FSL J1-208, a Virulent Uncommon Phylogenetic Lineage IV *Listeria monocytogenes* Strain with a Small Chromosome Size and a Putative Virulence Plasmid Carrying Internalin-Like Genes

Henk C. den Bakker, Barbara M. Bowen, Lorraine D. Rodriguez-Rivera, and Martin Wiedmann

Cornell University, Department of Food Science, Ithaca, New York, USA

The bacterial genus *Listeria* contains both saprotrophic and facultative pathogenic species. A small genome size has been suggested to be associated with the loss of pathogenic potential of *L. welshimeri* and *L. seeligeri*. In this paper we present data on the genome of *L. monocytogenes* strain FSL J1-208, a representative of phylogenetic lineage IV. Although this strain was isolated from a clinical case in a caprine host and has no decreased invasiveness in human intestinal epithelial cells, our analyses show that this strain has one of the smallest *Listeria* chromosomes reported to date (2.78 Mb). The chromosome contains 2,772 protein-coding genes, including well-characterized virulence-associated genes, such as *inlA*, *inlB*, and *inlC* and the full *prfA* gene cluster. The small genome size is mainly caused by the absence of prophages in the genome of *L. monocytogenes* FSL J1-208, and further analyses showed that the total size of prophage-related regions is highly correlated to chromosome size in the genus *Listeria*. *L. monocytogenes* FSL J1-208 carries a unique type of plasmid of approximately 80 kbp that does not carry genes annotated as being involved in resistance to antibiotics or heavy metals. The accessory genes in this plasmid belong to the internalin family, a family of virulence-associated proteins, and therefore this is the first report of a potential virulence plasmid in the genus *Listeria*.

Listeria monocytogenes is a primarily saprotrophic *Firmicutes* member that can be found in a wide variety of environments, such as soil, decaying plant material, and water. Once it invades a compatible host, it changes from a saprotrophic lifestyle into an intracellular pathogenic lifestyle. Extensive phylogenetic research based on both enzyme electrophoresis (34) and nucleotide sequence data (37, 52) has shown that *L. monocytogenes* consists of four divergent phylogenetic lineages, designated lineages I, II, III, and IV (52). Lineage IV was formerly classified as lineage IIIB, a subpopulation of lineage III (39), and only recently was considered divergent enough from the other lineages to warrant its own lineage (52). In a study that predates the current subdivision of lineage III into lineages III and IV, Jeffers et al. (20) showed that lineage III strains, while being generally rare, were found less frequently among human cases and were more common among animal cases. This observation suggests a putative difference in the ecology of lineage IV and III strains, with the possibility of host specificity of lineage III and IV for nonprimate mammals. Alternatively, the low frequency of lineage III and IV among human listeriosis cases can also be explained by the fact that lineage III and IV isolates show a low prevalence among food isolates (16, 51, 53) and thus a low exposure of lineage III and IV strains for the human host. This may be related to the fact that lineage III and IV strains are less adapted to stresses experienced in foods and food processing environments (9). Current genome sequencing efforts have mainly focused on genomes of strains representing lineages I and II (as of 26 July 2011, 27 out of 33 publicly available genome sequences belonged to lineage I and II), as these lineages represent the majority of isolates involved in human clinical cases, while lineage III is represented by 5 genomes (three virulence attenuated strains, HCC23, L99, and M7, as well as FSL J2-071 and FSL F2-208), and lineage IV is represented by 1 partially sequenced genome (FSL J1-208).

In this study we use a newly generated high-quality draft ge-

nome sequence of *L. monocytogenes* FSL J1-208 to further elucidate various aspects of genome evolution in *L. monocytogenes*. While a partial draft of FSL J1-208 has been released previously by the Broad Institute (*Listeria monocytogenes* Sequencing Project; Broad Institute of Harvard and MIT [<http://www.broadinstitute.org/>]; GenBank accession NZ_AARL000000000), we performed *de novo* sequencing to obtain a near-complete draft genome of this strain. FSL J1-208 was isolated from a goat that was part of one of two listeriosis outbreaks that occurred in two separate goat herds on the same farm during a 16-month period and were caused by the same subtype, as determined by ribotyping; the outbreak strain was hypothesized to have spread among these herds through a venereal route (55). This makes this strain highly unusual, since *L. monocytogenes* infections in goats usually occur through ingestion of contaminated feedstuff and subsequent entry of the host either via the intestine (54) or through small wounds in the lips and the oral cavity (5). Further characterization of this strain showed that it had a higher invasion efficiency in Caco-2 cells and a consistently high number of CFU in liver, spleen, small intestine, and lymph nodes in a guinea pig infection model compared to lineage I, II, and III strains, indicating that this may be a strain with a high virulence potential (31). The availability of a high-quality draft genome of a lineage IV strain should increase our understanding of genome evolution and putative host specificity in *L. monocytogenes*.

Received 21 September 2011 Accepted 2 January 2012

Published ahead of print 13 January 2012

Address correspondence to Henk C. den Bakker, hcd5@cornell.edu.

Supplemental material for this article may be found at <http://aem.asm.org/>.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/AEM.06969-11

MATERIALS AND METHODS

Whole-genome sequencing and *de novo* assembly. A low-coverage draft genome of *L. monocytogenes* FSL J1-208 was initially generated by the Broad Institute (GenBank accession NZ_AARL000000000). To obtain a near-full-length, high-coverage, and high-quality draft, this strain was resequenced using the Illumina platform.

High-molecular-weight DNA of *L. monocytogenes* strain FSL J1-208 was extracted using the method described by Flamm et al. (13). Whole-genome sequencing was performed with the Illumina (San Diego, CA) GAII sequencer at the Cornell University Life Sciences Core Laboratories Center (Ithaca, NY) with 60-bp paired end reads. *De novo* assembly was performed using Velvet version 0.7.55 (56) and VelvetOptimizer version 2.1.0 (S Gladman; <http://bioinformatics.net.au/software.shtml>) to optimize the assembly parameters. Contigs were aligned to *L. monocytogenes* EGD-e (GenBank accession number AL591824 [14]) by using the Move Contigs application in Mauve version 2.3.1 (6). Adjacent contigs were inspected for sub-Kmer overlap in Sequencher 4.10.1 (Gene Code Corp., Ann Arbor, MI), and contigs were joined if overlap was found. Remaining gaps between contigs (except for gaps containing rRNA clusters and two highly repetitive regions) were closed using traditional Sanger-based sequencing; in two cases, contigs could be joined using sequence data previously generated by the Broad Institute. To infer the size of the complete chromosome, the lengths of the gaps that included rRNA clusters was estimated based on whole-genome alignments with *L. monocytogenes* EGD-e and *L. monocytogenes* HCC23.

Genome annotation. The RAST annotation server (2) was used to obtain an initial automated annotation of the genome sequence; this was followed by manual annotation in Artemis release 11.22 (41). Putative prophages or prophage-derived regions were detected using ProPhinder (25). IslandViewer (24) was used to scan the genome for large regions introduced in the genome by horizontal gene transfer (HGT). In addition, the SIGI-HMM algorithm in Colombo version 3.8 (50) was used to detect putative imported genes.

Genome content analysis. The genome sequence of *L. monocytogenes* FSL J1-208 was compared to the sequence of *L. monocytogenes* HCC23. HCC23 is the first lineage III strain that was completely sequenced (47) and is extremely similar to lineage III strains L99 and M7 (GenBank accession numbers FM211688 and CP002816, respectively). Given the phylogenetic position of lineage III (10), HCC23 is most closely related to FSL J1-208 and, therefore, suitable for detailed comparative genomic analysis. Mauve version 2.3.1 (6) was used to create a whole-genome alignment of the *L. monocytogenes* FSL J1-208 and the genome of *L. monocytogenes* HCC23 using the progressive Mauve algorithm. The same software was used to create a list of orthologues and unique genes in both genomes, using an identity cutoff of $\geq 60\%$ and a coverage cutoff of $\geq 70\%$. Pseudogenes, putative pseudogenes, and genes that were inconsistently called between genomes were excluded from the list of accessory genes. In order to assess which type of “insertions” contributed most to the accessory genome, genes were categorized into three classes: (i) prophage-derived genes, (ii) “miniclust” insertions that consisted of two or more genes, and (iii) “singleton insertions” containing a single gene/ORF. All statistical analyses were performed in R version 2.13.0 (<http://www.R-project.org/>).

Southern blotting. The physical presence of the plasmid as inferred from the whole-genome sequence was confirmed using a Southern blotting approach (45) with digoxigenin (DIG)-labeled probes. Undigested and XhoI (New England BioLabs, Ipswich, MA)-digested genomic DNA was electrophoresed on an agarose gel by pulsed-field gel electrophoresis (PFGE) (15) and transferred onto nylon membranes (Roche, Branchburg, NJ) for hybridization with two plasmid-specific probes, a probe targeting a 205-bp sequence of the plasmid-borne internalin *inlP3* and a probe targeting 247 bp of the lytic transglycosylase gene (here called *P60*, because of the presence of a P60-like domain). The DIG-labeled probes were produced using the PCR DIG probe synthesis kit (Roche, Branchburg, NJ)

and PCR primers targeting these genes (see Table S1 in the supplemental material for primers).

Sequence comparison of virulence plasmid to other *Firmicutes* plasmids. To compare the plasmid of *L. monocytogenes* FSL J1-208 to plasmids found in other *Firmicutes*, a protein BLAST search was performed against the NCBI protein database and against proteins carried on plasmids of *Clostridium perfringens* (pCP13; GenBank accession NC_003042) and *Enterococcus faecium* (pGM1; GenBank accession AB206333).

Distribution of virulence plasmid in *L. monocytogenes*. A representative selection of *Listeria* spp. isolates (see Table S2 in the supplemental material) was screened by PCR for the presence of two virulence plasmid-specific genes. First, isolates were screened for *P60*, which is part of the conjugative operon, and when isolates were positive for this gene they were screened for *inlP3* and the plasmid's origin of replication region (see Table S1 of the supplemental material for primers). The *P60* and *inlP3* PCR products were sequenced for isolates that were found positive for these genes.

Phylogenetic classification of internalin-like proteins found in the genome of FSL J1-208. A phylogeny of internalin-like proteins found in FSL J1-208 and a representative collection of *L. monocytogenes* strains (EGDe, F2365, CLIP80459, HCC23, and FSL F2-208) was constructed as described by den Bakker et al. (11). Briefly, an amino acid alignment of the internalins was created using MAFFT (21), and a maximum likelihood (ML) phylogeny was created in PhyML 3.0 (17).

Multilocus sequence analysis of outbreak-related isolates. In addition to FSL J1-208, a number of other clinical isolates were available from the listeriosis outbreak from which FSL J1-208 was recovered (55). The phylogenetic relationship of five other isolates involved in the caprine outbreak and one isolate from a caprine listeriosis case on a nearby farm (Table 1) was inferred based on the 10-gene multilocus sequence typing (MLST) loci described by den Bakker et al. (10). RaxML version 7.0.4 (46) was used to infer an ML phylogeny and ML bootstrap values based on 100 bootstrap replicates.

Phylogenetic clustering based on replication protein. To infer the putative relationship of plasmid pLMIV (the name given here to the plasmid found in *L. monocytogenes* FSL J1-208) to previously published *Listeria* plasmids (23) and a wide variety of plasmids found in other *Firmicutes*, we performed a phylogenetic analysis of the *repA* gene (the main plasmid replication gene). Nucleotide sequences of 41 *repA* loci were aligned guided by their amino acid translation (1) by using Muscle (12). PhyML 3.0 (17) was used to infer a ML phylogeny by using a model of nucleotide evolution that was selected based on the results of JModeltest version 1.1 (35) using the Akaike information criterion. Nonparametric ML bootstrap support was based on 100 bootstrap replicates.

Plasmid curing and Caco-2 invasion assays. To determine if the presence of plasmid pLMIV in the genome of FSL J1-208 influences the invasion efficiency of this strain in human intestinal epithelial cells (a phenotype associated with the presence of *inlA*, which encodes internalin A), we compared the invasiveness of FSL J1-208 to the invasiveness of this strain previously cured of the plasmid pLMIV (the cured strain is designated FSL B2-294). FSL J1-208 was cured from its plasmid by repeated passage of the parental strain at high temperature (40°C) as described by Margolles and de los Reyes-Gavilán (26), without curing agent. The absence of pLMIV was confirmed by the absence of PCR products for *P60*, *inlP3*, and the plasmid's origin of replication. The Caco-2 assays were performed as previously described by Nightingale et al. (30) with three biological replications (with three technical replications per biological replication).

Conjugation experiments. Because pLMIV lacks genes involved in antibiotic resistance and thus lacks selective markers that could be used to select for putative transconjugants, we introduced a kanamycin resistance gene (*kanR*) into pLMIV. The *kanR* gene was introduced into the *inlP3* gene of pLMIV as previously described by Orsi et al. (33). Conjugation experiments were performed with the kanamycin-resistant FSL J1-208 mutant (FSL B2-322) as donor and the streptomycin-resistant *L. monocytogenes* 10403S (a lineage II strain) as recipient. Donor and recipient

TABLE 1 Overview of isolates associated with the 1997-1998 caprine outbreak^a as well as an additional isolate obtained from a separate farm in Georgia

Isolate	Date of first disease signs	Location	Isolate source	Serotype	Ribotype	10-Gene MLST ST ^e	Lineage	Result for PCR-based confirmation of ^b :			
								<i>P60</i>	<i>inIP3</i>	<i>cnaB</i>	ori
FSL J1-158	May 1997	Farm 1, GA	Vaginal swab	4b	DUP-10142	lmo22	IV	+	+	—	+
FSL J1-159	May 1997	Farm 1, GA	Rectal swab ^d	4b	DUP-10142	lmo22	IV	+	+	—	+
FSL J1-160	May 1997	Farm 1, GA	Fetus ^d	4b	DUP-10142	lmo22	IV	+	+	—	+
FSL J1-208	Jan. 1998	Farm 1, GA	Brain	4a	DUP-10142	lmo23	IV	+	+	+	+
FSL M1-001	Jan. 1998	Farm 1, GA	Brain	4b	DUP-10142	lmo22	IV	+	+	—	+
FSL M1-002	Mar. 1998	Farm 1, GA	Brain	4b	DUP-10142	lmo22	IV	—	—	—	—
FSL M1-003 ^c	Feb. 1998	Farm 2, GA	Brain	4b	DUP-10148	lmo24	IIIC	+	—	+	—

^a This outbreak has been described in detail by Wiedmann et al. (44).^b The PCR targeted *P60*, a lytic transglycosylase involved in conjugative transfer, *inIP3*, a plasmid-specific internalin, and *ori*, the origin of replication of the plasmid.^c This strain was not included in the report by Wiedmann et al. (44).^d Sample was obtained from the same goat as isolate FLJ1-158.^e ST based on the 10-gene MLST described by den Bakker et al. (10).

strains were grown overnight in liquid brain heart infusion (BHI) broth at 37°C with shaking. Aliquots (100 µl) of the donor and recipient strain cultures were spotted on top of each other on a BHI agar plate, left for 1 h at room temperature, and subsequently incubated overnight at 37°C. The next day the lawn, consisting of donor and recipient strains, was liquefied in BHI and plated on BHI agar plates containing 50 µg/ml of kanamycin and 200 µg/ml of streptomycin and on plates containing only one of the selective agents as a control. PCR amplification of *sigC*, a gene found in 10403S but not in FSL J1-208, was used to discriminate putative transconjugant 10403S colonies from FSL J1-208 colonies that had spontaneously acquired streptomycin resistance.

Nucleotide sequence accession numbers. The genome sequence of *L. monocytogenes* FSL J1-208 generated in this study has been deposited in GenBank. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [AEIS00000000](#). The version described in this paper is the first version, accession number [AEIS01000000](#).

RESULTS

***L. monocytogenes* FSL J1-208 has a small genome that lacks prophages and large genomic islands.** The total length of the draft chromosome measured 2,741,253 bp, containing 2,772 predicted protein-coding sequences. The total length of the rRNA clusters and the short unsequenced regions adjacent to these clusters is approximately 35,276 bp, which makes for an estimated total chromosome length of 2,776,529 bp. This size makes this chromosome the smallest among the currently sequenced *L. monocytogenes* chromosomes and even the smallest chromosome in the genus *Listeria*. This chromosome carries only 16 internalins, a small number compared to most other *L. monocytogenes* strains, such as F2365 (lineage I; 25 internalins [4]), EGD-e (lineage II; 25 internalins [4]), FSL F2-208 (lineage IIIC; 23 internalins [11]), but with internalins remarkably close to the number of chromosomal internalins found in virulence-attenuated lineage IIIA strains (HCC23 and M7 [17 internalins] and L99 [18 internalins]). However, most of the internalin genes currently proven to encode important virulence factors (*inlA*, *inlB*, and *inlC*) are present on the chromosome of FSL J1-208. In addition, the full *prfA* virulence cluster was present in FSL J1-208.

The FSL J1-208 genome contains a 10.73-kb region, which encodes genes associated with prophages, and putatively represents a remnant of a prophage or mononin region and therefore is not considered a prophage. No genomic islands (regions introduced

by HGT) were found with IslandViewer; however, a SIGI-HMM search indicated that 78 of the ORFs found in the FSL J1-208 genome were potentially introduced by HGT. A BLASTx search revealed that (i) 48 of these ORFs are also present in other *L. monocytogenes* genomes and therefore may represent either genes incorrectly identified as introduced into the genome by HGT or genes that have been acquired by the most recent common ancestor of *L. monocytogenes*, (ii) 10 ORFs were present in *L. monocytogenes* lineage III/IV but in no other *L. monocytogenes* genomes and may therefore have been acquired by the most recent common ancestor of lineages III and IV, (iii) 3 ORFs had no similarity to any protein in the NCBI database, and (iv) 14 ORFs that had only similarity to genes found in genomes of non-*L. monocytogenes* Firmicutes (e.g., *L. seeligeri*, *Bacillus cereus*, *Enterococcus faecium*). Although most of the 14 ORFs with similarity to non-*L. monocytogenes* genes have no known function (i.e., conserved hypothetical genes), one ORF (LMIV_2122) showed homology to a type II intron-related reverse transcriptase and is part of a cluster of three ORFs that was likely introduced into the genome by HGT.

Comparative analysis of the genome sequences of FSL J1-208 and HCC23 (lineage III; genome size, 2,976,212 bp) revealed 2,591 genes that were shared between both strains. FSL J1-208 contains 181 genes found in FSL J1-208 but not in HCC23, and HCC23 counts 414 genes that are found in HCC23 but not in FSL J1-208. The difference in genome size between FSL J1-208 and HCC23 is about 200 kb; 61.6% of this genome size difference (corresponding to about 124 kb) can be attributed to the complete absence of prophages in the FSL J1-208 genome (Table 2). Miniclusters account for 30.7% of the chromosomal length difference between HCC23 and FSL J1-208 (see Tables S3 and S4 in the supplemental material for an overview of miniclusters and prophage regions); while the number of miniclusters is approximately the same in HCC23 and FSL J1-208 (38 and 36, respectively), the average length of miniclusters in FSL J1-208 is shorter (3,168 bp) than in HCC23 (4,625 bp). Singleton insertions are responsible for 7.6% of the chromosomal length difference between FSL J1-208 and HCC23. The number of singleton insertions in FSL J1-208 is smaller than in HCC23 (29 and 43, respectively), and singleton insertions in FSL J1-208 are on average shorter than in HCC23 (726 bp and 845 bp, respectively). Since most of the difference in chromosome length between FSL J1-208 and HCC23 is accounted

TABLE 2 Approximate lengths of strain-specific insertions

Insertion	FSL J1-208		HCC23	
	Total length (bp) ^a	Avg length (bp) ^b	Total length (bp) ^a	Avg length (bp) ^b
Prophage	0	0	123,955	41,318
Miniclust	114,052	3,168	175,737	4,625
Singleton	21,351	736	36,753	855

^a The length of the insertion was calculated from the first base pair of the first ORF in the insertion/deletion region to the last base pair of the last ORF in the insertion/deletion region.

^b The average length of a genomic insertion (either a prophage, miniclust, or singleton insertion) was calculated by dividing the total length of the insertion by the number of insertions.

for by the absence of prophages in the genome of FSL J1-208, we decided to test if there was a correlation between chromosome size and the sum of the length of prophages in *Listeria* (see Table S5 in the supplemental material). We obtained data from nine fully sequenced *Listeria* genomes (*L. monocytogenes* F2365, *L. monocytogenes* EGD-e, *L. monocytogenes* 08-5578, *L. monocytogenes* 08-5923, *L. monocytogenes* HCC23, *L. innocua* CLIP11262, *L. welshimeri* SLCC5334, and *L. seeligeri* SLCC) and found a significant correlation between chromosome size (including prophages), and the total length of prophages (Fig. 1; Pearson's product-moment correlation $r = 0.81$; $P = 0.0082$). There was no significant correlation between the chromosome size without the prophage regions and the total prophage size (Pearson's product-moment correlation $r = 0.07$; $P = 0.85$).

***L. monocytogenes* FSL J1-208 contains a plasmid that carries virulence-related genes but no resistance genes.** Automated annotation of the contigs resulting from the initial *de novo* assembly of the Illumina reads showed that two contigs contained a number of plasmid-associated genes, such as genes involved in plasmid replication and conjugation. Closing of the gaps between the two contigs confirmed that these two contigs form a circular plasmid of 77,825 bp. Phylogenetic analysis based on *repA* sequences (Fig. 2) placed pLMIV in the same clade as other previously described *Listeria* plasmids (23); however, it is found in a distinct subclade.

The size and the presence of this plasmid were further confirmed by PFGE and Southern blotting with probes targeting the lytic transglycosylase gene involved in conjugation of plasmids and an internalin gene (*inlP3*) predicted to be on this plasmid. PFGE of undigested genomic DNA revealed a band of about 80 kbp, representing the plasmid (see Fig. S1 in the supplemental material). This band hybridized with the plasmid-specific probes and was absent in the negative control (*L. monocytogenes* 10403S).

The plasmid found in strain FSL J1-208 contains 82 ORFs; most of these ORFs have homology to genes involved in plasmid replication (homologues of *repA*, *repB*, and *parA*) and conjugation (Fig. 3A; Table 3). The position of the switch in GC skew coincides with the position of the intergenic region between *repB* and *repA*, which is typical for the origin of replication of theta-replicating plasmids (44). Within the conjugation region is a putative transfer operon (a type IV secretion-like system). ORFs in this operon encode homologues of transfer genes, such as *traE*, *traG*, and a lytic transglycosylase. Genes found in this operon show a slightly higher similarity to homologues found in plasmids of *Enterococcus* species (median 36% amino acid [aa] identity) than homologues found in plasmids in other *Listeria* species (median 28% aa identity).

The plasmid contains four genes belonging to the internalin family (named *inlP1* to *inlP4*) and one ORF that lacks the leucine-rich repeat (LRR) region characteristic for internalins but contains other internalin-specific conserved domains, such as an LRR-adjacent domain, two *Listeria/Bacteroides* repeats, and an LPXTG motif (Fig. 3B). A phylogenetic comparison of these five internalin-like genes to internalin-like genes found in previously sequenced *Listeria* genomes showed that four of these genes cluster in a clade that contains known virulence-associated internalins, like internalin A and internalins of the internalin C2/H cluster (Fig. 4). One of the plasmid-borne internalin-like genes (*inlP3*) is placed in a completely different part of the tree (Fig. 4), indicating that this internalin is not closely related to the other plasmid-borne internalins and may have been introduced into the plasmid on a separate occasion during the evolutionary history of the plasmid. The invasion efficiency in human epithelial Caco-2 cells of the plasmid-cured FSL B2-294 strain was not significantly different from its parental strain FSL J1-208 (Wilcoxon rank sum test, $P = 0.7$) (Fig. 5).

Besides the aforementioned internalin-like genes, a small cluster of accessory genes encoding a putative two-component regulatory system and an ABC transporter shows amino acid homology to a similar gene cluster in the genomes of several species of *Clostridium* (e.g., *Clostridium perfringens* strain 13, *Clostridium scindens* ATCC 35704). The putative function of this gene cluster is currently unknown.

Homologues of the *cna-B* repeat protein found on the FSL J1-208 plasmid are commonly found in ICE in *Firmicutes* and plasmids of Gram-positive bacteria and are possibly involved in the conjugation process. Upstream of the transfer operon of plasmid pLMIV, an ORF encoding a relatively large (968-aa) protein which contains six *cna-B*-type domains is found. *cna-B*-type domains are found in collagen binding proteins; however, these domains do not mediate collagen binding but are thought to form a stalk in the collagen binding protein (8). An LPXTG motif cell wall

FIG 1 Scatter plot of chromosome sizes of various *Listeria* strains versus total sizes of prophage regions. Values on the x axis indicate the total chromosome sizes (in bp), and values on the y axis indicate the total sizes of prophage regions. Abbreviations: *L.m.*, *Listeria monocytogenes*; *L.i.*, *Listeria innocua* CLIP11262; *L.w.*, *L. welshimeri* SLCC5334; *L.s.*, *L. seeligeri* SLCC3954.

FIG 2 Phylogenetic classification of pLMIV based on sequence information for *repA*. The tree was constructed using a maximum likelihood optimization criterion. Values on the branches are bootstrap values based on 100 maximum likelihood bootstrap replicates. Abbreviations: *Lactob.*, *Lactobacillus*; *Lactoc.*, *Lactococcus*; *Pedioc.*, *Pediococcus*; *Staph.*, *Staphylococcus*; *Strep.*, *Streptococcus*. The compositions of plasmids in *Listeria* groups 1 and 2 have been described by Kuenne et al. (23).

anchor domain and a signal peptide are also present in this protein, indicating that this protein is anchored to the peptidoglycan cell wall. Homologous *cna-B* domain-containing proteins were found in the genome sequences of *L. monocytogenes* FSL N3-165 (44% aa identity; six *cna-B* domains), *L. welshimeri* SLCC5334 (24% aa identity; four *cna-B* domains), *L. monocytogenes* EGD-e (26% aa identity; four *cna-B* domains), *L. monocytogenes* FSL F2-208 (46% aa identity; six *cna-B* domains); in all cases these proteins were found upstream of the transfer operon of integrated conjugative elements (ICE). The fact that this protein is found in both ICE and plasmids in *Clostridium* (3) in a similar position relative to the transfer operon seems to indicate that this gene plays a general role in the conjugation process in *Firmicutes*.

A PCR-based screen for the presence of a pLMIV-like plasmid among 21 *Listeria* species isolates and 50 *L. monocytogenes* isolates (see Table S2 for more information on the isolates) only revealed pLMIV-like plasmids to be present in *L. monocytogenes* isolates. Four isolates (FSL J1-158, FSL J1-159, FSL J1-160, and FSL M1-001) were found to be positive for *inlP3*, *P60*, and the plasmid-specific origin of replication; however, a fragment of the *cna-B* repeat protein could not be amplified from these isolates. This indicated that these isolates (all involved in the same caprine listeriosis outbreak) carry a plasmid that is very similar to pLMIV. Interestingly, an isolate of a sporadic caprine listeriosis case (FSL M1-003) from the same geographical region as the outbreak isolates tested positive for the *cna-B* repeat protein and *P60*; however,

FIG 3 (A) Map of pLMIV. Arrows, ORFs encoding proteins with similarity to proteins found in other Gram-positive bacteria; bars, hypothetical proteins without sequence similarity to currently known proteins. Green arrows, proteins with BLAST similarity to genes associated with plasmid replication and conjugation; red arrows, internalins or internalin-related proteins; blue arrows, transposases or integrases. Asterisks, proteins with >20% similarity to proteins found in plasmid pMG1 of *Enterococcus faecium*; dots, proteins with >20% similarity to proteins found in plasmid pCP13 of *Clostridium perfringens*. (Inner circle) Histogram of the GC skew. Yellow bars, relative overrepresentation of GC; purple bars, relative underrepresentation of GC. (B) Overview of conserved domains/repeats found in the four plasmid-borne internalins and the one internalin-like protein. Numbers within the regions represent the number of repeats; numbers behind the bars represent the total lengths of the given genes (in amino acid residues).

TABLE 3 Open reading frames in plasmid pLMIV

Locus tag	Strand orientation	Length (bp)	Function	General function	Highest BLAST match	Affiliation of BLAST match	E-value	% aa identity ^a
LMIV_p001	+	195	Hypothetical protein					
LMIV_p002	+	276	Hypothetical protein					
LMIV_p003	+	606	Integrase/recombinase		ref ZP_05644917.1 recombinase/integrase	<i>E. casseliflavus</i> EC30	2.00E-51	54
LMIV_p004	+	651	Hypothetical protein		ref YP_001468413.1 gp27	<i>Listeria</i> phage A500	1.00E-33	36
LMIV_p005	+	201	Hypothetical protein		ref YP_849011.1 hypothetical protein lwe0810	<i>L. welshimeri</i> serotype 6b SLCC5334	2.00E-12	60
LMIV_p006	+	306	Hypothetical protein		ref ZP_06684183.1 predicted protein	<i>L. monocytogenes</i> FSL J1-194	5.00E-32	73
LMIV_p007	+	1,152	Putative DNA polymerase IV, ImpB/MucB/SamB family protein	DNA repair	ref YP_850822.1 ImpB/MucB/SamB family protein	<i>L. welshimeri</i> serotype 6b SLCC5334	0.00E+00	85
LMIV_p008	-	480	Hypothetical protein		ref ZP_02418054.1 hypothetical protein ANACAC_00621	<i>Anaerostipes caccae</i> DSM 14662	6.00E-05	31
LMIV_p009	-	825	Putative plasmid partition protein ParA	Replication	gb AAF27301.1 AF154674_2 ParA	<i>L. lactis</i> subsp. <i>lactis</i>	4.00E-42	36
LMIV_p010	-	738	Plasmid replication initiator protein RepB	Replication	ref YP_003329296.1 hypothetical protein pCD02p7	<i>L. paracasei</i> subsp. <i>paracasei</i>	5.00E-35	40
LMIV_p011	+	1,650	Plasmid replication protein RepA	Replication	ref ZP_05231473.1 plasmid replication protein	<i>L. monocytogenes</i> FSL J1-194	5.00E-180	63
LMIV_p012	-	948	Transcriptional activator		ref YP_002349129.1 transcriptional activator, Rgg/GadR/ MutR family, C	<i>L. monocytogenes</i> HCC23	4.00E-35	30
LMIV_p013	+	186	Hypothetical protein					
LMIV_p014	-	900	Transcriptional activator		ref YP_015009.1 transcriptional activator, putative	<i>L. monocytogenes</i> 4b F2365	4.00E-22	28
LMIV_p015	+	576	Hypothetical protein					
LMIV_p016	+	1,077	Hypothetical protein signal peptide prediction		ref ZP_05232385.1 predicted protein	<i>Listeria monocytogenes</i> FSL N3-165	9.00E-15	27
LMIV_p017	+	2,904	CnaB protein	Conjugation	ref ZP_05232384.1 conserved hypothetical protein	<i>Listeria monocytogenes</i> FSL N3-165	0	46
LMIV_p018	+	369	Hypothetical protein					
LMIV_p019	+	780	Hypothetical protein		ref ZP_04807304.1 conserved hypothetical protein	<i>Clostridium cellulovorans</i> 743B	7.8E-2	25
LMIV_p020	+	246	Hypothetical protein					
LMIV_p021	+	156	Hypothetical protein					
LMIV_p022	+	513	Hypothetical protein		ref ZP_05266897.1 conserved hypothetical protein	<i>Listeria monocytogenes</i> HPB2262	7.00E-61	65
LMIV_p023	+	375	Hypothetical protein					
LMIV_p024	+	255	Hypothetical protein					
LMIV_p025	+	225	Hypothetical protein		ref ZP_03964062.1 conserved hypothetical protein	<i>Lactobacillus paracasei</i> subsp. <i>paracasei</i> ATCC 25302	9.00E-09	41
LMIV_p026	+	1,176	Hypothetical protein	Conjugation	ref YP_805821.1 hypothetical protein LSEI_0530	<i>Lactobacillus casei</i> ATCC 334	5.00E-06	19

(Continued on following page)

TABLE 3 (Continued)

Locus tag	Strand orientation	Length (bp)	Function	General function	Highest BLAST match	Affiliation of BLAST match	E-value	% aa identity ^a
LMIV_p027	+	1,209	Probable cell surface protein (LPXTG motif)	Conjugation	ref ZP_02078073.1 hypothetical protein EUBDOL_01887	<i>Eubacterium dolichum</i> DSM 3991	3.00E-52	46
LMIV_p028	+	783	LtrC protein	Conjugation	ref YP_001032676.1 LtrC protein	<i>Lactococcus lactis</i> subsp. <i>cremoris</i> MG1363	3.00E-18	30
LMIV_p029	+	201	Hypothetical protein	Conjugation				
LMIV_p030	+	375	Hypothetical protein	Conjugation				
LMIV_p031	+	1,062	Hypothetical protein	Conjugation	ref YP_001038243.1 chromosome segregation ATPase-like protein	<i>Clostridium thermocellum</i> ATCC 27405	8.00E-06	27
LMIV_p032	+	2,220	TraG, type IV secretion system	Conjugation	ref ZP_05423903.1 predicted protein	<i>Enterococcus faecalis</i> T1	7.00E-137	40
LMIV_p033	+	1,800	Transmembrane protein	Conjugation	emb CBL31234.1 hypothetical protein	<i>Enterococcus</i> sp. 7L76	1.00E-65	36
LMIV_p034	+	321	Hypothetical protein	Conjugation				
LMIV_p035	+	2,535	TraE, type IV secretion system	Conjugation	ref ZP_05423908.1 predicted protein	<i>E. faecalis</i> T1	2.00E-150	42
LMIV_p036	+	621	Putative lipoprotein LpqB	Conjugation	ref ZP_01174130.1 hypothetical protein B14911_28540	<i>Bacillus</i> sp. NRRL B-14911	1.00E-04	24
LMIV_p037	+	1,089	Lytic transglycosylase, P60 domain family	Conjugation	ref YP_001654023.1 putative lipoprotein	<i>B. thuringiensis</i> INTA-FR7-4	3.00E-54	36
LMIV_p038	+	816	Conserved hypothetical protein, plasmid related	Conjugation	ref ZP_05424150.1 predicted protein	<i>E. faecalis</i> T2	5.00E-14	22
LMIV_p039	+	249	Hypothetical protein	Conjugation				
LMIV_p040	+	426	Hypothetical protein	Conjugation	ref ZP_05423913.1 predicted protein	<i>E. faecalis</i> T1	4.00E-08	29
LMIV_p041	+	1,200	Hypothetical protein	Conjugation	emb CBL31241.1 hypothetical protein	<i>Enterococcus</i> sp. 7L76	3.00E-89	43
LMIV_p042	+	774	Transmembrane protein	Conjugation				
LMIV_p043	+	1,650	Zinc β -ribbon domain-containing protein	Conjugation	emb CBL31243.1 hypothetical protein	<i>Enterococcus</i> sp. 7L76	1.00E-83	34
LMIV_p044	+	855	Hypothetical protein	Conjugation	ref ZP_02631385.1 hypothetical protein AC3_A0232	<i>C. perfringens</i> serotype E JGS1987	1.00E-34	32
LMIV_p045	+	390	Hypothetical protein	Conjugation	ref ZP_03670544.1 hypothetical protein LmonFR_06904	<i>L. monocytogenes</i> FSL R2-561	4.00E-20	92
LMIV_p046	-	264	Hypothetical protein	Conjugation	ref NP_463994.1 hypothetical protein lmo0465	<i>L. monocytogenes</i> EGD-e	6.00E-21	89
LMIV_p047	+	183	Hypothetical protein	Conjugation				
LMIV_p048	+	525	Hypothetical protein	Conjugation				
LMIV_p049	+	273	Hypothetical protein	Conjugation				
LMIV_p050	+	732	Competence-specific nuclease	Conjugation	ref YP_003170382.1 DNA-entry nuclease	<i>L. rhamnosus</i> GG	1.00E-49	52
LMIV_p051	+	624	Hypothetical protein					
LMIV_p052	+	234	Hypothetical protein					
LMIV_p053	+	1,239	Cytosine-specific methyltransferase		ref YP_003445902.1 methyl transferase	<i>S. mitis</i> B6	3.00E-124	54
LMIV_p054	+	1,371	Hypothetical protein		ref ZP_04863688.1 conserved hypothetical protein	<i>Clostridium</i> phage D-1873	6.00E-09	22
LMIV_p055	-	261	Hypothetical protein		ref ZP_05260880.1 hypothetical protein LmonJ_14116	<i>L. monocytogenes</i> J0161	4.00E-30	80

(Continued on following page)

TABLE 3 (Continued)

Locus tag	Strand orientation	Length (bp)	Function	General function	Highest BLAST match	Affiliation of BLAST match	E-value	% aa identity ^a
LMIV_p056	—	453	Hypothetical protein		ref ZP_05268857.2 phage protein	<i>L. monocytogenes</i> F6900	2.00E-16	58
LMIV_p057	+	150	Hypothetical protein		ref NP_463996.1 hypothetical protein lmo0467	<i>L. monocytogenes</i> EGD-e	2.2E-2	44
LMIV_p058	+	1,482	Hypothetical protein					
LMIV_p059	+	114	Hypothetical protein					
LMIV_p060	+	363	Hypothetical protein					
LMIV_p061	+	114	Hypothetical protein					
LMIV_p062	+	1,443	inIP1		gb ABC26543.1 internalin C2	<i>L. monocytogenes</i> FSL F2-601	2.00E-86	42
LMIV_p063	+	267	Transposase OrfA, IS3 family, putative					
LMIV_p064	+	1,782	inIP2		ref YP_003465839.1 putative secreted protein	<i>L. seeligeri</i> serotype 1/2b SLCC3954	4.00E-132	44
LMIV_p065	+	645	Putative exported protein		ref YP_848511.1 hypothetical protein lwe0310	<i>L. welshimeri</i> serotype 6b SLCC5334	5.00E-108	89
LMIV_p066	+	735	Putative exported protein		ref YP_848511.1 hypothetical protein lwe0310	<i>L. welshimeri</i> serotype 6b SLCC5334	5.00E-114	80
LMIV_p067	+	3,819	inIP3		ref ZP_00234781.1 internalin protein lin1204	<i>L. monocytogenes</i> serotype 1/2a F6854	0.00	48
LMIV_p068	+	267	Transposase OrfA, IS3 family, putative		ref YP_013105.1 IS3 family transposase OrfA	<i>L. monocytogenes</i> serotype 4b F2365	4.00E-20	59
LMIV_p069	+	276	Transposase OrfA, IS3 family, putative		ref ZP_05389830.1 transposase	<i>L. monocytogenes</i> FSL J1-175	5.00E-15	49
LMIV_p070	+	918	Internalin-like protein		gb ABC26574.1 internalin D	<i>L. monocytogenes</i> FSL E1-123	2.00E-82	50
LMIV_p071	+	120	Hypothetical protein					
LMIV_p072	—	528	Transcriptional regulator, AcrR family		ref NP_347359.1 AcrR family transcriptional regulator	<i>C. acetobutylicum</i> ATCC 824	1.00E-21	32
LMIV_p073	+	954	Alpha/beta-hydrolase		ref ZP_02861586.1 hypothetical protein ANASTE_00793	<i>Anaerofustis stercorihominis</i> DSM 17244	1.00E-70	45
LMIV_p074	+	876	Cell wall endopeptidase, M23/M37 family		ref ZP_04322876.1 peptidase, M23/M37 family	<i>B. cereus</i> m1293	1.00E-55	41
LMIV_p075	—	804	Putative transposase OrfB		ref YP_849813.1 putative transposase	<i>L. welshimeri</i> serotype 6b SLCC5334	4.00E-145	94
LMIV_p076	—	552	Transposase IS3/IS911		ref YP_849812.1 putative transposase	<i>L. welshimeri</i> serotype 6b SLCC5334	6.00E-77	80
LMIV_p077	—	2,031	Putative antibiotic ABC transporter, membrane protein		ref ZP_02431192.1 hypothetical protein CLOSCI_01412	<i>C. scindens</i> ATCC 35704	1.00E-137	37
LMIV_p078	—	768	ABC transporter, ATP-binding protein		ref ZP_02431191.1 hypothetical protein CLOSCI_01411	<i>C. scindens</i> ATCC 35704	3.00E-84	62
LMIV_p079	—	795	Sensor histidine kinase		ref ZP_02431190.1 hypothetical protein CLOSCI_01410	<i>C. scindens</i> ATCC 35704	8.00E-54	39
LMIV_p080	—	669	Two-component response regulator		ref ZP_02431189.1 hypothetical protein CLOSCI_01409	<i>C. scindens</i> ATCC 35704	5.00E-60	51
LMIV_p081	+	123	Hypothetical protein					
LMIV_p082	+	2,331	inIP4		dbj BAI78336.1 internalin A	<i>L. monocytogenes</i> 55-4-3	0.00	45

^a The percent amino acid identity, based on a BLAST match to the ORF found in pLMIV.

it tested negative for the presence of *inlP3* and the plasmid-specific origin of replication. The absence of the plasmid-specific origin of replication suggests that this isolate may carry an ICE with pLMIV-like conjugation genes that integrated into the chromosome. Multilocus sequence analysis (Fig. 6) showed that all the isolates that contained pLMIV-like plasmids were found in lineage IV, while the isolate carrying the putative chromosomal ICE belonged to lineage III. Specifically, isolates FSL J1-158, FSL J1-159, FSL J1-160, and FSL M1-001 were found in a well-supported phylogenetic clade (100% bootstrap support) (Fig. 6) within lineage IV, while FSL J1-208 was found in a distinct clade within lineage IV (Fig. 6). The nucleotide sequences of the *inlP3* fragment (156 bp) and the *P60* fragment (209 bp) were identical in FSL J1-208 and FSL J1-158, FSL J1-159, FSL J1-160, and FSL M1-001, which suggests recent transfer of the plasmid between these strains. However, given the short length of these sequences, there is a high probability (7.9%) of observing identical sequences (based on a 0.7% divergence of FSL J1-208 versus FSL J1-158, FSL J1-159, FSL J1-160, and FSL M1-001 and a Poisson distribution for the probability of observing a single-nucleotide polymorphism per nucleotide site). Transfer of the plasmid with the kanamycin resistance gene from FSL J1-208 to *L. monocytogenes* 10403S was not observed during two replicates of the conjugation experiments.

DISCUSSION

In this study we produced and analyzed a near-complete draft genome of FSL J1-208, a representative of *L. monocytogenes* lineage IV. This genome, along with the genomes available for lineage III strains, will provide an improved understanding of *L. monocytogenes* evolution. Our analyses specifically showed that (i) the genome of *L. monocytogenes* FSL J1-208 is among the smallest in the genus *Listeria*, (ii) this small genome size can be largely attributed to the absence of prophages in the genome of FSL J1-208, (iii) FSL J1-208 contains a rare internalin-carrying plasmid, and (iv) this plasmid contains a *cna-B* repeat protein, which seems to be a hallmark for a certain class of integrative and conjugative elements and plasmids in *Firmicutes*.

The small genome size of *L. monocytogenes* FSL J1-208 can largely be explained by the absence of large prophage regions. Our estimate of the chromosome size of FSL J1-208 is 2,776,529 bp, which makes it the smallest *Listeria* chromosome currently sequenced. The previously reported smallest genomes in *Listeria* were the genomes of *L. welshimeri* SLCC5334 (19) and *L. seeligeri* SLCC3954 (48), which measure 2,814,130 and 2,797,636 bp, respectively. These small genome sizes have been attributed to genome reduction associated with loss of virulence in these species (19, 48); however, this is clearly not the case in *L. monocytogenes* FSL J1-208. The virulent nature of FSL J1-208 is not only demonstrated by its involvement in a listeriosis outbreak (55), but also with a guinea pig model (31), where it proved to be among the most virulent *L. monocytogenes* strains. Comparison of the genomes of FSL J1-208 and HCC23, the most closely related *L. monocytogenes* strain for which the genome has been fully sequenced, showed that 63% of the difference in size between these two strains could be attributed to the absence of prophages in FSL J1-208. Gene clusters that could only be found in FSL J1-208 and not in HCC23 were generally smaller than similar HCC23-specific clusters, suggesting a reduction in the genome of FSL J1-208 compared to HCC23.

Although genome reduction in bacteria has been shown to be commonly associated with a transition of a saprotrophic to a pathogenic or obligate symbiotic lifestyle (7, 28), it has also been shown to occur in the adaptation to nutrient-rich food environments in *Lactobacillus* species (43). This suggests that genome reduction is generally associated with niche specialization and may indicate that strains of pathogenic and nonpathogenic *Listeria* with a small genome are more specialized to a specific niche than strains with a larger genome. We showed here that genome size in *Listeria* is highly correlated with the presence of prophages, and it should be noted that prophages are highly mobile and seem to be easily lost or acquired by a strain (as shown by Orsi et al. [32]). The range in genome size between strains within the genus *Listeria* is very limited compared to the range in genome size of aforementioned *Lactobacillus* species and obligate endosymbionts. Genomes sizes in *Listeria* range from 2.8 to 2.9 Mb (without plasmid and prophages), while *Lactobacillus* species range in genome size from 1.8 Mb (*L. bulgaricus* ATCC 11842 [49]) to 3.3 Mb (*L. plantarum* [22]), which suggests that chromosome/genome sizes are more conserved among *Listeria* spp. than among *Lactobacillus* spp.

Total prophage region size is generally highly correlated with genome size among bacteria (29). We found that this was true for the genus *Listeria* when total prophage region size was included in the chromosome size; however, this correlation is absent when the total prophage region size in a chromosome is compared to the chromosome size with the prophage regions excluded. This indicates that, while total prophage size is a major contributor to the variation in total genome size in *Listeria*, larger genomes (measured without prophage regions) do not necessarily harbor a larger proportion of prophage-related regions. We hypothesize that *Listeria*, being highly conserved in gene synteny (18), has only a limited number of sites in the genome that allow for prophage insertion without deleterious effects for the host.

Ecological lifestyle has been shown to affect mobile DNA gene content (such as prophages), with obligate intracellular bacteria having the least mobile DNA, extracellular bacteria having more mobile DNA genes than obligate intracellular bacteria, and facultative intracellular bacteria having the most mobile DNA genes (29). *Listeria* is a facultative intracellular bacterium, and the absence of prophages in the genome of FSL J1-208 could therefore indicate that this particular strain has shifted in its ecology to having either a more saprotrophic (extracellular) lifestyle or a more intracellular (pathogenic) lifestyle. Additional research on the genomics and ecology of lineage IV strains is needed to confirm if there is really a difference in ecology compared to other *L. monocytogenes* lineages and to confirm that loss of prophage regions is a general trend in the genome evolution of lineage IV.

The plasmid found in *L. monocytogenes* FSL J1-208 represents a novel plasmid for *Listeria* and is characterized by the presence of several internalins. Our data revealed a new type of plasmid for *L. monocytogenes*, which is a putative virulence plasmid. In contrast, all plasmids described for the genus *Listeria* to this date are classified as resistance plasmids, i.e., they carry antibiotic and/or heavy metal resistance genes (see Kuenne et al. [23] for an overview), and no virulence plasmids have previously been described. The origin of plasmid pLMIV is unclear; however, different functional regions (e.g., regions encoding genes involved in replication or conjugation and the region containing accessory genes) show different levels of sequence similarity with plasmid-associated genes in other *Firmicutes* genera, supporting a possible

FIG 5 Invasion efficiencies in Caco-2 cells of *L. monocytogenes* strains 10403S, FSL B2-294 (FSL J1-208 strain cured of the plasmid pLMIV), and FSL J1-208. Invasion efficiency (the number of recovered cells per number of cells used for inoculation) was normalized to the invasion efficiency obtained for *L. monocytogenes* 10403S of the same biological replicate (hence, 10403S shows an invasion efficiency of 100% without an error bar in this figure). Three independent invasion assays (consisting of three technical replicates each) were performed for each strain tested.

chimeric origin of the plasmid. For instance, genes found in the accessory region of this plasmid show amino acid homology to genes found in *Listeria* genomes (both *L. monocytogenes* and *Listeria* spp.), and a small number of accessory genes shows amino acid similarity to genes found in *Clostridium*. ORFs found in the region involved in conjugational transfer show relatively low homology (median of 36% aa identity) to genes of various *Firmicutes* (mainly *Enterococcus*, *Lactobacillus*, and *Listeria*) currently found in GenBank, which most likely reflects the novelty of this plasmid and our limited knowledge of plasmid diversity in *Listeria* and other *Firmicutes*.

Plasmids similar to pLMIV identified previously in *Clostridium botulinum* (27) and *Clostridium perfringens* (40) have an extremely low transfer frequency, consistent with our observations that (i) pLMIV was found in a small portion of the *Listeria* population and (ii) conjugation experiments between FSL J1-208 (lineage IV) and *L. monocytogenes* 10403S (lineage II) did not yield any transconjugants. However, the fact that pLMIV or variants thereof were present in two divergent subtypes associated with the same caprine outbreak suggests that the plasmid can be transferred between strains in nature. Possibly, this plasmid can only be transferred between strains within a lineage and is unlikely to be transferred between lineages. Alternatively, this plasmid may have been acquired by the most recent common ancestor of lineage IV and subsequently lost in some subpopulations. The fact that we found identical partial sequences of *P60* and *inlP3* in FSL J1-208 and FSL

FIG 6 Maximum likelihood tree based on the concatenated sequences of the MLST scheme described by den Bakker et al. (10). Values on the branches are bootstrap values based on 100 maximum likelihood bootstrap replicates. The intraspecific phylogenies of *L. welshimeri*, *L. seeligeri*, *L. innocua*, and *L. marthii*, the subspecies of *L. ivanovii*, and lineages I and II of *L. monocytogenes* have been collapsed to triangles. Open circles indicate the presence of a pLMIV-like plasmid (as determined by PCR of the origin of replication of the plasmid and *inlP3*), and the arrow indicates the presence of a putative pLMIV-like integrated conjugative element (as determined by PCR of *P60* and *cnaB*).

J1-158, FSL J1-159, FSL J1-160, and FSL M1-001 seems to support the hypothesis of horizontal transfer of the plasmid; however, additional experiments involving lineage IV recipients and analysis of the sequence divergence of (preferably full) plasmid sequences are necessary to conclusively confirm whether this plasmid is transferred horizontally or vertically.

It is generally assumed that plasmids deliver an evolutionary advantage to their host to make up for the fitness costs of plasmid replication (36). We did not find a significant difference between the invasion efficiency of human intestinal epithelial Caco-2 cells of a plasmid-cured strain compared to its parental strain, suggesting no involvement of the plasmid-borne internalins in the invasion of human epithelial cells. In the case of pLMIV, this selective advantage might be a better compatibility to the caprine host, which may also explain the presence of this plasmid in three genetically different strains that were involved in the same outbreak (Table 1; see also Wiedmann et al. [55]). Another selective advantage for strain FSL J1-208 might be that the plasmid complements the low number of internalins found on the chromosome (only 16 internalins).

Based on a PCR assay, at least two variants of the virulence plasmid were present in the isolates associated with the caprine listeriosis outbreak. The presence of these different variants (a variant with the lytic transglycosylase gene and the *inlP3* gene and a variant with the lytic transglycolase gene, the *inlP3* gene, and the *cna-B* repeat protein) seems to indicate that the plasmid is not stable in gene content over time.

The *cna-B* repeat protein found on the plasmid pLMIV is commonly associated with plasmids and integrated and conjugative elements. The role of the *cna-B* repeat protein, though often annotated as a collagen binding protein (42), is unknown. It is often suggested that this protein is a potential virulence protein, because of the presence of *cna-B* domains that are homologous to similar domains found in collagen binding proteins in *Staphylococcus aureus* (8). It has been shown, however, that other domains mediate collagen binding (pfam05737), and instead the *cna-B* repeats are thought to form a stalk (8) that projects, in the case of *S. aureus*, the actual collagen binding domains from the bacterial surface. We found that this type of protein, the *cna-B* repeat protein, is often associated with ICE in *Listeria* and with conjugative plasmids and ICE in other *Firmicutes* (*Clostridium* [3] and *Enterococcus* [38]). This association with the conjugative plasmids and ICE suggests that this *cna-B* repeat protein is involved in the actual conjugation process and may putatively be involved in coaggregation of donor and receptor cells.

Conclusion. This work shows that sequencing of rare genotypes of isolates from within relatively well-known species such as *Listeria monocytogenes* can greatly improve our knowledge of the biology and genome evolution of the whole species. In particular, in *L. monocytogenes* our knowledge of genome diversity and mobile elements has been largely dominated by isolates from either food items or human clinical cases, which represents probably only a small part of the diversity of this predominantly saprotrophic organism.

ACKNOWLEDGMENTS

We acknowledge the Broad Institute for letting us use the preliminary sequence data of *L. monocytogenes* FSL J1-208. We thank Sherry Roof for her help with sequencing and Renato Orsi for his helpful comments on the manuscript.

This work was supported by USDA Special Research grants 2005-34459-15625 and 2006-34459-16952.

REFERENCES

- Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38:W7–W13.
- Aziz RK, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.
- Bannam TL, Teng WL, Bulach D, Lyras D, Rood JL. 2006. Functional identification of conjugation and replication regions of the tetracycline resistance plasmid pCW3 from *Clostridium perfringens*. *J. Bacteriol.* 188:4942–4951.
- Bierne H, Sabet C, Personnic N, Cossart P. 2007. Internalins: a complex family of leucine-rich repeat-containing proteins in *Listeria monocytogenes*. *Microbes Infect.* 9:1156–1166.
- Braun U, Stehle C, Ehrensperger F. 2002. Clinical findings and treatment of listeriosis in 67 sheep and goats. *Vet. Rec.* 150:38–42.
- Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403.
- Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA. 2009. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc. Natl. Acad. Sci. U. S. A.* 106:9063–9068.
- Deivanayagam CC, et al. 2000. Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* 8:67–78.
- De Jesus AJ, Whiting RC. 2003. Thermal inactivation, growth, and survival studies of *Listeria monocytogenes* strains belonging to three distinct genotypic lineages. *J. Food Prot.* 66:1611–1617.
- den Bakker HC, Bundrant BN, Fortes ED, Orsi RH, Wiedmann M. 2010. A population genetics-based and phylogenetic approach to understanding the evolution of virulence in the genus *Listeria*. *Appl. Environ. Microbiol.* 76:6085–6100.
- den Bakker HC, et al. 2010. Comparative genomics of the bacterial genus *Listeria*: genome evolution is characterized by limited gene acquisition and limited gene loss. *BMC Genomics* 11:688.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
- Flamm RK, Hinrichs DJ, Thomashow MF. 1984. Introduction of pAM beta 1 into *Listeria monocytogenes* by conjugation and homology between native *L. monocytogenes* plasmids. *Infect. Immun.* 44:157–161.
- Glaser P, et al. 2001. Comparative genomics of *Listeria* species. *Science* 294:849–852.
- Graves LM, Swaminathan B. 2001. PulseNet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. *Int. J. Food Microbiol.* 65:55–62.
- Gray M, et al. 2004. *Listeria monocytogenes* isolates from foods and humans form distinct but overlapping populations. *Appl. Environ. Microbiol.* 70:5833–5841.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307–321.
- Hain T, Steinweg C, Chakraborty T. 2006. Comparative and functional genomics of *Listeria* spp. *J. Biotechnol.* 126:37–51.
- Hain T, et al. 2006. Whole-genome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria innocua* as compared to *Listeria monocytogenes*. *J. Bacteriol.* 188:7405–7415.
- Jeffers G, et al. 2001. Comparative genetic characterization of *Listeria monocytogenes* isolates from human and animal listeriosis cases. *Microbiology* 147:1095–1104.
- Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinformatics* 9:286–298.
- Kleerebezem M, et al. 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc. Natl. Acad. Sci. U. S. A.* 100:1990–1995.
- Kuenne C, et al. 2010. Comparative analysis of plasmids in the genus *Listeria*. *PLoS One* 5:e12511.
- Langille MGI, Brinkman FSL. 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 25:664–665.
- Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. 2008. Prophin-

- der: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics* 24:863–865.
26. Margolles A, de los Reyes-Gavilán CG. 1998. Characterization of plasmids from *Listeria monocytogenes* and *Listeria innocua* strains isolated from short-ripened cheeses. *Int. J. Food Microbiol.* 39:231–236.
27. Marshall KM, Bradshaw M, Johnson EA. 2010. Conjugative botulinum neurotoxin-encoding plasmids in *Clostridium botulinum*. *PLoS One* 5:e11087.
28. Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108:583–586.
29. Newton ILG, Bordenstein SR. 2010. Correlations between bacterial ecology and mobile DNA. *Curr. Microbiol.* 62:198–208.
30. Nightingale KK, Windham K, Martin KE, Yeung M, Wiedmann M. 2005. Select *Listeria monocytogenes* subtypes commonly found in foods carry distinct nonsense mutations in *inlA*, leading to expression of truncated and secreted internalin A, and are associated with a reduced invasion phenotype for human intestinal epithelial cells. *Appl. Environ. Microbiol.* 71:8764–8772.
31. Oliver HF, Orsi RH, Wiedmann M, Boor KJ. 2010. *Listeria monocytogenes* sigma B has a small core regulon and a conserved role in virulence but makes differential contributions to stress tolerance across a diverse collection of strains. *Appl. Environ. Microbiol.* 76:4216–4232.
32. Orsi RH, et al. 2008. Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment. *BMC Genomics* 9:539.
33. Orsi RH, Bowen BM, Wiedmann M. 2010. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. *BMC Genomics* 11:102.
34. Piffaretti JC, et al. 1989. Genetic characterization of clones of the bacterium *Listeria monocytogenes* causing epidemic disease. *Proc. Natl. Acad. Sci. U. S. A.* 86:3818–3822.
35. Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25:1253–1256.
36. Rankin DJ, Rocha EPC, Brown SP. 2010. What traits are carried on mobile genetic elements, and why? *Heredity* 106:1–10.
37. Rasmussen OF, Skouboe P, Dons L, Rossen L, Olsen JE. 1995. *Listeria monocytogenes* exists in at least three evolutionary lines: evidence from flagellin, invasive associated protein and listeriolysin O genes. *Microbiology* 141:2053–2061.
38. Rice LB, Carias LL, Marshall SH, Hutton-Thomas R, Rudin S. 2007. Characterization of Tn5386, a Tn916-related mobile element. *Plasmid* 58:61–67.
39. Roberts A, et al. 2006. Genetic and phenotypic characterization of *Listeria monocytogenes* lineage III. *Microbiology* 152:685–693.
40. Rood JJ, Scott VN, Duncan CL. 1978. Identification of a transferable tetracycline resistance plasmid (pCW3) from *Clostridium perfringens*. *Plasmid* 1:563–570.
41. Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945.
42. Sayeed S, Li J, McClane BA. 2007. Virulence plasmid diversity in *Clostridium perfringens* type D isolates. *Infect. Immun.* 75:2391–2398.
43. Schroeter J, Klaenhammer T. 2009. Genomics of lactic acid bacteria. *FEMS Microbiol. Lett.* 292:1–6.
44. Snyder L, Champness W. 2007. *Molecular genetics of bacteria*, 3rd ed. ASM Press, Washington, DC.
45. Southern E. 2006. Southern blotting. *Nat. Protoc.* 1:518–525.
46. Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
47. Steele CL, et al. 2011. Genome sequence of lineage III *Listeria monocytogenes* strain HCC23. *J. Bacteriol.* 193:3679–3680.
48. Steinweg C, et al. 2010. Complete genome sequence of *Listeria seeligeri*, a nonpathogenic member of the genus *Listeria*. *J. Bacteriol.* 192:1473–1474.
49. van de Guchte M, et al. 2006. The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc. Natl. Acad. Sci. U. S. A.* 103:9274–9279.
50. Waack S, et al. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 7:142.
51. Ward T, et al. 2004. Intraspecific phylogeny and lineage group identification based on the *prfA* virulence gene cluster of *Listeria monocytogenes*. *J. Bacteriol.* 186:4994–5002.
52. Ward TJ, Ducey TF, Usgaard T, Dunn KA, Bielawski JP. 2008. Multi-locus genotyping assays for single nucleotide polymorphism-based subtyping of *Listeria monocytogenes* isolates. *Appl. Environ. Microbiol.* 74:7629–7642.
53. Ward TJ, et al. 2010. Molecular and phenotypic characterization of *Listeria monocytogenes* from U.S. Department of Agriculture Food Safety and Inspection Service surveillance of ready-to-eat foods and processing facilities. *J. Food Prot.* 73:861–869.
54. Wesley IV. 1999. Listeriosis in animals, p. 39–73. In Ryser ET, Marth EH (ed), *Listeria*, listeriosis and food safety, 2nd ed. Marcel Dekker, Inc., New York, NY.
55. Wiedmann M, et al. 1999. Molecular investigation of a listeriosis outbreak in goats caused by an unusual strain of *Listeria monocytogenes*. *J. Am. Vet. Med. Assoc.* 215:369–371.
56. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.