

## Genetic Markers Unique to *Listeria monocytogenes* Serotype 4b Differentiate Epidemic Clone II (Hot Dog Outbreak Strains) from Other Lineages

Matthew R. Evans,<sup>1†</sup> Bala Swaminathan,<sup>2</sup> Lewis M. Graves,<sup>2</sup> Eric Altermann,<sup>1</sup>  
Todd R. Klaenhammer,<sup>1</sup> Ryan C. Fink,<sup>1†</sup> Sheri Kernodle,<sup>1</sup>  
and Sophia Kathariou<sup>1\*</sup>

Department of Food Science, North Carolina State University, Raleigh, North Carolina 27695,<sup>1</sup> and Centers for Disease Control and Prevention, Atlanta, Georgia<sup>2</sup>

Received 14 July 2003/Accepted 20 December 2003

**A small number of closely related strains of *Listeria monocytogenes* serotype 4b, designated epidemic clone I (ECI), have been implicated in numerous outbreaks of food-borne listeriosis described during the past two decades in Europe and North America. In 1998 to 1999, a multistate outbreak traced to contaminated hot dogs involved a different strain type of serotype 4b, with genetic fingerprints rarely encountered before. In spite of the profound economic and public health impact of this outbreak, the implicated bacteria (designated epidemic clone II [ECII]) have remained poorly characterized genetically, and nucleotide sequences specific for these strains have not been reported. Using genome sequence information, PCR, and Southern blots, we identified DNA fragments which appeared to be either absent or markedly divergent in the hot dog outbreak strains but conserved among other serotype 4b strains. PCR with primers derived from these fragments as well as Southern blots with the amplicons as probes readily differentiated ECII from other serotype 4b strains. The serotype 4b-specific region harboring these fragments was adjacent to *inlA*, which encodes a well-characterized virulence determinant. The findings suggest that ECII strains have undergone divergence in portions of a serotype-specific region that is conserved in other serotype 4b strains. Although the mechanisms that drive this divergence remain to be identified, DNA-based tools from this region can facilitate the detection and further characterization of strains belonging to this lineage.**

*Listeria monocytogenes*, the etiological agent of the food-borne illness listeriosis, is a facultative intracellular pathogen of humans and animals. The organism is frequently found in the environment and can grow at low temperatures, thus having the potential to contaminate cold-stored, ready-to-eat foods. Listeriosis is associated with severe symptoms (abortion, stillbirths, meningitis, encephalitis, and septicemia) and has a high mortality rate (20 to 30%) (10, 24, 26). Cases of human listeriosis primarily involve strains of three serotypes, 1/2a, 1/2b, and 4b.

The population genetic structure of *L. monocytogenes* appears to have substantial clonality (2, 22). This is especially evident in the case of outbreaks of food-borne listeriosis, most of which have involved a small number of genetically related strains, commonly of serotype 4b (2, 3, 16, 22). A cluster of closely related strains, designated epidemic clone I (ECI), have been implicated in numerous outbreaks in Europe and North America, including those in Nova Scotia, Canada (coleslaw, 1981), California (Mexican-style cheese, 1985), France (pork tongue in aspic, 1992), and others (16). Genetic and phenotypic markers specific to ECI strains have been identified (14, 28, 29), and the genome sequencing of one ECI strain impli-

cated in the 1985 California outbreak has been undertaken ([www.tigr.org](http://www.tigr.org)), with support by the U.S. Department of Agriculture.

Between August 1998 and March 1999, *L. monocytogenes* of serotype 4b was implicated in a multistate outbreak of listeriosis that was linked to the consumption of contaminated hot dogs and was responsible for 101 human cases, including 21 deaths (4, 5). The genetic fingerprints of the implicated bacteria, determined by ribotyping and pulsed-field gel electrophoresis (PFGE), had only rarely been encountered among clinical isolates prior to this outbreak (4). The 1998 to 1999 hot dog outbreak strains thus appear to represent a novel epidemic-associated lineage, designated epidemic clone II (ECII).

The hot dog outbreak was associated with heavy disease and economic burdens and seriously compromised the perceived safety of ready-to-eat meats. However, the implicated strains (ECII) have remained poorly characterized, and simple means for their specific detection and differentiation from other strains are not yet available. Such tools would greatly facilitate the detection and monitoring of these strains in foods and food-processing environments.

Our laboratory has been pursuing the identification of genetic markers that are potentially unique to these strains. In this study, we show that the hot dog outbreak strains differed significantly in a portion of the genome that was otherwise highly conserved among strains of serotype 4b and specific to this serotype. DNA-based tools derived from this region readily differentiated these outbreak strains from other strains of serotype 4b.

\* Corresponding author. Mailing address: Department of Food Science, North Carolina State University, Campus Box 7624, Raleigh, NC 27695. Phone: (919) 513-2075. Fax: (919) 515-7124. E-mail: [skathar@unity.ncsu.edu](mailto:skathar@unity.ncsu.edu).

† Present address: Department of Microbiology, North Carolina State University, Raleigh, NC 27695.

TABLE 1. Strains of *L. monocytogenes* serotype 4b used in this study

Strain category (no. of strains)	PulseNet pattern designation <sup>a</sup> (no. of isolates)	PCR result <sup>b</sup>	
		4bSF7 and 4bSF18	4bSF149
ECI <sup>c</sup> (10)	GX6A16.0012,GX6A12.0007 (4) GX6A16.0291,GX6A12.0367 (1) GX6A16.0189,GX6A12.0697 (2) GX6A16.0003,GX6A12.0048 (1) GX6A16.0031,GX6A12.0426 (1) GX6A16.0031,GX6A12.0126 (1)	+	+
ECII <sup>d</sup> (8)	GX6A16.0002,GX6A12.0002 (2): E <sub>0</sub> GX6A16.0002,GX6A12.0003 (2): E <sub>1</sub> GX6A16.0002,GX6A12.0004 (2): E <sub>2</sub> GX6A16.0002,GX6A12.0057 (1): E <sub>3</sub> GX6A16.0016,GX6A12.0003 (1): E <sub>1</sub>	–	+
Pate outbreak (4)	GX6A16.0003,GX6A12.0048 (4)	+	+
Boston outbreak (1)	GX6A16.0003,GX6A12.0048 (1)	+	+
Other 4b <sup>e</sup> (16)	G6XA16.0012,GX6A12.0007 (1) GX6A16.0031,GX6A12.0126 (1) GX6A16.0031,GX6A12.0247 (2) GX6A16.0031,GX6A12.0426 (3) GX6A16.0031,GX6A12.0696 (2) GX6A16.0320,GX6A12.0388 (3) NA (4)	+	+

<sup>a</sup> PFGE patterns are reported with standard PulseNet nomenclature: GX6 = *L. monocytogenes*; A16 and A12 indicate AscI and ApaI restriction enzymes, respectively; five-digit number designations indicate sequentially assigned pattern numbers for each unique pattern in the PulseNet database. The number of isolates with a given PFGE pattern is given in parentheses. NA, not available.

<sup>b</sup> PCR results obtained with primers designed for fragments 4bSF7, 4bSF18, and 4bSF149 (Table 2). +, expected amplicon obtained; –, no PCR product obtained.

<sup>c</sup> Included three isolates each from the California (Mexican-style cheese), Canada (coleslaw), and Switzerland (soft cheese) outbreaks and one isolate from the pork tongue in aspic outbreak (France).

<sup>d</sup> E<sub>1</sub>, E<sub>2</sub>, and E<sub>3</sub> are minor variations of the pattern labeled E<sub>0</sub>.

<sup>e</sup> Serotype 4b strains without known implication in outbreaks.

## MATERIALS AND METHODS

**Bacterial strains and growth conditions.** The categories of serotype 4b strains of *L. monocytogenes* used in this investigation are listed in Table 1. These strains were derived from the *Listeria* strain collections at the Centers for Disease Control and Prevention and at North Carolina State University. The ECII strains had the PFGE type characteristic of the hot dog outbreak strains (4). E<sub>1</sub>, E<sub>2</sub>, and E<sub>3</sub> are minor variations of the pattern labeled E<sub>0</sub> and have PFGE patterns identical to that of one of the enzymes employed for PFGE determinations (commonly AscI) and minor differences with the other enzyme (commonly ApaI) (Table 1). In addition, other selected strains were used, including *L. monocytogenes* serotype 1/2a strains G2228, J0161, and G3984, serotype 1/2b strains G3978, G4027, G6005, and G4601, one strain each of serotypes 1/2c, 3b, and 4d (G3963, G3986, and G3987, respectively), and *Listeria innocua* 101A. The *Listeria* strains were grown and preserved as previously described (21).

**Isolation of genomic DNA and restriction enzyme analysis.** The DNeasy tissue kit (Qiagen, Valencia, Calif.) was used to extract and isolate DNA from 24-h *Listeria* cultures by the method suggested by the manufacturer. Determination of DNA resistance or sensitivity to Sau3AI digestion was performed as previously described (29). Restriction enzymes were purchased from New England Biolabs (Beverly, Mass.) or from Promega (Madison, Wis.) and were used according to the manufacturer's instructions.

**PCR and Southern blots.** The primers used in this study are listed in Table 2. Some primers were designed based on preliminary sequence data for *L. monocytogenes* serotype 4b, which were obtained from the Institute for Genomic Research through the website at <http://www.tigr.org>. Primers were designed with Primer3 software ([http://www.genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www.genome.wi.mit.edu/genome_software/other/primer3.html)) and purchased from Bio-Synthesis, Inc. (Lewisville, Tex.), or from Qiagen. PCR employed X-Taq DNA polymerase (Fisher). The reaction mixtures were subjected to a hot start (95°C for 5 min) prior to 30 cycles of amplification at 95°C for 1 min, 50°C for 1 min, and 72°C for 2 min, with a final extension at 72°C for 20 min, in a Progene thermocycler. Amplified products were separated by electrophoresis through a 1.0% (wt/vol) agarose gel with 1× Tris-acetate-EDTA (TAE) running buffer. The predicted sizes of the PCR products with the forward and reverse primers for the 4b-specific fragments (4bSF) 7, 18, 149, and 5B are listed in Table 2.

DNA probes for Southern blotting were obtained with the PCR products amplified from genomic DNA of the serotype 4b strain F2381, implicated in the same outbreak as the strain employed in the serotype 4b genome sequencing ([www.tigr.org](http://www.tigr.org)). The PCR products were excised from the gel (0.8% agarose gel with 1× TAE running buffer) and purified with the QIAquick gel extraction kit (Qiagen). DNA probes were labeled with digoxigenin (Genius kit; Roche) by following the manufacturer's instructions. Southern blots were performed as described previously (21), with hybridizations at 42°C and 30°C for high- and low-stringency conditions, respectively.

**Nucleotide sequence analysis.** For DNA and protein database searches and analyses, we used FASTA (University of Wisconsin GCG package; Genetics Computer Group, Madison, Wis.), BLAST algorithms (<http://www.ncbi.nlm.nih.gov/BLAST/>), and Artemis software (25; <http://www.sanger.ac.uk/Software/Artemis>). Preliminary sequence data for *L. monocytogenes* serotype 4b were obtained from the Institute for Genomic Research through the website at <http://www.tigr.org>. Global Annotation of Multiplexed On-Site Blasted DNA Sequences (GAMOLA; [www.cals.ncsu.edu:8050/food\\_science/KlaenhammerLab/GAMOLA](http://www.cals.ncsu.edu:8050/food_science/KlaenhammerLab/GAMOLA)) was used to automatically predict coding open reading frames (ORFs) and annotate the draft sequence in the genomic region that harbors 4bSF7 and 4bSF18. GAMOLA relies on the available software Glimmer2 (8), the NCBI toolkit (National Center for Biotechnology Information, Bethesda, Md.), and HMMER2.2g (9) and combines the results into functionally annotated DNA sequences in GenBank format. The annotated sequences of the serotype 4b genome, as well as the genomes of *L. monocytogenes* EGD (serotype 1/2a) and *L. innocua* (13), were visualized with Artemis (25). The location and number of EcoRI restriction sites (GAATTC) present within specific sequences were determined with Bioedit version 5.06 (Tom Hall, Department of Microbiology, North Carolina State University, Raleigh, N.C.).

## RESULTS

The hot dog outbreak strains (ECII) were of serotype 4b. The eight strains used in this study possessed a common ri-

TABLE 2. Oligonucleotide primers used for PCR amplification

Primer	Sequence	Accession no. <sup>a</sup>
<i>gltA</i> (P1) Forward	5' TCA TCG TAT CGC TTC TGT G 3'	AF033015
<i>gltA</i> (P2) Reverse	5' GTG CCA TTA CTA CAG GTG CA 3'	AF033015
4bSF7 Forward	5' CCT TGG CAT AAA ATT TAC TCT TGG 3'	AJ410352
4bSF7 Reverse	5' TGA TCT TTT TCT TTT CCG AGT C 3'	AJ410352
4bSF18 Forward	5' ACG GGC GTT TTA TAT TAA ATG GG 3'	AJ410359
4bSF18 Reverse	5' AAT ATC TCG AAA ACT CCG AGT C 3'	AJ410359
4bSF149 Forward	5' TAT GTG AAA TAA ACG ATG AAA TAA 3'	AJ410385
4bSF149 Reverse	5' ACA TCA ATG AGC GTT TCT AAA ATG 3'	AJ410385
1358-1359IGS F	5' ATC AAC CAG CAA CGC ATA CA 3'	
1358-1359IGS R	5' TCT CCT GCA TCC GAA AAA TC 3'	
1364-1365IGS F	5' TAG CCA AAT CAG AAG CCA GG 3'	
1364-1365IGS R	5' TTC ATG CCC TTC TAA CCT TTG 3'	
ORF1365-1366 F	5' CAT TAA AAG CAA GGC CGT TT 3'	
ORF1365-1366 R	5' TTG TTC TAC ACT GCA CCC GA 3'	
ORF1367 F	5' GTT TCA TTC CGC TGC TTC TC 3'	
ORF1367 R	5' AGA CTT TCC TTT ACG TTT ACT CCA 3'	
ORF1368 F	5' TTC TTA GTT GGT GGA AAT GCA A 3'	
ORF1368 R	5' ACT TCC GAT TTT CCG AAT TT 3'	
ORF1370-1371 F	5' TTG ACA CGG AAG CAG AAG TG 3'	
ORF1370-1371 R	5' TTG TGA GGA TGT TCT CGC TG 3'	
ORF1372-1373 F	5' CTG CGT GAG CAT AAG CAA GA 3'	
ORF1372-1373 R	5' TGC CTC ACA TCG TCT ACT GC 3'	
ORF1374 F	5' TTG GGC CGT ATG AGG AAT TA 3'	
ORF1374 R	5' CTC GTC GCC TAT GAT GTC AA 3'	
ORF1376 F	5' CCA GAA GAC CGT AAA ATC AGC 3'	
ORF1376 R	5' TAA ACG TCC AAT GAG AAG TGC 3'	
ORF1376-1379 F	5' AAA CCG AAC AAG GAA TGA CG 3'	
ORF1376-1379 R	5' ACG TGC TCC TTT TAA GCG AA 3'	
ORF1379F	5' ATT AAA GCG AGT TCA GGC GA 3'	
ORF1379R	5' GTT GAC GTC TTC GGG TTG AT 3'	

<sup>a</sup> When no accession number is indicated, the sequences were obtained from the Institute for Genomic Research through its website.

botype and several closely related PFGE patterns (Table 1). The genomic DNA of ECII strains could be readily digested with Sau3AI (data not shown), suggesting that these strains lacked the methylated cytosines at GATC sites reported for

ECI strains (29). The strains were tested in colony immunoblots with monoclonal antibodies that have been found to be specific for serotype 4b and for the less frequent serotypes 4d and 4e (17). All ECII isolates tested reacted positively with these serotype-specific antibodies, including antibody c74.22, which was unable to recognize a fraction of the serotype 4b strains implicated in certain earlier outbreaks (6). In addition, genomic DNA from ECII bacteria hybridized in Southern blots with the serogroup 4-specific gene *gtcA* (23), as well as with the serotype 4b-, 4d-, and 4e-specific genes *gltA* and *gltB* (21), and the expected PCR product was produced with primers derived from the *gltA* sequence (data not shown).

**Primers derived from two serotype 4b-specific genomic DNA fragments failed to produce the expected PCR product from ECII strains.** To further determine the presence and conservation of serotype 4b sequences in ECII strains, we employed PCR with primers derived from the sequences of an additional panel of six DNA fragments, designated 4b-specific fragments (4bSF) and identified in subtractive hybridization studies by other investigators (14). Primers specific to four of these fragments, 4bSF143, 4bSF149, 4bSF166, and 4bSF167, produced PCR products of the predicted sizes (200, 337, 539, and 305 bp, respectively) with ECII strains, as well as with all other strains of serotype 4b that we examined (data not shown). DNA sequence analysis revealed that 4bSF167 actually corresponded to an internal fragment of the serotype 4b-specific gene *gltA* identified earlier in our laboratory and mentioned above (21).

Interestingly, primers derived from the sequences of two other serotype 4b-specific fragments identified in subtractive hybridization studies (14), 4bSF7 and 4bSF18, failed to yield a PCR product when genomic DNA of ECII strains was used as the template. In contrast, the expected product was produced with DNA from other serotype 4b strains. Figure 1 shows the PCR results for a representative panel of strains with primers derived from 4bSF7, and the PCR results for all strains tested with primers specific for 4bSF7, 4bSF18, and 4bSF149 are summarized in Table 1.

**Sequences homologous to 4bSF7 and 4bSF18 could not be detected in the genome of ECII strains by Southern blots.** The negative PCR results with the ECII strains and primers derived from 4bSF7 and 4bSF18 suggested nucleotide sequence divergence in one or both primer regions. To determine whether there was a lack of sufficient overall homology of these DNA fragments with the genomes of the ECII strains, the PCR amplicons obtained with strain F2381 as the template were used as probes in Southern blots. When 4bSF7 was used as the probe, no signal was obtained with EcoRI-digested genomic DNA of any of the ECII isolates tested. In contrast, signals (ca. 3.1 kb) were produced with all other serotype 4b DNAs screened (Fig. 2A). When 4bSF18 was used as the probe in Southern blots, the ECII strains again failed to yield any signals (Fig. 2B). No hybridization bands were obtained with the genomic DNA from strains of serotype 1/2a or 1/2b or *L. innocua* with either 4bSF7 or 4bSF18 as the probe (Fig. 2A and B), in agreement with the previous description of these fragments as specific to serotype 4b (14).

When the Southern blots were repeated under low-stringency conditions, the ECII strains remained completely negative, as did strains of serotypes 1/2a and 1/2b and *L. innocua*, whereas strong bands were produced by all other serotype 4b

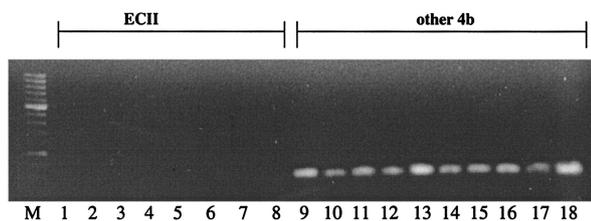


FIG. 1. Differentiation of the hot dog outbreak strains (ECII) from other strains of *L. monocytogenes* serotype 4b by PCR with primers derived from 4bSF7. Lanes: M, 100-bp molecular size marker XIV (Roche); 1 to 8, hot dog outbreak strains; 9 to 18, other serotype 4b strains from the categories shown in Table 1. PCR was done as described in Materials and Methods.

strains (data not shown). The data suggest that the sequences corresponding to 4bSF7 and 4bSF18 were either extensively divergent in ECII or absent from the genomes of these strains, whereas they were conserved among all other serotype 4b isolates tested. In addition, the fact that hybridizing bands of

the same size were obtained with both the 4bSF7 and 4bSF18 probes in Southern blots of EcoRI-digested DNA suggested that 4bSF7 and 4bSF18 were located in the same EcoRI fragment in the genome of serotype 4b strains of *L. monocytogenes*.

**Sequence analysis suggests that 4bSF7 and 4bSF18 are in close proximity to each other in the serotype 4b genome.** BLAST analysis of 4bSF7 and 4bSF18 against the genome of the sequenced serotype 4b strain (www.tigr.org) revealed that 4bSF7 and 4bSF18 were located on the same contig (contig 760) of the genome and that, in fact, these two fragments were separated from each other by only 1,378 nucleotides, suggesting that they were part of the same serotype 4b-specific genomic region. Screening of this region for EcoRI sites indicated that 4bSF7 and 4bSF18 were on a common 3.1-kb EcoRI fragment in the sequenced strain (www.tigr.org), as was also indicated by the Southern blot hybridization data described above.

The GAMOLA-based analysis of the genomic region harboring 4bSF7 and 4bSF18 revealed that these fragments were internal to two putative ORFs, 1363 and 1361, respectively (Fig. 3). The ORFs identified in this region in serotype 4b are listed in Table 3 and shown in Fig. 3, which also summarizes the Southern blot results obtained with different probes in this region. It can be seen that no ECII sequences with detectable homology were found with probes specific to most of the serotype 4b-specific ORFs. Only three of the serotype 4b-specific ORFs (1364, 1367, and 1376) were conserved between ECII and other serotype 4b strains. BLAST analysis of ORF 1364 revealed that it harbored the serotype 4b-specific fragment 4bSF149 identified previously (14). As was mentioned earlier, our PCR data indicated that 4bSF149 was present in all serotype 4b strains screened, including ECII strains (Table 1). ORF 1376 is very long (6,585 bp) and encodes a putative cell wall-associated protein (Table 3). Interestingly, one of the intergenic regions (between ORFs 1372 and 1373) in the serotype 4b-specific region was also conserved between ECII and other serotype 4b strains (Fig. 3).

Surprisingly, ORFs 1358 and 1355, located immediately downstream of ORF 1359, had 96 and 90% identity with *inlA* and *inlB*, respectively, of *L. monocytogenes* EGD, which encode well-characterized internalins of *L. monocytogenes*. Southern blots with a DNA fragment internal to *inlA* as the probe yielded strongly hybridizing bands with all serotype 4b strains tested, including ECII strains, as well as with all other *L. monocytogenes* strains of diverse serotypes that were screened (data not shown).

In summary, the Southern blot and PCR data indicated that the genomic region that encompasses 4bSF7 and 4bSF18 and that appears to be divergent in or absent from the genome of ECII strains was flanked on one side by the virulence-associated gene encoding the internalin *InlA*, which is present in all *L. monocytogenes* strains, including ECII strains. On the other side the region was flanked by the serotype 4b-specific (and, as shown in Fig. 4, *L. innocua*-specific) ORFs 1376 and 1377, which are conserved among all serotype 4b strains, including ECII strains.

**Comparative genomic analysis of the serotype 4b-specific region.** None of the probes in the region of ORFs 1359 to 1377 (these ORFs included) hybridized with sequences in the genome of *L. monocytogenes* strains of serotypes 1/2a, 1/2b, 1/2c,

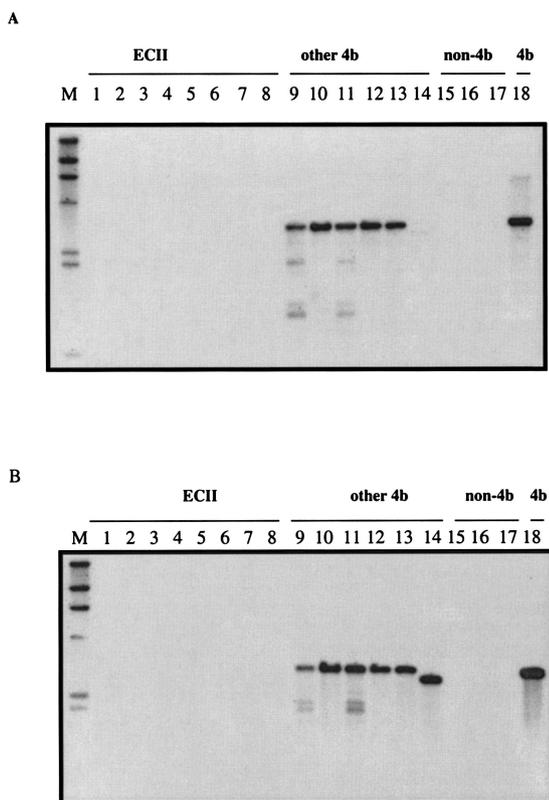


FIG. 2. Southern blot of EcoRI-digested chromosomal DNAs from the hot dog outbreak strains (ECII) and other representative strains of *L. monocytogenes*, with 4bSF7 (A) and 4bSF18 (B) as probes. Lanes: M, digoxigenin-labeled  $\lambda$  HindIII digest, used as molecular size markers (fragment sizes are [from top to bottom] 23, 9.4, 6.5, 4.3, 2.3, and 2.0 kb); 1 to 8, hot dog outbreak strains; 9 to 14 and 18, other serotype 4b strains from the categories shown in Table 1; 15, *L. monocytogenes* G2228 (serotype 1/2a); 16, *L. monocytogenes* G3978 (serotype 1/2b); 17, *L. innocua* 101A. Southern blots were done as described in Materials and Methods.

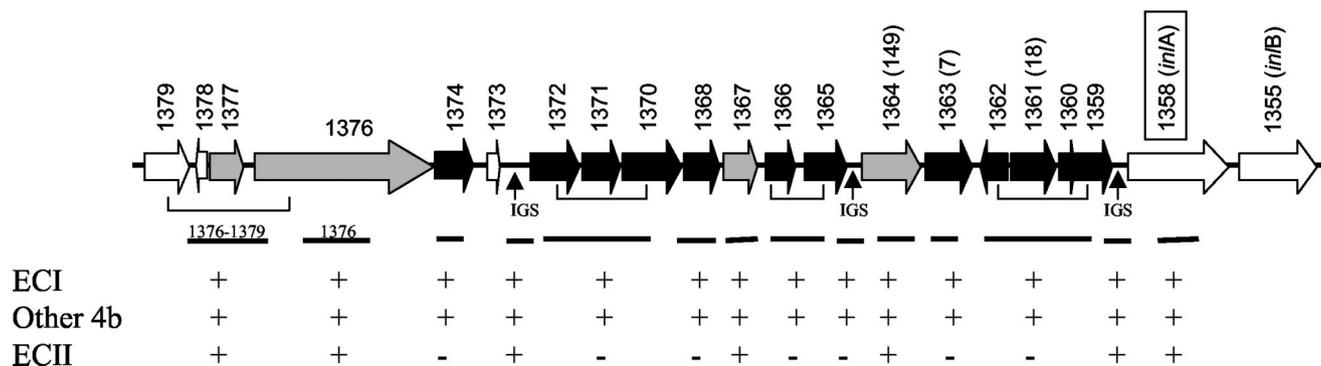


FIG. 3. Conservation of ORFs in the genomic region harboring 4bSF7 and 4bSF18 between ECII and other serotype 4b strains. DNA probes were generated with the primers listed in Table 2 and used in hybridizations as described in Materials and Methods. + and - indicate the presence and absence of a hybridization signal, respectively. *L. monocytogenes* strains of serotype 4d yielded results identical to those of other 4b strains. ORFs which lacked detectable homology in ECII strains (on the basis of the Southern blot data) are indicated in black, and serotype 4b-specific ORFs conserved in ECII strains are indicated in gray. The putative ORFs 1373 and 1378 (111 and 90 bp, respectively) were too short for adequate probe construction, and their conservation in ECII strains has not been determined. Arrows indicate the direction of transcription. IGS, intergenic regions. Note that *inlA* (ORFs 1358), *inlB* (ORF 1355), and ORF 1376 are much longer than the others, as detailed in Table 3, and are not drawn to scale.

and 3b (data not shown). The findings suggest that, within *L. monocytogenes*, only serotype 4b (and the closely related, albeit rare, serotype 4d) harbored these sequences, in agreement with the results from the Artemis-based genomic comparison between *L. monocytogenes* serotype 4b and EGD in this region. Interestingly, the long ORF 1376, which encodes a putative

wall-associated protein, along with small ORFs 1373 and 1377 were also harbored by *L. innocua* and in an apparently conserved genomic organization, albeit absent from *L. monocytogenes* strain EGD, which lacked the entire ca. 17-kb region harboring ORFs 1359 to 1378 (Fig. 4). The two genes flanking this region in strain EGD, ORFs 432 (homologous to serotype

TABLE 3. Putative ORFs in the *L. monocytogenes* serotype 4b-specific genomic region investigated in this study

ORF <sup>a</sup>	Location <sup>b</sup>	Size (bp)	% G+C	Comments
1358	1197148→1199580	2,433	36.86	<i>L. monocytogenes</i> EGD lmo0433 ( <i>inlA</i> ), 96% identity
1359	1200257→1200829	573	35.77	
1360	1200810→1200953	144	31.25	
1361	1200981→1201565	585	34.87	Harbors 4bSF18
1362	1201583→1201819	237	27.42	
1363	1202068→1202730	663	29.56	Harbors 4bSF7
1364	1202886→1203671	786	30.53	Harbors 4bSF149
IGS	1203672→1204268	597	33.11	
1365	1204269→1204760	492	27.84	
IGS	1204761→1205059	299	27.24	
1366	1205060→1205311	252	19.44	
1367	1205490→1205843	354	28.81	
1368	1205886→1206239	354	27.4	
1370	1206296→1206901	606	29.53	
1371	1206921→1207364	444	30.4	
1372	1207394→1207981	588	30.78	
IGS	1207982→1208992	1,011	30.57	
1373	1208993→1209103	111	38.73	<i>L. innocua</i> lin0455 (93% identity) <sup>c</sup>
IGS	1209104→1209376	273	29.93	
1374	1209377→1209856	480	30.41	
1376	1209859→1216443	6585	41.68	<i>L. innocua</i> lin0454(93% identity) <sup>d</sup>
IGS	1216444→1216701	258	26.74	
1377	1216702→1217124	423	34.27	<i>L. innocua</i> lin0453 (96% identity)
1378	1217136→1217225	90	31.11	
1379	1217323→1218066	744	42.2	<i>L. monocytogenes</i> EGD lmo0432; (91% identity); <i>L. innocua</i> lin0452 (92% identity) <sup>e</sup>

<sup>a</sup> ORFs identified by GAMOLA, based on preliminary sequence data for *L. monocytogenes* serotype 4b obtained from the Institute for Genomic Research through its website. IGS, intergenic sequence longer than 200 bp.

<sup>b</sup> Coordinates determined from the GAMOLA-based annotation of the region harboring 4bSF7 and 4bSF18.

<sup>c</sup> No significant homologs in the nucleotide and protein databases.

<sup>d</sup> Putative cell wall-associated protein.

<sup>e</sup> Similar to oxidoreductase.

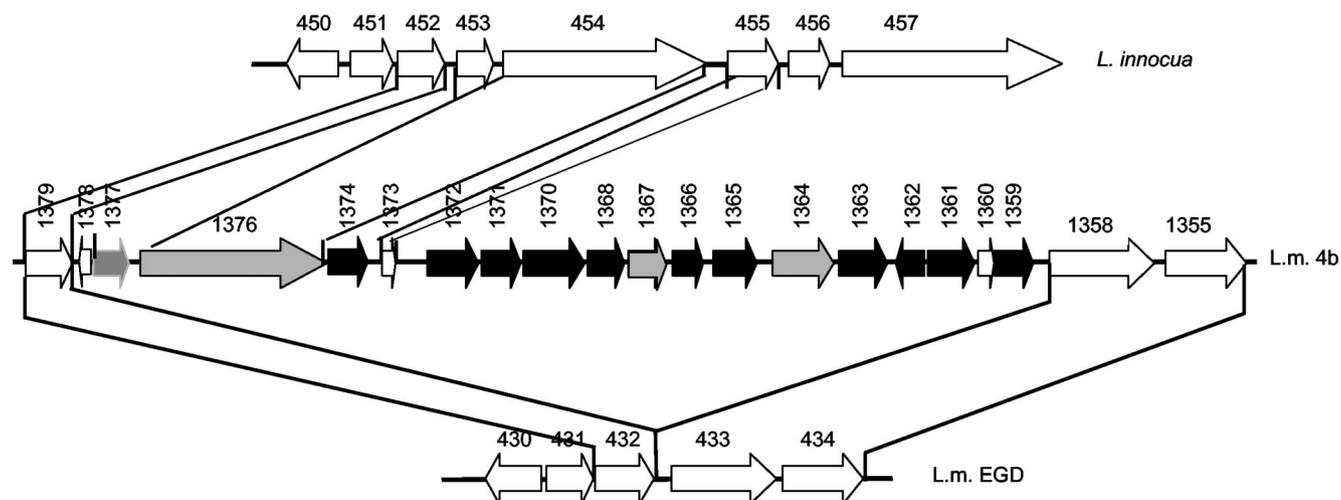


FIG. 4. Genomic organization in the genomic region harboring 4bSF7 and 4bSF18, and genomic counterparts of the region in the sequenced genomes of *L. monocytogenes* EGD (serotype 1/2a) and *L. innocua*. Artemis was used to visualize the location and direction of transcription of the putative ORFs. Black arrows indicate serotype 4b-specific ORFs not detected in ECII strains; gray arrows indicate ORFs that are serotype 4b specific (ORFs 1376 and 1377 are also conserved between serotype 4b strains and *L. innocua*) and conserved in ECII strains. Diagonal lines indicate ORFs conserved between different genomes. The putative ORFs 1373 and 1378 (111 and 90 bp, respectively) were too short for adequate probe construction, and their conservation in ECII strains has not been determined. Note that *inlA* (ORFs 1358 and 433), *inlB* (ORF 1355), and ORF 1376 are much longer than the others, as detailed in Table 3, and are not drawn to scale.

4b ORF 1379) and 433 (*inlA*, homologous to serotype 4b ORF 1358), were adjacent to each other in the EGD genome, without any intervening genes (Fig. 4).

With the exception of ORFs 1373 and 1376 (both of which were conserved between serotype 4b and *L. innocua*), the serotype 4b-specific ORFs had G+C contents ranging from 19.44 to 35.77% (Table 3), significantly lower than the average G+C content of *L. monocytogenes* (ca. 38%). With the exception of the aforementioned ORF 1376, no information on the functions of the serotype 4b-specific ORFs in this region currently exists, and no significant homologs were detected in the nucleotide and protein databases.

## DISCUSSION

Our basic understanding of the ecology, evolution, and virulence of epidemic-associated strains of *L. monocytogenes* remains limited. This is especially the case with ECII strains, which have been known for a relatively limited time. When first discovered during the 1998 to 1999 hot dog outbreak, ECII strains were found to have PFGE and ribotype fingerprints which were previously rarely encountered (4). However, information about whether ECII strains have potentially unique genetic, physiologic, and virulence-associated attributes that differentiate them from ECI and from other serotype 4b strains has been lacking. To date, only one RFLP that differentiates ECII from other serotype 4b strains has been identified. This polymorphism was detected with a DNA probe derived from the putative mannitol permease locus of *L. monocytogenes*, which differentiated among several lineages within serotype 4b (27).

In this study, we identified a serotype 4b-specific region which is either divergent in or absent from the genome of ECII strains. Interestingly, this region is adjacent to the internalin genes *inlA* and *inlB*, which are known to have important func-

tions in cell invasion and virulence in *L. monocytogenes* (7, 11, 15, 20).

In our experience with other divergent genes in *L. monocytogenes*, signals in Southern blots were very faint or absent under high stringency when DNA sequence identity was between 80 and 85% but could be obtained readily under low-stringency conditions (18). The absence of any signal with ECII under either high- or low-stringency conditions with several of the probes that we employed suggested that if the genomes of the ECII strains harbored sequences homologous to these probes, the level of nucleotide sequence identity would be below ca. 80%. Our current data do not permit us to conclude whether these sequences are absent or extensively divergent in the ECII genome. In addition, PCR with primers based on conserved sequences (e.g., ORFs 1364 and 1358, as well as 1376 and 1364) was not successful, even with PCR systems that would allow amplification of long (7 to 10 kb) products, which may indicate that these sequences have become rearranged within this serotype-specific region in the genome of the ECII strains in comparison to their organization in the genome of the sequenced serotype 4b ECI strain.

Comparative genomic analysis of the region studied here suggests that it may have been introduced into serotype 4b by an insertion in the genome between ORFs 1358 (*inlA*) and 1379. Alternatively, the region could have been present in an ancestral *L. monocytogenes* lineage and maintained only in serotype 4b, becoming lost from the genomes of strains of other serotypes. The former scenario appears more likely, since the G+C content of most ORFs in the region was significantly lower than the average for the genome (Table 3), suggesting that the corresponding genes in serotype 4b may have been acquired by horizontal transfer from another, unidentified source. The mechanisms mediating such possible transfer remain unclear, and the region lacked genetic features

frequently associated with horizontally transferred DNA (such as tRNA genes and transposases). Similar findings were obtained with other serotype-specific genes of *L. monocytogenes*, which also had G+C contents lower than the average for the genome (21, 23). Additional horizontal transfer events between *L. monocytogenes* serotype 4b and *L. innocua* may have resulted in the presence of *L. monocytogenes* serotype 4b-*L. innocua* conserved sequences such as ORFs 1373, 1376, and 1377 (Fig. 4). Evidence for horizontal transfer between *L. monocytogenes* serotype 4b and *L. innocua* has been described before (19). Alternatively, ORFs conserved in *L. monocytogenes* serotype 4b and *L. innocua* may represent ancestral sequences that have become eliminated from other *L. monocytogenes* lineages.

The question arises why this region, which was otherwise specific to serotype 4b strains, would harbor numerous ORFs that appeared to be markedly divergent (if not absent) in ECII strains. We noticed that most serotype 4b strains yielded the same hybridizing EcoRI fragment in Southern blots, with the size predicted by the analysis of this region in the sequenced strain. The dearth of polymorphisms suggested that the locations of the EcoRI sites in the probed region were conserved in the genomes of most *L. monocytogenes* strains of serotype 4b. This makes it all the more surprising that the ECII genome would have undergone extensive divergence or deletions in this region.

Although the mechanisms that have driven these genomic differences between ECII and other serotype 4b strains remain unknown at this time, the findings can be quite useful in the development of DNA-based tools to specifically detect strains of the ECII clonal group. For instance, serotype 4b strains that are negative by PCR or Southern blot with the primers and probes derived from ORFs such as 1361 and 1363 are likely to be of the ECII group, and, similarly, ECII status can be excluded if a strain is of serotype 4b and positive in PCR or in Southern blots with these primers and probes. It is noteworthy that we obtained identical results with all ECII strains isolated from the 1998 to 1999 hot dog outbreak in spite of minor differences in their PFGE patterns (Table 1). To our knowledge, these are the first genetic markers to be found capable of specifically identifying the ECII clonal group by PCR as well as by DNA hybridization-based approaches. They could be readily adapted to other technical formats, such as DNA microarrays.

Genetic studies with other serotype-specific genes have shown that the genes were involved in decoration of the cell wall polymer teichoic acid with serotype-specific sugars (21, 23) and may be involved in virulence in a murine model (1). The genes in the region described here may also be involved in the expression of cell surface components. The identification of ORF 1376, encoding a putative cell wall-associated protein and shared between serotype 4b *L. monocytogenes* and *L. innocua*, may suggest such putative functions.

Of special interest was the finding that the serotype 4b-specific region investigated here was immediately adjacent to *inlA*, which encodes internalin (InlA). InlA and InlB (encoded by *inlB*, immediately downstream of *inlA*) are important to the ability of *L. monocytogenes* to invade several classes of mammalian cells (7, 11, 12, 15), and *inlA* is essential for virulence in oral infections of transgenic mice that expressed a functional

InlA receptor (20). This suggested the possibility that the serotype 4b-specific genes in the region may also be implicated in host-pathogen interactions and virulence. Genetic studies that will allow the mutagenesis and functional analysis of the genes will be necessary to investigate this hypothesis and also to determine the possible involvement of the genes in this region in serotype-specific surface antigen composition. Such studies, along with the identification of the divergent sequences that the ECII genome appears to harbor in this region, will extend our understanding of the evolution of this region in serotype 4b *L. monocytogenes* and of genetic attributes that may be unique to ECII strains. The findings described in this study will facilitate the targeted analysis of such attributes. In addition, the results described here provide tools for enhanced specific detection and monitoring of these strains in foods, clinical samples, and the environment.

#### ACKNOWLEDGMENTS

This work was partially supported by U.S. Department of Agriculture Competitive National Research Initiative grant 2001-0969 and by the International Life Sciences Institute-North America (to S. Kathariou).

Preliminary sequence data for *L. monocytogenes* serotype 4b were obtained from the Institute for Genomic Research through the website, and we thank the Institute for Genomic Research for allowing us access to these data. Sequencing of the genome of *L. monocytogenes* serotype 4b by the Institute for Genomic Research was accomplished with support from the USDA. We thank Suleyman Yildirim for assistance with the use of Artemis software and Robin Siletsky for technical support in portions of this work.

#### REFERENCES

1. Autret, N., I. Dubail, P. Trieu-Cuot, P. Berche, and A. Charbit. 2001. Identification of new genes involved in virulence of *Listeria monocytogenes* by signature-tagged mutagenesis. *Infect. Immun.* **69**:2054–2065.
2. Bibb, W. F., B. G. Gellin, R. Weaver, B. Schwartz, B. D. Plikaytis, M. W. Reeves, R. W. Pinner, and C. V. Broome. 1990. Analysis of clinical and food-borne isolates of *Listeria monocytogenes* in the United States by multilocus enzyme electrophoresis and application of the method to epidemiologic investigations. *Appl. Environ. Microbiol.* **56**:2133–2141.
3. Buchrieser, C., R. Brosch, B. Catimel, and J. Rocourt. 1993. Pulsed-field gel electrophoresis applied for comparing *Listeria monocytogenes* strains involved in outbreaks. *Can. J. Microbiol.* **39**:395–401.
4. Centers for Disease Control. 1998. Multistate outbreak of listeriosis—United States, 1998. *Morb. Mortal. Wkly. Rep.* **47**:1085–1086.
5. Centers for Disease Control. 1999. Update: multistate outbreak of listeriosis—United States, 1998–1999. *Morb. Mortal. Wkly. Rep.* **47**:1117–1118.
6. Clark, E. E., I. Wesley, F. Fiedler, N. Promadej, and S. Kathariou. 2000. Absence of serotype-specific surface antigen and altered teichoic acid glycosylation among epidemic-associated strains of *Listeria monocytogenes*. *J. Clin. Microbiol.* **38**:3856–3859.
7. Cossart, P., and M. Lecuit. 1998. Interactions of *Listeria monocytogenes* with mammalian cells during entry and actin-based movement: bacterial factors, cellular ligands, and signaling. *EMBO J.* **17**:3797–3806.
8. Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**:4636–4641.
9. Eddy, S. R. 1998. Profile hidden Markov models. *Bioinformatics* **14**:755–763.
10. Farber, J. M., and P. L. Peterkin. 1991. *Listeria monocytogenes*, a food-borne pathogen. *Microbiol. Rev.* **55**:476–511.
11. Gaillard, J. L., P. Berche, C. Frehel, E. Gouin, and P. Cossart. 1991. Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci. *Cell* **65**:1127–1141.
12. Gaillard, J. L., F. Jaubert, and P. Berche. 1996. The *inlAB* locus mediates the entry of *Listeria monocytogenes* in hepatocytes in vivo. *J. Exp. Med.* **183**:359–369.
13. Glaser, P., L. Frangeul, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, et al. 2001. Comparative genomics of *Listeria* species. *Science* **294**:849–852.
14. Herd, M., and C. Kocks. 2001. Gene fragments distinguishing an epidemic-associated strain from a virulent prototype strain of *Listeria monocytogenes* belong to a distinct functional subset of genes and partially cross-hybridize with other *Listeria* species. *Infect. Immun.* **69**:3972–3979.
15. Ireton, K., and P. Cossart. 1997. Host-pathogen interactions during entry

- and actin-based movement of *Listeria monocytogenes*. *Annu. Rev. Genet.* **31**:113–138.
16. **Kathariou, S.** 2003. Foodborne outbreaks of listeriosis and epidemic-associated lineages of *Listeria monocytogenes*, p. 243–256. In M. E. Torrence and R. E. Isaacson (ed.), *Microbial food safety in animal agriculture*. Iowa State University Press, Ames, Iowa.
  17. **Kathariou, S., C. Mizumoto, R. D. Allen, A. K. Fok, and A. A. Benedict.** 1994. Monoclonal antibodies with a high degree of specificity for *Listeria monocytogenes* serotype 4b. *Appl. Environ. Microbiol.* **60**:3548–3552.
  18. **Lan, Z.** 2001. Genetic characterization of a *Listeria monocytogenes* serotype-specific genomic region. Ph.D. thesis. University of Hawaii, Honolulu.
  19. **Lan, Z., F. Fiedler, F., and S. Kathariou.** 2000. A sheep in wolf's clothing: *Listeria innocua* strains with teichoic acid-associated surface antigens and genes characteristic of *Listeria monocytogenes* serogroup 4. *J. Bacteriol.* **182**:6161–6168.
  20. **Lecuit, M., S. Vandormael-Pournin, J. Defort, M. Huerre, P. Gounon, C. Dupuy, C. Babinet, and P. Cossart.** 2001. A transgenic model for listeriosis: role of internalin in crossing the intestinal barrier. *Science* **292**:1722–1725.
  21. **Lei, X. H., F. Fiedler, Z. Lan, and S. Kathariou.** 2001. A novel serotype-specific gene cassette (*gltA-gltB*) is required for expression of teichoic acid-associated surface antigens in *Listeria monocytogenes* of serotype 4b. *J. Bacteriol.* **183**:1133–1139.
  22. **Piffaretti, J. C., H. Kressebuch, M. Aeschbacher, J. Bille, E. Bannerman, J. M. Musser, R. K. Selander, and J. Rocourt.** 1989. Genetic characterization of clones of the bacterium *Listeria monocytogenes* causing epidemic disease. *Proc. Natl. Acad. Sci. USA* **86**:3818–3822.
  23. **Promadej, N., F. Fiedler, P. Cossart, S. Dramsi, and S. Kathariou.** 1999. Cell wall teichoic acid glycosylation in *Listeria monocytogenes* serotype 4b requires *gtcA*, a novel, serogroup-specific gene. *J. Bacteriol.* **181**:418–425.
  24. **Rocourt, J., C. Jacquet, and J. Bille.** 1997. Human listeriosis. Publication WHO/FNU/FOS/97.1. World Health Organization, Geneva, Switzerland.
  25. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.-A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**:944–945.
  26. **Schuchat, A., B. Swaminathan, and C. V. Broome.** 1991. Epidemiology of human listeriosis. *Clin. Microbiol. Rev.* **4**:169–183.
  27. **Tran, H. L., and S. Kathariou.** 2002. Restriction fragment length polymorphisms detected with novel DNA probes differentiate among diverse lineages of serogroup 4 *Listeria monocytogenes* and identify four distinct lineages in serotype 4b. *Appl. Environ. Microbiol.* **68**:59–64.
  28. **Zheng, W., and S. Kathariou.** 1995. Differentiation of epidemic-associated strains of *Listeria monocytogenes* by restriction fragment polymorphism in a gene region essential for growth at low temperatures (4°C). *Appl. Environ. Microbiol.* **61**:4310–4314.
  29. **Zheng, W., and S. Kathariou.** 1997. Host-mediated modification of *Sau3AI* restriction in *Listeria monocytogenes*: prevalence in epidemic-associated strains. *Appl. Environ. Microbiol.* **63**:3085–3089.