Use of multi-locus sequencing typing as identification method for the food-borne pathogen *Listeria monocytogenes*: a review

Sonia Lamon, Domenico Meloni, Simonetta Gianna Consolati, Anna Mureddu, Rina Mazzette

Department of Veterinary Medicine, University of Sassari, Italy

Abstract

*Listeria monocytogenes* is an ubiquitous, intracellular pathogen which has been implicated within the past decade as the causative organism in several outbreaks of foodborne diseases. In this review, a new approach to molecular typing primarily designed for global epidemiology has been described: multi-locus sequencing typing (MLST). This approach is novel, in that it uses data that allow the unambiguous characterization of bacterial strains via the Internet. Our aim is to present the currently available selection of references on *Listeria monocytogenes* MLST detection methods and to discuss its use as gold standard to *Listeria monocytogenes* subtyping method.

Introduction

*Listeria monocytogenes*

The gram-positive bacterium *Listeria monocytogenes* is an ubiquitous, intracellular pathogen which has been implicated within the past decade as the causative organism in several outbreaks of foodborne diseases. It was described originally by Murray et al. and Pirie. At present, the genus *Listeria* includes fifteen species, *i.e.* L. monocytogenes, L. ivanovii, L. innocua, L. welshimeri, L. seeligeri, L. grayi, L. marthii, L. rocourtiae, L. fleichmannii, L. weihenstephanensis, L. floridensis, L. aquatilis, L. cornellensis, L. riparia and L. grandensis. Members of the genus *Listeria* are small rods, ranging between 0.5 and 4 μm in diameter and between 0.5 and 2 μm in length, non-spor-forming and facultative anaerobic. Peritrichous flagella give them a typical tumbling motility at room temperature (20-25°C). The different species of the genus *Listeria* have long been known to be genetically diverse and they can be distinguished on the basis of a few simple biochemical tests.

Listeriosis

Out of the fifteen species of the genus *Listeria*, *L. monocytogenes* is the major pathogen for humans. Very rare cases of infections associated with *L. ivanovii* and *L. seeligeri* have been described. Listeriosis is the bacterial infection caused by *L. monocytogenes* and two forms have been described in humans. Symptoms vary from febrile gastroenteritis in healthy people to life-threatening invasive infections, septicemia and meningocerebralitis in young, old, pregnant and immune-compromised (YOPI). In Europe, it is the fourth most common zoonotic disease and it has an annual incidence of 0.41 cases per 100,000 population. The highest notification rates were reported in persons aged 65 years and above and in persons aged below one year. Listeriosis has the highest hospitalization rate cases of all zoonoses in the EU (91.6%) and is the third leading cause of death in the EU, with an estimated case fatality rate of 17.8%. Evolution of the identification methods for bacterial pathogens

At present, epidemiological investigations use strain-typing procedures such as multi-locus enzyme electrophoresis (MLEE), pulsed-field gel electrophoresis (PFGE), random amplified polymorphic differences, and ribotyping. The characterization of pathogenic isolates plays a pivotal role in the information necessary to identify and track disease outbreaks. These most widely used molecular typing methods, are highly discriminatory and have added greatly to our knowledge on the epidemiology of many pathogenic bacteria in outbreak settings. These methods rely on uncharacterized genomic differences between isolates of bacterial species and the genetic variation indexed by these methods appears to accumulate relatively rapidly. The rapid accumulation of variation also leads to slight differences between DNA fragment patterns within a clone, which is necessary for fine-scale epidemiological investigations. Different methods may be appropriate for investigating local and global epidemiology, but in both cases they should be highly discriminatory such that isolates assigned to the same molecular type are likely to be descended from a recent common ancestor, and isolates that share a more distant common ancestor are not assigned to the same type. Methods that index rapidly evolving variation are useful for short term epidemiology but may be misleading for global epidemiology. Several studies have shown that techniques such as PFGE resolve isolates that are indistinguishable by multi-locus enzyme electrophoresis (MLEE). For example, MLEE studies of populations of *Salmonella enterica* have shown that isolates of serovar Typhi from typhoid fever belong to one of two closely related electrophoretic types (ETs). In contrast, isolates of serovar Typhi are relatively diverse according to PFGE. PFGE is therefore useful for studying individual outbreaks of typhoid fever because, unlike MLEE, it identifies the microvariation that is needed to distinguish between strains circulating within a geographic area. However, this technique is too discriminatory in long term epidemiology because it does not indicate that isolates that cause typhoid fever are members of a single globally distributed clonal lineage of *S. enterica*. To use a common metaphor, PFGE and other similar methods fail to see the forest for the trees. Out of the current techniques for long term epidemiology, the most appropriate for the identification of lineages that have an increased propensity to cause disease, is undoubtedly MLEE. This approach also has contributed most to our understanding of the global epidemiology and population structure of infectious agents. For many pathogens, MLEE has successfully identified clusters of closely related strains (clones or clonal complexes) that are particularly liable to cause disease. A major problem with MLEE, and all other current typing methods, is that the results obtained in different laboratories are difficult to compare. Maiden et al. in 1998 proposed multi-locus sequencing typing (MLST) as a nucleotide sequence-based approach to identify alleles directly from the nucleotide sequence.
sequences of internal fragments of housekeeping genes rather than by comparing the electrophoretic mobilities of the enzymes they encode. Housekeeping genes are involved in basic functions needed for the sustenance of the cell and are constitutively expressed. MLST was primarily designed and validated for *Neisseria meningitidis*. This modification has overwhelming advantages. First, far more variations can be detected, resulting in many more alleles per locus than are obtained with MLEE. Second, sequence data can be compared readily between laboratories, such that a typing method based on the sequences of gene fragments from a number of different housekeeping loci is fully portable and data stored in a single expanding central multi locus sequence database can be interrogated electronically via the Internet to produce a powerful resource for global epidemiology. Maiden et al. reported the development and validation of MLST for the identification of the virulent lineages of the bacterial pathogen *Neisseria meningitidis*. The MLST approach is, however, applicable to almost all pathogenic, or non-pathogenic, bacterial species and to many other haploid organisms.

**Multi-locus sequencing typing in Listeria monocytogenes**

The ubiquity of *L. monocytogenes* makes especially necessary the use of typing methods for the study of its epidemiology. Several methods have been used to differentiate *L. monocytogenes* strains: for example, based on somatic (O) and flagellar (H) antigens, 13 serotypes of *L. monocytogenes* have been recognized. These are identified alphanumerically: 1/2a, 1/2b, 1/2c, 3a, 3b, 3c, 4a, 4ab, 4b, 4c, 4d, 4e and 7.14 Over 95% of isolates in human listeriosis and in foods belong to serotypes 1/2a, 1/2b and 4b. This undermines the usefulness of serotyping in epidemiological investigations. A molecular typing method with a high discriminatory power could overcome this problem. Numerous molecular methods have been applied and compared for the typing of environmental and clinical *L. monocytogenes* isolates, such as multilocus enzyme electrophoresis (MLEE), random amplified polymorphic differences, ribotyping. Other fingerprinting typing methods described in the literature, such as PFGE, RAPD and AFLP have a comparable discriminatory power. RAPD is rapid and inexpensive but it has a low inter-laboratory reproducibility and PFGE is highly standardized but labor-intensive. AFLP presents several advantages in terms of reproducibility and throughput compared to other fingerprinting techniques. The use of capillary electrophoresis with an automatic sequencer for the accurate sizing of fragments and a specific software for data managing and rapid analysis of fingerprinting profiles makes this method a valuable tool for the characterization of microbial populations. The differences in the pathogenic potential of the *L. monocytogenes* strains, which are correlated with the flagellar antigen groups, have been demonstrated in *vitro* and *in vivo* and three distinct divisions or Lineages exist:

Division I: strains associated with epidemic outbreaks of listeriosis (serotypes 1/2b, 3b, 4b, 3c); Division II: strains isolated from sporadic cases of listeriosis (serotypes 1/2a, 1/2c, and 3a); Division III: strains rarely associated with cases of listeriosis (serotypes 4a and 4c). These three divisions are characterized by non-overlapping allelic variants of different genetic markers, suggesting strong linkage disequilibrium and an apparent lack of gene exchange between them. The non-random association of alleles at more loci suggested the hypothesis that the genetic structure of *L. monocytogenes* populations is basically clonal. There are three elements to take into account in the design of a new MLST *L. monocytogenes* system: the choice of the isolates to be used in the initial evaluation; the choice of the genetic loci to be characterized and the design of primers for gene amplification and nucleotide sequence determination. The optimization of the reaction conditions for amplification and sequencing is an important task often neglected.

**The choice of the isolates**

The choice of the isolates to be used is advisable to assemble a diverse isolate collection on the basis of existing typing information or epidemiological data. This should comprise around 100 isolates to ensure that the developed primers will be applicable to as many isolates as possible and to establish the levels of diversity present at each of the loci to be examined. Furthermore, the collection will ideally be representative of the bacterial population, rather than comprising a subset, such as human disease isolates. In the different studies about MLST applied to *L. monocytogenes*, the number and the source of strains is various. Salcedo (2003) examined a total of 62 strains, Revazishvili in 2004 compared MLST and PFGE examining a total of 175 strains, Ragon in 2008 used 360 strains and Parisi in 2009 examined 130 isolates. But in a 2007 study, Jiang et al. considered only 20 isolates to characterize the *in vitro* and *in vivo* pathogenicity of *L. monocytogenes* food-related isolates from eastern China and to define the phylogenetic diversity of *L. monocytogenes* isolates using MLST and PFGE.

**The choice of the genetic loci**

The choice of the genetic loci is an important factor to design a MLST system. Housekeeping genes flanked by genes of similar function, are good targets for MLST and the availability of complete genome sequences has greatly facilitated the identification of candidate loci. The number of the nucleotides that define a MLST allele is based on technical and financial constraints, principally the length of nucleotide sequence that can be readily determined with one sequencing extension reaction in each direction. In 1996, when work started on the first MLST project this was around 450 bp for most automated nucleotide sequencing instruments. Although longer sequences can now be routinely attained, experience with several bacterial species has indicated that fragments of housekeeping genes of this size are suitable. A candidate housekeeping gene, chosen on the basis of its predicted function in genome annotations, might be experiencing unexpected levels of recombination or selection and therefore proved to be unsuitable. A system that examines too few loci might be confused by the chance associations of alleles. The number of loci can be increased to improve resolution but there will come a point when, for epidemiological purposes, little additional information is attained for the expense and effort involved.

**The design of primers for gene amplification and nucleotide sequence determination**

The design of oligonucleotide primers for amplification and sequencing represents much of the work required for the development of a MLST system. It is highly recommended that a nested strategy is used, in which DNA fragments that are larger than required for the final sequence are initially amplified. The nucleotide sequences of these fragments are determined using distinct sequencing primers internal to the amplified fragment. Such primers generally produce a higher quality of nucleotide sequence data and the possibility of sequencing spurious amplification products is eliminated. Recent schemes that described MLST in *L. monocytogenes* were adapted from the study of Salcedo, using automated DNA sequencing to characterize the alleles present at different housekeeping genes. These surveys analyzed six and seven of respectively several housekeeping genes and virulence-associated gene loci.

**Multi-locus sequencing typing databases**

The aim of the original MLST scheme was to provide access to the data using the Internet, and to enhance clinical diagnosis, epidemiological monitoring, and population studies. MLST websites are virtual isolate collections. For *L. monocytogenes* for example, one of the most important reference database is: http://www.pasteur.fr/recherche/genopole/PF8/...
Organization into clonal complexes makes MLST data more amenable to epidemiological analysis. For example, a collection of 103 L. monocytogenes isolates contained a total of 66 STs and 8 main Clonal Complexes. The original conception of MLST used the allele number as the primary unit of analysis, which was appropriate for organisms where horizontal genetic exchange is common. MLST data can also, however, be interpreted by neighbor-joining tree analysis that was performed by all the authors mentioned in this review. Neighbor-joining tree is a bottom-up clustering method used to construct a phylogenetic tree for MLST analysis on the basis of nucleotide differences in the various gene fragments. This is more pertinent to bacteria where mutational change predominates over genetic exchange in the evolution of variants. Software such as the START (sequence type analysis and recombinational tests) package brings together many of the preliminary analyses that can be performed on MLST data at the present time. The START program requires that sequences of equal length will be analyzed in order to determine the numbers of alleles and polymorphic sites, mean G+C content, and perform recombination and selection (dnds) tests. Linkage analysis was carried out by using the index of association (IA). On the whole, MLST identified different allelic profiles (STs) in the different studies.

**Multi-locus sequencing typing vs. other typing methods for the identification of Listeria monocytogenes**

In 2003, Salcedo took the first step in the development of MLST methods for L. monocytogenes. Subsequently, the MLST scheme has been updated. Alternative sets of genes were used but were either not extensively validated or biased towards high levels of nucleotide diversity between two particular lineages, or based at least in part on virulence genes. Virulence associated genes generally provide improved discrimination among strains, but may reflect ecological adaptation and selection. In contrast, housekeeping genes are considered more appropriate to obtain an unbiased view of the population structure, as their polymorphisms can be considered nearly neutral and are less subject to horizontal transfer. Salcedo based his study on nine housekeeping genes: ABC transporter (abcz), histidine kinase (hikA), beta-glucosidase (bgldA), succinyl diaminopimelate desuccinylase (dape), catalase (cat), L-lactate dehydrogenase (lhd), D-amino acid aminotransferase (dat), superoxide dismutase (sod) and phosphoglucomutase (pym). In this study, similarities in the genome sequence between serotype 4b and 1/2b isolates might explain a higher prevalence of homologous recombination between these strains. As reported above, the schemes of the authors that subsequently described MLST in L. monocytogenes were adapted from this study. Also other authors have compared MLST to other molecular methods to differentiate strains of L. monocytogenes in epidemiological investigations and to try to understand which of the different methods have the highest discriminatory power. In the United States, Revazishvili et al. have compared serotyping, PFGE typing and MLST and in most cases, strains with the same ST and PFGE type belonged to the same serotype. However, there were exceptions. For example, L. monocytogenes strains 84 and 57 had ST121 and PFGE type P16 (both strains also had the same actA and hlyA alleles) but they had different serotypes: 1/2b and 3b, respectively. In some other cases, strains with identical genetic backgrounds (i.e., the same ST and PFGE type) also had the same actA and hlyA alleles. However, some strains with identical genetic backgrounds did not have the same actA and hlyA alleles. Several strains with the same PFGE type (i.e., they were indistinguishable by PFGE-typing) were differentiated and assigned distinct STs by MLST. In some cases, strains with the same PFGE type but distinct STs were not closely related based on the MLST dendrogram. In some instances, strains with the same ST had distinct PFGE types. Also no clear-cut correlation between serotypes and the source of isolation of the strains (environmental versus clinical) was observed in this study, although strains of the 1/2b and 4b serotypes seemed to predominate among the clinical specimens (75% of strains belonging to serotypes 1/2b and 4b were clinical isolates). Revazishvili et al. did not observe clusters...
(PFGE clusters were defined as groups of at least three distinct PFGE types) of clinical and environmental isolates in the PFGE-based dendrogram. However, START analysis of the MLST data revealed three predominantly clinical subclusters (MLST subclusters were defined as groups containing a minimum of five non-identical strains). The analyzed L. monocytogenes strains (a total of 175 isolates from environment, clinical cases and of unknown origin) were grouped in 8 serogroups, 57 PFGE types, and 122 STs, which suggests that the discriminatory abilities of the three methodologies differ, i.e., that serotyping is the least discriminatory and MLST is the most discriminatory of the three approaches. Serotyping is one of the oldest approaches for typing L. monocytogenes. However, the discriminatory power of serotyping is limited, and PFGE typing has been reported to be superior for differentiating L. monocytogenes strains. The observations that the number of STs generated by MLST was ca. 2-fold greater than the number of PFGE types (122 STs versus 57 PFGE types) and that several strains within the same PFGE type were differentiated by MLST support the idea that the discriminatory ability of MLST is greater than that of PFGE. Borucki et al. led a study in which the resolutions of four different subtyping methods (microarray, PFGE, ribotyping and MLST) were compared using a subset of strains. The study was conducted on 52 strains from bulk milk, food, human cases, soil and bovine from different countries in Europe and US. Microarray analysis and PFGE subtyping showed the highest resolution, MLST had moderate subtyping resolution, and ribotyping had the lowest resolution. The microarray analysis subtyping resolution was similar to that of PFGE with two enzymes, the current gold standard for molecular subtyping of L. monocytogenes strains. Nevertheless, occasionally the two techniques placed strains in different groups. This is not surprising, because the two techniques sample the genome differently. In eastern Chinese province Zhejiang food-related isolates were analyzed by Jiang et al. They compared MLST and PFGE. In this study, MLST analysis, based on partial nucleotide sequences of four housekeeping genes (betA, dat, recA and sigB) and three virulence genes (actA, inlA and inlB), provided a highly discriminatory subtyping method, which allowed to differentiate the L. monocytogenes isolates into 19 sequence subtypes. The virulence genes had better discriminatory power than the housekeeping genes (0.990 vs 0.895) and could be better for MLST subtyping, which is consistent with other observations. Interestingly, among the four housekeeping genes, recA was found to be the most variant gene but with the least level of discriminatory power (D.I.=0.690), indicating a high percentage of polymorphic sites occurring in the same type. Zhang et al. also demonstrated that the most polymorphic gene fragment dal was not the most discriminatory for reference strains. This finding may help choosing the housekeeping gene(s) to be used for MLST, such as the gene betL because of its higher D.I. Value. Findings of this study indicate that the discriminatory power of MLST (D.I.=0.990) was comparable to that of PFGE (D.I.=0.976), as described in another study (Zhang et al., 2004). In addition, Revazishvili et al. found MLST had greater discriminatory power than PFGE. This is because MLST detects all genetic variations within the amplified gene fragments, whereas PFGE only examines the variations in the cleavage sites for a particular restriction enzyme. In a French survey, Ragon et al. carried out a study on a total of 360 L. monocytogenes isolates from environment, food and animal over a very long period in which they analyzed the clonal structure, the evolution of serotypes and the phylogenetic structure of L. monocytogenes. Interestingly, some similarities can be detected when comparing this study with published MLST data on a Italian collection of L. monocytogenes isolates. For example, CC9 comprised all isolates of serotype 1/2c confirming the speculated monophyletic origin of this serotype; CC2, CC3, CC8 and CC9 were among the more prevalent CCs; CC1 and CC2 were dominated by isolates of serotype 4b/4c; CC3 comprised a large proportion of serotype 1/2b. On the other hand, some CCs frequently identified in the French survey were not identified in the Italian study. This underlines the differences in regional prevalence of single genotypes and the need for global studies to integrate information derived from different areas of the world. Moreover, Parisi et al. compared AFLP and MLST which exhibited comparable discriminatory powers (0.976 and 0.972 respectively) that were higher than that of serotyping (D.I.=0.739) and highlighted the low efficacy of this typing method. Based on this data, it is possible to make out how MLST give us main information about the phylogenetic structure of L. monocytogenes and provide a framework for the evolutionary history, epidemiology and virulence of this food-borne pathogen.

Conclusions

The data reported in literature and cited in this review, demonstrated that there are several similarities between the various studies conducted in different geographical areas. Except for same authors, it was found that MLST utilizing loci from housekeeping genes was more discriminatory than PFGE for typing of L. monocytogenes. Including additional loci in the analysis is likely to further increase the discriminatory power of MLST. At present, no established criteria for determining the minimal number of genes that must be analyzed to obtain a reliable MLST assignment have been established. Although the number is likely to be different for different species (depending on the level of clonality of the species), MLST utilizing seven housekeeping genes has been proposed to be a reasonable approach well-suited for most bacteria. The major advantage of MLST is the unambiguous nature of the data obtained and the ease with which these can be stored and transmitted electronically, meaning that any isolate that is typed using the method can be rapidly compared with all previously typed strains. MLST is the main method for characterizing hypervirulent/antibiotic-resistant clones of several pathogenic bacteria. National and international surveillance of bacterial clones can be performed using this resource. MLST could be a useful tool for the surveillance systems for listeriosis that might allow the identification and analysis of the distribution of these L. monocytogenes clones in the environment. MLST is increasingly applied as a routine typing tool that enables international comparison of isolates. MLST has been applied to problems as diverse as the emergence of antibiotic-resistant variants, the association of particular genotypes with virulence, or antigenic characteristics and the global spread of disease caused by novel variants. In addition to these medically motivated epidemiological analyses, MLST data have been exploited in evolutionary and population analyses, that estimate recombination and mutation rates and investigate evolutionary relationships among bacteria that are classified as belonging to the same genus. MLST can be used for epidemiological studies of any bacterial pathogen that exhibits variability in its housekeeping gene sequences. MLST may be of value for improving the differentiation of L. monocytogenes strains isolated during food-borne outbreaks of listeriosis and for tracing the outbreak-causing strains to their sources. Also, MLST may be used to determine the phylogenetic relatedness among L. monocytogenes strains, and its use should improve our understanding of the mechanisms involved in the emergence and divergence of various L. monocytogenes strains and serotypes, including serotypes primarily associated with human listeriosis. In order to improve the results of the MLST, some authors developed a multi-virulence-locus sequence typing (MVLS) study, in which a total of 14 strains were previously analyzed by sequencing the complete coding sequence of three housekeeping genes and six virulence genes. These housekeeping genes were found to be less discriminatory than some virulence genes, namely, actA and inlA.


