AEM
Journals.ASM.org

Genome Sequencing Identifies Two Nearly Unchanged Strains of Persistent *Listeria monocytogenes* Isolated at Two Different Fish Processing Plants Sampled 6 Years Apart

Anne Holch,[a]* Kristen Webb,[b] Oksana Lukjancenko,[c] David Ussery,[c] Benjamin M. Rosenthal,[d] Lone Gram[e]

National Food Institute, Technical University of Denmark, Kgs. Lyngby, Denmark[a]; Department of Biology, Allegheny College, Meadville, Pennsylvania, USA[b]; Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kgs. Lyngby, Denmark[c]; Animal Parasitic Diseases Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland, USA[d]; Department of Systems Biology, Technical University of Denmark, Kgs. Lyngby, Denmark[e]

***Listeria monocytogenes* is a food-borne human-pathogenic bacterium that can cause infections with a high mortality rate. It has a remarkable ability to persist in food processing facilities. Here we report the genome sequences for two *L. monocytogenes* strains (N53-1 and La111) that were isolated 6 years apart from two different Danish fish processers. Both strains are of serotype 1/2a and belong to a highly persistent DNA subtype (random amplified polymorphic DNA [RAPD] type 9). We demonstrate using *in silico* analyses that both strains belong to the multilocus sequence typing (MLST) type ST121 that has been isolated as a persistent subtype in several European countries. The purpose of this study was to use genome analyses to identify genes or proteins that could contribute to persistence. In a genome comparison, the two persistent strains were extremely similar and collectively differed from the reference lineage II strain, EGD-e. Also, they differed markedly from a lineage I strain (F2365). On the proteome level, the two strains were almost identical, with a predicted protein homology of 99.94%, differing at only 2 proteins. No single-nucleotide polymorphism (SNP) differences were seen between the two strains; in contrast, N53-1 and La111 differed from the EGD-e reference strain by 3,942 and 3,471 SNPs, respectively. We included a persistent *L. monocytogenes* strain from the United States (F6854) in our comparisons. Compared to nonpersistent strains, all three persistent strains were distinguished by two genome deletions: one, of 2,472 bp, typically contains the gene for *inlF*, and the other, of 3,017 bp, includes three genes potentially related to bacteriocin production and transport (*lmo2774*, *lmo2775*, and the 3′-terminal part of *lmo2776*). Further studies of highly persistent strains are required to determine if the absence of these genes promotes persistence. While the genome comparison did not point to a clear physiological explanation of the persistent phenotype, the remarkable similarity between the two strains indicates that subtypes with specific traits are selected for in the food processing environment and that particular genetic and physiological factors are responsible for the persistent phenotype.**

Listeria monocytogenes is a Gram-positive, food-borne, human-pathogenic bacterium that can cause listeriosis in humans. It affects predominantly immunocompromised individuals, the elderly, young babies, and fetuses *in utero* (1). Although listeriosis represents only 7.4% of all reported food-borne infections, the fatality rate (17%) and hospitalization rates (92.6%) are high (2).

The bacterium is common in food products and poses a special risk in ready-to-eat products that allow proliferation of the pathogen. It is not only a safety issue but also an economic concern, because 61% of food products recalled by the U.S. FDA between 1994 and 1998 were due to *L. monocytogenes* contamination (3). The bacterium is an intracellular human pathogen, and it also has a saprophytic life-style and can therefore be isolated from soil and decaying plant material (4). Although it can be present in raw food materials, the processing plant environment is typically the immediate source of *L. monocytogenes* contamination of food products (5–8). Even though food processing equipment and facilities are cleaned frequently, some molecular subtypes of *L. monocytogenes* may persist in the food processing environment for many years (7–9).

We have found that one specific molecular subtype of *L. monocytogenes* strains was dominant and persistent in several fish processing plants (8, 10, 11). Other subtypes were also isolated several times in the processing plants although not as frequently (8). We reasoned that if we could understand the physiological and genetic characteristics that enabled this persistence, we could develop targeted intervention strategies and improve food safety by reducing or eliminating the highly persistent subtypes. We have investigated a series of behavioral patterns that we hypothesized were likely to explain the strong persistence. However, these persistent strains are not particularly common in the outside environment (12); they do not grow better under food processing conditions, nor do they form better biofilms (13); and they do not appear to tolerate biocides (14) or desiccation (15) better than presumed nonpersistent strains.

Since strains of food processing plant persistent subtypes are likely contaminants of ready-to-eat products, it is important to determine the degree of risk to the consumer. In simple eukaryotic cell models and simple animal models (*Caenorhabditis elegans* and *Drosophila melanogaster*), the highly persistent strains were less

Received 30 November 2012 **Accepted** 16 February 2013
Published ahead of print 22 February 2013
Address correspondence to Lone Gram, gram@bio.dtu.dk.
* Present address: Anne Holch, CMC Biologics A/S, Copenhagen, Denmark.
A.H. and K.W. contributed equally to the study.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03715-12.
Copyright © 2013, American Society for Microbiology. All Rights Reserved.
doi:10.1128/AEM.03715-12

Downloaded from http://aem.asm.org/ on March 1, 2018 by ALLEGHENY COLLEGE

TABLE 1 *L. monocytogenes* strains used in the present study[b]

Strain	Serotype	Lineage	Source of isolation	Reference for isolation	Reference or source(s) for nucleotide database	Reference for genome sequence
N53-1	1/2a	II	Smokehouse environment	8	This study	This study
La111	1/2a	II	Cold-smoked salmon	10	This study	This study
F6854	1/2a	II	Turkey franks	28	J. Craig Venter Institute, TraceDB[a]	29
EGD-e	1/2a	II	Rabbit isolate, 1926		NCBI	30
F2365	4b	I	Mexican-style cheese	31	J. Craig Venter Institute	29

[a] Raw data.
[b] *L. monocytogenes* N53-1 and La111 have been sequenced in the present study, whereas DNA sequences from F6854, EGD-e, and F2365 were retrieved from online databases.

invasive than human clinical strains (13, 16–18). Surprisingly, in a more complex biological model (using oral dosing of pregnant guinea pigs), the strains infected placentas and fetuses just as efficiently as the clinical strains (18). Hence, this particular subtype is of key interest since it is a recurrent contaminant and may be a risk, especially to pregnant women.

The genomes of several strains of *L. monocytogenes* have been sequenced in recent years (9, 19–22). At present, there are 34 *L. monocytogenes* genomes publicly available, of which 16 are finished and 18 are available as draft sequences. This rapid expansion in publicly available genome sequences is key to understanding the evolutionary history of *L. monocytogenes* and to elucidating virulence regulation. Our intent here was to harness genome-based analyses to better understand the basis of this organism's persistence in particular food processing environments.

In this work, we initially addressed the discriminatory power of subtyping by comparing the genome sequences and predicted proteomes of two strains of *L. monocytogenes* isolated from different plants at different times but which share the same molecular subtype. These two strains were representative of the above-mentioned large group of strains that were isolated repeatedly from fish processing environments over many years and that were indistinguishable by molecular subtyping (8). Subsequently, we searched for features uniquely shared by these and another (previously sequenced) persistent strain in order to identify genes that may contribute to, or detract from, persistence in such environments.

MATERIALS AND METHODS

***Listeria monocytogenes* strains.** Two *L. monocytogenes* strains, representing a highly persistent molecular subtype, were sequenced for this study. Strain La111 was isolated from a package of cold-smoked salmon in 1996 (11), whereas strain N53-1 was isolated from a processing environment in 2002 (8). These isolates derived from different plants. Both strains were determined to be serotype 1/2a and lineage II strains. The strains were deemed identical based on random amplified polymorphic DNA (RAPD), pulsed-field gel electrophoresis (PFGE), and amplified fragment length polymorphism (AFLP) typing and similar to a large cluster of molecular subtypes that are often isolated from Danish fish smokehouses (8). The strains were isolated following a selective enrichment, streaking onto Oxford agar, and restreaking onto brain heart infusion (BHI) agar. Stock cultures were stored at −80°C in a medium containing 4% (wt/vol) glycerol, 2% (wt/vol) skim milk powder, and 3% (wt/vol) tryptone soya broth (TSB) (catalog number CM0129; Oxoid). Growth in the present study was performed with TSB at 37°C.

DNA purification. Genomic DNA was purified with a Fast DNA kit (catalog number 116540-400; MP Biomedicals), with modifications. Cells were harvested after growth for 24 h in TSB (catalog number CM0129; Oxoid), and the pellet was resuspended in 210 μl buffer 1 (0.58 M sucrose, 0.01 M Na-P, 10 μg/ml lysozyme). The suspension was heated for 1.5 h at 37°C, followed by washing. The pellet was resuspended in demineralized water, and the procedure for the Fast DNA kit was followed. RNA was removed by using Ambion RNase Cocktail (catalog number AM2286; Invitrogen).

Genome sequencing. *L. monocytogenes* N53-1 and La111 were sequenced by using second-generation methods on the Illumina Genome Analyzer II (GAII). Approximately 1 μg of total genomic DNA from each strain was used to generate a short-read library. Library preparation, DNA sequencing, and raw data processing via the Illumina Genome Analyzer Analysis Pipeline were carried out in accordance with the manufacturer's protocols for single-end 36-bp reads (Illumina, San Diego, CA). The only exceptions involved the random fractionation of the genomic DNA via sonication (rather than nebulization) and the use of 5 μl (rather than 1 μl) of template for the final PCR amplification of the library. The GAII was employed for 36 cycles to generate the nucleotide data. Each strain was sequenced in one lane containing 2 pM template and in a second lane containing 3 pM template.

Assembly of genomes. Prior to assembly, sequences were filtered to remove those reads that contained one or more ambiguous base calls. The N53-1 and La111 sequences were assembled separately by using the *de novo* assembler Velvet version 1.1.04 (23), with parameters determined by Velvet Optimizer 2.1.7 (S. Gladman and T. Seeman). A high-resolution, ordered, and oriented restriction map (optical map) was generated for the N53-1 genome by using the OpGen system (OpGen Technologies, Madison, WI) and the NcoI endonuclease. This physical evidence was subsequently used to constrain genome assembly of N53-1 contigs using Mapsolver software (OpGen) based on *in silico* digestion and comparison of restriction cut site patterns of each contig to the genome. The optical map of N53-1 was considered dispositive as evidence in placing contigs generated from the N53-1 isolate. We subsequently explored the applicability of the N53-1 physical evidence for its potential to assist in the assembly of La111, premised on the hypothesis that genomes so similar in sequence content would also share syntenic organization. A minimum score for the local alignment was set initially to 3 and then reduced to 2. Only unambiguous alignments were accepted. For both strains, contigs were concatenated in the order and orientation determined by the optical map alignment. Between each contig, the sequence 5′-NNNNNCATTCCATTCATTAATTAATTAATGAATGAATGNNNNN-3′ was inserted (24). This sequence was designed such that it introduces a stop codon in all six reading frames as well as a start codon in all reading frames, encouraging proper annotation of those genes residing near contig junctions (24).

Genome annotation. The predicted proteomes of all analyzed strains were extracted by using Prodigal software (25), which is able to recognize prokaryotic genes and identify translational initiation sites. tRNA-encoding sequences were located by using the tRNAscan-SE 1.21 server (26). Genome comparisons were made by using Mauve v 2.3.1 (27) and BLAST via the NCBI website.

Genome sequences from online databases. The genome of *L. monocytogenes* EGD-e (GenBank accession number NC_003210.1), which is a lineage II, serotype 1/2a strain, was downloaded from the NCBI website (http://www.ncbi.nlm.nih.gov/) and used as the reference strain (Table 1). Assembled genomes of *L. monocytogenes* F6854 and *L. monocytogenes*

Downloaded from http://aem.asm.org/ on March 1, 2018 by ALLEGHENY COLLEGE

F2365 were downloaded from the J. Craig Venter Institute website (http://www.jcvi.org/). F6854 belongs to the same ribotype (DUP-1053A) as two other strains isolated 12 years later and linked to the same food processing facility (8) and is, hence, a highly persistent subtype. The raw sequence data of *L. monocytogenes* F6854 from TraceDB (ftp://ftp.ncbi.nih.gov/pub/TraceDB/) was included in the data set for the single-nucleotide polymorphism (SNP) analysis.

BLAST Ring Image Generator. Visual comparison of genome homology was done by using BRIG (BLAST Ring Image Generator) (32; http://sourceforge.net/projects/brig/). BRIG is capable of generating circular comparison images for prokaryote genomes and displays similarity between a reference genome in the center and other query sequences. EGD-e was used as the reference genome and was compared to the genomes of N53-1, La111, F6854, and F2365. As the similarity is calculated from the respective reference, regions that are absent from the reference genome but present in one or more of the query sequences will not be displayed. The BRIG method uses the software BLASTALL v 2.2.25+ for the searches. The comparisons were done with default settings.

BLAT and BLAT matrices. The similarity between N53-1 and La111, and the similarity to the other strains of *L. monocytogenes*, was also assessed by a pairwise genome comparison. A matrix showing the fraction of genome-specific genes was constructed. For each gene in one genome, a BLAST-Like Alignment (BLAT) was performed against the second genome. BLAT rapidly searches for relatively short *k*-mers and extends these to high-scoring pairs (HSPs) (33). A given gene was considered to be specific if there were no HSPs satisfying the 50/50 rule, meaning that no sequence in the queried genome was at least 50% identical to the gene over at least 50% of its length.

SNP analysis. For SNP detection, the raw data sequences from N53-1, La111, and F6854 were mapped to the reference strain EGD-e. N53-1 and La111 were mapped to both F6854 and F2365. Also, raw data sequences from N53-1 were mapped to the *de novo*-assembled La111 genome, and the raw data sequences of La111 were mapped to the *de novo*-assembled N53-1 genome. After mapping the raw data, open reading frames were identified, and the read mappings were analyzed for the presence of SNPs. All steps of the SNP analysis were conducted by using CLC Genomics Workbench v 4.8 (CLC, Aarhus, Denmark) with the default settings, except for the minimum variant frequency, which was set at 85%. A list of the identified SNPs was exported to an Excel spreadsheet. All SNPs coding for silent mutations were deleted, and further analysis was conducted with the remaining nonsynonymous SNPs.

***In silico* MLST analysis.** Multilocus sequence typing (MLST) was used to analyze nucleotide variations in seven housekeeping genes (*acbZ*, *bglA*, *cat*, *dapE*, *dat*, *ldh*, and *lhkA*) spread across the bacterial chromosome (34; http://www.pasteur.fr/recherche/genopole/PF8/mlst/Lmono.html). An *in silico* PCR analysis was conducted on the N53-1 and La111 genomes by using CLC DNA Workbench v 6.5 with default settings. The obtained *in silico* PCR products were trimmed and uploaded to the *L. monocytogenes* MLST database (http://www.pasteur.fr/recherche/genopole/PF8/mlst/Lmono.html) for determination of the sequence type (ST).

Nucleotide sequence accession numbers. Genome sequences have been submitted to the EMBL database at the EBI website and can be found under accession numbers HE999704 (strain La111) and HE999705 (strain N53-1).

RESULTS AND DISCUSSION

General genome features. The next-generation sequencing of *L. monocytogenes* N53-1 generated over 70.8 million reads, of which 69.3 million reads were retained after removing those containing ambiguous base calls within their sequence. The N53-1 reads assembled into 314 contigs (N50 [a statistic measuring assembly quality] = 100,675). For La111, over 57 million reads were generated, with 54.8 million reads subsequently analyzed after removing sequences containing ambiguous base calls. *De novo* assembly of the La111 short reads formed 279 contigs (N50 = 106,240).

TABLE 2 Genomic assembly data for *L. monocytogenes* N53-1 and La111

Chromosome parameter	Value for indicated strain	
	N53-1	La111
Total length of all *de novo*-assembled contigs (bp)[a]	3,103,912	3,017,238
Total G+C content (%)	37.9	37.9
Assembled length excluding gap sequence (bp)[b]	2,553,709	2,534,555
Assembled G+C content (%)	37.9	37.9
No. of predicted tRNAs[c]	94	86
No. of predicted proteins[d]	3,323	3,302
No. of plasmids[e]	1	1

[a] Data were obtained after analysis by Velvet 1.1.04.
[b] Data were obtained after analysis by Mapsolver.
[c] Data were obtained after analysis by tRNAScan-SE 1.21.
[d] Data were obtained after analysis by Prodigal.
[e] Data were obtained after analysis by BLAST.

By using Mapsolver software, the *in silico* digestions of the *de novo* N53-1 assembled contigs were compared to the optical map. In total, 25 contigs were placed, representing 82.3% of the sequence data generated for N53-1 (assembled length excluding gaps divided by total length of all *de novo*-assembled contigs) (Table 2). Using BLAST, we found that of the three remaining large contigs (>30 kb), two unplaced contigs aligned well to other published *L. monocytogenes* nuclear genomes, and one aligned to the plasmid sequence of *L. monocytogenes* 08-5578. Of the 279 *de novo*-assembled contigs of La111, 19 aligned to the optical map of N53-1 under the strict default parameters representing 78.5% of the genome (assembled length excluding gaps divided by total length of all *de novo*-assembled contigs) (Table 2). An alignment by using BLAST revealed that five of the six unmapped, large contigs ranging in size from 34.9 kb to 54.6 kb aligned closely with the genomes of *L. monocytogenes* 08-5578, 08-5923, and/or EGD-e, and one contig (37.7 kb) aligned to the plasmid sequence from *L. monocytogenes* 08-5578. A second BLAST alignment showed that four of the six large contigs showed a very high level of similarity (>99%) to the assembled N53-1 genome and, as such, were added to the La111 alignment based on this similarity. The final La111 assembly consisted of 23 contigs representing 84% of the sequence data generated.

Excluding the inserted gap sequences (24), the N53-1 genome assembly was 2,553,709 bp in length, while La111 totaled 2,534,555 bp, and both strains had a G+C content of 37.9% (Table 2). These genome sizes are similar to the sizes of other sequenced *L. monocytogenes* genomes, which have been estimated to be between 2.87 Mb (*L. monocytogenes* Finland 1988 [GenBank accession number CP002004.1]) and 3.02 Mb (*L. monocytogenes* Scott A [GenBank accession number AFGI00000000.1]). Ninety-four and 86 tRNAs were predicted within the N53-1 and La111 genome sequences, respectively (Table 2). Using Prodigal for the protein BLAST matrix, N53-1 was predicted to have 3,323 proteins, and La111 was predicted to have 3,302 proteins (Table 2). The differences between the two strains likely derive from missing data in the La111 assembly. Differences in the number of predicted proteins and predicted tRNAs were observed when using

Downloaded from http://aem.asm.org/ on March 1, 2018 by ALLEGHENY COLLEGE

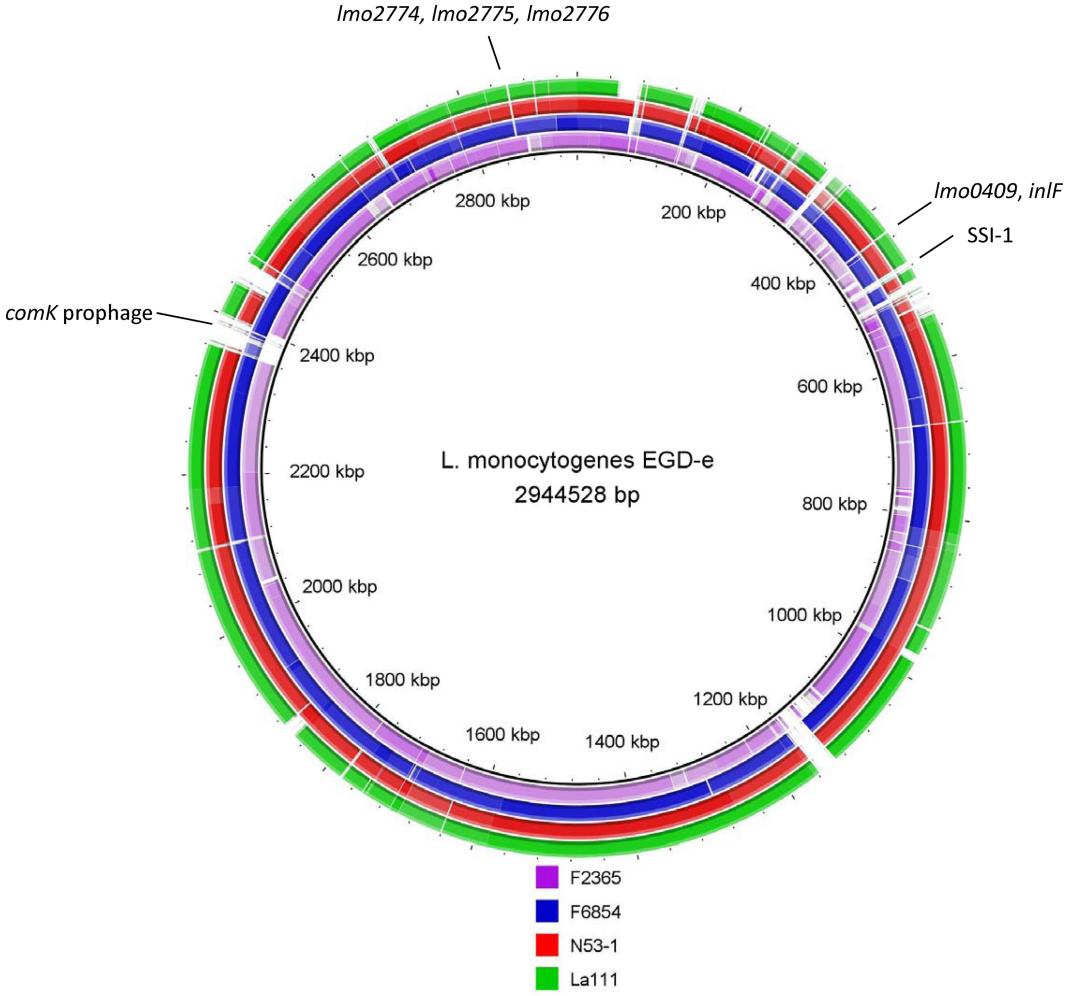

FIG 1 Circular map of *L. monocytogenes* N53-1, La111, F6854, and F2365 using EGD-e as a reference genome. The inner ring denotes the reference EGD-e genome with corresponding genetic coordinates. The next four rings denote the coding regions for the four queried strains, F6854 (blue), F2365 (purple), N53-1 (red), and La111 (green).

different programs. These differences are due to different algorithms and cutoff values used in the different programs.

Comparative genomics. N53-1 and La111 are very similar based on DNA subtyping (8), virulence gene sequencing (16), and phenotypic behavior (13, 16, 17). However, a whole-genome comparison of these two strains had not yet been attempted. Strains that are persistent might share genetic features that are not present in nonpersistent strains. This could include the presence or absence of entire genes, SNPs, or different patterns of gene expressions relative to presumably nonpersistent strains.

Conservation and variation in gene content between genomes were visualized by BRIG. The two newly sequenced genomes of N53-1 and La111 and the two downloaded genomes (F6854 and F2365) were included in the comparison, and EGD-e was used as a reference (Fig. 1). It should be noted that the F6854, N53-1, and La111 genomes are draft genomes and are not completely closed. Therefore, regions that are not included in the BRIG alignment most likely represent regions not sequenced in one or more genomes, deletions/insertions, or genome fragments replaced by a nonhomologous sequence.

A gap of 2,472 bp occurred in all three persistent strains (N53-1, La111, and F6854) relative to EGD-e and F2365, beginning at bp position 429629 and containing the *inlF* gene in F2365 (Fig. 1). InlF is a surface-anchored protein with unknown function; however, it plays a role in increased infection of L25 murine fibroblast cells (35) and is present in a large number of strains. Jia et al. (36) did not find any *inlF*-specific PCR products in lineage I strains, and Tsai et al. (37) found *inlF* in all tested lineage II strains and not in lineage I strains using gene sequencing. Doumith et al. (38) reported *inlF* in a least two-thirds of both lineage I and lineage II strains using a DNA microarray. Further studies of strains from highly persistent subtypes are required to determine if the absence of *inlF* promotes persistence.

A stretch of DNA of 3,017 bp was absent in N53-1, La111, and F6854 but present in EGD-e (at bp position 2857618) and F2365. The area covers *lmo2774*, *lmo2775*, and the 3′-terminal part of *lmo2776*. *lmo2774* encodes a homologue of a putative bacteriocin export ABC transporter, *lmo2775* a homologue of a bacteriocin-associated integral membrane protein, and *lmo2776* a homologue of lactococcin_972. The genes encoding these proteins are not well described, and no further information is available.

Downloaded from http://aem.asm.org/ on March 1, 2018 by ALLEGHENY COLLEGE

At bp position 2360713 in EGD-e, a large sequence of approximately 40,000 bp is not present in N53-1, La111, or F2365, whereas it is present in EGD-e and F6854. In F6854, the sequence has been identified as *comK* (major competence transcription factor). A prophage was previously shown to be inserted into *comK* in F6854 at this position (9, 39). Orsi et al. (9) used whole-genome sequence comparison to analyze four strains from the same processing plant: a food and outbreak pair from 1988 and a food and outbreak pair from 2000. These four strains differed by only 11 SNPs in the backbone sequence (excluding *comK* and the Thr-4 prophage) by an interstrain comparison. In all four sequenced strains (9), *comK* contained a prophage insertion of approximately 40,000 bp. In spite of the near uniformity of the backbone sequences, the prophage insert contained 1,274 SNPs that differentiated the pair from 1988 from the pair from 2000.

Recently, it was found that the presence of a prophage in *comK* could be a marker for rapid niche-specific adaptation, biofilm formation, and persistence (39); however, the two processing-persistent strains used in the present study may lack an intact prophage insertion in *comK* (gap of around 40 kbp in N53-1, La111, and F2365) (Fig. 1). We searched the La111 and N53-1 draft genomes for intact prophages using software described previously by Bohlin et al. (40) and found none. However, as our genome assemblies contain gaps representing regions where assembly of sequence data was not achieved, it is difficult to determine whether the full-length 42-kbp prophage is inserted into the *comK* gene within these two *Listeria* strains. We explored the possibility that the prophage is not present as one contiguous piece in our assemblies. Using nucleotide BLAST, portions (approximately 0.9 kbp) of the 28.5-kb *comK* prophage sequence from F6854 aligned well to the La111 and N53-1 assembled contigs. The most significant alignments occurred in the same area of the scaffold, and some of the alignments ended because of a gap in the sequences. Using MAQ (Mapping and Assembly with Qualities) (http://maq.sourceforge.net/), we found significant alignment of the raw sequence data from both strains across approximately 50% of the *comK* prophage reference sequence. Hence, there is strong evidence that at least a portion of a prophage is present in the La111 and N53-1 draft genomes. However, we are unsure as to whether the prophage, in its entirety, persists. This may be attributed to limitations in the assembly of repetitive regions and/or the inability to map reads that differ by more than 2 bases (a parameter of MAQ). Alternatively, the results may represent a relic of a previous phage insertion and subsequent deletion event. If the two *Listeria* strains do contain a prophage in *comK*, it could potentially be involved in the persistence mechanism (39).

At bp position 473841 in EGD-e, there is a gap of 7,500 bp in N53-1 and La111, whereas the gap size in F6854 and F2365 is 8,625 bp. The genes present in this region in EGD-e (*lmo0444*, *lmo0445*, *lmo0446* [*pva*], *lmo0447* [*gadD1*], and *lmo0448* [*gadT1*]), designated stress survival islet 1 (SSI-1), are responsible for growth at low pH and at high salt concentrations and the ability to survive and grow in model food systems (41). The size of the gap is larger in F6854 and F2365, as the islet in those strains contains only one gene (*LMOf2365_0481* homologue), whereas the islet in N53-1 and La111 contains genes homologous to *lin0464* and *lin0465*. A more detailed description of SSI-1 is presented below.

Comparative proteomics. The gene content of strains was compared in a BLAT matrix (Fig. 2). It displays the frequency of genes found in the "row" genome that are not also found in the "column" genome, as a proportion of the total number of genes in the row genome. Strains N53-1 and La111 are extremely similar, with only 2 (0.06%) of the predicted proteins in N53-1 not present in La111. In contrast, 144 and 143 (5%) of the predicted proteins in EGD-e were not present in N53-1 and La111, respectively. The genomes of both N53-1 and La111 are not fully sequenced, which could explain the missing predicted proteins in these two strains compared to EGD-e.

Of two predicted proteins present in N53-1 but absent in La111, one with unknown function has NACHT and WD repeat domain-containing protein 1. The WD40 domain is found in a number of eukaryotic proteins that cover a wide variety of functions, including adaptor/regulatory modules of signal transduction, pre-mRNA processing, and cytoskeleton assembly (http://www.ncbi.nlm.nih.gov/protein/308736994). An uncharacterized protein, YdeI, is the only predicted protein present in N53-1 and absent from EGD-e, while the glutamate synthase (NADPH) large chain and glycine betaine/carnitine/choline transport ATP-binding protein OpuCA are present in La111 but absent from EGD-e. As none of these proteins are present in both N53-1 and La111, none can independently suffice as a cause for persistence each of these strains. When turning to the 5% of predicted proteins that are unique to EGD-e, they cover a broad range of protein functions (see Table S1 in the supplemental material).

Single-nucleotide analysis. Although some SNPs may derive from sequencing errors, true SNPs may be either silent or nonsynonymous, in which case they may change the function of the transcribed protein or result in a truncated protein. An example of the latter possibility entailed the listerial surface protein InlA, where a nucleotide substitution from C to T results in a stop codon and production of a truncated protein (16, 42, 43). In the present study, SNP analyses assessing only nonsynonymous changes were carried out after mapping raw reads of the queried strain against a reference strain (Fig. 3). First, the three persistent strains (N53-1, La111, and F6854) were mapped against EGD-e, and between 3,471 and 5,037 SNPs were detected (Fig. 3A). Among these, 1,980 SNPs were shared between the three strains.

The numbers of SNPs detected between F6854 and our two newly sequenced strains were 3,829 and 3,819 for N53-1 and La111, respectively (Fig. 3B). All three strains belong to serotype 1/2a. Comparing our two newly sequenced strains to F2365, a serotype 4b strain, identified 5,848 and 5,840 SNPs, respectively (Fig. 3C).

In contrast, testing of N53-1 and La111 against each other identified no SNPs when using N53-1 as the reference; using La111 as the reference suggested only 18 SNPs, substantiating an extraordinarily close relationship between the two strains in spite of the 6-year interval separating their dates of isolation. The complete lack of SNPs between strains N53-1 and La111, despite being isolated 6 years apart from two different factories, may indicate that this genome type is especially well adapted to persisting in this environment.

***In silico* MLST analysis.** Seven *in silico* PCR products of between 458 and 702 bp were obtained from N53-1 and La111. After trimming, the sequences were uploaded to the *L. monocytogenes* MLST database, and both were identified as belonging to ST121: *abcZ-7*, *bglA-6*, *cat-8*, *dapE-8*, *dat-6*, *ldh-37*, and *lhkA-1*. F6854 belongs to ST11, which corresponds to *abcZ-7*, *bglA-6*, *cat-10*, *dapE-6*, *dat-1*, *ldh-2*, and *lhkA-1*, and EGD-e belongs to ST35 (*abcZ-6*, *bglA-5*, *cat-6*, *dapE-20*, *dat-1*, *ldh-4*, and *lhkA-1*) (44). A

Downloaded from http://aem.asm.org/ on March 1, 2018 by ALLEGHENY COLLEGE

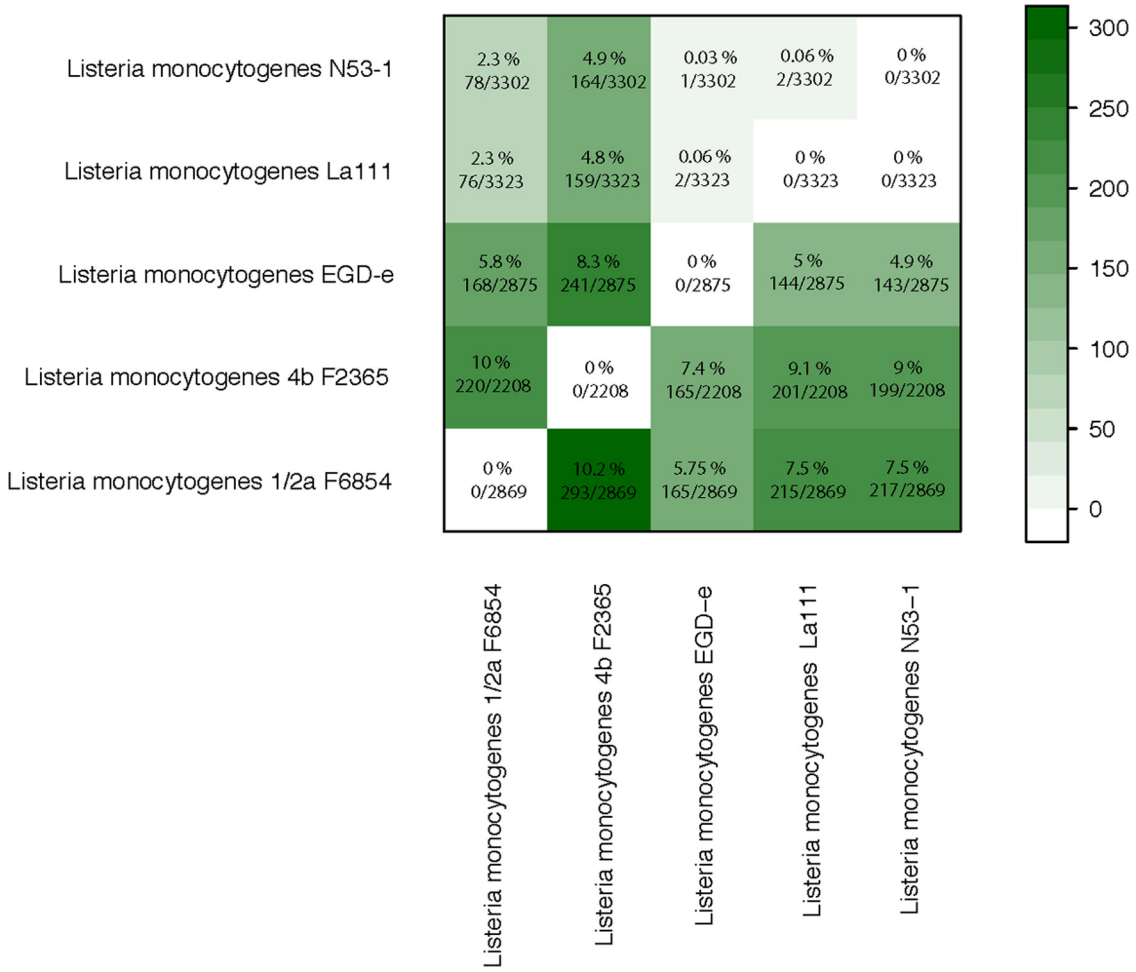

FIG 2 BLAT matrix of pairwise genome analysis between *L. monocytogenes* strains N53-1, La111, EGD-e, F2365, and F6854. Each row represents the specific genome of one genome compared to another, while the diagonal shows comparison to itself. In the matrix cells, the numbers of nonshared protein-coding genes are given both as a number and as a percentage, based on the ratio of the specific genome and total number of predicted genes of the query genome, as indicated. The cells in the matrix are colored darker as the fraction of similarity decreases.

recent study by Hein et al. (45) described ST121 strains isolated in Austria and Belgium from different ecological niches, including food, food processing facilities, and human cases, over several years. Two of the strains were isolated from the same dairy plant over a course of at least 3 years. *L. monocytogenes* ST121 strains have also been reported in France, Italy, and Spain (34, 46, 47). By PCR, Hein et al. (45) showed that the ST121 strain had a 2.2-kbp fragment (in N53-1 and La111), whereas the majority of serotype 1/2a strains had a 9.7-kbp fragment. The 9.7-kbp fragment is described as a five-gene stress survival islet (SSI-1) and contributes to growth under suboptimal conditions (41). A BLAST search of the 2.2-kbp fragment showed 95% identity with the two genes *lin0464* and *lin0465* from *Listeria innocua* CLIP 11262 (GenBank accession number AL596165.1). Hein et al. (45) speculated that the two *L. innocua* genes *lin0464* and *lin0465* both contribute to fitness of the ST121 strains in the environment. Furthermore, the ST121 strains also had the same premature stop codon in *inlA*, leading to a truncated InlA, as in our two processing-persistent strains. Altogether, we can conclude that the ST121 strains described in a variety of studies are identical to our two processing-persistent strains, whose genomes we have now sequenced. Evidence that this group of strains persists in the processing environment is mounting, and the basis for this attribute warrants investigation.

Conclusions. Several studies have reported the ability of particular molecular subtypes of *Listeria monocytogenes* to persist in food processing plants (7–9), where they constitute a recurrent source of product contamination. In the Danish fish processing industry, strains belonging to one particular subtype of *L. monocytogenes* have been isolated over several years in different processing plants. Strains of this subtype were isolated from four out of eight different processing plants and were the persistent and dominant type in three plants over a period of 6 years (8, 11). These data indicate that certain subtypes of *L. monocytogenes* may be specifically adapted to processing plant environments and are able to persist over long periods of time. However, our data do not allow us to conclude on the underlying ecology and evolution. Thus, we cannot say if a particular subtype at random enters the processing environments and, due to low growth rates, remains unchanged for years or if the conditions in the environment select for particular mutational changes over time. The use of genome sequencing of strains isolated repeatedly from a plant over longer

Downloaded from http://aem.asm.org/ on March 1, 2018 by ALLEGHENY COLLEGE

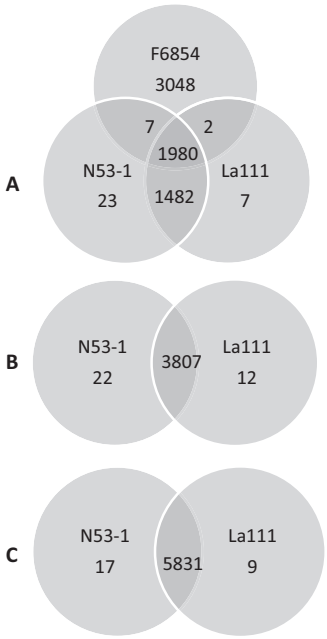

FIG 3 Detection of SNP by CLC genomics workbench between *L. monocytogenes* strains. Raw sequencing data were analyzed against either annotated genomic sequences (EGD-e) or genomic sequences with predicted open reading frames (F6854 and F2365). Only those SNPs resulting in an amino acid change are included in the figure. (A) N53-1, La111, and F6854 raw sequences are assembled versus EGD-e; (B) N53-1 and LA111 are assembled versus F6854; (C) N53-1 and La111 are assembled versus F2365.

periods of time could potentially unravel this. Such approaches have recently been used to analyze the changes in persistent *Pseudomonas aeruginosa* in cystic fibrosis lungs (48).

The present study is the first to sequence the genomes of two persistent food processing *L. monocytogenes* strains belonging to the same DNA subtype and isolated from two different processing environments that do not have any intertrade relationship. We demonstrate that the two persistent food processing strains are almost identical, as their predicted proteomes differ by only 2 proteins. One would expect that the food processing environment would impose strong selective pressures on the growth and survival of bacteria and that these, coupled with chance events, would result in establishment of different subtypes. Our data indicate that despite such differences, very specific genetic and physiological traits may enable long-term persistence in food processing factories.

We did find genes and proteins that were uniquely shared or absent in La111 and N53-1 (compared to the other strains). However, because the number of strains investigated is relatively limited, and the genomes of N53-1, La111, and F6854 are draft genomes, we cannot conclude which genes or mutations best explain persistence in this instance. Persistence may likely result from a combination of genetic and environmental characteristics. It is likely that other ST121 strains originating from other countries and other food product environments are highly homologous to the two newly sequenced ST121 strains, and comparing genomic and proteomic homology between a collection of ST121 strains could likely point to key persistence markers. Even though this study did not result in a clear explanation of the persistent phenotype of the subgroup of strains isolated in the Danish fish processing industry, the remarkable similarity between the two strains indicates that subtypes with specific traits are selected for in food processing environments and that particular genetic and physiological factors are responsible for the persistent phenotype.

ACKNOWLEDGMENTS

Anne Holch was supported by the Danish Research Council for Technology and Production Sciences (project 274-08-042). Kristen Webb was supported by the ARS Research Associate Program.

We thank Alicia Beavers, Tony Capuco, Chris Clover, Detiger Dunams, Monica Santin-Duran, Garrett Gleeson, Steve Schroeder, and Tad Sonstegard for support toward the completion of the project. We thank Jon Bohlin for help with prophage analyses.

REFERENCES

1. **Gillespie IA, McLauchlin J, Grant KA, Little CL, Mithani V, Penman C, Lane C, Regan M.** 2006. Changing pattern of human listeriosis, England and Wales, 2001–2004. Emerg. Infect. Dis. **12**:1361–1366.
2. **CDC.** 2010. Foodborne Diseases Active Surveillance Network (FoodNet): FoodNet surveillance report for 2008 (final report). CDC, Atlanta, GA.
3. **Wong S, Street D, Delgado SI, Klontz KC.** 2000. Recalls of foods and cosmetics due to microbial contamination reported to the US Food and Drug Administration. J. Food Prot. **63**:1113–1116.
4. **MacGowan AP, Bowker K, McLauchlin J, Bennett PM, Reeves DS.** 1994. The occurrence and seasonal changes in the isolation of *Listeria* spp. in shop bought food stuffs, human feces, sewage and soil from urban sources. Int. J. Food Microbiol. **21**:325–334.
5. **Eklund MW, Poysky FT, Paranjpye RN, Lashbrook LC, Peterson ME, Pelroy GA.** 1995. Incidence and sources of *Listeria monocytogenes* in cold-smoked fishery products and processing plants. J. Food Prot. **58**:502–508.
6. **Keto-Timonen R, Tolvanen R, Lunden J, Korkeala H.** 2007. An 8-year surveillance of the diversity and persistence of *Listeria monocytogenes* in a chilled food processing plant analyzed by amplified fragment length polymorphism. J. Food Prot. **70**:1866–1873.
7. **Norton DM, McCamey MA, Gall KL, Scarlett JM, Boor KJ, Wiedmann M.** 2001. Molecular studies on the ecology of *Listeria monocytogenes* in the smoked fish processing industry. Appl. Environ. Microbiol. **67**:198–205.
8. **Wulff G, Gram L, Ahrens P, Vogel BF.** 2006. One group of genetically similar *Listeria monocytogenes* strains frequently dominate and persist in several fish slaughter and smokehouses. Appl. Environ. Microbiol. **72**:4313–4322.
9. **Orsi R, Borowsky M, Lauer P, Young S, Nusbaum C, Galagan J, Birren B, Ivy R, Sun Q, Graves L, Swaminathan B, Wiedmann M.** 2008. Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment. BMC Genomics **9**:539. doi:10.1186/1471-2164-9-539.
10. **Vogel BF, Jorgensen LV, Ojeniyi B, Huss HH, Gram L.** 2001. Diversity of *Listeria monocytogenes* isolates from cold-smoked salmon produced in different smokehouses as assessed by random amplified polymorphic DNA analyses. Int. J. Food Microbiol. **65**:83–92.
11. **Vogel BF, Huss HH, Ojeniyi B, Ahrens P, Gram L.** 2001. Elucidation of *Listeria monocytogenes* contamination routes in cold-smoked salmon processing plants detected by DNA-based typing methods. Appl. Environ. Microbiol. **67**:2586–2595.
12. **Hansen CH, Vogel BF, Gram L.** 2006. Prevalence and survival of *Listeria monocytogenes* in Danish aquatic and fish-processing environments. J. Food Prot. **69**:2113–2122.
13. **Jensen A, Larsen MH, Ingmer H, Vogel BF, Gram L.** 2007. Sodium chloride enhances adherence and aggregation and strain variation influences invasiveness of *Listeria monocytogenes* strains. J. Food Prot. **70**:592–599.
14. **Kastbjerg VG, Gram L.** 2009. Model systems allowing quantification of sensitivity to disinfectants and comparison of disinfectant susceptibility of persistent and presumed nonpersistent *Listeria monocytogenes*. J. Appl. Microbiol. **106**:1667–1681.
15. **Vogel BF, Hansen LT, Mordhorst H, Gram L.** 2010. The survival of *Listeria monocytogenes* during long term desiccation is facilitated by sodium chloride and organic material. Int. J. Food Microbiol. **140**:192–200.
16. **Holch A, Gottlieb CT, Larsen MH, Ingmer H, Gram L.** 2010. Poor invasion of trophoblastic cells but normal plaque formation in fibroblastic cells despite *actA* deletion in a group of *Listeria monocytogenes* strains

Downloaded from http://aem.asm.org/ on March 1, 2018 by ALLEGHENY COLLEGE

persisting in some food processing environments. Appl. Environ. Microbiol. **76**:3391–3397.
17. **Jensen A, Thomsen LE, Jørgensen RL, Larsen MH, Roldgaard BB, Christensen BB, Vogel BF, Gram L, Ingmer H.** 2008. Processing plant persistent strains of *Listeria monocytogenes* appear to have a lower virulence potential than clinical strains in selected virulence models. Int. J. Food Microbiol. **123**:254–261.
18. **Jensen A, Williams D, Irvin EA, Gram L, Smith MA.** 2008. A processing plant persistent strain of *Listeria monocytogenes* crosses the feto-placental barrier in a pregnant guinea pig model. J. Food Prot. **71**:1028–1034.
19. **Briers Y, Klumpp J, Schuppler M, Loessner MJ.** 2011. Genome sequence of *Listeria monocytogenes* Scott A, a clinical isolate from a food-borne listeriosis outbreak. J. Bacteriol. **193**:4284–4285.
20. **Chen J, Xia Y, Cheng C, Fang C, Shan Y, Jin G, Fang W.** 2011. Genome sequence of the nonpathogenic *Listeria monocytogenes* serovar 4a strain M7. J. Bacteriol. **193**:5019–5020.
21. **Chen Y, Strain EA, Allard M, Brown EW.** 2011. Genome sequences of *Listeria monocytogenes* strains J1816 and J1-220, associated with human outbreaks. J. Bacteriol. **193**:3424–3425.
22. **Steele CL, Donaldson JR, Paul D, Banes MM, Arick T, Bridges SM, Lawrence ML.** 2011. Genome sequence of lineage III *Listeria monocytogenes* strain HCC23. J. Bacteriol. **193**:3679–3680.
23. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.
24. **Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Ros IMY, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou LW, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM.** 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome." Proc. Natl. Acad. Sci. U. S. A. **102**:13950–13955.
25. **Hyatt D, Chen GL, LoCascio P, Land M, Larimer F, Hauser L.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119. doi:10.1186/1471-2105-11-119.
26. **Schattner P, Brooks AN, Lowe TM.** 2005. The tRNAscan-SE, snoscan and snoGPS Web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. **33**:W686–W689. doi:10.1093/nar/gki366.
27. **Darling ACE, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**:1394–1403.
28. **CDC.** 1989. Listeriosis associated with consumption of turkey franks. MMWR Morb. Mortal. Wkly. Rep. **38**:267–268.
29. **Nelson KE, Fouts DE, Mongodin EF, Ravel J, Deboy RT, Kolonay JF, Rasko DA, Angiuoli SV, Gill SR, Paulsen IT, Peterson J, White O, Nelson WC, Nierman W, Beanan MJ, Brinkac LM, Daugherty SC, Dodson RJ, Durkin AS, Madupu R, Haft DH, Selengut J, Van Aken S, Khouri H, Fedorova N, Forberger H, Tran B, Kathariou S, Wonderling LD, Uhlich GA, Bayles DO, Luchansky JB, Fraser CM.** 2004. Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species. Nucleic Acids Res. **32**:2386–2395.
30. **Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, Charbit A, Chetouani F, Couve E, de Daruvar A, Dehoux P, Domann E, Dominguez-Bernal G, Duchaud E, Durant L, Dussurget O, Entian KD, Fsihi H, Garcia-Del Portillo F, Garrido P, Gautier L, Goebel W, Gomez-Lopez N, Hain T, Hauf J, Jackson D, Jones LM, Kaerst U, Kreft J, Kuhn M, Kunst F, Kurapkat G, Madueno E, Maitournam A, Vicente JM, Ng E, Nedjari H, Nordsiek G, Novella S, de Pablos B, Perez-Diaz JC, Purcell R, Remmel B, Rose M, Schlueter T, Simoes N, Tierrez A, Vazquez-Boland JA, Voss H, Wehland J, Cossart P.** 2001. Comparative genomics of *Listeria* species. Science **294**:849–852.
31. **Linnan MJ, Mascola L, Lou XD, Goulet V, May S, Salminen C, Hird DW, Yonekura ML, Hayes P, Weaver R, Audurier A, Plikaytis BD, Fannin SL, Kleks A, Broome CV.** 1988. Epidemic listeriosis associated with Mexican-style cheese. N. Engl. J. Med. **319**:823–828.
32. **Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA.** 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics **12**:402. doi:10.1186/1471-2164-12-402.
33. **Kent WJ.** 2002. BLAT—the BLAST-like alignment tool. Genome Res. **12**:656–664.
34. **Ragon M, Wirth T, Hollandt F, Lavenir R, Lecuit M, Le Monnier A, Brisse S.** 2008. A new perspective on *Listeria monocytogenes* evolution. PLoS Pathog. **4**:e1000146. doi:10.1371/journal.ppat.1000146.
35. **Kirchner M, Higgins DE.** 2008. Inhibition of ROCK activity allows InlF-mediated invasion and increased virulence of *Listeria monocytogenes*. Mol. Microbiol. **68**:749–767.
36. **Jia YM, Nightingale KK, Boor KJ, Ho A, Wiedmann M, McGann P.** 2007. Distribution of internalin gene profiles of *Listeria monocytogenes* isolates from different sources associated with phylogenetic lineages. Foodborne Pathog. Dis. **4**:222–232.
37. **Tsai YHL, Orsi RH, Nightingale KK, Wiedmann M.** 2006. *Listeria monocytogenes* internalins are highly diverse and evolved by recombination and positive selection. Infect. Genet. Evol. **6**:378–389.
38. **Doumith M, Cazalet C, Simoes N, Frangeul L, Jacquet C, Kunst F, Martin P, Cossart P, Glaser P, Buchrieser C.** 2004. New aspects regarding evolution and virulence of *Listeria monocytogenes* revealed by comparative genomics and DNA arrays. Infect. Immun. **72**:1072–1083.
39. **Verghese B, Lok M, Wen J, Alessandria V, Chen Y, Kathariou S, Knabel S.** 2011. *comK* prophage junction fragments as markers for *Listeria monocytogenes* genotypes unique to individual meat and poultry processing plants and a model for rapid niche-specific adaptation, biofilm formation, and persistence. Appl. Environ. Microbiol. **77**:3279–3292.
40. **Bohlin J, Skjerve E, Ussery DW.** 2008. Reliability and applications of statistical methods based on oligonucleotide frequencies in bacterial and archaeal genomes. BMC Genomics **9**:104. doi:10.1186/1471-2164-9-104.
41. **Ryan S, Begley M, Hill C, Gahan CGM.** 2010. A five-gene stress survival islet (SSI-1) that contributes to the growth of *Listeria monocytogenes* in suboptimal conditions. J. Appl. Microbiol. **109**:984–995.
42. **Jonquieres R, Bierne H, Mengaud J, Cossart P.** 1998. The *inlA* gene of *Listeria monocytogenes* L028 harbors a nonsense mutation resulting in release of internalin. Infect. Immun. **66**:3420–3422.
43. **Olier M, Pierre F, Lemaitre JP, Divies C, Rousset A, Guzzo J.** 2002. Assessment of the pathogenic potential of two *Listeria monocytogenes* human faecal carriage isolates. Microbiology **148**:1855–1862.
44. **Gilmour M, Graham M, Van Domselaar G, Tyler S, Kent H, Trout-Yakel K, Larios O, Allen V, Lee B, Nadon C.** 2010. High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. BMC Genomics **11**:120. doi:10.1186/1471-2164-11-120.
45. **Hein I, Klinger S, Dooms M, Flekna G, Stessl B, Leclercq A, Hill C, Allerberger F, Wagner M.** 2011. Stress survival islet 1 (SSI-1) survey in *Listeria monocytogenes* reveals an insert common to *Listeria innocua* in sequence type 121 *L. monocytogenes* strains. Appl. Environ. Microbiol. **77**:2169–2173.
46. **Parisi A, Latorre L, Normanno G, Miccolupo A, Fraccalvieri R, Lorusso V, Santagada G.** 2010. Amplified fragment length polymorphism and multi-locus sequence typing for high-resolution genotyping of *Listeria monocytogenes* from foods and the environment. Food Microbiol. **27**:101–108.
47. **Salcedo C, Arreaza L, Alcala B, De La Fuente L, Vazquez JA.** 2003. Development of a multilocus sequence typing method for analysis of *Listeria monocytogenes* clones. J. Clin. Microbiol. **41**:757–762.
48. **Yang L, Jelsbak L, Marvig RL, Damkiaer S, Workman CT, Rau MH, Hansen SK, Folkesson A, Johansen HK, Ciofu O, Hoiby N, Sommer MOA, Molin S.** 2011. Evolutionary dynamics of bacteria in a human host environment. Proc. Natl. Acad. Sci. U. S. A. **108**:7481–7486.

Downloaded from http://aem.asm.org/ on March 1, 2018 by ALLEGHENY COLLEGE