Complete Genome Sequences of 12 Isolates of *Listeria monocytogenes* Belonging to Serotypes 1/2a, 1/2b, and 4b Obtained from Food Products and Food-Processing Environments in Canada

Walid Mottawea,[a,b] Shu Chen,[c] Saleema Saleh-Lakha,[c] Sebastien Belanger,[d] Dele Ogunremi[d]

Department of Food Science and Agricultural Chemistry, Faculty of Agricultural and Environmental Sciences, McGill University, Montreal, Quebec, Canada[a]; Department of Microbiology and Immunology, Faculty of Pharmacy, Mansoura University, Mansoura, Egypt[b]; Laboratory Services Division, University of Guelph, Guelph, Ontario, Canada[c]; Ottawa Laboratory Fallowfield, Canadian Food Inspection Agency, Ottawa, Ontario, Canada[d]

ABSTRACT *Listeria monocytogenes* is the etiological agent for an often fatal food-borne illness known as listeriosis. Here, we present the complete genome sequences of 12 *L. monocytogenes* isolates representing the three most common serotypes of this pathogen (1/2a, 1/2b, and 4b), collected in Canada from different food products and environmental sources.

Address correspondence to Dele Ogunremi, dele.ogunremi@inspection.gc.ca.

Listeriosis, a life-threatening infection caused by *Listeria monocytogenes*, results in one of the highest mortality rates among foodborne illnesses in Canada, with an average of 44 deaths each year (90% probability intervals: 31-76 deaths) (1). In order to reduce the risk of listeriosis, studies aimed at understanding the nature and behavior of the causative bacterial organism at the genomic level are necessary. *L. monocytogenes* is a ubiquitous, Gram-positive bacterium commonly found in natural environments such as plants, soil, and surface water, from where it contaminates agricultural products, food-processing environments, and foods such as fresh produce and ready-to-eat meat and milk, leading to possible human exposure (2). More than 95% of human clinical cases are associated with three serotypes of *L. monocytogenes*: 1/2a, 1/2b, and 4b (3). Here, we report the complete genome sequences of 12 *L. monocytogenes* isolates belonging to these serotypes. The strains were collected from foods and food-processing environments in Canada during the period from 2002 to 2009 as part of routine testing procedures carried out by the Agriculture and Food Laboratory in Guelph, Ontario, Canada. Strains were selected to represent both predominant and unique genotypes among over 2,400 strains tested during a comprehensive genotypic analysis (4).

Genomic DNA samples were extracted from overnight cultures of the 12 *L. monocytogenes* strains using the Wizard Genomic DNA purification Kit (Promega, Madison, WI, USA). Each isolate was sequenced using PacBio (Pacific BioSciences Inc., Menlo Park, CA, USA) and Illumina MiSeq (Illumina Inc., San Diego, CA, USA) technologies, as per standard protocols in 2013, and assembled independently. The PacBio reads were assembled with the Celera assembler using the HGAP/Quiver protocol (5) and corrected with Illumina sequence data. The Illumina reads, on the other hand, were assembled using the hybrid approach as previously described (6). The final corrected genome for each chromosome was developed by aligning an *in silico* map of the Illumina read-

TABLE 1 GenBank accession numbers of 12 *L. monocytogenes* isolates from food products and food-processing environments

Isolate	Identification source	Serotype	GenBank accession no.
10-092876-1559	Chicken breast	1/2a	CP019614
10-092876-0168	Sprouts	1/2b	CP019615
10-092876-1063	Baby spinach	4b	CP019616
10-092876-0055	Environmental swab	1/2a	CP019617
10-092876-0731	Environmental sponge	1/2a	CP019618
10-092876-1155	Environmental swab	4b	CP019619
10-092876-1547	Environmental swab	4b	CP019620
10-092876-1235	Baby spinach	1/2a	CP019621
10-092876-0145	Ground beef	1/2b	CP019622
10-092876-1763	Chicken breast	1/2a	CP019623
10-092876-1016	Ready-to-eat fermented meat	1/2b	CP019624
10-092876-0769	Environmental sponge	4b	CP019625

corrected PacBio assembly with the map of the corresponding, independently assembled Illumina sequence against an optical, *de novo* whole-genome map. The map consisted of *Nhe*I restriction sites and was developed using the Argus optical mapping system (OpGen Inc., Gaithersburg, MD, USA), and the analysis was done using the MapSolver software. The nucleotide sequences of all detected gaps and misalignments were corrected using Clone Manager software (Professional edition, Scientific and Educational Software, Cary, NC, USA).

The generated genomes range from 2,913,947 to 3,072,093 bp with a GC content of $37.95 \pm 0.02\%$ (average $\pm$ standard error of the mean [SEM]). We used the Rapid Annotations using Subsystem Technology server (7–9) to annotate each genome. We identified $2,944 \pm 21.6$ (average $\pm$ SEM) protein-coding sequences and 64 ± 1.29 (average $\pm$ SEM) tRNAs per genome. Alignment of all 12 genome sequences using the Mauve system of multiple genome alignment (10) demonstrated clustering based on serotypes. Average nucleotide identity analysis (http://enve-omics.ce.gatech.edu/ani) confirmed similarities within each serotype (1/2a: 98.60 to 98.99%, $n = 5$; 1/2b: 99.44 to 99.94%, $n = 3$; 4b: 99.47 to 99.60%, $n = 4$) and was able to differentiate between serotype 1/2a and the other two serotypes (versus 1/2b: 94.32 to 94.47%; versus 4b: 94.19 to 94.35%); however, serotypes 1/2b and 4b were very similar (99.42 to 99.57%). A complete comparative genomic study of the 12 strains will be presented elsewhere.

Accession number(s). The complete genome sequences of these 12 *L. monocytogenes* isolates were deposited in GenBank under BioProject no. PRJNA 371291. Accession numbers are shown for each isolate in Table 1 (CP019614 to CP019625).

REFERENCES

1. Thomas MK, Murray R, Flockhart L, Pintar K, Fazil A, Nesbitt A, Marshall B, Tataryn J, Pollari F. 2015. Estimates of foodborne illness—related hospitalizations and deaths in Canada for 30 specified pathogens and unspecified agents. Foodborne Pathog Dis 12:820–827. https://doi.org/10.1089/fpd.2015.1966.

2. Zhang CX, Brooks BW, Huang H, Pagotto F, Lin M. 2016. Identification of surface protein biomarkers of *Listeria monocytogenes* via bioinformatics and antibody-based protein detection tools. Appl Environ Microbiol 82:5465–5476. https://doi.org/10.1128/AEM.00774-16.

3. Orsi RH, den Bakker HC, Wiedmann M. 2011. *Listeria monocytogenes* lineages: genomics, evolution, ecology, and phenotypic characteristics. Int J Med Microbiol 301:79–96. https://doi.org/10.1016/j.ijmm.2010.05.002.

4. Saleh-Lakha S, Allen V, Li J, Pagotto F, Odumeru J, Taboada E, Blais B,

Ogunremi D, Downing G, Lee S, Gao A, Chen S. 2013. Subtyping of a large collection of historical *Listeria monocytogenes* strains from Ontario, Canada by an improved multilocus variable-number tandem repeat analysis (MLVA). Appl Environ Microbiol 79:6472–6480.

5. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.

6. Ogunremi D, Devenish J, Amoako K, Kelly H, Dupras AA, Belanger S, Wang LR. 2014. High resolution assembly and characterization of genomes of Canadian isolates of *Salmonella Enteritidis*. BMC Genomics 15:713. https://doi.org/10.1186/1471-2164-15-713.

7. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL,

Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.

8. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.

9. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.

10. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.