Check for updates

Long-term persistence of diverse clones shapes the transmission landscape of invasive *Listeria monocytogenes*

Odion O. Ikhimiukor[1] , Lisa Mingle[2] , Samantha E. Wirth[2] , Damaris V. Mendez-Vallellanes[2], Hannah Hoyt[2], Kimberlee A. Musser[2] , William J. Wolfgang[2] and Cheryl P. Andam[1*]

Abstract

Background The foodborne bacterium *Listeria monocytogenes* (*Lm*) causes a range of diseases, from mild gastroenteritis to invasive infections that have high fatality rate in vulnerable individuals. Understanding the population genomic structure of invasive *Lm* is critical to informing public health interventions and infection control policies that will be most effective especially in local and regional communities.

Methods We sequenced the whole draft genomes of 936 *Lm* isolates from human clinical samples obtained in a two-decade active surveillance program across 58 counties in New York State, USA. Samples came mostly from blood and cerebrospinal fluid. We characterized the phylogenetic relationships, population structure, antimicrobial resistance genes, virulence genes, and mobile genetic elements.

Results The population is genetically heterogenous, consisting of lineages I–IV, 89 clonal complexes, 200 sequence types, and six known serogroups. In addition to intrinsic antimicrobial resistance genes (*fosX*, *lin*, *norB*, and *sul*), other resistance genes *tetM*, *tetS*, *ermG*, *msrD*, and *mefA* were sparsely distributed in the population. Within each lineage, we identified clusters of isolates with ≤ 20 single nucleotide polymorphisms in the core genome alignment. These clusters may represent isolates that share a most recent common ancestor, e.g., they are derived from the same contamination source or demonstrate evidence of transmission or outbreak. We identified 38 epidemiologically linked clusters of isolates, confirming eight previously reported disease outbreaks and the discovery of cryptic outbreaks and undetected chains of transmission, even in the rarely reported *Lm* lineage III (ST3171). The presence of animal-associated lineages III and IV may suggest a possible spillover of animal-restricted strains to humans. Many transmissible clones persisted over several years and traversed distant sites across the state.

Conclusions Our findings revealed the bacterial determinants of invasive listeriosis, driven mainly by the diversity of locally circulating lineages, intrinsic and mobile antimicrobial resistance and virulence genes, and persistence across geographical and temporal scales. Our findings will inform public health efforts to reduce the burden of invasive listeriosis, including the design of food safety measures, source traceback, and outbreak detection.

Keywords *Listeria monocytogenes*, Antimicrobial resistance, Genome, Mobile genetic element, Single nucleotide polymorphism, Outbreak

*Correspondence:
Cheryl P. Andam
candam@albany.edu
Full list of author information is available at the end of the article

Ikhimiukor *et al. Genome Medicine*   (2024) 16:109

Page 2 of 14

Background

The bacterium *Listeria monocytogenes* (*Lm*) is an opportunistic foodborne pathogen responsible for listeriosis [1]. It is ubiquitous in nature and can be found in soil, water, vegetation, and animal feces [2]. It is notable for its ability to survive and even grow under refrigeration and other food preservation measures, such as low pH, high salt concentration, and low water activity [3]. Listeriosis usually results from the ingestion of food products contaminated with *Lm* [1]. The number of listeriosis cases is small, but foodborne outbreaks often presenting as febrile gastroenteritis are frequently reported [4, 5]. *Lm* can spread beyond the intestines and cause more serious infections such as septicemia, meningitis or meningoencephalitis, and pregnancy-associated and neonatal listeriosis manifesting as miscarriage, stillbirth, or neonatal sepsis [5]. Invasive listeriosis has a high mortality rate (16–30%) and poses a significant risk to pregnant women, the elderly, or individuals with a weakened immune system (e.g., those with HIV, leukemia, cancer, or undergoing kidney transplant) [5, 6]. The high morbidity and case fatality rates associated with invasive listeriosis therefore make it a significant public health concern.

Lm is remarkably heterogeneous and is represented by numerous sequence types (STs) and clonal complexes (CCs) [7–9]. This ancient species has diversified into four major lineages designated I to IV [8, 10]. However, only a few clones have been implicated in disease outbreaks worldwide and have emerged in epidemic proportions of infections [4, 8, 11]. In a comprehensive analysis of 6,633 strains, CCs 1, 2, 4, and 6 (all belonging to lineage I) were overrepresented in human listeriosis cases affecting the central nervous system or maternal-neonatal infections [10]. CC1 is particularly notable because of its global distribution, mainly due to the transatlantic livestock trade, growth of cattle farming, and food industrialization [9]. The success of CC1 lies in its hypervirulent nature [10], efficiency in colonizing the gut [12], and prolonged survivability in primary infection foci [13]. ST6 (CC6) is increasingly reported as the causal agent in recent listeriosis outbreaks in Europe and Africa [14–17]. Clonal differences in *Lm* distribution, prevalence, and virulence features suggest unique adaptive strategies to specific habitats.

Large-scale, long-term population genomic investigations of *Lm* are needed to elucidate the genetic basis of invasive listeriosis, the emergence of new high-risk clones, and the changing dynamics of the disease. Here, we sequenced the genomes of 936 *Lm* isolates obtained from human clinical samples (mostly from blood and cerebrospinal fluid) in a two-decade active surveillance program across 58 counties in the State of New York, USA. Our findings revealed the bacterial determinants of invasive listeriosis, driven mainly by the diversity of locally circulating lineages, intrinsic and mobile antimicrobial resistance (AMR) and virulence genes, and persistence across geographical and temporal scales.

Methods

Selection of bacterial isolates

New York State mandates that all clinical *Lm* samples are reported and submitted to the Wadsworth Center, the public health laboratory of the New York State Department of Health. Hence, it is likely that the proportion of samples sent to the Wadsworth Center represent very close to all cases detected in the state. The Wadsworth Center carries out species confirmation and source traceback of clinical *Lm* as part of New York's bacterial foodborne outbreak detection and surveillance program. These isolates were received from New York health care providers and were recovered primarily from blood specimens collected from individuals diagnosed with *Lm* infection. A total of 964 *Lm* isolates in this study were received from clinical sources in 58 counties from 2000 to 2021 (Additional file 1: Fig. S1; Additional file 2: Table S1). Seven isolates do not have county information but were also included in our analyses. We also included an isolate collected in 1987 in New York from a blood sample but with unknown county information (Accession number SRR15305598; National Center for Biotechnology Information [NCBI]). All isolates were stored in glycerol solution at −80 °C.

DNA extraction, library preparation, and whole genome sequencing

Methods for total DNA extraction follow the standard operating procedure used by participating laboratories in PulseNet [18] (https://www.cdc.gov/pulsenet/pdf/pnl33-dna-extraction-and-quality-508.pdf). Briefly, overnight cultures were lysed in 180 μl of enzymatic lysis buffer (ELB) (PNL33) and 1.5 μl lysozyme (100 mg/ml, Sigma) for 30 min on a shaking thermomixer at 56 °C before 20 μl Proteinase K (50 μg/μl) was added for the remaining 30 min of incubation. They were then placed in the QIAcube or QIAcubeHT (Qiagen, Germantown, MD) and genomic DNA extracted using the standard QIAamp DNA Blood Minikit protocol or QIAamp 96 DNA QIAcube HT kit, respectively. Library preparation and sequencing were carried out in the Advanced Genomic Technologies Cluster (AGTC) at the Wadsworth Center. Library preparation followed standard Illumina protocols for Nextera XT or Nextera DNA Flex kits. Sequencing was done on either a MiSeq system using 2×250 chemistry and version 2 kits or on a NextSeq system using 2×150 chemistry and version 2 kits. NextSeq reads were demultiplexed using the Illumina BCL2FASTQ script

Ikhimiukor *et al. Genome Medicine*   (2024) 16:109

Page 3 of 14

(Illumina, San Diego, CA). Read quality was assessed to ensure that minimum quality thresholds established by the Center for Food Safety and Applied Nutrition (CFSAN) were met using MicroRunQC implemented on the GalaxyTrakr platform of Galaxy [19] or thresholds established by PulseNet using Bionumerics version 7.6.3. The mean and median sequence coverages for all reads were $91\times$ and $84\times$, respectively (range: $26\times$ to $202\times$). Q scores for all reads exceeded 32.5, and estimated genome sizes ranged from 2,724,566 to 3,778,059 bp (Additional file 1: Table S1). Genome sequences were submitted to the NCBI Pathogen Detection database (https://www.ncbi.nLm.nih.gov/pathogens/) and the Centers for Disease Control and Prevention (CDC) database in real time.

Genome assembly, sequence quality check, and annotation

We used Shovill v.1.1.0 (https://github.com/tseemann/shovill) to assemble the paired-end reads de novo. We employed the –trim flag for trimming of adapter sequences. Shovill uses the SPAdes assembly algorithm [20] but alters several pre- and post-assembly steps to rapidly produce comparable and high-quality assemblies. We used QUAST v.5.0.2 [21] and CheckM v.1.1.3 [22] to assess the quality of assembled genomes. Genomes with < 90% completeness and > 5% contamination were excluded from downstream analysis. We also excluded assemblies with > 200 contigs and an N50 < 40,000 bp to obtain high-quality genomes. After filtering low-quality genomes, we obtained a total of 936 genomes which were used for all downstream analyses (Additional file 1: Table S1; Additional file 2: Fig. S2). Genome completeness ranged from 98.73 to 99.85% (median = 99.44%) and genome contamination ranged from 0 to 2.79% (mean = 0.06%), which were all within the genome quality standards recommended by CheckM [22]. The number of contigs in this dataset ranged from 7 to 137 (median = 15) and N50 ranged from 41.7 to 1579.9 Kb (mean = 582.6 Kb). Median GC content of the *Lm* was 37.89%. Genomes sizes ranged from 2.75 to 3.17 Mbp. The resulting contigs were annotated using Prokka v.1.14.6 [23].

In silico sequence typing and serogroup identification

Prior to 2019, *Lm* isolates were initially typed by pulsed-field gel electrophoresis (PFGE) using the enzymes AscI and ApaI following the procedures set by CDC PulseNet [18] (http://www.cdc.gov/pulsenet/PDF/listeria-pfge-protocol-508c.pdf). Since implementing whole genome sequencing, the Wadsworth Center has used sequence variation in draft genomes to type *Lm* isolates in real time as they are received from healthcare providers. In our study, the *Lm* genome assemblies were uploaded to

the Institut Pasteur BIGSdb *Listeria* database (https://bigsdb.pasteur.fr/listeria/) [24] for curation to determine the identity of major clones defined using the 7-gene multi-locus sequence typing (MLST), PCR-serogroups, CCs, core genome MLST (cgMLST) types, sublineages, and lineages. The seven housekeeping genes used in MLST were *abcZ*, *bglA*, *cat*, *dapE*, *dat*, *ldh*, and *lhkA* [7]. The cgMLST classification is based on 1748 loci with a cutoff of seven allelic mismatches from at least another member of the group [8]. Genomes belonging to different sublineages differ by 150 allele mismatches, whereas lineages differ by $\geq$ 1500 out of 1748 loci [8].

Pan-genome analysis

We used Panaroo v.1.2.7 [25] to characterize the collective set of genes present in all genomes in our dataset, i.e., pan-genome [26]. We used the flag –strict option to ensure that only high-quality gene sequences were identified and clustered. The pan-genome consisted of 7500 orthologous gene families. These genes were categorized as core genes ($n = 2421$; genes present in 99% of genomes), softcore genes ($n = 103$; genes present in 95 to < 99% of genomes), shell genes (n = 593; genes present in 15 to < 95% of genomes), and cloud genes ($n = 4383$; genes present in < 15% of genomes) (Additional file 1: Table S2; Additional file 2: Fig. S3). The number of genes per genome ranged from 2683 (Accession number SRR14300122; isolate from blood from Westchester County) to 3169 (Accession number SRR14524738; isolate from cerebrospinal fluid from Westchester County). The mean number of genes per genome was 2869 ± 80. Nucleotide sequences were aligned using MAFFT v.7.471 [27].

Phylogenetic tree reconstruction

Single nucleotide polymorphisms (SNP) were extracted from the 2.39 Mbp concatenated alignment of 2421 core genes using SNP-sites [28]. The core genome alignment consisted of 418,360 SNPs, which was used as input in IQ-TREE v.2.1.4 [29] to build a maximum likelihood phylogeny. We used the ModelFinder algorithm to determine the best-fit model for ascertaining rate heterogeneity to improve accuracy of phylogenetic estimates [30]. Based on the output of ModelFinder, we used the general time reversible nucleotide substitution model [31] with an ascertainment bias correction and FreeRate heterogeneity [32] (GTR + ASC + R6). Branch support was assessed using 1000 bootstrap replicates implemented using the built-in ultrabootstrap UFBoot software [33]. Phylogenetic trees were rooted at the midpoint. Trees were visualized and annotated using figtree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and the Interactive Tree of Life [34].

Ikhimiukor *et al. Genome Medicine*   (2024) 16:109

Page 4 of 14

In silico detection of genetic determinants for AMR and virulence and mobile genetic elements

We used two databases to screen for AMR genes: the Institut Pasteur BIGSdb *Listeria* database [24] and the NCBI's AMRFinderPlus v.3.11.4 and its accompanying AMR database [35]. We also screened for the presence of virulence determinants, genomic islands, and pathogenicity islands on the BIGSdb *Listeria* platform. The presence of *Listeria* genomic islands (LGI) in the genomes was determined if they contained ≥ 90% of genes belonging to each genomic island [36]. The presence of *Listeria* pathogenicity islands (LIPI) was determined by the presence of LIPI-specific genes, and includes LIPI-1 (*prfA*, *plcA*, *hly*, *mpl*, *actA*, *plcB*), LIPI-2 (*inlB1- inlL*, *smlcL*, *surF3*), LIPI-3 *(llsAGHXBYDP)*, and LIPI-4 (LM9005581_70009 to LM9005581_70014) [36].

We used the PlasmidFinder database (implemented as of November 22, 2022) [37] on ABRicate v.1.0.1 (https://github.com/tseemann/abricate) to determine the presence of plasmids by detecting the replicon gene (*rep*) that encode the plasmid replicon initiator protein [37]. The mob-recon tool available on the MOB-suites package was used to reconstruct putative plasmid sequences from the draft genome assemblies [38]. We used VirSorter2 v.2.2.4 software to detect the presence and diversity of phage elements [39]. Sequences of putative phage and plasmid sequences detected in each genome were respectively concatenated to determine the totality of phage and plasmid DNA per genome. These concatenated sequences were used as query in AMRFinderPlus to detect the presence of AMR genes associated with these mobile genetic elements. We used Bakta [40] to polish the annotation of the five genomes harboring tetracycline resistance genes to obtain a better description of gene function and the genetic environment of tetracycline resistance genes. Their genomic neighborhood was visualized using clinker (https://github.com/gamcil/clinker).

Transmission and outbreak detection

Earlier methods of outbreak cluster identification by CDC PulseNet used PFGE analysis and epidemiological data to determine if genetically similar isolates were associated with disease outbreaks [18]. In our study, we used a conservative estimate of ≤ 20 core genome SNPs between *Lm* genome pairs to define a cluster. This 20 SNP threshold has previously been reported to reflect *Lm* isolates with an epidemiological link [41–43]. Within each lineage, we identified clusters of isolates with ≤ 20 SNPs in the core genome alignment (Additional file 1: Tables S9–S11; Additional file 2: Fig. S10–S12). These clusters may represent isolates that share a most recent common ancestor, e.g., they are derived from the same contamination source or demonstrate evidence of transmission or outbreak. We identified persistent clusters as three or more *Lm* genomes with pairwise distance of ≤ 20 SNPs isolated from the same or multiple counties ≥ 6 months apart. We also distinguish outbreak clusters as three or more *Lm* genomes with a pairwise distance of ≤ 20 SNPs isolated from the same county or state within ≤ 6 months apart. The recombination-free core genome alignment generated by Gubbins [44] was used as input in GrapeTree v.2.1 to create minimum spanning trees of the core SNP clusters [45].

Statistical information

We used the ggpubr v.0.4.0 package in RStudio v.2022.02.1 + 461 [46] to carry out the Wilcoxon Rank Sum test for comparing pairwise accessory and SNP (≤ 20) distances between *Lm* genomes, the number of AMR genes per genome, number of plasmid replicons per genome, and number of phages per genome between any two lineages. To evaluate the gene content diversity of accessory genes, we calculated the Jaccard pairwise distance. A Jaccard distance = 1 indicates that the two groups are entirely different in terms of their gene content, whereas 0 indicates that the two groups are highly similar. Mantel test was used to assess the correlation between genetic and geographical distance matrices for every pair of genomes. The longitude and latitude of each county were converted to Haversine distance using the R package geosphere v.1.5 (https://github.com/rspatial/geosphere). The Mantel test was run using Spearman's rank correlation on 1000 permutations on the R package vegan v.2.6 (https://cran.r-project.org/web/packages/vegan/index.html). We used a *p*-value threshold ≤ 0.05 to consider the significance of our results.

Results

The clinical *Lm* population is genetically diverse

We retrieved high-quality short-read genome sequences of 936 *Lm* isolates collected from human clinical samples in 58 counties across the State of New York (Fig. 1A; Additional file 1: Table S1; Additional file 2: Fig. S1). The counties with the highest number of isolates were Nassau (*n* = 133), Suffolk (*n* = 124), Westchester (*n* = 85), Erie (*n* = 68), and Monroe (*n* = 48), which altogether made up 48.93% of the entire dataset. Four counties were not represented in the population owing to limited or lack of samples submitted by healthcare providers from those counties. The dataset included isolates sampled from 2000 to 2021 (mean = 43 isolates per year; range = 21–62) (Fig. 1B; Additional file 2: Fig. S4). An isolate from 1987 was also included in the dataset for historical comparison. Most of the isolates were derived from blood (*n* = 760; 81.19%) and cerebrospinal fluid (*n* = 109; 11.64%) (Additional file 2: Fig. S5).
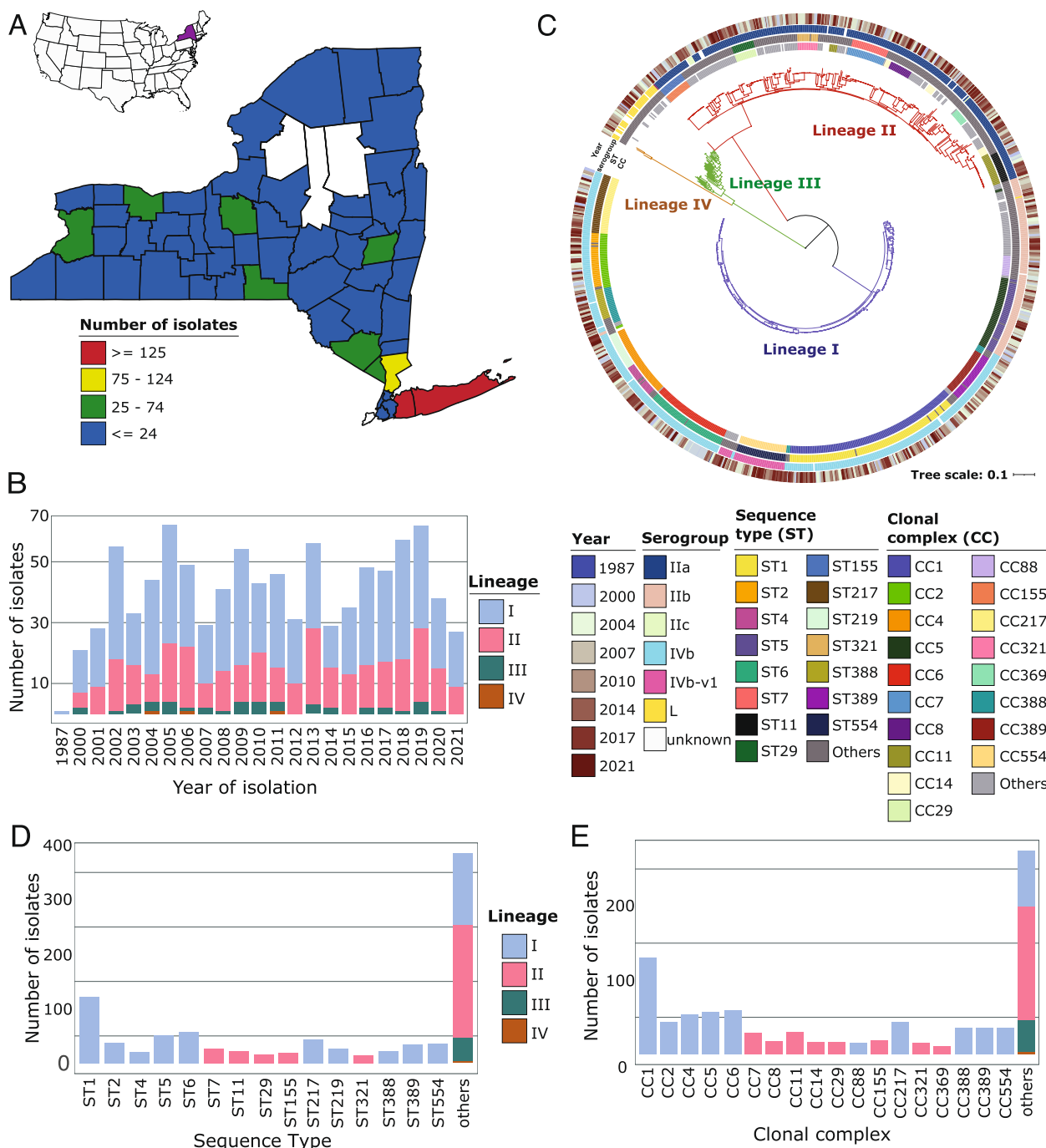
Fig. 1 Distribution of the 936 clinical *Lm* in New York. **A** A map of the USA showing the location of New York (colored in purple) and the counties in New York State. Counties where *Lm* was sampled in this study are colored according to the number of isolates. New York State has a total area of 54,556 sq. miles (141,300 km^2), a maximum length of 330 miles (530 km), and a maximum width of 285 mi (455 km). **B** Number of *Lm* isolates according to lineages per year. **C** Midpoint-rooted maximum likelihood phylogenetic tree based on the sequence alignment of 2421 core genes. Tree scale represents the number of nucleotide substitutions per site. Colored branches represent the four lineages I–IV. Colored outer rings representing the clonal complex (CC) and sequence types (STs) show only those with ≥ 10 representative genomes. CCs and STs with ≤ 10 representative genomes are denoted as others. **D** Number of isolates per major ST. **E** Number of isolates per major CC

Ikhimiukor *et al. Genome Medicine*　　(2024) 16:109

Page 6 of 14

The New York *Lm* population was derived from different genetic groups (Fig. 1C; Additional file 1: Table S3). Classification schemes based on the 7-gene MLST [7] revealed a total of 200 STs and 89 known CCs. The most frequent STs in this study were ST1 (122 genomes; 13.03%), ST6 (58 genomes; 6.19%), ST5 (52 genomes; 5.55%), and ST217 (43 genomes; 4.59%) (Fig. 1D). The most common CCs were CC1, CC6, and CC5 (Fig. 1E). All known four monophyletic lineages of *Lm* I, II, III, and IV [8, 47] were detected in our study, comprising 584, 306, 43, and 3 genomes, respectively. Genomes belonging to lineage I consisted of 62 STs and 29 known CCs, whereas 97 STs and 50 CCs made up lineage II. Lineage III consisted of 38 STs and 10 known CCs. Lastly, the three genomes in lineage IV represented three STs (ST3131, ST3163, and ST3173). Lineages I and II were detected every year throughout the sampling period (Fig. 1B). The two major lineages I and II have widespread distribution across New York State and were detected in 54 and 53 counties, respectively.

Variation in the somatic (O) and flagellar (H) antigens is also used to distinguish *Lm* because serogroup designation tends to be associated with virulence potential [48]. In silico prediction of PCR-serogroups showed genomes from this study belonged to six previously known PCR-serogroups (IIa, IIb, IIc, IVb, IVb-v1, L). PCR-serogroups were differentially distributed across the *Lm* phylogeny. PCR-serogroups IIa, IIb, and IIc were detected in 290, 127, and 5 genomes, respectively, whereas PCR-serogroups IVb, IVb-v1, and L were detected in 403, 45, and 36 genomes, respectively. We detected a total of 30 genomes with no match in the BIGSdb *Listeria* PCR-serogroup database [24] (Fig. 1C; Additional file 1:Table S3B). The unknown serogroups belonged to genomes in lineages I ($n = 8$ genomes), II ($n = 11$ genomes), and III ($n = 11$ genomes).

Lm has multiple AMR, some of which are carried by mobile genetic elements

The three lineages (I, II, III) differ in their accessory gene content ($p < 0.0001$, Wilcoxon rank sum test; Additional file 2: Fig. S6). We detected a total of 12 AMR genes representing eight antimicrobial classes in the entire *Lm* population (Additional file 1: Table S4). The most prevalent were the intrinsic AMR genes *fosX*, *norB*, *sul*, and *lin* (Fig. 2A). The first three genes were detected in all genomes, whereas the lincosamide resistance gene *lin* was present in 935 (99.89%) genomes. Acquired AMR genes in this study were detected at much lower frequencies. These included genes conferring resistance to tetracycline (either *tetM* or *tetS* present in five genomes from ST5, ST59, ST199, ST1039, and ST2928 recovered from blood; lineages I and II),

macrolide-lincosamide-streptogramin B (*ermG* in two ST2 genomes from cerebrospinal fluid; lineage I), macrolide (*msrD* and *mefA* in three ST2 genomes from cerebrospinal fluid, $n = 2$ and blood, $n = 1$; lineage I), biocides (*emrC* in one from ST8 from blood; lineage II), and aminoglycosides (*aacA4* in one ST155 genome from blood; lineage II). No significant difference in the number of AMR genes per genome was detected among lineages I–III (all p values > 0.05 for every pair of lineages, Wilcoxon rank sum test).

Mobile genetic elements contribute considerable diversity and functionality in bacterial cells, including the mobility of AMR genes [49]. First, we took a closer look at the flanking DNA of acquired AMR genes detected in this study to better understand their genetic environment. The tetracycline resistance genes *tetM* and *tetS* were flanked by conjugal transfer proteins and DNA-binding proteins (Additional file 2: Fig. S7). The genetic environment of contigs (> 120 kbp) harboring *ermG*, *msrD*, and *mefA* were frequently flanked by phage-like structures and were further identified as phage genetic material (Additional file 1: Table S5).

We sought to determine the presence and diversity of putative plasmids and phages. A total of 12 plasmid replicon types were detected (Additional file 1: Table S4). At least one plasmid replicon was identified in 171 genomes (or 18.26% of the population), of which 98, 70, and three genomes came from lineages I, II, and II, respectively. The genome with the highest number of plasmid replicons ($n = 4$) was detected in an ST5 isolate from a blood sample in Onondaga County (Accession no. SRR5451748). Another genome harbored three plasmid replicons (ST371) and was isolated from a blood sample in Nassau County. The most frequently detected plasmid replicon types were rep25_2_M640p00130(J1776plasmid) (lineage I = 62 genomes, lineage II = 20), rep26_2_repA(pLGUG1) (lineage I = 28, lineage II = 22), and rep26_4_repA(pLM 5578) (lineage I = 3, lineage II = 26). We found significant differences in the number of plasmids per genome between lineages I and II (p value = 0.027) and between lineages II and III (0.016), but not between lineages I and III (0.088) (Wilcoxon rank sum test).

At least one phage DNA element was detected in 919 genomes (Additional file 1: Table S5). All four lineages contained phage DNA, with a higher number of phage DNA elements per genome detected in lineages I and II (Fig. 2B). We found significant differences in the number of phage DNA per genome between lineages I and II ($p = 0.026$), lineages II and III ($p = 7.2$ e − 07), and lineages I and III ($p = 3.1$ e − 05) (Wilcoxon rank sum test). Because phage DNA may occupy a substantial portion of the bacterial chromosome [50], we estimated the combined sizes of all phage DNA per genome. Isolates from
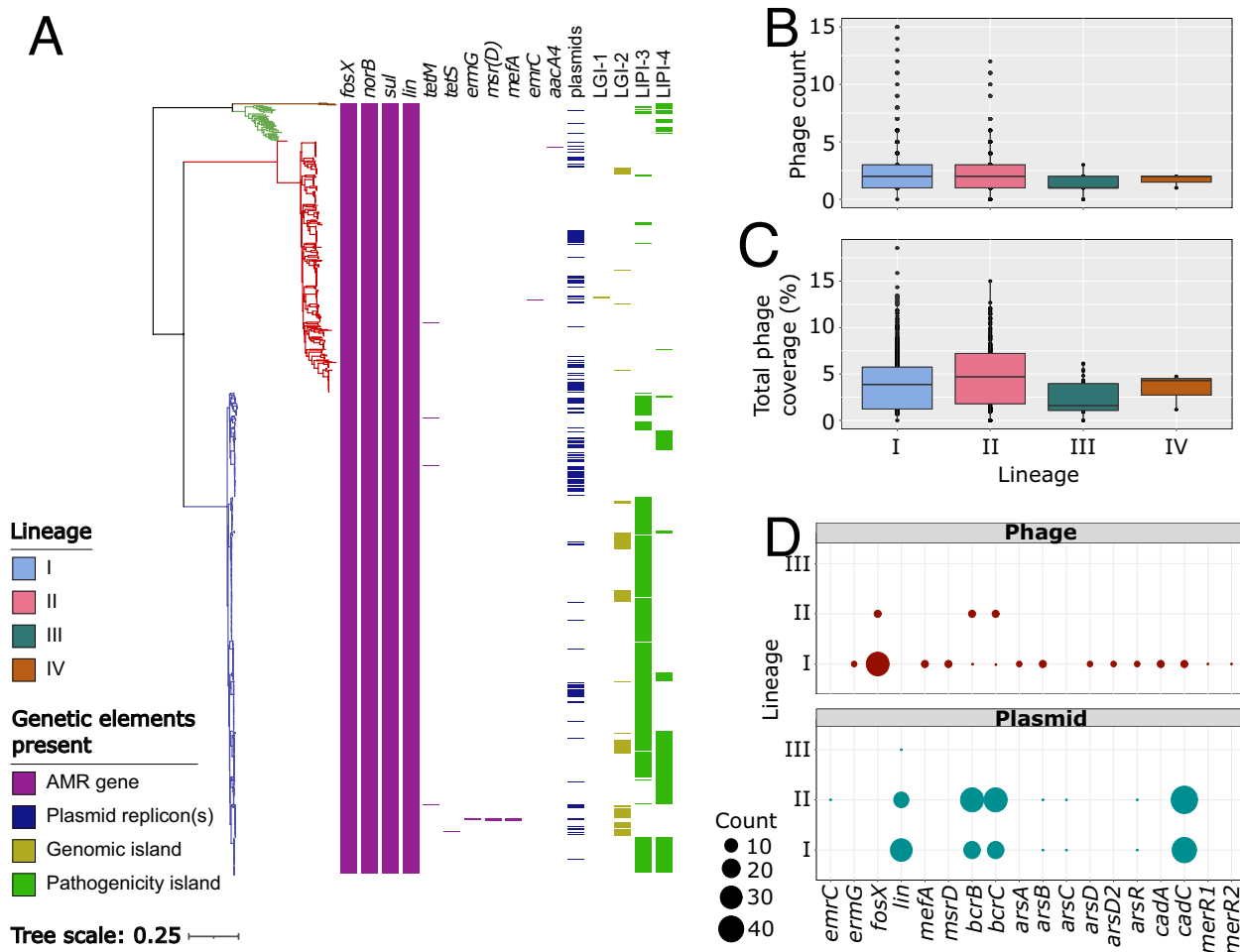
Ikhimiukor *et al. Genome Medicine*   (2024) 16:109

Page 7 of 14

Fig. 2 Antimicrobial resistance (AMR) and mobile genetic elements in clinical *Lm*. **A** Midpoint-rooted maximum likelihood phylogenetic tree showing the four *Lm* lineages and the distribution of AMR genes, plasmid replicons, *Listeria* genomic islands (LGI), and *Listeria* pathogenicity islands (LIP). Colored blocks represent the presents of these genetic elements. The tree is identical to that in Fig. 1B. Comparison of phage count (**B**) and phage coverage (**C**) among the four lineages (**B**). For panels **B** and **C**, the boxplot shows the 25th, 50th, and 75th percentiles and black dots show data points outside the interquartile range. **D** Comparison of the number of genomes per lineage that harbor AMR genes in phage and plasmids. The size of the circles is proportional to the number of genomes

the four lineages contain total phage DNA of less than 5% (median) of their genome (Fig. 2C). Two ST5 isolates (accession no. SRR14404494 and SRR14214577; lineage I) with eight and five identified phage elements harbored the largest combined phage DNA per genome of 576 kbp and 484 kbp, respectively. These were recovered from blood samples in Erie and Suffolk counties, respectively. An ST9 isolate (accession no. SRR3277646; lineage II) obtained from a blood sample in Suffolk also carried 466 kbp of total phage DNA.

We next screened the putative plasmid and phage DNA for the presence of genes conferring resistance to antimicrobials, biocides, and heavy metals. We identified four AMR genes in phage DNA (Fig. 2D; Additional file 1: Table S5). Phage-associated *fosX* was the most frequently

detected, which we identified in isolates belonging to lineage I (*n* = 33 genomes from eight STs) and lineage II (*n* = 3 genomes from three STs). Other resistance genes that we identified in our study such as *ermG* (two genomes), *msrD* (three genomes), and *mefA* (three genomes) were detected in phage DNA. Also present but less commonly detected were the heavy metal resistance genes *arsABDD2R* (arsenic), *cadAC* (cadmium), regulatory proteins encoded by *merR1* and *merR2* (mercury), and biocide resistance genes *bcrBC*. We were able to reconstruct putative plasmids from 152 genome assemblies, which were subsequently used as templates to predict the presence of the AMR genes they carry (Additional file 1: Table S6). The putative plasmids carried the AMR genes *lin* (n = 45 genomes) and *emrC* (one

Ikhimiukor *et al. Genome Medicine*   (2024) 16:109

Page 8 of 14

genome). The *lin* gene (lincosamide resistance) is predominantly chromosome-borne in *Lm* [51, 52]; however, we identified an *Lm* plasmid (NCBI accession number: NZ_LR134399.1) isolated from human blood harboring the *lin* gene. Mechanisms surrounding its mobility are unclear and require further investigation. In this study, some plasmids were also associated with genes conferring resistance to arsenic (*arsBCR*) in two genomes and cadmium (*cadC* in 83 genomes) as well as the disinfectant tolerance genes *bcrBC* in 51 genomes.

Lm lineages have multiple genomic and pathogenicity islands

Genomic islands are large syntenic blocks of genes that are integrated into the bacterial chromosome, often carrying genes conferring a selective advantage for the host bacterium and can be mobilized via horizontal gene transfer [36, 53]. We identified *Listeria* genomic islands LGI-1 (present in two lineage II genomes) and LGI-2 (present in 93 genomes belonging to lineage I [$n=82$] and lineage II [$n=11$]) (Fig. 2A, Additional file 1: Table S7).

Pathogenicity islands are a subset of genomic islands that carry virulence determinants and promote an infection cycle to enable the invasion of host cells, evasion of host's defenses through phagocytosis, and dissemination to nearby cells to re-initiate the infection cycle [36, 53]. We detected the *Listeria* pathogenicity islands LIPI-1, LIPI-3, and LIPI-4 in our dataset (Additional file 1: Table S8). LIPI-3 and LIPI-4, which are associated with hypervirulence, were detected in 435 and 195 genomes, respectively (Fig. 2A). LIPI-3 encodes listeriolysin S that functions both as a bacteriocin and hemolytic cytotoxic factor [54]. Previous studies report that LIPI-3 is commonly associated with epidemic outbreaks and is reported to be present primarily in lineage I and only in certain serogroups (I/IIb and IVb) [9, 10, 55]. The presence of LIPI-3 and LIPI-4 in lineage II genomes is rarely reported. In our dataset, we detected LIPI-3 in lineages I ($n=414$ genomes), II ($n=7$ genomes), and III ($n=6$ genomes) spanning multiple serogroups and STs. LIPI-3 is present in lineage II genomes belonging to ST380 (CC380; 2 genomes) and one genome each representing ST938, ST1867, ST1921, and ST3175 and are all members of CC938; and ST768 (CC768). LIPI-3 gene clusters were present on large chromosomal contigs (> 1.19 Mbp) in these genomes except in ST768 (~ 22 kbp).

LIPI-4 is a cluster of six genes implicated in neurological and placenta infections [10]. We detected LIPI-4 in all four lineages in our dataset (n = 170, 2, 20, and 3 genomes in lineages I–IV, respectively). LIPI-4 gene clusters were detected on 113 kbp and 548 kbp contigs belonging to ST1072 (SL1072) and ST1864 (CC1864, SL1864) in genomes isolated from Albany and New York counties,

respectively. Other virulence genes of various functions were distributed across the breadth of the *Lm* phylogeny and among the four lineages (Additional file 1: Table S8; Additional file 2: Fig. S8).

Geographical dissemination of epidemiologically linked *Lm* isolates in New York

Previous molecular studies of *Lm* established a threshold of ≤ 20 SNPs in a core genome alignment to define epidemiological linkages [41–43]. First, we used this threshold to determine the impact of geographical location to the genetic relationships of *Lm* isolates. The core genetic distance between every pair of isolates in lineage I is significantly higher between isolates from different counties than between isolates from the same county. Similar results were observed in lineage II (Additional file 2: Fig. S9; $p < 0.0001$ for both lineages I and II, Wilcoxon rank sum test).

Based on previously described criteria for cluster identification using SNP thresholds in the core genome alignment (see methods), we identified 23 and 14 core genome SNP clusters in lineages I and II, respectively (Figs. 3 and 4), and a single cluster in lineage III (Additional file 2: Fig. S13). In lineage I, five core genome SNP clusters (labeled 2, 3, 9, 16, and 18 in Fig. 3) corresponded to previously reported multistate outbreaks from the CDC PulseNet program, a national laboratory surveillance network of foodborne diseases [18]. Cryptic outbreaks, undetected transmission events, and shared contamination sources likely explain the remaining 18 core SNP clusters in lineage I. We identified 20 persistent clusters (from STs 1, 2, 4, 5, 6, 217, 382, and 55) and three outbreak clusters (from STs 1 and 5). Here, we highlight a few notable sequence clusters in lineage I (Fig. 3). Cluster 3 consisted of isolates from blood sampled in 2001 ($n=13$) and 2004 ($n=1$) that spanned seven counties on both eastern and western parts of New York State (approximately 191 miles or 468 km). They belonged to ST6 (cgMLST CT12957 and serogroup IVb). Pairwise SNP difference between genomes ranged from 0 to 3. Within this cluster, 10 genomes were reported by PulseNet to be associated in a multi-state outbreak. All genomes from this cluster harbored the pathogenicity island LIPI-3. The presence of an isolate collected in 2004 with identical genetic characteristics suggests the multi-year persistence of this cluster. The largest core SNP cluster in lineage I was cluster 16 ($n=35$ genomes) with pairwise core SNP difference ranging from 2 to 20. Isolates were derived from multiple body sources (blood = 27, cerebrospinal fluid = 5, placenta = 1, others = 2) between 2000 and 2021 from 24 counties. This cluster belonged to CC1, ST217, and serogroup IVb. All genomes harbor the

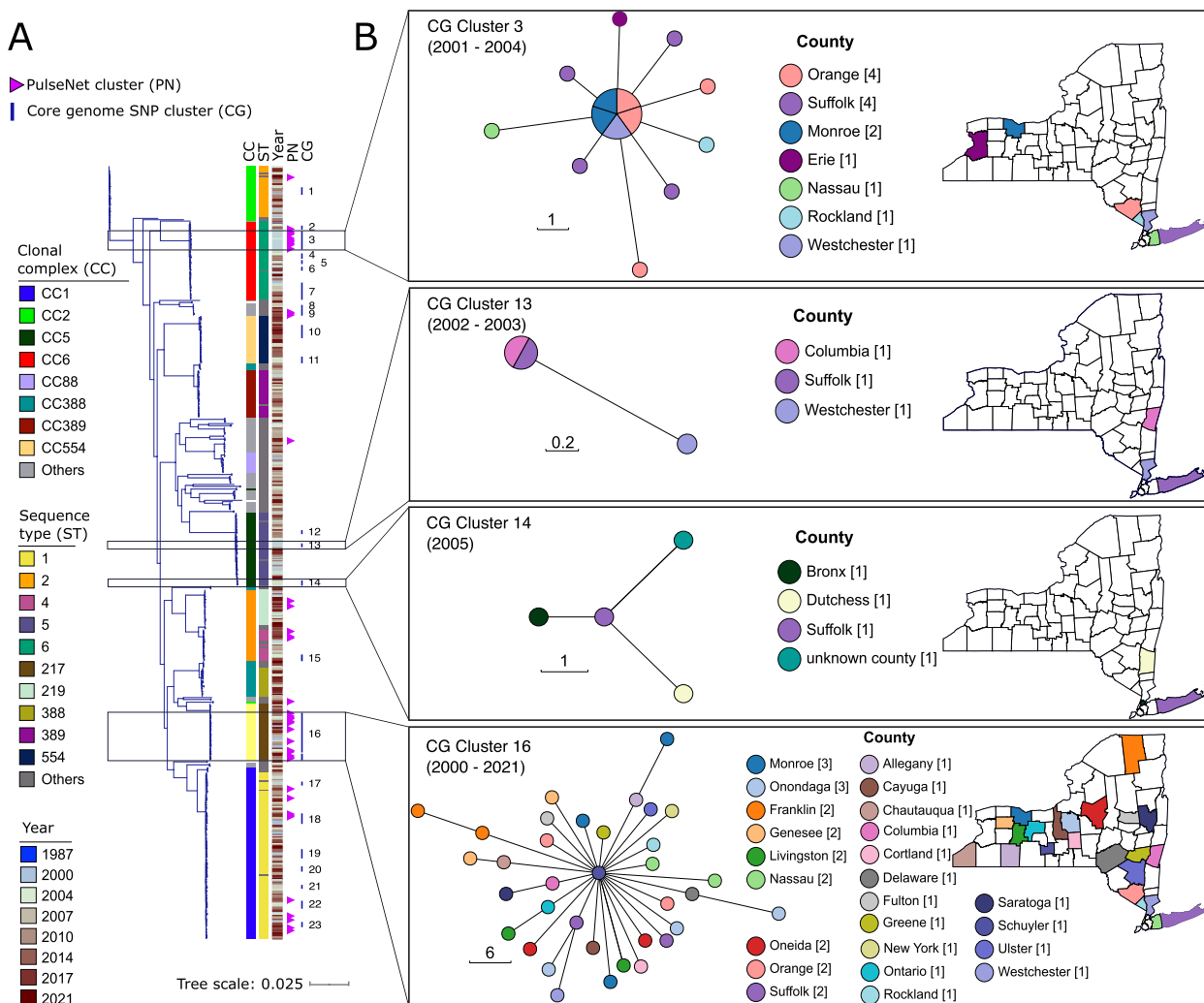Ikhimiukor *et al. Genome Medicine* (2024) 16:109

Page 9 of 14

Fig. 3 Phylogenetic relationship and core genome SNP clusters in *Lm* lineage I. **A** Maximum likelihood phylogenetic tree of Lineage I based on sequence alignment of 2610 core genes. The columns of colored blocks next to the tree show the clonal complexes (CC), sequence types (ST), and year of isolation. Outbreaks reported by PulseNet (PN_Outbreak) are represented by pink arrows, while clusters defined using a threshold of ≤ 20 core single nucleotide polymorphisms (SNP) are represented by a blue bar and numbered 1–23 (CG_Clusters). CCs and STs with ≥ 10 representative genomes are colored, whereas those with ≤ 10 representative genomes are denoted as others. **B** Minimal spanning grape trees representing select core genome SNP clusters colored by county of isolation. The scale represents the number of SNPs and the length of the scale is proportional to the number of SNP differences. The number in brackets next to the county name indicates the number of genomes

pathogenicity islands LIPI-3 and LIPI-4. A total of 13 genomes in this cluster were reported by PulseNet to be associated with a multi-state outbreak.

In lineage II, three core genome SNP clusters (labeled 3, 10, 13) corresponded to outbreaks reported by PulseNet and were also part of multi-state outbreaks (Fig. 4; Additional file 1: Table S10). All 14 core SNP clusters in lineage II persisted for ≥ 6 months and included members of STs 7, 11, 21, 29, 155, 204, 321, 360, 378, 573, and 635. LIPI-3 and LIPI-4 were not detected in the genomes from these clusters. Similar to lineage I clusters, there were clusters in lineage II that

also spanned multiple geographically distant counties from across the entire length of the State and were detected for many years.

In lineage III, we identified one core genome SNP cluster consisting of three isolates derived from blood (*n* = 2) and cerebrospinal fluid (*n* = 1) in Suffolk County between January and March 2010 (Additional file 2: Fig. S13). Genomes in this cluster were identical (i.e., zero SNPs apart), belonged to ST3171 (cgMLST type CT13941, serogroup L), and harbored LIPI-3 and LIPI-4.

We also sought to determine if the distribution of *Lm* isolates in New York is associated with the distance
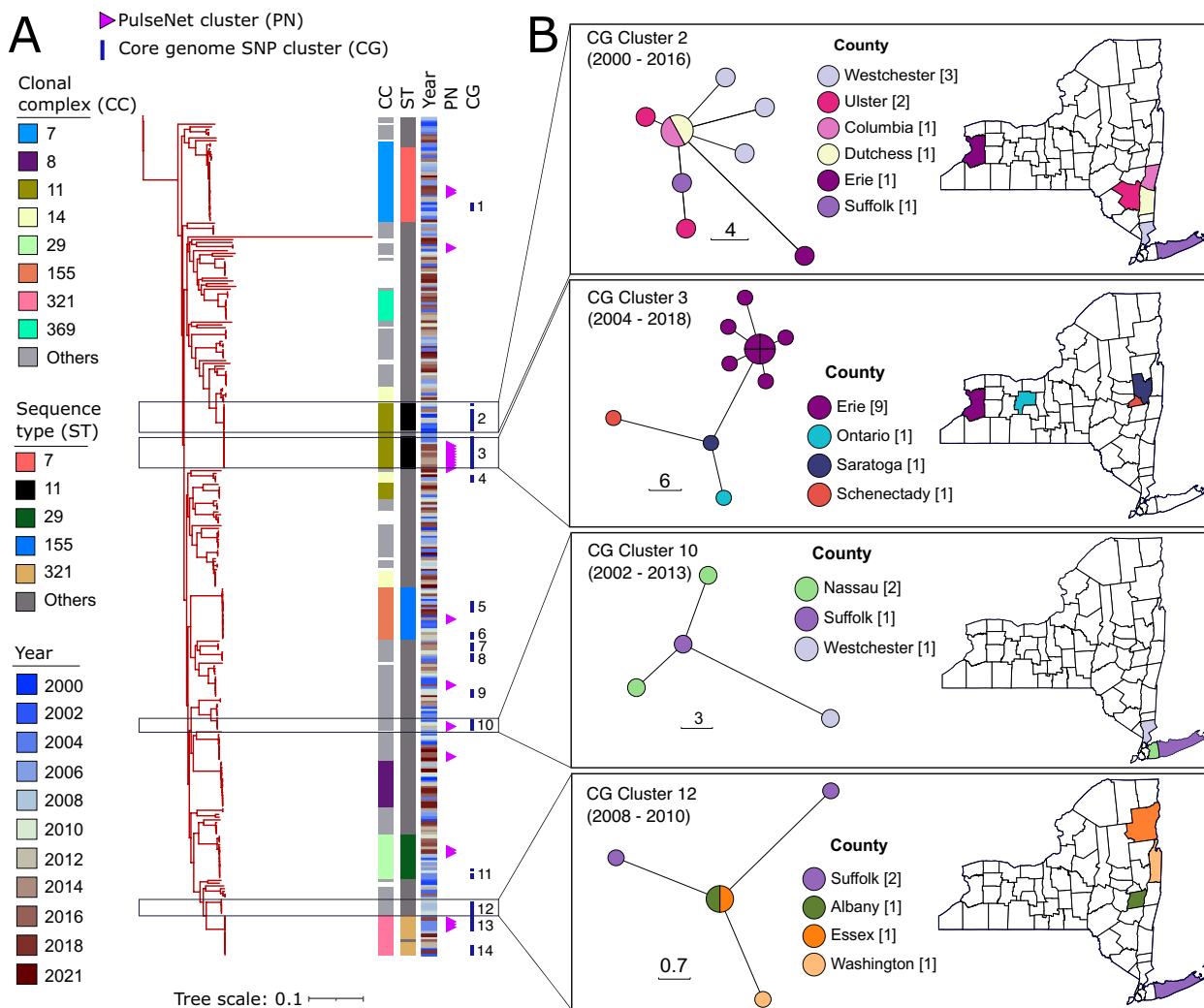
Ikhimiukor *et al. Genome Medicine*   (2024) 16:109

Page 10 of 14

Fig. 4 Phylogenetic relationship and core genome SNP clusters in *Lm* lineage II. **A** Maximum likelihood phylogenetic tree of lineage II based on sequence alignment of 2585 core genes. The columns of colored blocks next to the tree show the clonal complexes (CC), sequence types (ST), and year of isolation. Outbreaks reported by PulseNet (PN_Outbreak) are represented by pink arrows, while clusters defined using a threshold of ≤ 20 core single nucleotide polymorphisms (SNP) are represented by a blue bar and numbered 1–14 (CG_Clusters). CCs and STs with ≥ 10 representative genomes are colored, whereas those with ≤ 10 representative genomes are denoted as others. **B** Minimal spanning grape trees representing select core genome SNP clusters colored by county of isolation. The scale represents the number of SNPs and the length of the scale is proportional to the number of SNP differences. The number in brackets next to the county name indicates the number of genomes

between counties of isolation. We carried out a Mantel test of pairwise *Lm* genetic distances (based on pairwise core genome SNPs) and geographical distances between counties. When considering the entire *Lm* dataset, results revealed a significant but very weak correlation between genetic and geographic distances ($R = 0.03612$, $p = 0.006$) (Additional file 2: Fig. S14). We also carried out a Mantel test for only the outbreak genomes identified in lineages I and II (i.e., genomes labeled as CG in Figs. 3 and 4). We detected a significant but very weak correlation between genetic and geographical distances in lineage II ($R = 0.1238$,

$p = 0.004$), but not in lineage I ($R = -0.07276$, $p = 0.984$) (Additional file 2: Fig. S15).

Overall, these results show that invasive *Lm* associated with disease transmission and outbreaks were derived from multiple genetic backgrounds in lineages I–III. Many of the *Lm* clones that have epidemiological linkages can persist over many years and traverse geographically distant sites.

Discussion

Invasive listeriosis is life-threatening, particularly to pregnant and immunocompromised individuals [5, 6]. Understanding the population genomic structure of

Ikhimiukor *et al. Genome Medicine*   (2024) 16:109

Page 11 of 14

invasive *Lm* is critical to informing public health interventions and infection control policies that will be most effective in local and regional communities. Analysis of 936 clinical *Lm* isolates from New York State revealed three important findings. First, invasive clones harbored multiple resistance and virulence genes and transmission potential can emerge from different genetic backgrounds, including the rarely reported lineages III and IV. Our results are consistent with previous studies of CCs 1, 2, 4, and 6 in lineage I [4, 8, 10, 11] and also expand to other lineages, STs, and CCs. Mobile genetic elements such as plasmids, phages, and pathogenicity islands mediate the emergence and confluence of clinically relevant traits within the same strain. Second, epidemiologically linked clusters of isolates identified using core genome SNPs and pathogenicity islands confirmed previous outbreaks reported by PulseNet [18]. We also discovered clusters of isolates that are likely part of undetected transmission events, shared contamination sources, or cryptic outbreaks. Lastly, years-long persistence is an important feature of invasive *Lm*, and geographical distance does not appear to influence the spread of *Lm* at local scales.

Our work underscores the importance of longstanding routine surveillance of invasive *Lm* using whole genome sequencing at the state and county levels. Whole genome sequencing in disease surveillance systems provides critical granular output and identification of outbreaks that are often missed by the traditional PFGE approach used by PulseNet, and our results demonstrate this. A focus on local geographical scales can uncover tremendous genetic diversity in terms of lineages, CCs, STs, and serogroups, which may be obscured in global-scale studies. In New York, *Lm* lineages (I and II) and numerous CCs within each lineage can cause outbreaks and transmission events that may remain unnoticed using traditional surveillance and contact tracing methods. Our findings should form the basis for more intensive investigations of the causes of these linked isolates, e.g., through follow-up interviews with patients and source traceback. A broader and systematic sampling campaign that includes environmental and food sources is required to further uncover *Lm* transmission routes and reservoirs. These results also greatly expand previous findings that sublineages of the globally important *Lm* CC1 are country-specific but localized persistence occurs [9]. Equally important is that whole genome sequencing provides fine-scale resolution and critical insights on the genetic determinants of AMR present in a population. For example, notable in our dataset is the high prevalence of the *sul* gene. Beta-lactams are the first-line therapy for listeriosis; however, trimethoprim-sulphamethoxazole is the drug of choice for treating listeriosis in patients allergic to beta-lactams [56, 57]. Although not observed in our study, acquired

trimethoprim resistance in *Listeria* has been reported in previous studies [58, 59], which can be problematic in antimicrobial therapy of patients with beta-lactam allergy.

Furthermore, it is noteworthy that lineages III and IV were detected in our study. They are known to be rare and are associated with animal cases of listeriosis, with lineage IV more exclusively associated with ruminants [60, 61], which may suggest that human infection in *Lm* is not strictly lineage-specific. The rare human *Lm* lineage III (ST3171) detected in our study included a potential outbreak cluster and contained LIPI-3 and LIPI-4. We found no prior record of ST3171 on the BIGSbd *Lm* database. Outbreak reports associated with *Lm* lineage III are primarily reported in animals, causing syndromes like abortions and neurolisteriosis [61–63]. The presence and proliferation of these rare *Lm* lineages in humans may represent increased zoonotic spillover of previously animal-restricted strains to become human-adapted strains. Future work should explore whether and how lineages III and IV acquired the capacity to infect the human host (e.g., through horizontal transfer of host-adaptive genes, allelic variation through homologous recombination, and/or gene loss), opportunities for animal-human transmission, and changes in their prevalence in both human and animal hosts.

The long-term persistence of *Lm* lineages and CCs that we observed in our study and also reported elsewhere [64–67] can be attributed to both environmental and genetic factors. Certain characteristics possessed by *Lm*, such as the capacity to replicate at adverse temperatures, pH, salt concentrations, and low water activity, increase the risk of contamination in food, food products, and the environment [3], and their persistence is often fueled by sanitation shortfalls [68]. Hidden reservoirs of infection in the environment, e.g., food products, soil, farms, and other sources where *Lm* are known to thrive, can act as a vast well of rare lineages and genetic variants that have the potential to be clinically relevant. The presence and diversity of mobile genetic elements carrying AMR, virulence, and other adaptive genes that enhance their survival in a variety of stressful environments further contribute to their persistence. Particularly notable are the pathogenicity islands LIPI-3 and LIPI-4 associated with hypervirulence and are considered to be present mainly in serogroup IV lineage I strains [9, 10, 55]. In our study, we detected LIPI-3 and LIPI-4 in less common lineages, such as in ST3171 genomes belonging to serogroup L lineage III that were involved in a likely unnoticed outbreak in Suffolk County. LIPI-3 and LIPI-4 have been reported in non-pathogenic *Listeria innocua* [55] and hence, they may have been horizontally acquired by some *Lm* clones from outside the species.

Ikhimiukor *et al. Genome Medicine* (2024) 16:109

Page 12 of 14

Our study is not without limitations. Sampling was inconsistent across New York State, reflecting disparities in disease reporting from local health providers. Clinical outcomes of listeriosis are often self-reported and individuals who do not visit their local health providers are overlooked in surveillance. Inconsistencies in sampling may therefore miss other STs, clinically relevant genetic elements, and past and current transmission chains within and between those counties. Our study does not include phenotypic data from antimicrobial susceptibility tests because the New York State Department of Health does not incorporate phenotypic resistance in *Listeria* surveillance. As such, the prevalence of AMR genes in our dataset cannot be corroborated and does not interpret to resistant phenotype in the strains. We also lacked information on clinical cases, such as the forms of listeriosis, age ranges, and comorbidities. We hope that our findings provide a strong impetus to further improve the surveillance system of *Lm* in the state that will include more associated metadata. Moreover, despite the robustness of the plasmid reconstruction approach we used in this study, we acknowledge the limitations associated with accurately inferring plasmid genomes and their content from short-read sequencing data. This is particularly important in accurately inferring the location of AMR genes. For example, we detected the presence of the *lin* gene (lincosamide resistance) in putative reconstructed plasmids in 45 genomes. Long-read sequencing and plasmid mobility experimentation of these 45 genomes will shed critical insights on its dissemination across the population. Nonetheless, our dataset provides a comprehensive picture of the standing genetic diversity that will be most important as a baseline for future surveillance efforts and as basis for targeting underrepresented areas in New York State.

Conclusions

Our analysis of *Lm* genomes that were obtained from routine surveillance in New York spanning over two decades reveals the bacterial determinants of invasive listeriosis, including the diversity of locally circulating lineages, mobile genetic elements, and patterns of geographical and temporal spread. Our findings will inform public health efforts to reduce the burden of invasive listeriosis, including the use of effective antimicrobials, design of food safety measures, source traceback, and outbreak detection.

Abbreviations

AMR	Antimicrobial resistance
CC	Clonal complex
LGI	*Listeria* Genomic island
LIPI	*Listeria* Pathogenicity island
Lm	*Listeria monocytogenes*
MLST	Multilocus sequence typing
SL	Sublineage
SNP	Single nucleotide polymorphism
ST	Sequence type

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-024-01379-4.

> Additional file 1. Description of supplementary Tables S1-S6. The tables are presented as worksheets in one xls file.
>
> Additional file 2. Supplementary figures S1–S15 and respective figure legends.

Authors' contributions
C.P.A., O.O.I and W.J.W. designed and guided the work. O.O.I. carried out all bioinformatics analyses. K.A.M., L.M., and W.J.W. oversaw bacterial sampling, surveillance, sequencing, and deposition of sequence data in NCBI and CDC databases. S.W, D.V.M.V., and H.H. carried out subculturing, DNA extraction, and curation of sequence data and metadata. C.P.A. and O.O.I. wrote the manuscript. All authors have read and approved the final manuscript.

Availability of data and materials
The dataset supporting the conclusions of this article is included within the article and its supplementary files. Genome sequence data of Lm isolates are available in NCBI Sequence Read Archive (PRJNA514286: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA514286/; PRJNA212117: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA212117; PRJNA215355: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA215355; PRJNA483181: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA483181). BioProject and BioSample accession numbers for each genome are listed in Supplementary Table S1. Genomes were also curated in the BIGSdb Lm database and are publicly available (https://bigsdb.pasteur.fr/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_listeria_isolates&l=1&page=query).

Declarations

Ethics approval and consent to participate
Samples used in the study were subcultured bacterial isolates that had been archived in the routine course of the state surveillance program as part of public health. No patient specimens were used and patient-protected health information was not collected. Therefore, informed consent was not required. This work has been determined to be exempt from human subjects research by the Institutional Review Board of the Wadsworth Center.

Consent for publication
Not applicable.

Ikhimiukor *et al. Genome Medicine*   (2024) 16:109

Page 13 of 14

Author details
[1]University at Albany, State University of New York, Albany, NY, USA. [2]New York State Department of Health, Wadsworth Center, Albany, NY, USA.

References

1. Koopmans MM, Brouwer MC, Vázquez-Boland JA, van de Beek D. Human listeriosis. Clin Microbiol Rev. 2023;36:e0006019.
2. Félix B, Sevellec Y, Palma F, Douarre PE, Felten A, Radomski N, et al. A European-wide dataset to uncover adaptive traits of Listeria monocytogenes to diverse ecological niches. Sci Data. 2022;9:190.
3. Osek J, Lachtara B, Wieczorek K. Listeria monocytogenes - how this pathogen survives in food-production environments? Front Microbiol. 2022;13: 866462.
4. Chen Y, Gonzalez-Escalona N, Hammack TS, Allard MW, Strain EA, Brown EW. Core genome multilocus sequence typing for identification of globally distributed clonal groups and differentiation of outbreak strains of Listeria monocytogenes. Appl Environ Microbiol. 2016;82:6258–72.
5. Schlech WF. Epidemiology and clinical manifestations of Listeria monocytogenes infection. Microbiol Spectr 2019;7:GPP3-0014-2018.
6. Pohl AM, Pouillot R, Bazaco MC, Wolpert BJ, Healy JM, Bruce BB, et al. Differences among incidence rates of invasive listeriosis in the U. S. FoodNet population by age, sex, race/ethnicity, and pregnancy status, 2008–2016. Foodborne Pathog Dis. 2019;16:290–7.
7. Salcedo C, Arreaza L, Alcalá B, de la Fuente L, Vázquez JA. Development of a multilocus sequence typing method for analysis of Listeria monocytogenes clones. J Clin Microbiol. 2003;41:757–62.
8. Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, et al. Whole genome-based population biology and epidemiological surveillance of Listeria monocytogenes. Nat Microbiol. 2016;2:16185.
9. Moura A, Lefrancq N, Wirth T, Leclercq A, Borges V, Gilpin B, et al. Emergence and global spread of Listeria monocytogenes main clinical clonal complex. Sci Adv. 2021;7:eabj9805.
10. Maury MM, Tsai Y-H, Charlier C, Touchon M, Chenal-Francisque V, Leclercq A, et al. Uncovering Listeria monocytogenes hypervirulence by harnessing its biodiversity. Nat Genet. 2016;48:308–13.
11. Painset A, Björkman JT, Kiil K, Guillier L, Mariet J-F, Félix B, et al. LiSEQ - whole-genome sequencing of a cross-sectional survey of Listeria monocytogenes in ready-to-eat foods and human clinical cases in Europe. Microb Genom. 2019;5:e000257.
12. Maury MM, Bracq-Dieye H, Huang L, Vales G, Lavina M, Thouvenot P, et al. Hypervirulent Listeria monocytogenes clones' adaption to mammalian gut accounts for their association with dairy products. Nat Commun. 2019;10:2488.
13. Vázquez-Boland JA, Wagner M, Scortti M. Why are some Listeria monocytogenes genotypes more likely to cause invasive (brain, placental) infection? mBio. 2020;11:e03126-20.
14. Halbedel S, Wilking H, Holzer A, Kleta S, Fischer MA, Lüth S, et al. Large nationwide outbreak of invasive listeriosis associated with blood sausage, Germany, 2018–2019. Emerg Infect Dis. 2020;26:1456–64.
15. Thomas J, Govender N, McCarthy KM, Erasmus LK, Doyle TJ, Allam M, et al. Outbreak of listeriosis in South Africa associated with processed meat. N Engl J Med. 2020;382:632–43.
16. Nüesch-Inderbinen M, Bloemberg GV, Müller A, Stevens MJA, Cernela N, Kollöffel B, et al. Listeriosis caused by persistence of Listeria monocytogenes serotype 4b sequence type 6 in cheese production environment. Emerg Infect Dis. 2021;27:284–8.
17. McLauchlin J, Aird H, Amar C, Barker C, Dallman T, Lai S, et al. An outbreak of human listeriosis associated with frozen sweet corn consumption: Investigations in the UK. Int J Food Microbiol. 2021;338:108994.
18. Kubota KA, Wolfgang WJ, Baker DJ, Boxrud D, Turner L, Trees E, et al. PulseNet and the changing paradigm of laboratory-based surveillance for foodborne diseases. Public Health Rep. 2019;134:22S-28S.
19. Gangiredla J, Rand H, Benisatto D, Payne J, Strittmatter C, Sanders J, et al. GalaxyTrakr: a distributed analysis tool for public health whole genome sequence data accessible to non-bioinformaticians. BMC Genomics. 2021;22:114.
20. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
21. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.
22. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
23. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
24. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595. http://bigsdb.pasteur.fr/
25. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020;21:180.
26. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. The microbial pan-genome. Curr Opin Genet Dev. 2005;15:589–94.
27. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
28. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom. 2016;2:e000056.
29. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
30. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
31. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. American mathematical society: lectures on mathematics in the life sciences. Am Math Soc 1986;17:57–86.
32. Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst Biol. 2001;50:913–25.
33. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22.
34. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.
35. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep. 2021;11:12728.
36. Wiktorczyk-Kapischke N, Skowron K, Wałecka-Zacharska E. Genomic and pathogenicity islands of Listeria monocytogenes-overview of selected aspects. Front Mol Biosci. 2023;10:1161486.
37. Carattoli A, Hasman H. PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). Methods Mol Biol. 2020;2075:285–94.
38. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom. 2018;4:e000206.
39. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome. 2021;9:37.
40. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom. 2021;7:000685.
41. Wang YU, Pettengill JB, Pightling A, Timme R, Allard M, Strain E, et al. Genetic diversity of Salmonella and Listeria isolates from food facilities. J Food Prot. 2018;81:2082–9.
42. Allard MW, Strain E, Rand H, Melka D, Correll WA, Hintz L, et al. Whole genome sequencing uses for foodborne contamination and compliance: Discovery of an emerging contamination event in an ice cream facility using whole genome sequencing. Infect Genet Evol. 2019;73:214–20.

Ikhimiukor *et al. Genome Medicine*   (2024) 16:109

Page 14 of 14

43. Castro H, Douillard FP, Korkeala H, Lindström M. Mobile elements harboring heavy metal and bacitracin resistance genes are common among Listeria monocytogenes strains persisting on dairy farms. mSphere. 2021;6:e0038321.

44. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43:e15.

45. Zhou Z, Alikhan N-F, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res. 2018;28:1395–404.

46. RStudio Team. RStudio: Integrated Development for R. http://www.rstudio.com/. PBC, Boston; 2020.

47. Orsi RH, den Bakker HC, Wiedmann M. Listeria monocytogenes lineages: Genomics, evolution, ecology, and phenotypic characteristics. Int J Med Microbiol. 2011;301:79–96.

48. Quereda JJ, Morón-García A, Palacios-Gorba C, Dessaux C, García-Del Portillo F, Pucciarelli MG, et al. Pathogenicity and virulence of Listeria monocytogenes: A trip from environmental to medical microbiology. Virulence. 2021;12:2509–45.

49. Weisberg AJ, Chang JH. Mobile genetic element flexibility as an underlying principle to bacterial evolution. Annu Rev Microbiol. 2023;77:603–24.

50. Hatfull GF, Hendrix RW. Bacteriophages and their genomes. Curr Opin Virol. 2011;1:298–303.

51. Hanes RM, Huang Z. Investigation of antimicrobial resistance genes in Listeria monocytogenes from 2010 through to 2021. Int J Environ Res Public Health. 2022;19:5506.

52. Manqele A, Adesiyun A, Mafuna T, Pierneef R, Moerane R, Gcebe N. Virulence potential and antimicrobial resistance of Listeria monocytogenes isolates obtained from beef and beef-based products deciphered using whole-genome sequencing. Microorganisms. 2024;12:1166.

53. Vázquez-Boland JA, Kuhn M, Berche P, Chakraborty T, Domínguez-Bernal G, Goebel W, et al. Listeria pathogenesis and molecular virulence determinants. Clin Microbiol Rev. 2001;14:584–640.

54. Cotter PD, Draper LA, Lawton EM, Daly KM, Groeger DS, Casey PG, et al. Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I Listeria monocytogenes. PLoS Pathog. 2008;4:e1000144.

55. Lee S, Parsons C, Chen Y, Dungan RS, Kathariou S. Contrasting genetic diversity of Listeria pathogenicity islands 3 and 4 harbored by nonpathogenic Listeria spp. Appl Environ Microbiol. 2023;89:e0209722.

56. Swaminathan B, Gerner-Smidt P. The epidemiology of human listeriosis. Microbes Infect. 2007;9:1236–43.

57. Ma Y, Hu W, Song W. A case report of oral sulfamethoxazole in the treatment of posttransplant Listeria monocytogenes meningitis. Transl Androl Urol. 2023;12:524–9.

58. Granier SA, Moubareck C, Colaneri C, Lemire A, Roussel S, Dao T-T, et al. Antimicrobial resistance of Listeria monocytogenes isolates from food and the environment in France over a 10-year period. Appl Environ Microbiol. 2011;77:2788–90.

59. Morvan A, Moubareck C, Leclercq A, Hervé-Bazin M, Bremont S, Lecuit M, et al. Antimicrobial resistance of Listeria monocytogenes strains isolated from humans in France. Antimicrob Agents Chemother. 2010;54:2728–31.

60. Jeffers GT, Bruce JL, McDonough PL, Scarlett J, Boor KJ, Wiedmann M. Comparative genetic characterization of Listeria monocytogenes isolates from human and animal listeriosis cases. Microbiology (Reading). 2001;147:1095–104.

61. Whitman KJ, Bono JL, Clawson ML, Loy JD, Bosilevac JM, Arthur TM, et al. Genomic-based identification of environmental and clinical Listeria monocytogenes strains associated with an abortion outbreak in beef heifers. BMC Vet Res. 2020;16:70.

62. Bundrant BN, Hutchins T, den Bakker HC, Fortes E, Wiedmann M. Listeriosis outbreak in dairy cattle caused by an unusual Listeria monocytogenes serotype 4b strain. J Vet Diagn Invest. 2011;23:155–8.

63. Zaitsev SS, Khizhnyakova MA, Feodorova VA. Retrospective investigation of the whole genome of the hypovirulent Listeria monocytogenes strain of ST201, CC69, lineage III, isolated from a piglet with fatal neurolisteriosis. Microorganisms. 2022;10:1442.

64. Van Walle I, Björkman JT, Cormican M, Dallman T, Mossong J, Moura A, et al. Retrospective validation of whole genome sequencing-enhanced surveillance of listeriosis in Europe, 2010 to 2015. Euro Surveill. 2018;23:1700798.

65. Scaltriti E, Bolzoni L, Vocale C, Morganti M, Menozzi I, Re MC, et al. Population structure of Listeria monocytogenes in Emilia-Romagna (Italy) and implications on whole genome sequencing surveillance of listeriosis. Front Public Health. 2020;8:519293.

66. Coipan CE, Friesema IHM, van Hoek AHAM, van den Bosch T, van den Beld M, Kuiling S, et al. New insights into the epidemiology of Listeria monocytogenes - A cross-sectoral retrospective genomic analysis in the Netherlands (2010–2020). Front Microbiol. 2023;14:1147137.

67. Friesema IHM, Verbart CC, van der Voort M, Stassen J, Lanzl MI, van der Weijden C, et al. Combining whole genome sequencing data from human and non-human sources: tackling Listeria monocytogenes outbreaks. Microorganisms. 2023;11:2617.

68. Carpentier B, Cerf O. Review-Persistence of Listeria monocytogenes in food industry equipment and premises. Int J Food Microbiol. 2011;145:1–8.

Publisher's Note