



“Epidemic Clones” of *Listeria monocytogenes* Are Widespread and Ancient Clonal Groups

Thomas Cantinelli, Viviane Chenal-Francisque, Laure Diancourt, Lise Frezal, Alexandre Leclercq, Thierry Wirth, Marc Lecuit, Sylvain Brisse

► To cite this version:

Thomas Cantinelli, Viviane Chenal-Francisque, Laure Diancourt, Lise Frezal, Alexandre Leclercq, et al.. “Epidemic Clones” of *Listeria monocytogenes* Are Widespread and Ancient Clonal Groups. *Journal of Clinical Microbiology*, 2013, 51 (11), pp.3770-3779. 10.1128/JCM.01874-13 . pasteur-02869829

HAL Id: pasteur-02869829

<https://pasteur.hal.science/pasteur-02869829>

Submitted on 16 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Page 1, version 26/08/2013

Epidemic clones of *Listeria monocytogenes* are widespread and ancient clonal groups

Thomas Cantinelli¹, Viviane Chenal-Francisque¹, Laure Diancourt², Lise Frezal³,
Alexandre Leclercq¹, Thierry Wirth^{4,5}, Marc Lecuit^{1,6,7,*} and Sylvain Brisse^{8,9*}

¹Institut Pasteur, National Reference Center for *Listeria* and World Health Organisation
Collaborating Center for *Listeria*, Biology of Infection Unit, Paris, France

²Institut Pasteur, Genotyping of Pathogens and Public Health, Paris, France

³Ecole Normale Supérieure, Department of Biology, 46 rue d'Ulm, Paris, France

⁴Laboratoire de Biologie Integrative, Ecole Pratique des Hautes Etudes, Paris, France

⁵Muséum National d'Histoire Naturelle, UMR-CNRS 7205, Département Systématique et
Evolution, 16 rue Buffon, F-75231 Paris cedex 05, France

⁶Inserm U1117

⁷ Université Paris Descartes, Assistance Publique-Hôpitaux de Paris, Service des Maladies
Infectieuses et Tropicales, Hôpital Universitaire Necker-Enfants Malades, Institut Imagine, Paris,
France.

⁸ Institut Pasteur, Microbial Evolutionary Genomics, Paris, France

⁹ CNRS, UMR3525, Paris 75015, France

* Corresponding authors:

S. Brisse. Institut Pasteur, Microbial Evolutionary Genomics, 28 rue du Dr Roux, F-75724
Paris, France. E-mail: sylvain.brisse@pasteur.fr; Phone +33 1 40 61 36 58.

M. Lecuit. Institut Pasteur, Biology of Infection Unit, 28 rue du Dr Roux, F-75724 Paris,
France. E-mail: marc.lecuit@pasteur.fr; Phone: + 33 1 40 61 30 29.

Page 1, 26/08/2013

29

30

31

Abstract

32

33

34 The food-borne pathogen *Listeria monocytogenes* is genetically heterogeneous. Although
35 some clonal groups were implicated in multiple outbreaks, there is currently no consensus on
36 how 'epidemic clones' should be defined. The objectives of this work were to compare the
37 patterns of sequence diversity on two sets of genes that are widely used to define
38 *L. monocytogenes* clonal groups: multilocus sequence typing (MLST) and multi-virulence
39 locus sequence typing (MvLST). Further, we evaluated the diversity within clonal groups by
40 pulsed-field gel electrophoresis (PFGE). Based on 125 isolates of diverse temporal,
41 geographical and source origins, MLST and MvLST genes (i) had similar patterns of
42 sequence polymorphism, recombination and selection, (ii) provided concordant phylogenetic
43 clustering and (iii) had similar discriminatory power, which was not improved when
44 combining both datasets. Inclusion of representative strains of previous outbreaks
45 demonstrated the correspondence of 'epidemic clones' with previously recognized MLST
46 clonal complexes. PFGE analysis demonstrated heterogeneity within major clones, most of
47 which were isolated decades before their involvement in outbreaks. We conclude that the
48 'epidemic clones' denominations represent a redundant but largely incomplete nomenclature
49 system for MLST-defined clones, which must be regarded as successful genetic groups that
50 are widely distributed across time and space.

51

52 Introduction

53 *Listeria monocytogenes* is a foodborne pathogen that can cause listeriosis, a severe
54 invasive infection with a particularly high (20-30%) case fatality rate in persons at risk.
55 Listeriosis is currently regarded as increasing in incidence in Europe, especially in the elderly
56 (13, 23), and can occur in large outbreaks as illustrated in recent years (22, 38). To assist
57 epidemiological surveillance and outbreak investigation, different strain typing methods are
58 used, including serotyping and pulsed-field gel electrophoresis (17, 28). Population diversity
59 studies at the global scale have revealed that *L. monocytogenes* is a genetically heterogeneous
60 species (10, 46, 47, 56) and a variety of strain genotyping methods are used to characterize
61 and classify isolates into four major lineages and clonal groups thereof (7, 12, 37, 47, 50, 55,
62 56). The precise delineation of lineages and clonal groups is a prerequisite to characterize the
63 links between within-species genetic variation and important characteristics such as
64 pathogenic potential, virulence or epidemiology.

65 Given the facultative nature of genetic exchange in bacteria, reproduction is
66 predominantly clonal. In evolutionary biology terms, clones can be defined as groups of
67 isolates that descend from a common ancestor and accumulate differences among themselves
68 by a predominantly mutational process. As *L. monocytogenes* is one of the bacterial species
69 with the lowest rate of homologous recombination (11, 47), clones are expected to evolve
70 slowly and to be recognizable over large temporal and geographic scales. As a matter of fact,
71 the discovery of genetically similar isolates involved either in geographically and temporally
72 distant outbreaks or in large, single outbreaks led to the definition of *L. monocytogenes*
73 epidemic clones (EC) (6, 11, 27, 28, 46). Although initial ECs have been defined primarily
74 based on PFGE, MLEE and ribotyping (27, 28), subsequent ECs were mostly defined on the
75 basis of multi virulence locus sequence typing (MvLST) (6, 29, 38). MvLST, which is based
76 on the analysis of six to eight genes, has also been used to redefine previously described
77 epidemic clones (5, 6).

78 In many bacterial species, multilocus sequence typing (MLST) is used as a reference
79 method for clonal group definition (18, 39). An MLST scheme based on seven housekeeping
80 genes was developed for *L. monocytogenes* (47, 49). Using this approach, we recognized
81 highly prevalent clones (8, 47), which were defined using a simple and flexible operational
82 definition: clones are clonal complexes, i.e. groups that share 6 out of 7 allelic sequences with
83 at least one other member of the group. Currently, it is unknown how MLST clonal
84 complexes and epidemic clones correspond to each other.

85 The objectives of this study were (i) to determine the phylogenetic position of
86 reference strains of previously defined epidemic clones within the MLST framework; (ii) to
87 compare the patterns of diversity, recombination and selection of MLST and MvLST genes;
88 and (iii) to estimate the amounts of PFGE diversity within clonal groups or epidemic clones.
89 We demonstrate that MLST and MvLST define largely concordant clonal groups, and that the
90 current approach to define 'epidemic clones' based on MvLST is redundant with the MLST
91 nomenclature.
92

Materials and Methods

Strain selection

A total of 125 *Listeria monocytogenes* isolates were included (**Table S1**). First, in order to establish the position of outbreak strains or other reference strains in the international MLST scheme maintained at Institut Pasteur (www.pasteur.fr/mlst), a group of 50 strains was assembled. This reference set included (i) 27 reference strains from well documented outbreaks, including 19 strains of the outbreak set included in the ILSI collection (20) and 8 genome sequence reference strains corresponding to outbreak strains; (ii) fifteen lineage I and II strains from the diversity set of the ILSI collection; (iii) seven additional reference strains representing available genome sequences, and reference strain EGD. This reference set included the intensely studied laboratory reference strains EGD-e, LO28, 10304S, F2365 and Scott A, as well as strains representing 'epidemic clones' ECI, ECII, ECIII and ECIV/ECIV.

Second, with the aim of evaluating the ability of PFGE and MvLST to discriminate within and among MLST-defined clonal complexes (CC), we selected 75 isolates (the clone diversity set) representing multiple isolates of the major MLST-defined CCs, i.e. CC1, CC2, CC6, CC7, CC8 and CC9 (8, 47). These isolates were selected from our previous global MLST study of *L. monocytogenes* (8) to represent geographically and temporally diverse isolates. They originate from the Institut Pasteur *Listeria* Collection (CLIP) and from the Seeliger's *Listeria* Culture Collection (SLCC) (Rocourt and Seeliger, 1985).

Identification and serotyping

Isolates were identified as *L. monocytogenes* using API *Listeria* strips (BioMerieux, La Balme Les Grottes, France) and their serotype determined by classical serotyping and PCR serogrouping (15). Genomic DNA used as MLST or MvLST PCR template was extracted using the Promega Wizard Genomic DNA purification kit (Promega, Madison, WI, USA).

Multilocus sequence typing (MLST)

For the purposes of this study, MLST was performed as described (47). Novel alleles and profiles were incorporated into the international MLST database at www.pasteur.fr/mlst. MLST gene sequences from 16 publicly available genome sequences were extracted from the public sequence repositories.

Multi-virulence locus sequence typing (MvLST)

In this study, eight virulence-associated genes of *L. monocytogenes* were sequenced for 111 strains, while for 16 strains, the gene sequences were extracted from the available genome sequences. Six of the MvLST genes (*PrfA*, *inlB*, *inlC*, *clpP*, *dal*, *lisR*) were taken from S. Knabel's group MvLST scheme (59). The two remaining genes, *inlA* and *actA*, were included as they were used to complement the six previous ones in a subsequent MvLST analysis (6). Note that we sequenced the entire length of the *inlA* gene (2,400 nt), as we did in a previous study (47), whereas only 458 nt of this gene were sequenced in Chen et al. (2007). For *actA*, because we experienced difficulties in PCR amplification using previously described primers, we sequenced 450 nt from the 3' part of the gene, whereas the 582 nt template of Chen et al. (2007) corresponds to the 5' region of the *actA* gene. PCR primers (Table 1) were taken from Chen et al. (6) or designed in this study with EPRIMER3 (<http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::eprimer3>) based on the published genome sequence of *Listeria monocytogenes* strain EGD-e (NCBI RefSeq NC_003210.1). PCR amplification conditions as follows. For gene *dal*, 10 min at 95°C followed by 30 x (95°C 30 s, 57.6°C 30 s, 72°C 36 s) and 5 min at 72°C. For *prfA*, *inlB*, *inlC*, *clpP* and *lisR*, 15 min at 95°C followed by 25 x (94°C 30 s, 55°C 30 s, 72°C 1 min) and 7 min at 72°C. For *actA*, 10 min at 95°C followed by 30 x (95°C 30 s, 54.7°C 30 s, 72°C 36 s) and 5 min at 72°C. Finally for *inlA*, 5 min at 94°C followed by 35 x (94°C 30 s, 55.2°C 30 s, 72°C 90 s) and 10 min at 72°C. The PCR products were purified by ultrafiltration (Millipore, France) and were sequenced on both strands with Big Dye v.1.1 chemistry on an ABI3730XL sequencer (Applied BioSystems). As for MLST, each nucleotide was sequenced in both directions and validated by at least two independent chromatogram traces. MvLST gene sequences from published genome were extracted from GenBank entries.

Pulsed-field gel electrophoresis (PFGE)

Each isolate was typed by PFGE according to PulseNet standardized procedures with *AscI* and *Apal* restriction enzymes (24). Data analysis was performed using BIONUMERICS version 6.5 (Applied Maths, Sint-Martens Latem, Belgium). *Apal* and *AscI* PFGE types were defined as differing from other types by at least two bands for each individual enzyme.

Phylogenetic reconstructions

Clonal complexes were defined based on MLST data as groups of allelic profiles sharing 6 out of 7 genes with at least one other member of the group (47). For phylogenetic

analyses, gene sequences were concatenated independently for the MLST and MvLST schemes, and neighbour-joining trees were obtained using BIONUMERICS based on the concatenated sequences using the p -distance (i.e., uncorrected percent of nucleotide mismatches). Nucleotide polymorphism and summary statistics were calculated using DNASP v5 (36). Minimum spanning trees were constructed using BIONUMERICS. Simpson's index and adjusted Rand coefficient were computed using the website www.comparingpartitions.info (3).

Recombination and selection analyses

We tested for recombination within phylogenetic lineages I and II for all loci independently, as well as for the concatenated MvLST and MLST loci using the LDHAT v2.2 software (40). LDHAT employs a coalescent-based method to estimate the population-scaled mutation ($\theta = 2N_e\mu$) and recombination ($\rho = 2N_er$) rates, where N_e is the effective population size, r the rate at which recombination events separate adjacent nucleotides and μ is the mutation rate per nucleotide. The ratio r/μ were calculated as $\rho/(\theta/L)$, where L is the gene length (sequence length). This r/μ ratio ranges from 0, which indicates full clonal reproduction, to $\gg 1$, which is expected under free recombination. Significance of the evidence for recombination was tested using nonparametric, permutation-based tests implemented in LDHAT.

To detect signatures of positive selection, flat ω (d_N/d_S) ratios (as generated with DNASP) might fail to detect on-going selection on a fraction of the molecule. We therefore tested the presence of both recombination and positive selection using OMEGAMAP (57), which is able to disentangle the confounding effects of recombination and selection in a Bayesian framework. We chose a model with variable blocks of the ratio of non-synonymous to synonymous rates, ω , and the population recombination rate, ρ . A first run was performed for all loci using a block size set at 30 codons. A second run was performed using a block size set at 10 codons for the loci where a posterior probability of selection greater than 0.80 was detected in the first run. For all runs, an inverse distribution of range [0.01, 100] was used for ω and ρ . For the other parameters μ , κ and Φ InDel we used improper inverse distributions with starting values 0.1, 3.0 and 0.1, respectively. All genes were run with 100,000 iterations and 10 re-orderings, with the first 8,000 iterations discarded as burn-in period.

Results and Discussion

Phylogenetic positions of outbreak and reference strains based on MLST

To position reference strains, including isolates previously assigned to epidemic clones I to IV, within the diversity of *L. monocytogenes*, the sequences of internal portions of the seven multilocus sequence typing (MLST) genes (3,288 nucleotides in total) were gathered from 14 available genome sequences, 34 strains of the ILSI reference collection analyzed previously (7) and 76 isolates of clones belonging to CC1, CC2, CC6, CC7, CC8 and CC9 (7). In addition, we sequenced the 7 MLST genes from strain EGD, which is used as a reference strain (31, 41) but has not been genotyped by MLST to our knowledge. For these 125 isolates or reference strains, alignment of the sequences of six genes revealed no insertion/deletion event, whereas the *ldh* sequence of strain EGD revealed a 2-codons insertion at alignment position 423. There were 20 to 61 variable nucleotide sites per gene, corresponding to 7.5% of variable sites on average (**Table 2**). This variation allowed distinguishing 10 to 20 alleles per genes, resulting in 51 distinct sequence types (ST).

As previously described (7), clonal complex (CC) 1 included the reference strains of the 1981 Nova Scotia coleslaw outbreak, the 1983-1987 Switzerland Vacherin Mont d'Or outbreak, the 1986-1987 California Jalisco soft cheese outbreak and the 1992 French pork tongue in jelly outbreak (**Figure 1**). These strains were previously attributed to epidemic clone ECI (28). Clone CC2 comprised reference strains of the 1983 Massachusetts pasteurized milk outbreak (strain Scott A), the 1987-1988 UK and Ireland pâté outbreak, and the 1997 Italy gastroenteritis outbreak (**Figure 1**). These strains have been attributed to ECIIa (27), later renamed ECIV with a more restricted definition that excluded the 1983 Massachusetts milk outbreak (6). All these reference strains including Scott A belonged to ST2, which is the central and most frequent genotype of CC2 (**Figure 1**). These results show that reference strains of ECI, ECII and ECIIa/ECIV fall in MLST-defined CC1, CC6 and CC2, respectively. CC3 included the reference strain of the 1994 Illinois chocolate milk outbreak. CC4 included the 1999-2000 France pork rillettes outbreak. CC6 included the 1998-1999 USA multistate hotdog outbreak and the 2002 USA multistate delicatessen turkey outbreak, previously attributed to ECII. Finally, the 2000-2001 North Carolina Mexican-style fresh cheese outbreak reference strain belong to ST558 (**Figure 1**). The unique genotype of this outbreak strain is consistent with its characterization as ST24 (12). Note that ST5 included the isolates of the recently described ECVI clone (38).

Lineage II comprised reference strains of fewer outbreaks. ST11 included the 1988 Oklahoma turkey franks case and the two strains from the 2000 USA multistate sliced turkey deli meat outbreak, which was traced to the same food processing facility as the 1988 isolate. All three ST11 strains were previously attributed to ECIII (28). Note that whereas ECI, ECII and ECIV have been implicated in multiple outbreaks (28, 46), this was not the case to our knowledge for ECIII. CC8 comprised the 2008 Canada ready-to-eat meat products outbreak, consistent with previous report (22). Strains of this outbreak were recently proposed to represent ECV (29). Clone ECVII (involved in the 2011 US Cantaloupe outbreak) was shown to belong to CC7 (38).

The four widely used laboratory reference strains EGD, EGD-e, LO28 and 10403S belonged to two distant clonal complexes of lineage II, CC9 (EGD-e and LO28) and CC7 (10403S and EGD). These results indicate that despite their similar names, strains EGD and EGD-e appear to be phylogenetically distant (**Figure 1**), excluding direct laboratory descent between them.

The genome reference strains of serotype 1/2b FSL J2-064 and FSL N1-017 fell in major clones CC5 and CC3, respectively, thus also representing important clonal groups of *L. monocytogenes*. In contrast, the three genome reference strains FSL J1-194 (ST88), FSL J1-175 (ST87) and FSL N3-165 (ST90) represented rare STs not previously encountered, except for ST87, which was described for one food strain in Colombia (8).

To illustrate the position of the above studied strains within the diversity of *L. monocytogenes*, a minimum spanning tree (MStree) including all currently published MLST strains was constructed (**Figure 1**). Whereas outbreak strains mostly fell into the central and most frequent ST of clonal complexes, the laboratory reference strains EGD, EGD-e, LO28 and 10403S represented variant STs that were not commonly found among clinical or food isolates (**Figure 1**). Clearly, the diversity of *L. monocytogenes* is only partly represented by the laboratory reference strains that have been analysed so far in studies of host-pathogen interactions, virulence or other characteristics.

The distribution of MLST alleles showed that most CCs can be differentiated based on one or a few genes (**Table 3**). For example, allele *lhkA*-3 was uniquely observed for isolates of CC1, *abcZ*-1 was specific for CC2, and *bglA*-9 was shared by isolates of CC6. These CC-specific alleles allow a rapid screening of isolates to identify those belonging to CCs of interest. However, because recombination can occur, MLST based on the 7 genes sequences should be used for confirmation on selected isolates.

Diversity of MvLST genes and comparison with MLST

It has been suggested that virulence-associated genes included in MvLST studies evolve faster than MLST genes, which code for central, housekeeping functions of the cell (59). To test this hypothesis, the 125 study isolates were sequenced at eight MvLST genes used for epidemic clone definition (6, 29, 38). The genetic variation recorded at MvLST genes (**Table 2**) represented 6.62% of nucleotide sites on average, ranging from 3.06 for *inlC* to 12.6% for gene *dal*. This level of variation was therefore slightly lower than the amount observed for MLST genes (7.5%). The number of distinct alleles per gene varied from 4 alleles for gene *clpP* to 28 for gene *inlA*. When considering the eight MvLST genes, there were 364 nucleotide polymorphisms out of the 5,496 sites that were sequenced, resulting in 40 distinct MvLST types. 51 distinct STs were found based on MLST. When excluding the atypically variable gene *ldh* from the MLST data (see below), 41 sequence types were distinguished, very similar to the number of MvLST types. These results show that MvLST does not provide more discrimination than MLST even when including 8 MvLST genes, among which the entire *inlA* gene sequence. In previous work, MvLST was found to be more discriminatory than MLST (59), but this suggestion was based on different MLST genes (*prs*, *sigB* and *recA*) and only 14 strains.

Similarly, the variation of MvLST types within the major MLST clones was limited (**Table 4**). No more than 4 distinct MvLST types were observed within CCs. Besides, most isolates within a CC were identical at the eight MvLST genes, as the other MvLST types were represented by only one or few isolates. Remarkably, the homogeneity of MvLST gene sequences within CCs was observed across large geographical and time scales (**Table S1**). For example, the most frequent MvLST type in CC1 was observed for isolates collected between 1963 and 2000 from 13 distinct countries in North America, Europe, Oceania and Asia. These observations show that MvLST gene sequences are stable through time and space, resulting in the same MvLST type being observed for isolates that are considered unrelated from an epidemiological point of view. From a practical standpoint, the low variation of MvLST within single STs or CCs implies that adding the analysis of virulence genes to MLST does not improve the discriminatory power significantly: Simpson's index of discrimination was 0.946 (59 types, 95% confidence interval 0.925-0.967) for MLST+MvLST combined, whereas it was 0.932 (0.908-0.956) for MLST and 0.898 (0.871-0.925) for MvLST.

InlA evolves by in-frame deletions as well as variation leading to truncated forms of internalin

Internalin, encoded by *inlA*, plays a critical role in *L. monocytogenes* invasion of cultured epithelial cells (21), and is a critical virulence determinant, mediating *L. monocytogenes* crossing of the intestinal and placental barriers (14, 35). Alignment of the MvLST gene sequences revealed that the gene *inlA* differed from the seven other genes by the presence of insertion-deletion events. First, all strains of CC6 had a deletion of 9 nucleotides corresponding to positions 2,212-2,220 in the *inlA* sequence of the other strains which result in the predicted loss of 3 amino acids in the pre-anchor region of InlA. Second, ILSI strain FSL C1-122 had a 30-nt deletion in *inlA* (positions 2,141-2,170), which results in the predicted loss of 10 amino acids (723-732) in the pre-anchor region of InlA. These two in-frame changes are not expected to impair InlA function, as the functional domain of InlA is the LRR-IR region (34), and the CC6 and FSL C1-122 InlA variants harbour an intact anchor region (32). Third, three strains of CC9 had a G to A change at position 2,054 resulting in a stop codon at amino-acid position 694. Fourth, seven other strains of CC9 including LO28 had a one nt deletion in *inlA* at position 1635, resulting in the first premature stop codon described for *inlA* (26). Fifth, three other CC9 isolates had a G to A change at position 1,380, resulting in a stop codon at position 460. Six, strain LM05-01099 had a G to T change at position 976, resulting in a stop codon at position 326. Finally, strain CLIP 11308 had a one nt deletion at position 5 in gene *inlA*, resulting in a predicted stop codon at amino-acid position 9. Functional, non-truncated InlA has been associated with the clinical origin of *L. monocytogenes* isolates (25), and premature stop codons in *inlA* leading to the secretion of a non-functional truncated protein have been associated with reduced pathogenicity (25, 26, 42, 47).

Pattern of recombination, selection and phylogenetic relationships

Homologous recombination and selection both influence the evolutionary rate and the diversity pattern of genes that they affect. To characterize the evolutionary forces that act of the diversification of clonal groups, we searched for evidence of recombination and selection in MLST and MvLST genes. Using LDHAT, evidence for recombination was detected at five out of the 15 loci (**Suppl. Table S2**). Within lineage I, recombination was only detected at the *inlB* locus. Within lineage II, recombination was detected at the three MvLST genes *actA*, *dal* and *inlA*. Among the MLST genes, recombination was detected only at gene *dapE* in lineages

I and II. These results confirm that recombination is infrequent in *L. monocytogenes* and indicate that MvLST genes are more affected by recombination than MLST genes.

To detect positive selection, the software OMEGAMAP (57) was used. This software separates the effects of recombination and positive selection, thus avoiding false positives for selection in the presence of recombination. Using a posterior probability threshold of 0.95, the results revealed positive selection within lineage I at MvLST genes *actA* and *inlC*, two major virulence factors, as well as at MLST loci *dapE* and *ldh* (**Suppl. Figure 1**). No evidence for positive selection was found at the other loci within lineage I, including at the recombining gene *inlB*. However, positive selection is highly likely to occur at gene *inlA*, which had a posterior probability of 0.94. No positive selection was detected within lineage II at any of the 15 genes. In this lineage, the highest posterior probability (0.84) was observed for gene *prfA*, coding for the master regulator of virulence genes. In protein ActA, the positively selected amino acids within lineage I were at codons 124 and 130 (88 and 94 of our sequenced region, **Suppl. Figure 1**), located within the actin tail formation domain (52). Overall, these results indicate contrasting impacts of recombination and selection between lineages: whilst lineage II was more affected by recombination, consistent with previous findings (11), lineage I was more affected by positive selection.

The phylogenetic analysis of the concatenated gene sequences from the MLST gene set (**Figure 2 left**) and the MvLST gene set (**Figure 2 right**) gave highly similar results. As expected, the two major lineages of *L. monocytogenes* corresponded to the deepest branches and were clearly separated based on both gene sets. Based on MvLST genes, both lineages showed unequally long branches, consistent with a higher impact of recombination on MvLST genes. Nevertheless, both gene sets grouped strains into the same shallow branches, and MLST-defined CCs were recovered as monophyletic groups based on both gene sets (**Figure 2**). We also classified isolates into MvLST clonal complexes using their 8-digit allelic profile. These groupings were highly concordant with MLST clonal complexes, as illustrated by the adjusted Rand coefficient (0.988, 95% confidence interval: 0.976-1.000; and 0.981 (0.964-1.000) when excluding gene *ldh*, see below). These results show that groupings of *L. monocytogenes* isolates obtained based on either MLST or MvLST gene sets are nearly identical.

Source-sink selection at gene *ldh* and impact on CC definitions

For most genes, all isolates of a given clonal complex shared the same allelic sequence (**Table 3**). However, gene *ldh* was a notable exception, as this gene was highly variable

361 within CCs, resulting in a large proportion of STs differing from the central genotype of each
362 CC solely by gene *ldh*. We observed that almost all *ldh* nucleotide differences within CCs
363 implied an amino-acid change in the deduced protein sequence (**Table S1**). Further, all
364 amino-acid changes were localized in the 3'-end (codon numbers 93 to 147) of the sequenced
365 *ldh* region, consistent with the results obtained with OMEGAMAP (**Suppl. Figure 1**). In
366 addition, we observed the same amino-acid substitutions in two pairs of isolates from
367 different CCs: change Y95C was observed both in CC1 (ST252) and in CC8 (ST289),
368 whereas change S120R was found both in CC1 (ST248) and CC7 (ST85). These evolutionary
369 parallelisms represent strong evidence for positive selection on *ldh* and contribute to the
370 atypically high diversity of this gene within CCs. Interestingly, the 93 – 137 region within
371 which changes were observed, includes a variable flexible loop of enzyme lactate
372 dehydrogenase in which amino acid substitutions were associated with adaptation to cold
373 temperatures in teleost fishes (19). In striking contrast to the observed pattern of intra-clone
374 variation, the amino acid sequence of lactate dehydrogenase was unchanged among the
375 central genotype of most *L. monocytogenes* clonal complexes and was even conserved
376 between lineages I and II. This absence of diversification at the amino-acids level in the long
377 term indicates strong purifying selection on *ldh*. Together, these results indicate that *ldh*
378 evolves in a source-sink fashion, whereby variants are selected in the short term because they
379 provide a selective advantage in a sink environment, but these variants are counter-selected in
380 the source environment, which sustains the population in the long run. Strikingly, the time
381 since initial isolation of strains with amino acid substitution in lactate dehydrogenase was
382 significantly larger than for non-mutated isolates ($p = 0.024$, two-sided Mann-Whitney test).
383 No substitution was observed in isolates collected after year 2000 (**Table S1**). We
384 hypothesize that non-synonymous mutations in gene *ldh* could have been selected during
385 long-term storage in the laboratory and would not represent naturally occurring variation.

386 The positive selection pattern at the 3' end of the *ldh* sequence led us to evaluate the
387 impact of this gene on the assignment of isolates to clonal complexes. We selected the 863
388 isolates and reference strains previously analyzed herein and in our previous studies (2, 7, 8,
389 33, 47). Classification of the isolates into CCs was performed using either the 7 MLST genes
390 (**Figure 1**) or the six genes after excluding *ldh* (MLST-6). For the 515 isolates of lineage I,
391 there was a total correspondence between CCs obtained based on 7 or 6 genes. Similarly, in
392 lineage III, there was no impact of exclusion of gene *ldh* on the grouping of the 25 isolates.
393 However, when analyzing the 318 isolates of lineage II based on MLST-6, a few
394 discrepancies were observed compared to MLST-7. First, CC8 and CC16 became merged into

a single CC, as these two CCs differed by *ldh* (alleles *ldh*-3 and *ldh*-2, respectively). Similarly, CC11 (*ldh*-2) and CC19 (*ldh*-24) became merged using MLST-6, and ST226 (*ldh*-96) became a member of CC177 (*ldh*-26). Finally, ST126 (*ldh*-65) became part of CC7 (*ldh*-2). Among the *ldh* alleles that affected CC membership, only *ldh*-65 has a non-synonymous change (H121R) that could potentially have been selected. Therefore, the positive selection on the 3' region of the *ldh* sequence potentially impacted the classification of only a single isolate (SLCC875, of ST126) out of the 863 isolates. The adjusted Rand coefficient between MLST-7 and MLST-6 was 0.992 (95% confidence interval: 0.979-1.000). We conclude that the use of gene *ldh* in MLST does not impact importantly the classification of isolates into clonal complexes.

Minimal age of clones and heterogeneity of pulsed-field gel electrophoresis (PFGE)

Although we believe it was not the intention of the author who coined the name 'epidemic clone' (28), this denomination might suggest that such groups undergo recent emergence and have diffused within epidemiological timescales. To provide a minimal age of clones, we searched the public MLST database (as of July 8th, 2013 – 2,586 isolates) for the oldest isolate of the central ST of each major clone. ST1, ST2, ST3, ST4, ST5, ST6, ST7, ST8 and ST9 were isolated as early as 1936, 1954, 1955, 1965, 1992, 1990, 1927, 1991 and 1949, respectively. These dates are several decades older than the first described outbreak of the corresponding 'epidemic clone', and it is well possible that the central STs existed much earlier.

In order to estimate the diversity within clones, 110 isolates (all except the genome references) were analyzed by PFGE using two enzymes separately. A total of 65 *Apa*I types, 48 *Asc*I types and 76 combined PFGE types were distinguished (Suppl. Figure 2). PFGE patterns within an MLST-defined CC were more similar among themselves than with patterns of other clones. This was illustrated by cluster analysis of combined *Apa*I and *Asc*I PFGE patterns (Suppl. Figure 2), which grouped isolates of each clone within a single branch. One exception was the grouping of the two CC558 strains from the 2000 Mexican-style fresh cheese outbreak, which were clustered within the CC1 branch (Figure S2). As ST1 and ST558 are not closely related (differing by five out of seven MLST genes), this observation might suggest an evolutionary convergence of the PFGE profiles of clones CC1 and CC558.

Although they clustered together, the PFGE patterns within a given clonal group showed a high level of heterogeneity (Table 5). First, the seven clonal groups represented by multiple isolates displayed 2 to 21 distinct PFGE types, when combining profiles obtained

429 with *ApaI* and *AscI* enzymes. CC1 and CC2, which both comprised 25 analyzed isolates, were
430 the most heterogeneous, with 21 and 17 distinct PFGE profiles, respectively. Furthermore, the
431 most unrelated profiles within these CCs differed by more than 20 bands when considering
432 both *ApaI* and *AscI* patterns (**Table 5**). These results show that PFGE has higher
433 discriminatory power than MLST, and that CCs exhibit high levels of PFGE diversity. As a
434 matter of fact, isolates that differ by three or more PFGE bands are generally considered as
435 being epidemiologically unrelated (51). Therefore, the PFGE diversity observed within *L.*
436 *monocytogenes* clones confirms the observation based on date and country of origin, that most
437 isolates that share the same or closely related MLST or MvLST type, are epidemiologically
438 unrelated. As there is no evidence for epidemiological links between most outbreaks of
439 *L. monocytogenes* caused by the same EC, the 'epidemic clone' denomination is potentially
440 misleading. Instead, these results favour the view that 'epidemic clones' are evolutionary
441 successful lineages that are widely distributed, and caused multiple independent outbreaks
442 due to their large distribution (8, 12, 28, 46).

443 It is possible that 'epidemic clones' might in fact represent genetically distinct
444 subgroups within MLST defined CCs (8), even though the MLST and MvLST genes cannot
445 discriminate such subgroups. Comparison of PFGE patterns showed that among isolates of
446 CC1, the strains representative of the 1981 Nova Scotia coleslaw outbreak were clearly
447 distinct from the representatives of the two other CC1 outbreaks (1987 Vacherin Mont d'Or
448 and 1985 Jalisco outbreaks), which were more similar among each other. Within CC2, the
449 1987 UK and Ireland pâté outbreak and the 1983 Massachusetts milk outbreak strains were
450 undistinguishable. Finally within CC6, the strains from the 2002 multistate ready-to-eat meat
451 outbreak and from the 1998 hot-dog outbreak were distinguished by a conspicuous doublet
452 when *AscI* enzyme was used (**Suppl. Figure 2**). These results show that within a single CC,
453 strains with distinct PFGE patterns were implicated in outbreaks. Clearly, the within-clone
454 phylogenetic relationships need to be established more precisely, using the high-resolution of
455 whole genome sequencing (48, 60), to determine whether they are structured into subgroups
456 with distinctive involvements in outbreaks.

457 **Conclusions**

458 Research into the pathogenicity or epidemiology differences among *L. monocytogenes*
459 clonal groups would benefit from a widely accepted way to define clonal groups, as well as a
460 reference system to name them. Serotyping has provided a widely shared language, but
461 unfortunately this approach lacks discrimination and is unreliable as a classification system.
462

463 Similarly, the recognition of 'epidemic clones' as groups of similar isolates causing multiple
464 outbreaks, has been invaluable for research on *L. monocytogenes* epidemiology and biology.
465 However, the concept of epidemic clone was not designed to become a common
466 nomenclature system for *L. monocytogenes*. First, it only includes clones involved in
467 outbreaks, whereas most listeriosis cases are sporadic, thus excluding an important fraction of
468 *L. monocytogenes* diversity. Second, there is no consensual definition of an epidemic clone.
469 For example, authors either accept gene sequence variation within clones (12), or not (5, 6,
470 29, 37, 38). In particular, ECIV to ECVII were defined based on complete identity at the six
471 MvLST genes. Strict application of this definition led, for example, to the exclusion of strain
472 Scott A, based on a single nucleotide change in *lisR*, from ECIV (6). As shown here, this
473 strain is otherwise identical to other strains of CC2, EC1a and ECIV at all other MvLST
474 genes, as well as at MLST genes, suggesting that its exclusion from ECIV is too restrictive.
475 An early definition of clone in a medical microbiology context was given by Orskov as
476 genetically similar (but not identical) isolates from unrelated places and times that share many
477 common features (44). The MLST-based definition of clones as CCs (18, 39) is in line with
478 this definition, as it allows one isolate to belong to a CC even if it differs at one out of six
479 genes from another member of the clone. The flexible definition of CCs thus accommodates
480 random accumulation of variation within clones. In fact, MLST has initially been devised to
481 provide a reliable way to recognize and define clonal groups (18, 39).

482 In this work, we have shown that reference isolates of distinct epidemic clones belong
483 to distinct CCs. However, whereas all ECs have a corresponding group based on MLST, only
484 a small fraction of MLST diversity is covered by the EC nomenclature (**Figure 1**). We
485 demonstrated that MvLST genes evolve at a similar rate as MLST genes, implying that both
486 approaches will delimit equivalent groups of isolates. Finally, we showed that CCs represent
487 ancient, diverse and globally distributed genetic groups. This result is in line with the view of
488 CCs in many bacterial species (1, 18, 39).

489 Given the above, we regard EC denominations as a redundant but partial and less
490 flexible denomination system for MLST groups. Given the close correspondence between
491 MLST CCs and ECs, the identification methods developed previously for specific ECs (4, 9,
492 16, 17, 29, 55, 58) should apply readily to the distinction of MLST clones. Unlike MvLST,
493 MLST data are publicly available for comparison through a dedicated web site, providing a
494 continuously growing overview of *L. monocytogenes* diversity. The MLST database
495 (<http://www.pasteur.fr/mlst>) currently contains 2,586 isolates and is being used by an
496 increasing number of laboratories (2, 7, 8, 29, 30, 38, 43, 45, 47, 49, 53, 54) for epidemiology

497 and population biology of this pathogen, thus providing *de facto* a common language for the
498 denomination of clonal groups of *L. monocytogenes*.
499

500 **Acknowledgments**

501 We thank Martin Wiedmann for providing the ILSI reference strain collection. Estimations of
502 the minimal age of clonal complexes was improved thanks to deposition by Jana Haase and
503 Mark Achtman of multilocus sequence typing data of ancient isolates into the Institut Pasteur
504 MLST database. The National Reference Center for *Listeria* and Biology of Infection Unit
505 acknowledge support by Institut Pasteur, Institut de Veille Sanitaire, Institut National de la
506 Santé et de la Recherche Médicale, Fondation pour la Recherche Médicale, Ville de Paris,
507 Fondation BNP Paribas, the Listress EU program, LabEx IBEID and the European Research
508 Council.

509

510 **Funding**

511 This work was supported financially by Institut Pasteur and Institut de Veille Sanitaire.

512

Figure legends**Figure 1. Minimum spanning tree of MLST data for 863 *L. monocytogenes* isolates.**

Each circle represents one sequence type (ST), the size of which is related to the number of isolates with this ST, as indicated (111 represents ST1, the central ST of clonal complex CC1). Red sectors represent the 125 isolates or reference strains included in this work, whereas white sectors represent previously published strains. Links between circles are represented according to the number of allelic mismatches between STs, as indicated. Grey zones surrounding groups of STs represent clonal complexes (CC). Most CC numbers are indicated; those CCs that correspond to 'epidemic clones' (EC) are written in bold with corresponding EC number in parentheses. Laboratory reference strains are indicated in blue. STs of outbreak reference strains are indicated by colored triangles as given on the right. Top, lineage II (starting with CC37); bottom, lineage I (starting with CC195).

Figure 2. Compared phylogenies based on MLST (left) and MvLST (right) gene sequences of the 125 study isolates.

Phylogenies were obtained based on concatenated gene sequences using the neighbour joining method based on uncorrected p-distance. Branches corresponding to clonal complexes are colored according to the central legend. The position of reference strains is indicated.

References

1. **Achtman, M.** 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* **62**:53-70.
2. **Adgamov, R., E. Zaytseva, J. M. Thiberge, S. Brisse, and S. Ermolaeva.** 2012. Genetically related *Listeria monocytogenes* strains isolated from lethal human cases and wild animals, p. 235-250. *In* M. Caliskan (ed.), Genetic Diversity in Microorganisms. InTech.
3. **Carrico, J. A., C. Silva-Costa, J. Melo-Cristino, F. R. Pinto, H. de Lencastre, J. S. Almeida, and M. Ramirez.** 2006. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. *J Clin Microbiol* **44**:2524-32.
4. **Chen, Y., and S. J. Knabel.** 2008. Prophages in *Listeria monocytogenes* contain single-nucleotide polymorphisms that differentiate outbreak clones within epidemic clones. *J Clin Microbiol* **46**:1478-84.
5. **Chen, Y., W. Zhang, and S. J. Knabel.** 2005. Multi-virulence-locus sequence typing clarifies epidemiology of recent listeriosis outbreaks in the United States. *J Clin Microbiol* **43**:5291-4.
6. **Chen, Y., W. Zhang, and S. J. Knabel.** 2007. Multi-virulence-locus sequence typing identifies single nucleotide polymorphisms which differentiate epidemic clones and outbreak strains of *Listeria monocytogenes*. *J Clin Microbiol* **45**:835-46.
7. **Chenal-Francisque, V., L. Diancourt, T. Cantinelli, V. Passet, C. Tran-Hykes, H. Bracq-Dieye, A. Leclercq, C. Pourcel, M. Lecuit, and S. Brisse.** 2013. An optimized MLVA assay and its complementarity with PFGE and MLST for *Listeria monocytogenes* clone identification and surveillance. *J Clin Microbiol* **61**.
8. **Chenal-Francisque, V., J. Lopez, T. Cantinelli, V. Caro, C. Tran, A. Leclercq, M. Lecuit, and S. Brisse.** 2011. Worldwide distribution of major clones of *Listeria monocytogenes*. *Emerg Infect Dis* **17**:1110-2.
9. **Cheng, Y., J. W. Kim, S. Lee, R. M. Siletzky, and S. Kathariou.** 2010. DNA probes for unambiguous identification of *Listeria monocytogenes* epidemic clone II strains. *Appl Environ Microbiol* **76**:3061-8.
10. **den Bakker, H. C., B. N. Bundrant, E. D. Fortes, R. H. Orsi, and M. Wiedmann.** 2010. A population genetics-based and phylogenetic approach to understanding the evolution of virulence in the genus *Listeria*. *Appl Environ Microbiol* **76**:6085-100.
11. **den Bakker, H. C., X. Didelot, E. D. Fortes, K. K. Nightingale, and M. Wiedmann.** 2008. Lineage specific recombination rates and microevolution in *Listeria monocytogenes*. *BMC Evol Biol* **8**:277.
12. **den Bakker, H. C., E. D. Fortes, and M. Wiedmann.** 2010. Multilocus sequence typing of outbreak-associated *Listeria monocytogenes* isolates to identify epidemic clones. *Foodborne Pathog Dis* **7**.
13. **Denny, J., and J. McLauchlin.** 2008. Human *Listeria monocytogenes* infections in Europe--an opportunity for improved European surveillance. *Euro Surveill* **13**.
14. **Disson, O., S. Grayo, E. Huillet, G. Nikitas, F. Langa-Vives, O. Dussurget, M. Ragon, A. Le Monnier, C. Babinet, P. Cossart, and M. Lecuit.** 2008. Conjugated action of two species-specific invasion proteins for fetoplacental listeriosis. *Nature* **455**:1114-8.
15. **Doumith, M., C. Buchrieser, P. Glaser, C. Jacquet, and P. Martin.** 2004. Differentiation of the major *Listeria monocytogenes* serovars by multiplex PCR. *J Clin Microbiol* **42**:3819-22.

- 583 16. **Ducey, T. F., B. Page, T. Usgaard, M. K. Borucki, K. Pupedis, and T. J. Ward.**
584 2007. A single-nucleotide-polymorphism-based multilocus genotyping assay for
585 subtyping lineage I isolates of *Listeria monocytogenes*. *Appl Environ Microbiol*
586 **73**:133-47.
- 587 17. **Evans, M. R., B. Swaminathan, L. M. Graves, E. Altermann, T. R.**
588 **Klaenhammer, R. C. Fink, S. Kernodle, and S. Kathariou.** 2004. Genetic markers
589 unique to *Listeria monocytogenes* serotype 4b differentiate epidemic clone II (hot dog
590 outbreak strains) from other lineages. *Appl Environ Microbiol* **70**:2383-90.
- 591 18. **Feil, E. J.** 2004. Small change: keeping pace with microevolution. *Nat. Rev.*
592 *Microbiol.* **2**:483-95.
- 593 19. **Fields, P. A., and G. N. Somero.** 1998. Hot spots in cold adaptation: localized
594 increases in conformational flexibility in lactate dehydrogenase A4 orthologs of
595 Antarctic notothenioid fishes. *Proc Natl Acad Sci U S A* **95**:11476-81.
- 596 20. **Fugett, E., E. Fortes, C. Nnoka, and M. Wiedmann.** 2006. International Life
597 Sciences Institute North America *Listeria monocytogenes* Strain Collection:
598 Development of Standard *Listeria monocytogenes* Strain Sets for Research and
599 Validation Studies. *J. Food Prot.* **69**:2929-2938.
- 600 21. **Gaillard, J. L., P. Berche, C. Frehel, E. Gouin, and P. Cossart.** 1991. Entry of
601 *Listeria monocytogenes* into cells is mediated by internalin, a repeat protein
602 reminiscent of surface antigens from gram-positive cocci. *Cell* **65**:1127-41.
- 603 22. **Gilmour, M. W., M. Graham, G. Van Domselaar, S. Tyler, H. Kent, K. M. Trout-**
604 **Yakel, O. Larios, V. Allen, B. Lee, and C. Nadon.** 2010. High-throughput genome
605 sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne
606 outbreak. *BMC Genomics* **11**:120.
- 607 23. **Goulet, V., C. Hedberg, A. Le Monnier, and H. de Valk.** 2008. Increasing
608 incidence of listeriosis in France and other European countries. *Emerg Infect Dis*
609 **14**:734-40.
- 610 24. **Graves, L. M., and B. Swaminathan.** 2001. PulseNet standardized protocol for
611 subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel
612 electrophoresis. *Int J Food Microbiol* **65**:55-62.
- 613 25. **Jacquet, C., M. Doumith, J. I. Gordon, P. M. Martin, P. Cossart, and M. Lecuit.**
614 2004. A molecular marker for evaluating the pathogenic potential of foodborne
615 *Listeria monocytogenes*. *J Infect Dis* **189**:2094-100.
- 616 26. **Jonquieres, R., H. Bierne, J. Mengaud, and P. Cossart.** 1998. The *inlA* gene of
617 *Listeria monocytogenes* LO28 harbors a nonsense mutation resulting in release of
618 internalin. *Infect Immun* **66**:3420-2.
- 619 27. **Kathariou, S.** 2003. Foodborne outbreaks of listeriosis and epidemic-associated
620 lineages of *Listeria monocytogenes*, p. 243-256. *In* M. E. Torrence and R. E. Isaacson
621 (ed.), *Microbial food safety in animal agriculture*. Iowa State University Press, Ames,
622 IA.
- 623 28. **Kathariou, S.** 2002. *Listeria monocytogenes* virulence and pathogenicity, a food
624 safety perspective. *J Food Prot* **65**:1811-29.
- 625 29. **Knabel, S. J., A. Reimer, B. Verghese, M. Lok, J. Ziegler, J. Farber, F. Pagotto,**
626 **M. Graham, C. A. Nadon, and M. W. Gilmour.** 2012. Sequence typing confirms
627 that a predominant *Listeria monocytogenes* clone caused human listeriosis cases and
628 outbreaks in Canada from 1988 to 2010. *J Clin Microbiol* **50**:1748-51.
- 629 30. **Koopmans, M. M., M. C. Brouwer, M. W. Bijlsma, S. Bovenkerk, W. Keijzers, A.**
630 **van der Ende, and D. van de Beek.** 2013. *Listeria monocytogenes* Sequence Type 6
631 and Increased Rate of Unfavorable Outcome in Meningitis: Epidemiologic Cohort
632 Study. *Clin Infect Dis* **57**:247-53.

- 633 31. **Kuhn, M., S. Kathariou, and W. Goebel.** 1988. Hemolysin supports survival but not
634 entry of the intracellular bacterium *Listeria monocytogenes*. *Infect Immun* **56**:79-82.
- 635 32. **Lebrun, M., J. Mengaud, H. Ohayon, F. Nato, and P. Cossart.** 1996. Internalin
636 must be on the bacterial surface to mediate entry of *Listeria monocytogenes* into
637 epithelial cells. *Mol Microbiol* **21**:579-92.
- 638 33. **Leclercq, A., V. Chenal-Francisque, H. Dieye, T. Cantinelli, R. Drali, S. Brisse,
639 and M. Lecuit.** 2011. Characterization of the novel *Listeria monocytogenes* PCR
640 serogrouping profile IVb-v1. *Int J Food Microbiol* **147**:74-7.
- 641 34. **Lecuit, M., H. Ohayon, L. Braun, J. Mengaud, and P. Cossart.** 1997. Internalin of
642 *Listeria monocytogenes* with an intact leucine-rich repeat region is sufficient to
643 promote internalization. *Infect Immun* **65**:5309-19.
- 644 35. **Lecuit, M., S. Vandormael-Pournin, J. Lefort, M. Huerre, P. Gounon, C. Dupuy,
645 C. Babinet, and P. Cossart.** 2001. A transgenic model for listeriosis: role of
646 internalin in crossing the intestinal barrier. *Science* **292**:1722-5.
- 647 36. **Librado, P., and J. Rozas.** 2009. DnaSP v5: a software for comprehensive analysis of
648 DNA polymorphism data. *Bioinformatics* **25**:1451-2.
- 649 37. **Lomonaco, S., Y. Chen, and S. J. Knabel.** 2008. Analysis of additional virulence
650 genes and virulence gene regions in *Listeria monocytogenes* confirms the
651 epidemiologic relevance of multi-virulence-locus sequence typing. *J Food Prot*
652 **71**:2559-66.
- 653 38. **Lomonaco, S., B. Verghese, P. Gerner-Smidt, C. Tarr, L. Gladney, L. Joseph, L.
654 Katz, M. Turnsek, M. Frace, Y. Chen, E. Brown, R. Meinersmann, M. Berrang,
655 and S. Knabel.** 2013. Novel epidemic clones of *Listeria monocytogenes*, United
656 States, 2011. *Emerg Infect Dis* **19**:147-50.
- 657 39. **Maiden, M. C.** 2006. Multilocus sequence typing of bacteria. *Annu Rev Microbiol*
658 **60**:561-88.
- 659 40. **McVean, G., P. Awadalla, and P. Fearnhead.** 2002. A coalescent-based method for
660 detecting and estimating recombination from gene sequences. *Genetics* **160**:1231-41.
- 661 41. **Murray, E. G. D., R. A. Webb, and M. B. R. Swann.** 1926. A disease of rabbits
662 characterised by a large mononuclear leucocytosis, caused by a hitherto undescribed
663 bacillus *Bacterium monocytogenes* (n.sp.). *The Journal of Pathology and Bacteriology*
664 **29**:407-439.
- 665 42. **Nightingale, K. K., R. A. Ivy, A. J. Ho, E. D. Fortes, B. L. Njaa, R. M. Peters, and
666 M. Wiedmann.** 2008. inlA premature stop codons are common among *Listeria*
667 *monocytogenes* isolates from foods and yield virulence-attenuated strains that confer
668 protection against fully virulent strains. *Appl Environ Microbiol* **74**:6570-83.
- 669 43. **Nilsson, R. E., T. Ross, J. P. Bowman, and M. L. Britz.** 2013. MudPIT profiling
670 reveals a link between anaerobic metabolism and the alkaline adaptive response of
671 *Listeria monocytogenes* EGD-e. *PLoS One* **8**:e54157.
- 672 44. **Orskov, F., and I. Orskov.** 1983. From the national institutes of health. Summary of
673 a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the
674 enterobacteriaceae and other bacteria. *J Infect Dis* **148**:346-57.
- 675 45. **Parisi, A., L. Latorre, G. Normanno, A. Miccolupo, R. Fraccalvieri, V. Lorusso,
676 and G. Santagada.** 2010. Amplified Fragment Length Polymorphism and Multi-
677 Locus Sequence Typing for high-resolution genotyping of *Listeria monocytogenes*
678 from foods and the environment. *Food Microbiol* **27**:101-8.
- 679 46. **Piffaretti, J. C., H. Kressebuch, M. Aeschbacher, J. Bille, E. Bannerman, J. M.
680 Musser, R. K. Selander, and J. Rocourt.** 1989. Genetic characterization of clones of
681 the bacterium *Listeria monocytogenes* causing epidemic disease. *Proc Natl Acad Sci U*
682 *S A* **86**:3818-22.

- 683 47. **Ragon, M., T. Wirth, F. Hollandt, R. Lavenir, M. Lecuit, A. Le Monnier, and S.**
684 **Brisse.** 2008. A new perspective on *Listeria monocytogenes* evolution. *PLoS Pathog*
685 **4:e1000146.**
- 686 48. **Roetzer, A., R. Diel, T. A. Kohl, C. Ruckert, U. Nubel, J. Blom, T. Wirth, S.**
687 **Jaenicke, S. Schuback, S. Rusch-Gerdes, P. Supply, J. Kalinowski, and S.**
688 **Niemann.** 2013. Whole genome sequencing versus traditional genotyping for
689 investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular
690 epidemiological study. *PLoS Med* **10:e1001387.**
- 691 49. **Salcedo, C., L. Arreaza, B. Alcala, L. de la Fuente, and J. A. Vazquez.** 2003.
692 Development of a multilocus sequence typing method for analysis of *Listeria*
693 *monocytogenes* clones. *J Clin Microbiol* **41:757-62.**
- 694 50. **Sperry, K. E., S. Kathariou, J. S. Edwards, and L. A. Wolf.** 2008. Multiple-locus
695 variable-number tandem-repeat analysis as a tool for subtyping *Listeria*
696 *monocytogenes* strains. *J Clin Microbiol* **46:1435-50.**
- 697 51. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H.**
698 **Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction
699 patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain
700 typing. *J Clin Microbiol* **33:2233-9.**
- 701 52. **Travier, L., S. Guadagnini, E. Gouin, A. Dufour, V. Chenal-Francisque, P.**
702 **Cossart, J. C. Olivo-Marin, J. M. Ghigo, O. Disson, and M. Lecuit.** 2013. ActA
703 promotes *Listeria monocytogenes* aggregation, intestinal colonization and carriage.
704 *PLoS Pathog* **9:e1003131.**
- 705 53. **Wang, P., H. Yang, Y. Hu, F. Yuan, G. Zhao, Y. Zhao, and Y. Chen.** 2012.
706 Characterization of *Listeria monocytogenes* isolates in import food products of China
707 from 8 provinces between 2005 and 2007. *J Food Sci* **77:M212-6.**
- 708 54. **Wang, Y., A. Zhao, R. Zhu, R. Lan, D. Jin, Z. Cui, Z. Li, J. Xu, and C. Ye.** 2012.
709 Genetic diversity and molecular typing of *Listeria monocytogenes* in China. *BMC*
710 *Microbiol* **12:119.**
- 711 55. **Ward, T. J., T. Usgaard, and P. Evans.** 2010. A targeted multilocus genotyping
712 assay for lineage, serogroup, and epidemic clone typing of *Listeria monocytogenes*.
713 *Appl Environ Microbiol* **76:6680-4.**
- 714 56. **Wiedmann, M., J. L. Bruce, C. Keating, A. E. Johnson, P. L. McDonough, and C.**
715 **A. Batt.** 1997. Ribotypes and virulence gene polymorphisms suggest three distinct
716 *Listeria monocytogenes* lineages with differences in pathogenic potential. *Infect*
717 *Immun* **65:2707-16.**
- 718 57. **Wilson, D. J., and G. McVean.** 2006. Estimating diversifying selection and
719 functional constraint in the presence of recombination. *Genetics* **172:1411-25.**
- 720 58. **Yildirim, S., W. Lin, A. D. Hitchins, L. A. Jaykus, E. Altermann, T. R.**
721 **Klaenhammer, and S. Kathariou.** 2004. Epidemic clone I-specific genetic markers
722 in strains of *Listeria monocytogenes* serotype 4b from foods. *Appl Environ Microbiol*
723 **70:4158-64.**
- 724 59. **Zhang, W., B. M. Jayarao, and S. J. Knabel.** 2004. Multi-virulence-locus sequence
725 typing of *Listeria monocytogenes*. *Appl Environ Microbiol* **70:913-20.**
- 726 60. **Zhou, Z., A. McCann, E. Litrup, R. Murphy, M. Cormican, S. Fanning, D.**
727 **Brown, D. S. Guttman, S. Brisse, and M. Achtman.** 2013. Neutral genomic
728 microevolution of a recently emerged pathogen, *Salmonella enterica* serovar Agona.
729 *PLoS Genet* **9:e1003471.**
- 730
- 731

Figure 1

The figure displays a phylogenetic tree of *E. coli* O157 strains, with nodes representing clinical cases (CC) and branches representing the genetic relationships between them. The tree is rooted at the bottom left with a node labeled '111'. The tree is divided into several clusters, each associated with a specific outbreak strain. The clusters are labeled as follows:

- CC199, CC177, CC20, CC193, CC19, CC155, CC16, CC21, CC26, CC121, CC101, CC90:** This cluster is associated with the outbreak strain 08-5578, Canada, 2008 (represented by a purple triangle). Other labels include EGD-e, L028, and FSLN3-165.
- CC8 (ECV), CC7 (ECVII), CC11 (ECIII), CC9, CC37, CC195, CC2 (ECIa or ECIV), CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain TS4/F.6854, USA 1989 (represented by a yellow triangle). Other labels include 10403S, EGD, and FSLJ2-064, FSLJ1-194, FSLJ1-175.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain G6054, USA 1994 (represented by a green triangle). Other labels include ST558 and FSLJ2-064.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain CLIP 80459, France 1999 (represented by a purple triangle). Other labels include ST558 and FSLJ2-064.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain H7550, USA 1998-99 (represented by a yellow triangle). Other labels include ST558 and FSLJ2-064.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain J0211, USA 2000-2001 (represented by a yellow triangle). Other labels include ST558 and FSLJ2-064.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain Scott A, USA, 1983 (represented by a purple triangle). Other labels include ST558 and FSLJ2-064.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain Aureli97, Italy 1987 (represented by a purple triangle). Other labels include ST558 and FSLJ2-064.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain TS38/L.3306, UK & Ireland 1987-89 (represented by a purple triangle). Other labels include ST558 and FSLJ2-064.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain TS27/L.4738, Canada 1981 (represented by a green triangle). Other labels include ST558 and FSLJ2-064.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain TS60/L.4486b, Switzerland, 1983-87 (represented by a green triangle). Other labels include ST558 and FSLJ2-064.
- CC195, CC4, CC1 (ECI), CC6 (ECII), CC3, CC5 (ECVI), CC315:** This cluster is associated with the outbreak strain F2365, California, 1986-1987 (represented by a green triangle). Other labels include ST558 and FSLJ2-064.

Legend:

- 111
- 62
- 17
- 1
- 0

Line styles:

- 1
- 2
- 3
- 4
- > 4

Outbreak strains:

- 08-5578, Canada, 2008
- TS4/F.6854, USA 1989
- J0161, USA 2000
- G6054, USA 1994
- CLIP 80459, France 1999
- H7550, USA 1998-99
- J1735, USA, 2002
- J0211, USA 2000-2001
- Scott A, USA, 1983
- Aureli97, Italy 1987
- TS38/L.3306, UK & Ireland 1987-89
- TS27/L.4738, Canada 1981
- TS60/L.4486b, Switzerland, 1983-87
- F2365, California, 1986-1987

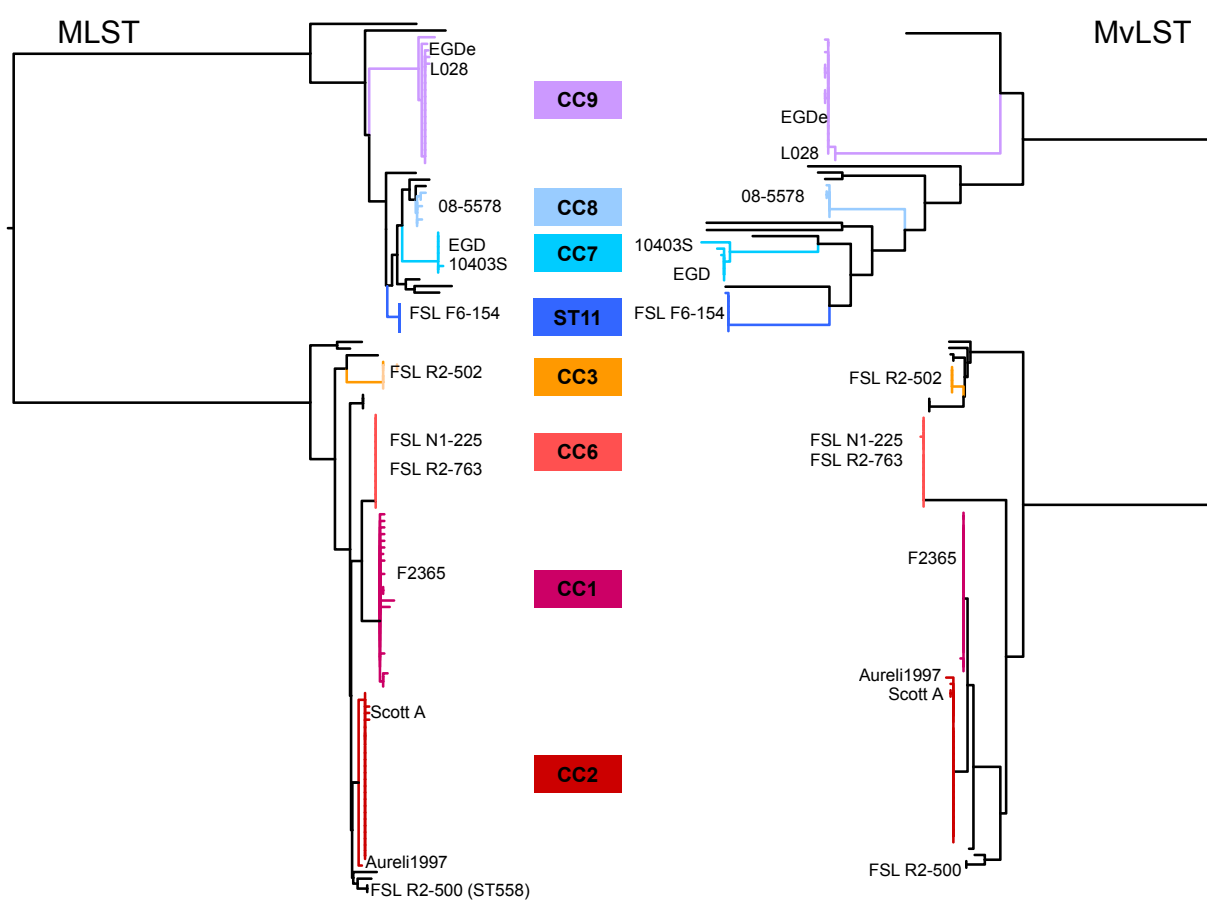


Table 1. Primers used for MvLST.

Gene	Forward primer	Reverse primer	Location (EGD-e)	Size of PCR product (bp)	Size of template (bp)	Reference
<i>actA</i>	CGA CAT AAT ATT TGC AGC GAC A	GAA TCT AAG TCA CTT TCA GAA GCA T	209,541 to 210,041	500	450	This study
<i>clpP</i>	CCA ACA GTA ATT GAA CAA ACT AGC C	GAT CTG TAT CGC GAG CAA TG	2,542,026 to 2,542,524	498	419	Zhang & al (2004)
<i>daf</i>	GAA GGT ATC TAC ACG CAT TTT GC	GCC AAT TAT CGT TAC TTT TGA ACC	925,391 to 925,909	518	428	This study
<i>iniB</i>	CAT GGG AGA GTA ACC CAA CC	GCG GTA ACC CCT TTG TCA TA	457,954 to 458,453	499	432	Zhang & al (2004)
<i>iniC</i>	AAC CAT CTA CAT AAC TCC CAC	CGG GAA TGC AAT TTT TCA CTA	1,860,322 to 1,860,822	500	457	Zhang & al (2004)
<i>lisR</i>	CGG GGT AGA AGT TTG TCG TC	ACG CAT CAC ATA CCC TGT CC	1,402,726 to 1,403,224	498	447	Zhang & al (2004)
<i>prfA</i>	TGC GAT GCC ACT TGA ATA TC	AAC GGG ATA AAA CCA AAA CCA	203,811 to 204,311	500	460	Zhang & al (2004)
<i>iniA</i>	CGGATGCAGGAGAAAATCC	CTTTCACACTATCCTCTCC	454,463 to 457,027	2,564	2,403	Ragon & al (2008)

Table 2. Polymorphism of the MLST and MvLST genes, based on the 125 study isolates

	Length	No. alleles	No. Polymorphic sites	% Polymorphic sites	Ks	Ka	Ka/Ks	π (%)
MvLST genes								
<i>actA</i>	450	14	44	9.78	0.07	0.02	0.35	3.21
<i>clpP</i>	419	4	14	3.34	0.06	0.001	0.02	1.44
<i>dal</i>	428	14	54	12.62	0.15	0.002	0.01	3.88
<i>inlA</i>	2403	28	140	5.83	0.06	0.06	1.01	1.76
<i>inlB</i>	432	15	40	9.26	0.03	0.04	1.38	3.38
<i>inlC</i>	457	10	14	3.06	0.03	0.01	0.25	1.18
<i>lisR</i>	447	5	38	8.50	0.02	0	0	0.51
<i>prfA</i>	460	9	20	4.35	0.06	0.00008	0.001	1.37
Concatenate	5,496	40	364	6.62	0.06	0.03	0.53	1.96
MLST genes								
<i>abcZ</i>	537	14	37	6.89	0.08	0.001	0.01	1.96
<i>bglA</i>	399	15	21	5.26	0.04	0.0006	0.01	0.91
<i>cat</i>	486	17	26	5.35	0.08	0.002	0.03	1.93
<i>dapE</i>	462	19	40	8.66	0.11	0.01	0.08	3.27
<i>dat</i>	471	10	61	12.95	0.20	0.01	0.07	5.72
<i>ldh</i>	453	20	39	8.61	0.06	0.0008	0.01	1.37
<i>lhkA</i>	480	10	20	4.17	0.06	0.0026	0.04	1.55
Concatenated	3,294	51	244	7.41	0.09	0.004	0.05	2.41

Ks: No. of synonymous changes per synonymous site. Ka: No. of non-synonymous changes per non-synonymous site.

 π : nucleotide diversity.

Table 3. Distribution of MLST alleles within clones, observed based on the 125 study isolates.											
Lineage	CC	No. isolates	No. of distinct alleles (alleles found)							No. ST	No. ST without <i>ldh</i>
			<i>abcZ</i>	<i>bglA</i>	<i>cat</i>	<i>dapE</i>	<i>dat</i>	<i>ldh</i>	<i>lhkA</i>		
I	CC1	25	1 (3)	2 (1;56)	3 (1;20;60)	4 (1;34;42;44)	4 (3;14;18;42)	9 (1;11;12;61;90;101;103;104;105)	1 (3)	16	10
	CC2	25	1 (1)	2 (1;38)	1 (11)	2 (3;11)	1 (2)	1 (1)	2 (5;8)	4	4
	CC3	5	1 (3)	1 (4)	1 (4)	1 (4)	1 (2)	1 (1)	1 (5)	1	1
	CC6	15	1 (3)	1 (9)	1 (9)	1 (3)	1 (3)	1 (1)	1 (5)	1	1
	Others	10	5 (1;2;11;12;34)	4 (1;2;12;39)	3 (4;11;12)	6 (1;3;14;15;16;38)	2 (2;3)	3 (1;5;39)	5 (2;3;4;5;7)	7	7
II	CC7	6	1 (5)	1 (8)	1 (5)	1 (7)	1 (6)	1 (2;22;38)	1 (1)	3	1
	CC8	6	5 (5;57)	1 (6)	1 (2)	2 (9;29)	1 (5)	3 (3;89;121)	1 (1)	5	3
	CC9	19	2 (6;33)	1 (5)	1 (6)	2 (4;20)	1 (1)	3 (4;57;62)	1 (1)	5	3
	Others	14	5 (5;6;7;21;53)	6 (5;6;10;13;27;52)	9 (2;3;8;10;16;17;24;26;39)	6 (4;6;7;8;21;67)	5 (1;5;6;13;39)	4 (2;3;24;32)	4 (1;6;14;41)	10	10

Table 4. MvLST allelic diversity within clones, observed based on the 125 study isolates											
Lineage	CC	No. strains	No. of distinct alleles (allele number)								No. MVLST types
			<i>actA</i>	<i>clpP</i>	<i>dal</i>	<i>inlA</i>	<i>inlB</i>	<i>inlC</i>	<i>lisR</i>	<i>prfA</i>	
I	CC1	25	2 (14;15)	1 (4)	1 (5)	1 (28)	1 (3)	2 (1;4)	1 (1)	1 (3)	3
	CC2	25	2 (10;13)	1 (4)	1 (5)	1 (28)	1 (4)	1 (1)	3 (1;2;5)	1 (2)	4
	CC3	5	1 (10)	1 (4)	1 (5)	1 (22)	1 (5)	1 (4)	1 (1)	1 (2)	1
	CC6	15	2 (10;12)	1 (4)	1 (5)	1 (29)	1 (3)	1 (4)	1 (3)	2 (2;5)	3
	Others	10	4 (9;10;11;14)	1 (4)	1 (4;5;6;7)	6 (21;23;24;25;26;27)	4 (1;2;6;7)	3 (2;3;4)	1 (3)	4 (1;2;4;6)	7
II	CC7	6	2 (6;8)	1 (2)	2 (11;13)	2 (9;10)	2 (10;13)	2 (10;11)	1 (4)	2 (7;9)	4
	CC8	6	1 (4)	1 (1)	1 (8)	2 (14;15)	2 (9;12)	1 (10)	1 (4)	1 (8)	2
	CC9	19	1 (1)	1 (2)	2 (1;2)	6 (1;2;3;4;5;6)	1 (8)	1 (5)	1 (4)	1 (8)	3
	Others	14	7 (1;2;3;4;5;6;7)	3 (1;2;3)	7 (1;3;8;9;10;12;14)	11 (4;7;8;11;12;13;16;17;18;19;20)	6 (8;9;10;11;14;15)	6 (5;6;7;8;9;10)	2 (4;6)	1 (8)	10

Table 5. Pulsed field gel electrophoresis diversity per clone

Lineage	Clone	No. Isolates	No. of <i>AscI</i> profiles (Max. No. of different bands)	No. of <i>ApaI</i> profiles (Max. No. of different bands)	No. of <i>AscI</i> + <i>ApaI</i> profiles
I	CC1	25	11 (13)	12 (9)	21
	CC2	25	7 (7)	13 (15)	17
	CC3	3	1 (0)	2 (6)	2
	CC6	15	2 (5)	5 (10)	6
	Others	6	5 (10)	5 (12)	5
II	CC7	5	4 (5)	4 (7)	5
	CC8	4	3 (8)	2 (3)	4
	CC9	20	6 (9)	4 (16)	7
	Others	11	9 (12)	8 (13)	9