Prophages in *Listeria monocytogenes* Contain Single-Nucleotide Polymorphisms That Differentiate Outbreak Clones within Epidemic Clones

Yi Chen* and Stephen J. Knabel

*Department of Food Science, The Pennsylvania State University, University Park, Pennsylvania 16802*

Received 20 September 2007/Returned for modification 12 November 2007/Accepted 28 January 2008

A fragment end ligation-mediated PCR strategy was used to analyze the Ascl pulsed-field gel electrophoresis patterns of *Listeria monocytogenes* epidemic clone II (ECII), which led to the identification of single-nucleotide polymorphisms (SNPs) in prophage regions that differentiated the two ECII outbreak clones. SNPs in prophages that differentiated the outbreak clones of ECIII and -IV were also identified.

*Listeria monocytogenes* is an opportunistic intracellular foodborne pathogen which causes listeriosis, a severe disease with a high case fatality rate (20% to 30%), especially in immunocompromised populations. Previous molecular subtyping studies identified four major epidemic clones (ECs) of *L. monocytogenes* (ECI, ECII, ECIII, and ECIV) which caused multiple outbreaks worldwide (3, 8). An EC can be defined as a group of isolates that are part of the same chain of transmission in one epidemic and thus are the descendants of the source strain (2, 13). An outbreak clone can be defined as a group of asexually transmitted isolates associated with one outbreak (this study). Among the ECs of *L. monocytogenes*, ECI is a serotype 1/2a cluster which was associated with outbreaks linked to the consumption of contaminated coleslaw (Nova Scotia, 1981), soft cheese (Switzerland, 1983–1987; California, 1985), and pork tongue (France, 1992) and has possibly been linked to some other outbreaks (8). ECII was associated with an outbreak linked to contaminated hot dogs (United States, 1998–1999) and an outbreak linked to contaminated turkey deli meats (United States, 2002) (2, 9). ECIV is another serotype 4b cluster which caused an outbreak linked to contaminated pâté (United Kingdom, 1988) and an outbreak linked to contaminated vegetables (Boston, 1979) (3). ECIII isolates are serotype 1/2a isolates which caused outbreaks linked to contaminated hot dogs (United States, 1989) and turkey deli meats (United States, 2000) manufactured in the same food processing plant. Isolates within an EC are closely related and often have close or indistinguishable molecular subtypes, as determined by various subtyping techniques such as ribotyping and pulsed-field gel electrophoresis (PFGE) (8). Chen et al. (3) demonstrated that isolates within each EC had identical sequence types, as determined by a multi-virulence-locus sequence typing strategy which targeted internal regions of *prfA*, *inlB*, *inlC*, *clpP*, *dal*, and *lisR* (15). Tracking ECs of *L. monocytogenes* is important for understanding the long-term transmission of this pathogen and establishing efficient surveillance systems for its control. In addition, the differentiation of outbreak clones in the same EC is often needed for short-term epidemiology to identify the source of specific outbreaks and to study the evolution and transmission of different outbreak clones within the same EC. Ducey et al. (4) sequenced seven genomic regions covering 29 kbp of the *L. monocytogenes* genome and identified three single-nucleotide polymorphisms (SNPs) in *hly*, *dnaE*, and *inlB* that could distinguish the three outbreak clones of ECI. Recently completed whole-genome sequences of four ECIII isolates allowed the identification of SNPs that could separate different ECIII outbreak clones (*Listeria monocytogenes* Sequencing Project, Broad Institute of Harvard and MIT [http://www.broad.mit.edu]). However, there have been no reports validating the epidemiological relevance of SNPs in ECIII. In addition, there have been no reports of SNPs that can differentiate the outbreak clones of ECII and ECIV. S. Lomonaco, Y. Chen, and S. J. Knabel (unpublished data) identified hypervariable regions of five housekeeping genes (*dnaE*, *traB*, *purM*, *hisC*, and the phosphotransferase system gene and eight virulence genes (*lplA1*, *pgaA*, *srtA*, *inlJ*, *bsh*, *hly*, *actA*, and *inlB*) and sequenced those regions in ECII and ECIV isolates. However, no sequence variations in these genes were observed in different isolates within ECII or ECIV. Therefore, the primary objective of the present study was to identify and validate novel SNPs that can differentiate outbreak clones within ECs II, III, and IV. These SNPs could then be incorporated into sequence-based or SNP-based subtyping schemes for investigating the epidemiology of *L. monocytogenes*.

For the present study, the identification of SNPs started with ECII because this EC caused two significant U.S. multistate listeriosis outbreaks recently (2). Kathariou et al. (9) utilized whole-genome macroarray analysis to subgroup ECII isolates and found that hybridization array signals based on prophage and internalin regions could subgroup ECII isolates into three different clusters. However, their whole-genome macroarray strategy could not differentiate the 1998 outbreak isolates as a group from the 2002 outbreak isolates (9). Graves et al. (6) and Kathariou (9) reported that ApaI PFGE could subgroup ECII isolates into four closely related clusters, with some clusters containing both the 1998 outbreak isolates and the 2002 outbreak isolates. In summary, isolates within each ECII outbreak

---

* Corresponding author. Mailing address: Center for Food Safety and Applied Nutrition, FDA, HFS-712, 5100 Paint Branch Parkway, College Park, MD 20740. Phone: (301) 436-2783. Fax: (301) 436-2644. E-mail: yi.chen@fda.hhs.gov.

† Published ahead of print on 6 February 2008.
had different genomic macroarray profiles and ApaI PFGE patterns; thus, the isolates within the same outbreak would be considered different strains according to the definition of strain given by the European Study Group on Epidemiological Markers (14). Therefore, we propose “outbreak clone” as a better term than the traditional term “outbreak strain” to describe different strains involved in the same outbreak (6). Previously, AscI PFGE was the only subtyping method that could distinguish the 1998 outbreak isolates as a group from the 2002 outbreak isolates (Fig. 1). However, the molecular mechanism(s) that caused the AscI PFGE banding pattern difference between the 1998 outbreak clone and the 2002 outbreak clone was unknown because the whole-genome sequence of the 2002 outbreak clone was not available. The whole-genome sequences of the 1998 outbreak clone (H7858) and the 2002 outbreak clone (H7550) were obtained from PulseNet. The original isolate identifiers from the CDC were used (courtesy of Bala Swaminathan and PulseNet).

A fragment that can be amplified by genome-wide analysis using oligonucleotide arrays (G-WAONA) was used to amplify the end segments of the 320-kbp and 330-kbp AscI PFGE macrorestriction fragments from J1816, a reference isolate of the 2002 outbreak clone. The AscI PFGE analysis of J1816 was performed by following the standardized protocol of the U.S. Centers for Disease Control and Prevention (CDC) (7). The 320-kbp and 330-kbp fragments were excised together from the PFGE gel, digested in the gel by using the restriction enzyme MspI (New England BioLabs, MA) and following the manufacturer’s instructions, and extracted using a gel extraction kit (Qiagen, Valencia, CA). Extracted DNA was incubated with MspI again for 1 hour to ensure complete digestion. An AscI adapter was generated by annealing two oligonucleotides, 5′-GAATACGGTACTAACCCTGCTA-3′ and 5′-CGCGTAGCAGGT-3′, and the primer that targeted the AscI adapter (referred to as the AscI primer) was 5′-TACCGTACTAACCCTGCTA-3′. An MspI adapter was generated by annealing two oligonucleotides, 5′-GCTGGAAGACGTGAAG-3′ and 5′-CGCTTCAGTCTCT-3′, and the primer that targeted the MspI adapter (referred to as the MspI primer) was 5′-TAGCTGGAAGACGTGAAG-3′. All oligonucleotides were synthesized at the Nucleic Acid Facility at The Pennsylvania State University, and their 5′ ends were not phosphorylated. The adapters were designed such that the ligation between restriction fragments and adapters “killed” the restriction site. The digestion mixture was then mixed with 1 μl of AscI adapter (10 μM) and 1 μl of MspI adapter (10 μM) and incubated with T4 DNA ligase overnight at room temperature by following the manufacturer’s instructions. After ligation, proteins in the reaction mix were precipitated and removed by phenol-chloroform, and DNA was further purified using a MoBio DNA purification kit (MoBio Laboratories, CA). Purified DNA was then mixed with 1 μl of AscI primer (10 μM), 1 μl of MspI primer (10 μM), and PCR master mix (Qiagen, CA) to a final volume of 50 μl. PCR was then performed using a Mastercycler thermocycler (Eppendorf Scientific, Hamburg, Germany) with an initial extension step at 72°C for 3 min, 94°C for 1 min, 45°C for 1 min, and 72°C for 3 min prior to 30 cycles of 1 min at 94°C, 1 min at 45°C, and 1 min at 72°C, followed by a final extension step at 72°C for 8 min. The PCR mixture was then mixed with 6× loading buffer and separated on a 1.5% agarose gel in 0.5× Tris-borate-EDTA buffer. FELM-PCR products were visualized by ethidium bromide staining, excised from the gel, and extracted using a Qiagen gel extraction kit for sequencing. DNA cycle-sequencing reactions were performed at the Pennsylvania State University Shared Nucleic Acid Facility by using an MJ Research Tetrad thermocycler, 3′ BigDye-labeled dyeoxyribonucleotide triphosphates (v3.1 dye terminators), and protocol number 43032337 (Applied Biosystems, Foster City, CA). Cycle-sequencing reaction products were separated and analyzed on an ABI 3730xl DNA analyzer by using the ABI Data Collection program (v2.0). Data were analyzed with ABI Sequencing Analysis software (v5.1.1). Both the MspI primer and the AscI primer were used as sequencing primers in separate runs. The DNA sequences of the FELM-PCR products were compared with the whole-genome sequence of a reference isolate of the 1998 outbreak clone (H7858) by Web-based BLAST, provided by The Institute for Genomic Research (http://www.tigr.org).

FELM-PCR results are shown in Fig. 3. Figure 4 is a sche-
matic of the ~320-kbp and ~330-kbp AscI macrorestriction fragments based on FELM-PCR, sequencing, and BLAST analysis. The discussion below is based on the forward strand only and on the whole-genome sequence of H7858, the reference isolate of the 1998 outbreak clone. Four ends (A, B, C, and D) of these two macrorestriction fragments could have potentially been amplified by FELM-PCR; however, only two of them were amplified (Fig. 3). Sequencing and BLAST analysis revealed that the ~320-bp amplicon (GenBank accession no. EU159666) was homologous to an internal region of ilvA around coordinate 2328434 in H7858 and the ~150-bp amplicon (GenBank accession no. EU159664) was homologous to an internal region of LMOh7858_2426 around coordinate 2649160. LMOh7858_2426 putatively encodes protein gp13 of bacteriophage A118 (11) and is part of the putative prophage ΦH7858.2, the only reported prophage in H7858 (12). The sequences of both amplicons generated by using the AscI primer corresponded to the forward strand of the whole-genome sequence of H7858. Therefore, the AscI restriction sites are upstream of the two amplicons in the forward strand. Given the above findings, the ~320-bp amplicon must correspond to end A and the ~150-bp amplicon must correspond to end C (Fig. 4). In addition, the relative distance between ends A and C in the chromosome is 321 kbp (2,649,160 kbp minus 2,328,434 kbp) (Fig. 4), which corresponds to the ~320-kbp PFGE macrorestriction fragment in Fig. 1. Therefore, ends B and C must be adjacent to each other in the chromosome, and the size of fragment I is 321 kbp (Fig. 4). Ends B and D were not amplified (Fig. 3), perhaps because they were either too long or too short for PCR amplification. The sequence of end C also revealed the point mutation (from 5′-GGCGCCGCGC-3′ to 5′-GGTGCGCGC-3′; mutation indicated in bold) in the AscI restriction site at the beginning of end C (Fig. 4), which caused the AscI PFGE banding pattern difference between J1816 and H7858 (Fig. 1). As sequencing at the end of a fragment is not always reliable, additional PCR and sequencing analyses of the region harboring this AscI restriction site (between ends B and C) were performed to confirm the point mutation. Interest-
ingly, to obtain a relatively short amplicon (<800 bp) for rapid sequencing, two different PCR primer pairs were needed to specifically amplify this region in each isolate, indicating that there were extensive sequence variations in the priming regions between H7858 and J1816 (Fig. 5). Primer pair 2426_F (5'-CAACCGGTGATGGAGTATT-3') and 2426_R (5'-AAA CGTCATTTTTAACCGATG-3') specifically amplified an internal region of LMOh7858_2426 in the 1998 outbreak clone; primer pair 2664_F (5'-CACCTGTACCCGCGCTAT-3') and 2664_R (5'-AGTTTCCGGGAGGGTCTAAAT-3') specifically amplified an LMOh7858_2426-homologous region in the 2002 outbreak clone (Fig. 5). PCR was performed with an activation step at 95°C for 15 min prior to 15 cycles of a touchdown program (94°C for 30 seconds, annealing for 1 min, and extension at 72°C for 30 seconds, with the annealing temperature decreased by 1°C every 3 cycles from 56°C to 52°C) and then 15 cycles of 94°C for 30 seconds, 51°C for 1 min, and 72°C for 30 seconds, followed by one final cycle of 72°C for 8 min. A total of 1 µl of purified DNA (15 ng/µl) was mixed with 1 µl of each primer (10 µM) and PCR master mix (Qiagen) to achieve a final volume of 50 µl for PCR amplification. The sequence comparison of the amplicons revealed multiple SNPs that could differentiate H7858 and J1816, including the point mutation between ends B and C (Table 1; Fig. 4). BLAST analysis and calculation of the size of fragment I, described above, utilized the whole-genome sequence of H7858. Therefore, an underlying assumption of the above analysis is that there was no genomic rearrangement involving fragment I between H7858 (a reference isolate of the 1998 outbreak clone) and J1816 (a reference isolate of the 2002 outbreak clone); otherwise, the distance between ends A and C in J1816 might not be 321 kbp (Fig. 4). To address this issue, we performed whole-genome sequence analysis using 20 *L. monocytogenes* isolates (5, 12; Listeria monocytogenes Sequencing Project, Broad Institute of Harvard and MIT [http://www.broad.mit.edu]) and found that a 306-kbp region between housekeeping genes *ilvA* (harboring end A in Fig. 4) and *yhzC* (upstream of prophage H7858.2) is relatively conserved among all isolates and contains no major genomic rearrangements. Therefore, it is reasonable to conclude that the distance between *ilvA* and *yhzC* in J1816 is also around 306 kbp. However, as part of a prophage, end C might be involved in genomic rearrangements. Therefore, the region between *yhzC*...
and end C in J1816 was sequenced and determined to be 15 kbp long, which confirmed that the distance between ends A and C was indeed 321 kbp (306 plus 15) (Fig. 4). Given the above findings, the sequencing of the FELM-PCR products correctly revealed the SNP that caused the AscI PFGE banding pattern differences between J1816 and H7858 (Table 1; Fig. 4). The region between yhzC and end C in J1816 has a genomic structure that is very similar to the region between yhzC and LMOh7858_2426 in H7858, with both conserved and variable open reading frames.

Although prophage regions in ECII contain SNPs that can distinguish the two ECII outbreak isolates (J1816 and H7858) (Fig. 4), it was unknown whether or not these SNPs could differentiate all 1998 outbreak isolates as a group from all 2002 outbreak isolates. Therefore, the epidemiological relevance of prophage SNPs within each outbreak clone was evaluated. An LMOh7858_2426-homologous region and an arbitrarily chosen prophage locus (ORF2422) were analyzed in 9 outbreak isolates from 1998 (H7355, H7969, H7550, H7557, H7569, H7738, H7762, H7961, and H7962) and 11 outbreak isolates from 2002 (J1703, J1735, J1736, J1776, J1816, J1817, J1815, J1925, J1927, J1926, and J1928) obtained from the CDC. The LMOh7858_2426-homologous region was amplified and sequenced using primer pairs 2426_F/2426_R and 2664_F/2664_R can be used for the rapid differentiation of the 1998 and 2002 outbreak clones, and subsequent sequencing and SNP identification are not necessary for the differentiation (Fig. 5). Interestingly, two SNPs in ORF2422 were not identical among isolates in the 1998 outbreak clone (data not shown) and thus separated isolates within this outbreak clone. This is not surprising, because whole-genome macroarray analysis (9) and ApaI PFGE (6) both revealed that there might be DNA polymorphisms within the 1998 outbreak clone, and therefore, different strains in the same outbreak clone may have different sequences in certain prophage regions. In summary, epidemiologically relevant SNPs in certain prophage regions differentiated the 1998 outbreak clone from the 2002 outbreak clone. In contrast, differential macroarray signals in the prophage regions failed to differentiate these two outbreak clones (9), perhaps because the macroarray signals were affected by DNA polymorphisms that were not stable within the same outbreak clone. Alternatively, the macroarrays might not be sensitive enough to detect the epidemiologically relevant SNPs identified in the present study.

The ability of prophage sequences to differentiate the 1998 and 2002 outbreak clones led us to hypothesize that SNPs in prophage regions may also be useful for differentiating closely related outbreak clones of ECIII and ECIV. Therefore, the second objective of the present study was to identify and validate these SNPs. The sequences of putative prophage regions in four ECIII isolates (F6854, F6900, J0161, and J2818) were obtained from the Broad Institute of Harvard and MIT (http://www.broad.mit.edu). Sequence comparisons revealed multiple SNPs in the putative prophage ΦF6854.2 (12) that could differentiate the two outbreak clones of ECIII (data not shown). LMOfl6854_2363, a putative phage terminase gene, was arbitrarily selected, and two internal regions were PCR amplified from two additional 1989 outbreak isolates (FSL F6854.2) that could differentiate these two prophage clones (9), perhaps because the macroarray signals were affected by DNA polymorphisms that were not stable within the same outbreak clone. Alternatively, the macroarrays might not be sensitive enough to detect the epidemiologically relevant SNPs identified in the present study.

Table 1. Epidemiologically relevant SNPs that differentiated the outbreak clones of ECII, ECIII, and ECIV in the present study

<table>
<thead>
<tr>
<th>EC</th>
<th>OC date and location</th>
<th>Position of SNP at indicated locus&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECII</td>
<td>1998, United States</td>
<td>LMOh7858_2426</td>
</tr>
<tr>
<td></td>
<td>2002, United States</td>
<td>ORF2422</td>
</tr>
<tr>
<td></td>
<td></td>
<td>289</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LMOf6854_2363</td>
</tr>
<tr>
<td></td>
<td></td>
<td>445</td>
</tr>
<tr>
<td>ECIII</td>
<td>1989, United States</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>2000, United States</td>
<td>G</td>
</tr>
<tr>
<td>ECIV</td>
<td>1998, United Kingdom</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>1979, Boston</td>
<td>T</td>
</tr>
</tbody>
</table>

<sup>a</sup> Due to space limitations, not all identified SNPs are shown. OC, outbreak clone.

<sup>b</sup> Position numbers are based on the entire sequence of each locus.
GAA-3')/ECIII_{SNP1}_R (5'-CGGGACTGTAATAGTACGT GT-3') and ECIII_{SNP2}_F (5'-AGAACCGGTTGGGATTAG G-3')/ECIII_{SNP2}_R (5'-TATTTCGGGTGTTCCTCGAT-3'). PCR was performed under the same conditions as those used for amplifying ORF2422. All amplicons were subsequently sequenced (GenBank accession numbers EU159669 and EU159670). The results from eight ECIII isolates revealed that the sequences of the two internal regions of LMO16854_2363 were identical within each outbreak clone but different between the two outbreak clones. Therefore, the SNPs in LMO16854_2363 are accurate epidemiological markers for distinguishing the two outbreak clones of ECIII (Table 1).

The prophage analysis of ECIV was more difficult because no whole-genome sequences of ECIV were available. However, bacteriophage PSA was reported to be integrated into the outbreak strain (ScottA) of the 1983 Boston outbreak (16), which is closely related to the ECIV isolates (3). Therefore, we hypothesized that the ECIV isolates might contain prophage regions derived from bacteriophage PSA and the sequences of those prophage-related regions might be used to distinguish the two outbreak clones of ECIV. Lauer et al. (10) reported a primer pair, NC16/PL95, for the detection of genomic regions related to prophage PSA in L. monocytogenes. The primer NC16 targets phosphoesterase, the left flanking region of prophage PSA, and the primer PL95 targets the phage integrase (10). Positive amplification using NC16/PL95 indicates the presence of the integrase (10). Three United Kingdom outbreak isolates (FSL J1-129, FSL J1-116, and FSL N3-013) and one Boston vegetable outbreak isolate (FSL J1-220) were obtained from Cornell University. PCR was performed under conditions that were similar to those used for amplifying ORF2422, except that the annealing temperature was 45°C and 2 μl of each primer was used. The primer pair NC16/PL95 generated positive PCR results, suggesting the presence of PSA-related regions in the ECIV isolates (Fig. 5); however, without complete sequencing of the prophage, it remains unknown if the entire PSA prophage is present. The NC16/PL95 amplicons of all ECIV isolates were sequenced, and SNPs that could differentiate the two outbreak clones of ECIV were subsequently identified (Table 1) (GenBank accession numbers EU159671 and EU159672). The NC16/PL95 amplicon spanned the partial sequences of the phosphoesterase and phage integrase genes and an intergenic region in between, and all SNPs were in the intergenic region (Table 1). The sequences of the NC16/PL95 amplicons were identical within the United Kingdom outbreak clone and different between the two outbreak clones. The stability of these SNPs in the Boston vegetable outbreak clone could not be evaluated because only one isolate of this clone was available. However, the data presented here strongly suggest that SNPs in prophage PSA-related regions might be good epidemiological markers for identifying and differentiating ECIV outbreaks.

In the present study, SNPs in certain prophage regions in L. monocytogenes proved to be epidemiologically relevant for the differentiation of closely related outbreak clones within ECII, ECIII, and possibly also ECIV. These unique SNPs possessed increased discriminatory power beyond the previously developed multi-virulence-locus sequence typing strategy. While the functions of prophage genes in L. monocytogenes are poorly understood, the majority of the genes are probably not in use after integration into the host chromosome (1). Therefore, unlike housekeeping and virulence genes, Listeria prophage genes may not be under selective pressure, which may explain why certain prophage genes are more variable than housekeeping or virulence genes within ECs. It is also intriguing how prophages evolved between the 1998 and 2002 outbreak clones within ECII. The fact that the first 15-kbp part of the prophage in the 1998 and 2002 outbreak clones has both conserved and variable regions supports the idea that the evolution of this prophage might result from a combination of vertical evolution and horizontal gene transfer. Complete sequencing of the prophage in the 2002 outbreak clone and comparison of prophage sequences in different L. monocytogenes isolates may shed further light on the evolution and diversity of these prophages.

Identification of the sequence variations that cause PFGE pattern differences has been difficult because (i) single bands in PFGE banding patterns may actually represent two restriction fragments with nearly identical sizes, (ii) very small restriction fragments may run out of the gel, and/or (iii) in silico PFGE banding patterns are different from the actual PFGE patterns in many cases. PFGE has been a gold standard for studying the epidemiology of major food-borne pathogens since 1996, and tens of thousands of isolates have been subtyped by PFGE. Analyzing PFGE patterns by utilizing the FELM-PCR approach described in the present study may allow the identification of additional markers with high epidemiological relevance and lead to a better understanding of epidemiology and the evolution of bacterial pathogens.

This study was supported by a U.S. Department of Agriculture Special Milk Safety grant to The Pennsylvania State University. We thank Martin Wiedmann and Bala Swaminathan for providing the bacterial strains used in this research. We also thank Bala Swaminathan for providing the CDC PFGE patterns of ECII isolates.

REFERENCES