AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

# Draft Genome Sequence of *Legionella* Species Isolated from Drinking Water in an Italian Industry

Luna Girolamini,[a] Silvano Salaris,[a] Massimiliano Orsini,[b] Maria Rosaria Pascale,[a] Marta Mazzotta,[a] Antonella Grottola,[c] Sandra Cristino[a]

[a]Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy
[b]Laboratory of Microbial Ecology and Genomics of Microorganisms, Italian Health Authority and Research Organization for Animal Health and Food Safety, Legnaro, Italy
[c]Regional Reference Laboratory for Clinical Diagnosis of Legionellosis, Unit of Molecular Virology and Microbiology, Modena University Hospital, Modena, Italy

Luna Girolamini and Silvano Salaris contributed equally to this work. Author order was determined by drawing straws.

**ABSTRACT** We report the draft genome sequences of an environmental *Legionella* strain isolated from an industrial water distribution system in Italy. Macrophage infectivity potentiator (*mip)* and $\beta$-subunit of RNA polymerase (*rpoB*) genes were used to perform the species identification. Whole-genome sequencing (WGS) and average nucleotide identity (ANI) identified the isolate as belonging to a presumptive novel *Legionella* species, with a genome length of 3,281,851 bp.

*Legionella* spp. are pathogenic Gram-negative bacteria that are ubiquitous in water and soil. *Legionella* includes more than 66 species and some of them are potentially able to cause a severe form of pneumonia, called Legionnaires' disease (1).

The *Legionella* sp. strain 31fl33 was isolated from a drinking water in a company located in the Emilia-Romagna region (Italy) during a routine *Legionella* surveillance program.

Water samples and *Legionella* isolation were performed according to ISO 19458:2006 and ISO 11731:2017, respectively (2, 3). Samples were seeded onto selective medium with glycine-vancomycin-polymyxin B-cycloheximide (GVPC) (Thermo Fisher Diagnostics, Basingstoke, UK) and incubated until 15 days at 35 $\pm$ 2°C in 2.5% $CO_2$. Suspected colonies were subcultured on buffered charcoal yeast extract (BCYE) with and without L-cysteine (Thermo Fisher).

DNA isolation was performed by InstaGene Matrix (Bio-Rad, Hercules, CA, USA) and the identification of isolate was performed by macrophage infectivity potentiator (*mip*) and RNA polymerase $\beta$ subunit (*rpoB*) genes sequencing (4, 5). A BigDye kit was used for the sequencing reaction and DNA sequences were analyzed on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Sequences obtained were analyzed on the BLAST platform by the National Center for Biotechnology Information (NCBI) and European Working Group for *Legionella* Infections (EWGLI) databases. The best match returned was *L. feeleii*, reference strain ATCC 35072 (GenBank accession no. GCA_001648615.1), with similarities of 98.2% and 95.1% for *mip* and *rpoB,* respectively.

An Illumina Nextera XT DNA Library Preparation kit (Illumina, New England Biolabs, Ipswich, MA, USA) was used to perform next-generation sequencing (NGS) library preparation using 100 ng of DNA. Subsequently, the Illumina NextSeq 500 platform (2 $\times$ 250 paired-end reads) was used for the sequencing.

The bioinformatics workflow for the whole-genome sequencing (WGS) analysis included the following steps: through TORMES v.1.2.0 (6), an automated pipeline for analysis of the whole bacterial genome, raw reads were subjected to sequence quality filtering (PRINSEQ v.0.20.4) (7) and *de novo* genome assembly (SPAdes v.13.4.1) (8). The

**Editor** David A. Baltrus, University of Arizona



Address correspondence to Sandra Cristino, sandra.cristino@unibo.it.

The authors declare no conflict of interest.

**Received** 30 November 2021
**Accepted** 25 January 2022
**Published** 10 February 2022

generated contigs were passed to CSAR v.1.1.1 (9) in order to build the scaffolds. Scaffolding was performed using the genomes of different evolutionarily related organisms based on taxonomic identification: *Legionella hackeliae* strain ATCC 35250 (GenBank accession no. LN681225.1) by using Kraken2 v.2.0.9 (10), and *Legionella feeleii* strain ATCC 35072 based on *mip* and *rpoB* identification. The best scaffolding result was obtained for *L. hackeliae*.

To close or reduce the gaps contained in the CSAR output, a remapping of the reads using the scaffolds as a reference sequence was performed with Geneious Prime v.2021.2.2 software (http://www.geneious.com) (11). The obtained draft genome was submitted to GenBank requiring the annotation, performed by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP v.4.3) (12) with the following accession numbers: SRR16560654 and JAJHHJ000000000.

The results, summarized in Table 1, represent the assembling and annotation by the PGAP and the completeness of the genome assembly determined by Benchmarking Universal Single-Copy Orthologs (BUSCO) v.5.0.0 (13).

A DFAST_QC (14) analysis was carried out using FastANI (15) to calculate average nucleotide identity (ANI) for a taxonomic identity of the genome by querying against 13,000 reference genomes from NCBI type strains. FastANI identified the *L. feeleii* WO-44C (ATCC 35072) (GenBank accession no. GCA_900639755.1) as the closest relative strain for our isolate 31fl33, with a similarity of 93.99%. Therefore, we can consider this strain as a new *Legionella* species due to the assumption that two strains belonging to different species show pairwise ANI values below a 95% identity threshold (16).

**Data availability.** The draft genome assembly is available in the GenBank database and can be accessed with SRA and assembly accession numbers SRR16560654 and JAJHHJ000000000.

## ACKNOWLEDGMENTS

This research received no specific grant from any funding agency in the public, commercial, or non-profit sectors.

S.C. is the corresponding author. L.G., S.S., and S.C. conceived and designed the experiments and wrote the paper. L.G., M.R.P., and M.M. performed sample collection and culture experiments. M.O. performed the whole-genome sequencing. A.G. performed gene sequencing. L.G., S.S., and M.O. performed the bioinformatics analysis.

**TABLE 1** Genome statistics data obtained from NCBI and BUSCO quality analyses

| Attribute | Data for strain 31fl33 |
|---|---|
| No. of raw reads | 1,321,792 |
| Avg read length (bp) | 253 |
| Coverage (×) | 99 |
| Total length (bp) | 3,281,851 |
| No. of contigs | 3 |
| GC content (%) | 41.3 |
| $N_{50}$ (bp) | 1,369,339 |
| No. of coding sequences | 2,968 |
| No. of rRNAs | 1 |
| No. of tRNAs | 41 |
| | |
| BUSCO results (% [no. of genes]) | |
| Complete | 100 (124) |
| Single-copy complete | 100 (124) |
| Duplicated complete | 0.0 (0) |
| Fragmented | 0.0 (0) |
| Missing | 0 (0) |
| Total no. of BUSCO genes | 124 |

## REFERENCES


1. Jomehzadeh N, Moosavian M, Saki M, Rashno M. 2019. *Legionella* and legionnaires' disease: an overview. J Acute Dis 8:221–232. https://www.jadweb.org/article.asp?issn=2221-6189;year=2019;volume=8;issue=6;spage=221;epage=232;aulast=Jomehzadeh.
2. International Organization for Standardization (ISO). 2006. ISO 19458:2006: Water quality: sampling for microbiological analysis. International Organization for Standardization, Geneva, Switzerland.
3. International Organization for Standardization (ISO). 2017. ISO 11731:2017: Water quality: enumeration of Legionella. International Organization for Standardization, Geneva, Switzerland.
4. Ratcliff RM, Lanser JA, Manning PA, Heuzenroeder MW. 1998. Sequence-based classification scheme for the genus *Legionella* targeting the *mip* gene. J Clin Microbiol 36:1560–1567. https://doi.org/10.1128/JCM.36.6.1560-1567.1998.
5. Ko KS, Lee HK, Park MY, Lee KH, Yun YJ, Woo SY, Miyamoto H, Kook YH. 2002. Application of RNA polymerase $\beta$-subunit gene (*rpoB*) sequences for the molecular differentiation of *Legionella* species. J Clin Microbiol 40:2653–2658. https://doi.org/10.1128/JCM.40.7.2653-2658.2002.
6. Quijada NM, Rodríguez-Lázaro D, Eiros JM, Hernández M. 2019. TORMES: an automated pipeline for whole bacterial genome analysis. Bioinformatics 35:4207–4212. https://doi.org/10.1093/bioinformatics/btz220.
7. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
9. Chen KT, Liu CL, Huang SH, Shen HT, Shieh YK, Chiu HT, Lu CL. 2018. CSAR: a contig scaffolding tool using algebraic rearrangements. Bioinformatics 34:109–111. https://doi.org/10.1093/bioinformatics/btx543.
10. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:257. https://doi.org/10.1186/s13059-019-1891-0.
11. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199.
12. 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.
13. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol 1962:227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.
14. Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 34:1037–1039. https://doi.org/10.1093/bioinformatics/btx713.
15. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:1–8. https://doi.org/10.1038/s41467-018-07641-9.
16. Kim M, Oh HS, Park SC, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346–351. https://doi.org/10.1099/ijs.0.059774-0.