designated type LmMT1:N1-017 since they share their LmMT1 sequence with database strain N1-017 (Figure 7). (If there are multiple strains with that sequence in the database, as is the case for N1-017, the first one deposited is given priority). On the other hand, strain 57066 (and 32 additional USDA-ARS strains) had a unique LmMT1 sequence (deposited in NCBI with accession numbers KT626013-KT626044), and hereafter strains with matching sequence will be designated type LmMT1:57066. These designations are intrinsically informative, since the strain name (e.g., N1-017) can be used to search sequence, literature, and internet databases. Of course, the utility of this type designation system would be enhanced by consistent and comprehensive annotation of NCBI sequence files.

The development and validation of microbial typing systems has traditionally relied on strain collections that were sufficiently large and diverse [16,18,20,31]. However, access to these collections can be limited, and their laboratory analysis is costly, time consuming, and prone to experimental error. In contrast, the development and validation of LmMT1 and LmMT2 relied on publicly available genome sequences, whose number and diversity have expanded considerably in recent years. Indeed, one justification for this expansion in genome sequencing has been its potential application to strain typing. Of course, genome sequences, analyzed exclusively for SNPs, can and have been applied directly to investigations of foodborne outbreaks [35,49-54]. These applications were primarily retrospective, however, and this limitation will likely remain since the true costs for genome

sequencing (equipment, reagents, and personnel) remain high. As demonstrated here, conventional, inexpensive sequence analysis of a single, maximally informative genomic locus such as LmMT1 that combines indels and SNPs can provide strain resolution comparable to PFGE, which has proven to be more than adequate in nearly all cases of outbreak detection and investigation.

References

1. Cartwright EJ, Jackson KA, Johnson SD, Graves LM, Silk BJ, et al. (2013) Listeriosis outbreaks and associated food vehicles, United States, 1998-2008. *Emerg Infect Dis* 19: 1-9.
2. Centers for Disease Control and Prevention (CDC) (2013) Vital signs: Listeria illnesses, deaths, and outbreaks—United States, 2009-2011. *MMWR Morb Mortal Wkly Rep* 62: 448-452.
3. Carpentier B, Cerf O (2011) Review—Persistence of *Listeria monocytogenes* in food industry equipment and premises. *Int J Food Microbiol* 145: 1-8.
4. Ferreira V, Wiedmann M, Teixeira P, Stasiewicz MJ (2014) *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. *J Food Prot* 77: 150-170.
5. Olsen SJ, Patrick M, Hunter SB, Reddy V, Kornstein L, et al. (2005) Multistate outbreak of *Listeria monocytogenes* infection linked to delicatessen turkey meat. *Clin Infect Dis* 40: 962-967.
6. Simmons C, Stasiewicz MJ, Wright E, Warchocki S, Roof S, et al. (2014) *Listeria monocytogenes* and *Listeria* spp. contamination patterns in retail delicatessen establishments in three U.S. states. *J Food Prot* 77: 1929-1939.
7. Allerberger F, Wagner M (2010) Listeriosis: a resurgent foodborne infection. *Clin Microbiol Infect* 16: 16-23.
8. Camejo A, Carvalho F, Reis O, Leitão E, Sousa S, et al. (2011) The arsenal of virulence factors deployed by *Listeria monocytogenes* to promote its cell infection cycle. *Virulence* 2: 379-394.
9. Reddy S, Lawrence ML (2014) Virulence characterization of *Listeria monocytogenes*. *Methods Mol Biol* 1157: 157-165.
10. Graves LM, Swaminathan B (2001) PulseNet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. *Int J Food Microbiol* 65: 55-62.
11. Gerner-Smidt P, Hise K, Kincaid J, Hunter S, Rolando S, et al. (2006) PulseNet USA: a five-year update. *Foodborne Pathog Dis* 3: 9-19.
12. Hyytiä-Trees EK, Cooper K, Ribot EM, Gerner-Smidt P (2007) Recent developments and future prospects in subtyping of foodborne bacterial pathogens. *Future Microbiol* 2: 175-185.
13. Nightingale K (2010) *Listeria monocytogenes*: knowledge gained through DNA sequence-based subtyping, implications, and future considerations. *J AOAC Int* 93: 1275-1286.
14. Datta AR, Laksanalamai P, Solomotis M (2013) Recent developments in molecular sub-typing of *Listeria monocytogenes*. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* 30: 1437-1445.
15. Murphy M, Corcoran D, Buckley JF, O'Mahony M, Whyte P, et al. (2007) Development and application of Multiple-Locus Variable number of tandem repeat Analysis (MLVA) to subtype a collection of *Listeria monocytogenes*. *Int J Food Microbiol* 115: 187-194.
16. Chenal-Francoise V, Diancourt L, Cantinelli T, Passet V, Tran-Hykes C, et al. (2013) Optimized multilocus variable-number tandem-repeat analysis assay and its complementarity with pulsed-field gel electrophoresis and multilocus sequence typing for *Listeria monocytogenes* clone identification and surveillance. *J Clin Microbiol* 51: 1868-1880.
17. Li X, Huang B, Eglezos S, Graham T, Blair B, et al. (2013) Identification of an optimized panel of variable number tandem-repeat (VNTR) loci for *Listeria monocytogenes* typing. *Diagn Microbiol Infect Dis* 75: 203-206.
18. Lindstedt BA, Tham W, Danielsson-Tham ML, Vardund T, Helmersson S, et al. (2008) Multiple-locus variable-number tandem-repeats analysis of *Listeria monocytogenes* using multicolour capillary electrophoresis and comparison with pulsed-field gel electrophoresis typing. *J. Microbiol Methods* 72: 141-148.
19. Miya S, Kimura B, Sato M, Takahashi H, Ishikawa T, et al. (2008) Development of a multilocus variable-number of tandem repeat typing method for *Listeria monocytogenes* serotype 4b strains. *Int J Food Microbiol* 124: 239-249.
20. Saleh-Lakha S, Allen VG, Li J, Pagotto F, Odumeru J, et al. (2013) Subtyping of a large collection of historical *Listeria monocytogenes* strains from Ontario, Canada, by an improved multilocus variable-number tandem-repeat analysis (MLVA). *Appl Environ Microbiol* 79: 6472-6480.
21. Sperry KE, Kathariou S, Edwards JS, Wolf LA (2008) Multiple-locus variable-number tandem-repeat analysis as a tool for subtyping *Listeria monocytogenes* strains. *J Clin Microbiol* 46: 1435-1450.
22. Zhou K, Aertsen A, Michiels CW (2014) The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiol Rev* 38: 119-141.
23. De Cesare A, Bruce JL, Dambaugh TR, Guerzoni ME, Wiedmann M (2001) Automated ribotyping using different enzymes to improve discrimination of *Listeria monocytogenes* isolates, with a particular focus on serotype 4b strains. *J Clin Microbiol* 39: 3002-3005.
24. Van Kessel JS, Karns JS, Gorski L, Perdue ML (2005) Subtyping *Listeria monocytogenes* from bulk tank milk using automated repetitive element-based PCR. *J Food Prot* 68: 2707-2712.
25. Zunabovic M, Domig KJ, Pichler I, Kneifel W (2012) Monitoring transmission routes of *Listeria* spp. in smoked salmon production with repetitive element sequence-based PCR techniques. *J Food Prot* 75: 504-511.
26. Fonnesbech Vogel B, Fussing V, Ojeniyi B, Gram L, Ahrens P (2004) High-resolution genotyping of *Listeria monocytogenes* by fluorescent amplified fragment length polymorphism analysis compared to pulsed-field gel electrophoresis, random amplified polymorphic DNA analysis, ribotyping, and PCR-restriction fragment length polymorphism analysis. *J Food Prot* 67: 1656-1665.
27. Roussel S, Félix B, Grant K, Dao TT, Brisabois A, Amar C (2013) Fluorescence amplified fragment length polymorphism compared to pulsed field gel electrophoresis for *Listeria monocytogenes* subtyping. *BMC Microbiol* 13: 14.
28. Stessl B, Ruckerl I, Wagner M (2014) Multilocus sequence typing (MLST) of *Listeria monocytogenes*. *Methods Mol Biol* 1157: 73-83.
29. Chen Y, Zhang W, Knabel SJ (2007) Multi-virulence-locus sequence typing identifies single nucleotide polymorphisms which differentiate epidemic clones and outbreak strains of *Listeria monocytogenes*. *J Clin Microbiol* 45: 835-846.
30. Ward TJ, Ducey TF, Usgaard T, Dunn KA, Bielawski JP (2008) Multilocus genotyping assays for single nucleotide polymorphism-based subtyping of *Listeria monocytogenes* isolates. *Appl Environ Microbiol* 74: 7629-7642.
31. Laksanalamai P, Jackson SA, Mammel MK, Datta AR (2012) High density microarray analysis reveals new insights into genetic footprints of *Listeria monocytogenes* strains involved in listeriosis outbreaks. *PLoS One* 7: e32896.
32. Gilmour MW, Graham M, Van Domselaar G, Tyler S, Kent H, et al. (2010) High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. *BMC Genomics* 11: 120.
33. Holch A, Webb K, Lukjancenko O, Ussery D, Rosenthal BM, Gram L (2013) Genome sequencing identifies two nearly unchanged strains of persistent *Listeria monocytogenes* isolated at two different fish processing plants sampled 6 years apart. *Appl Environ Microbiol* 79: 2944-2951.
34. Orsi RH, Borowsky ML, Lauer P, Young SK, Nusbaum C, et al. (2008) Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment. *BMC Genomics* 9: 539.
35. Schmid D, Allerberger F, Huhulescu S, Pietzka A, Amar C, et al. (2014) Whole genome sequencing as a tool to investigate a cluster of seven cases of listeriosis in Austria and Germany, 2011-2013. *Clin Microbiol Infect* 20: 431-436.
36. Malachowa N, Sabat A, Gniadkowski M, Krzyszton-Russjan J, Empel J, et al. (2005) Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-field gel electrophoresis, spa typing, and multilocus sequence typing for clonal characterization of *Staphylococcus aureus* isolates. *J Clin Microbiol* 43: 3095-3100.
37. Beall B, Facklam R, Thompson T (1996) Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci. *J Clin Microbiol* 34: 953-958.
38. Jolley KA, Brehony C, Maiden MC (2007) Molecular typing of meningococci: recommendations for target choice and nomenclature. *FEMS Microbiol Rev* 31: 89-96.

39. Kittl S, Heckel G, Korczak BM, Kuhnert P (2013) Source attribution of human *Campylobacter* isolates by MLST and *fla*-typing and association of genotypes with quinolone resistance. PLoS One 8: e81796.
40. Miya S, Takahashi H, Kamimura C, Nakagawa M, Kuda T, Kimura B (2012) Highly discriminatory typing method for *Listeria monocytogenes* using polymorphic tandem repeat regions. J Microbiol Methods 90: 285–291.
41. Doumith M, Buchrieser C, Glaser P, Jacquet C, Martin P (2004) Differentiation of the major *Listeria monocytogenes* serovars by multiplex PCR. J Clin Microbiol 42: 3819-3822.
42. Liu Y, Ceruso M, Gunther NW, Pepe T, Cortesi ML, Fratamico P (2012) Construction of *Listeria monocytogenes* mutants with in-frame deletions in putative ATP-Binding Cassette (ABC) transporters and analysis of their growth under stress conditions. J Microb Biochem Technol 4: 141-146.
43. Orsi RH, den Bakker HC, Wiedmann M (2011) *Listeria monocytogenes* lineages: Genomics, evolution, ecology, and phenotypic characteristics. Int J Med Microbiol 301: 79-96.
44. Dalton CB, Austin CC, Sobel J, Hayes PS, Bibb WF, et al. (1997) An outbreak of gastroenteritis and fever due to *Listeria monocytogenes* in milk. N Engl J Med 336: 100-105.
45. Gottlieb SL, Newbern EC, Griffin PM, Graves LM, Hoekstra RM, et al. (2006) Multistate outbreak of Listeriosis linked to turkey deli meat and subsequent changes in US regulatory policy. Clin Infect Dis 42: 29-36.
46. Chen Y, Zhang W, Knabel SJ (2005) Multi-virulence-locus sequence typing clarifies epidemiology of recent listeriosis outbreaks in the United States. J Clin Microbiol 43: 5291-5294.
47. Chen Y, Strain EA, Allard M, Brown EW (2011) Genome sequences of *Listeria monocytogenes* strains J1816 and J1-220, associated with human outbreaks. J Bacteriol 193: 3424-3425.
48. Cronquist AB, Mody RK, Atkinson R, Besser J, Tobin D'Angelo M, et al. (2012) Impacts of culture-independent diagnostic practices on public health surveillance for bacterial enteric pathogens. Clin Infect Dis 54 Suppl 5: S432-S439.
49. Ashton PM, Peters T, Ameh L, McAleer R, Petrie S, et al. (2015) Whole Genome Sequencing for the Retrospective Investigation of an Outbreak of *Salmonella typhimurium* DT 8. PLoS Curr 7.
50. den Bakker HC, Switt AI, Cummings CA, Hoelzer K, Degoricija L, et al. (2011) A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. *enterica* serovar Montevideo pulsed-field gel electrophoresis type. Appl Environ Microbiol 77: 8648-8655.
51. Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, et al. (2011) Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med 364: 981-982.
52. Octavia S, Wang Q, Tanaka MM, Kaur S, Sintchenko V, et al. (2015) Delineating community outbreaks of *Salmonella enterica* serovar *typhimurium* using whole genome sequencing: insights into genomic variability within an outbreak. J Clin Microbiol 53: 1063-1071.
53. Turabelidze G, Lawrence SJ, Gao H, Sodergren E, Weinstock GM, et al. (2013) Precise dissection of an *Escherichia coli* O157:H7 outbreak by single nucleotide polymorphism analysis. J Clin Microbiol 51: 3950-3954.
54. Underwood AP, Dallman T, Thomson NR, Williams M, Harker K, et al. (2013) Public health value of next-generation DNA sequencing of enterohemorrhagic *Escherichia coli* isolates from an outbreak. J Clin Microbiol 51: 232-237.