



# Draft Genome Sequences of Four Clinical *Legionella pneumophila* Isolates from Ontario, Canada

Alexander Fortuna,<sup>a</sup> Ricardo Ramnarine,<sup>a</sup> Aimin Li,<sup>a</sup> Nahuel Fittipaldi,<sup>a,b</sup> Christine Frantz,<sup>a</sup> Gustavo V. Mallo<sup>a,b</sup>

<sup>a</sup>Public Health Ontario, Toronto, Canada

<sup>b</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

**ABSTRACT** *Legionella pneumophila* outbreak investigations require the development of reliable typing methods to better understand the genetic relationships of the isolates involved. Here, we report the draft genome sequences of four clinical *Legionella pneumophila* isolates obtained between 2000 and 2012 in Ontario, Canada.

*Legionella pneumophila* is a bacterial pathogen responsible for Legionnaire's disease, a severe pneumonia-like illness with a high mortality rate (1). *L. pneumophila* may be found in natural and man-made environments and is transmitted through the inhalation of aerosolized water droplets containing the bacterium (2). The *L. pneumophila* isolates in this study were collected from patients in Ontario, Canada, between 2000 and 2012.

Whole-genome sequencing (WGS) is transforming the way infectious disease bacterial research is performed, providing increased resolution in outbreak scenarios (3). As part of an outbreak investigation, it is important to develop reliable methods for the typing of *L. pneumophila* isolates. The draft genome sequences from this study (of isolates LG57, LG59, LG61, and LG63) will be instrumental in developing new strategies for outbreak investigations in Ontario.

Genomic DNA was extracted from cultured *L. pneumophila* isolates using the automated easyMAG extraction system (bioMérieux Canada, Inc., Canada) and then quantified using a Qubit spectrophotometer (Thermo Fisher Scientific, USA). DNA sequencing libraries were prepared using the Nextera XT kit (Illumina, USA). Individually tagged libraries were checked on an Agilent 2100 Bioanalyzer for library quality and sequenced as a part of a flow cell using Illumina V2 chemistry with 2 × 150-bp paired-end reads in a MiSeq platform. Quality control was performed using FastQC (4), and low-quality reads were removed prior to assembly with Sickle (quality score, 30; minimum length, 50) (5). Reads were *de novo* assembled using SPAdes version 3.9.1, with k-mer lengths of 21, 33, 55, and 77 (6). The resulting contigs (>1,000 bp) were outputted to Pilon version 1.22 to correct single-nucleotide polymorphisms (SNPs) and indels (7). The resulting contigs were then ordered against a closely related complete *L. pneumophila* genome with Mauve aligner (8). Open reading frame identification and genome annotation were performed using NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 4.4 (9). Genome quality, size, and G+C content were estimated using Quast version 4.6.2 (10). For the four assemblies, the genome size and G+C contents varied from 3.43 to 3.59 Mb and 38.2 to 38.3%, respectively (Table 1).

Separately, short reads were assembled using plasmidSPAdes (11) to identify potential plasmids. We identified a plasmid contig in the LG57 strain. For confirmation, a BAM file including all paired-end reads and the SPAdes assembly graphs were submitted to Recycler (12), confirming a 76,262-bp plasmid with high sequence similarity (97%) to a 31.9-kb fragment of the *L. pneumophila* subsp. *pneumophila* strain Lorraine plasmid pLELO (GenBank accession no. FQ958212), which we termed pLG57.

Received 7 March 2018 Accepted 19 March 2018 Published 12 April 2018

**Citation** Fortuna A, Ramnarine R, Li A, Fittipaldi N, Frantz C, Mallo GV. 2018. Draft genome sequences of four clinical *Legionella pneumophila* isolates from Ontario, Canada. Genome Announc 6:e00295-18. <https://doi.org/10.1128/genomeA.00295-18>.

**Copyright** © 2018 Fortuna et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Gustavo V. Mallo, [gustavo.mallo@oahpp.ca](mailto:gustavo.mallo@oahpp.ca).

**TABLE 1** Summary of statistics for *L. pneumophila* draft genome assemblies

Isolate	Type	Yr	Sequence type	No. of contigs	G+C content (%)	Size (bp)	$N_{50}$ (bp)	No. of coding genes	Accession no.
LG57	Chromosome	2005	222	49	38.24	3,590,292	163,641	3,172	PQXA00000000
LG59	Chromosome	2012	37	21	38.3	3,475,634	300,825	3,087	PQWZ00000000
LG61	Chromosome	2000	154	21	38.19	3,429,439	289,473	2,966	PQWY00000000
LG63	Chromosome	2001	1	41	38.27	3,447,361	175,373	3,017	PQWX00000000
pLG57	Plasmid	2005		1	37.36	76,262		81	PQXA01000049

**Accession number(s).** This whole-genome project has been deposited at DDBJ/EMBL/GenBank under the accession numbers listed in Table 1. The versions described in this paper are the first versions.

## ACKNOWLEDGMENT

This work was supported by Public Health Ontario.

## REFERENCES

- Fields BS, Benson RF, Besser RE. 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev* 15:506–526. <https://doi.org/10.1128/CMR.15.3.506-526.2002>.
- Prussin AJ, Schwake DO, Marr LC. 2017. Ten questions concerning the aerosolization and transmission of *Legionella* in the built environment. *Build Environ* 123:684–695. <https://doi.org/10.1016/j.buildenv.2017.06.024>.
- Jolley KA, Maiden MCJ. 2014. Using MLST to study bacterial variation: prospects in the genomic era. *Future Microbiol* 9:623–630. <https://doi.org/10.2217/fmb.14.24>.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. <https://github.com/najoshi/sickle>.
- Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clin-genpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* 20:714–737. <https://doi.org/10.1089/cmb.2013.0084>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44: 6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner P. 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32:3380–3387. <https://doi.org/10.1093/bioinformatics/btw493>.
- Rozov R, Kav AB, Bogumil D, Shterzer N, Halperin E, Mizrahi I, Shamir R. 2017. Recycler: an algorithm for detecting plasmids from *de novo* assembly graphs. *Bioinformatics* 33:475–482. <https://doi.org/10.1093/bioinformatics/btw651>.