Draft Genome Sequence of *Leptospira yasudae* Strain BJ3, Isolated from the Soil of an *Ex Situ* Wild Animal Conservation Area

Nadia Aqilla Shamsusah,a ⓘHani Kartini Agustar,a Fairuz Amran,b Rozita Hodc

aDepartment of Earth Sciences and Environmental, Universiti Kebangsaan Malaysia, Bangi, Malaysia
bBacteriology Unit, Institute for Medical Research, Kuala Lumpur, Malaysia
cDepartment of Community Health, Universiti Kebangsaan Malaysia, Cheras, Malaysia

ABSTRACT Previously, a novel *Leptospira* strain (BJ3) was isolated from the soil of an *ex situ* wild animal conservation area in Perak, Malaysia. Molecular identification via whole-genome sequencing confirmed that the strain was *Leptospira yasudae*. Here, we report the draft genome sequence of *L. yasudae* strain BJ3.

Leptospirosis is a zoonotic disease caused by helical bacteria from the genus *Leptospira* (1). This disease has spread throughout the world, especially in the tropical and subtropical regions (2). Leptospirosis can be transmitted either through direct contact with the urine of infected animals or indirectly through contaminated soil/water (3). Previously, a strain of *Leptospira* (BJ3) was isolated from the soil of an *ex situ* wild animal conservation area in Perak, Malaysia (4). Based on the classification of its 16S rRNA gene, the isolate was deemed a novel species named *Leptospira yasudae* and was classified under the pathogenic group. Infection in wildlife might result in the introduction of new serovars to humans and endangered species, hence hampering the success of wildlife conservation. In the present study, we conducted whole-genome sequencing to further characterize the strain. Here, we present the draft genome sequence of *L. yasudae* strain BJ3.

The bacteria were first cultured and maintained in liquid Ellinghausen-McCullough-Johnson-Harris (EMJH) medium at 30°C and observed weekly under a darkfield microscope. To obtain a pure single isolate, one small drop of the EMJH culture was spread onto an EMJH agar plate. After 7 days, a single colony was picked and cultured in fresh liquid EMJH medium. Once the culture reached its maximum density of $\sim 1 \times 10^8$ leptospires/ml, genomic DNA (gDNA) was extracted using a Presto mini gDNA bacteria kit (Geneaid Biotech, Taiwan), based on the manufacturer's protocol. Sequencing was performed at 1st BASE Laboratories, Malaysia. Sequencing libraries were generated using a NEBNext Ultra DNA library prep kit for Illumina (NEB, USA) and sequenced on a HiSeq 4000 system (Illumina, USA) to obtain 2×150-bp reads. Sequence adaptors and reads with low-quality scores were removed using BBDuk in the BBTools package (5). The read quality was then assessed using FastQC v0.11.3 (6). Quality control (QC) reads were assembled *de novo* using SPAdes v3.11.1 (7). The resulting >1,000-bp contigs (contiguous sequences) then underwent a search using BLAST against the National Center for Biotechnology Information (NCBI) nucleotide database (https://blast.ncbi.nlm.nih.gov/). For certain scaffolds with low coverage (<100× coverage) and the best BLAST result for nontarget (non-*Leptospira*) organisms, the contigs were removed because they were likely sequence contaminants. The genome statistics were assessed using QUAST v4.6.0 (8); finally, the contigs were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) as per the GenBank submission (9). The tRNAs and rRNAs were predicted using tRNAscan-SE v1.3.1 (10) and RNAmmer v1.2 (11).

Default parameters were used for all software. To characterize the serogroup/serovar of the isolate, multilocus sequence typing (MLST) analysis scheme 1 (12) was performed and compared with the PubMLST database (https://pubmlst.org/). Moreover, traditional serotyping (microscopic agglutination test) was performed at the Institute for Medical Research (IMR), Malaysia, and the Pasteur Institute, France, against 46 local and global antiserum samples.

The data set comprised 6,343,849 paired-end reads and a raw coverage depth of $213\times$. *De novo* assembly resulted in 4,537,672 bases with 26 contigs and an N_{50} value of 642,216 bp. The *L. yasudae* strain BJ3 draft genome comprises $\sim$4.54 Mbp, with an overall GC content of 45.23%. In total, there were 4,134 coding sequences (CDS), 2 5S rRNAs, 1 16S rRNA, 1 23S rRNA, and 59 pseudogenes. Partial sequencing of the 16S rRNA gene indicated that *L. yasudae* strain BJ3 is closely related to *L. kmetyi* strain Bejo-Iso9. For MLST analysis, *L. yasudae* strain BJ3 was assigned a new sequence type (ST) number (ST262), as no match was found for its serogroup/serovar in the PubMLST database, suggesting that it might contain a new kind of serogroup/serovar. Traditional serotyping also showed negative reactions against the 46 local and global antiserum samples tested, thus supporting our MLST findings. A detailed report of the genomic data analysis, in the context of virulence and unique characteristics, will be provided in future publications. The availability of the *L. yasudae* strain BJ3 genome sequence could yield much information, thus enhancing our understanding of its role as a new local pathogen of humans and wildlife.

Data availability. This whole-genome project has been deposited in the NCBI GenBank database under accession number JACCKC000000000. The BioProject accession number is PRJNA624226, and the BioSample accession number is SAMN14572009. The Illumina HiSeq raw data have been deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP332200.

REFERENCES

1. Dietrich M, Mühldorfer K, Tortosa P, Markotter W. 2015. *Leptospira* and bats: story of an emerging friendship. PLoS Pathog 11:e1005176. https://doi.org/10.1371/journal.ppat.1005176.

2. Jara M, Escobar LE, Rodriges RO, Frias-De-Diego A, Sanhueza J, Machado G. 2019. Spatial distribution and spread potential of sixteen *Leptospira* serovars in a subtropical region of Brazil. Transbound Emerg Dis 66:2482–2495. https://doi.org/10.1111/tbed.13306.

3. Haake DA, Levett PN. 2015. Leptospirosis in humans. Curr Top Microbiol Immunol 387:65–97. https://doi.org/10.1007/978-3-662-45059-8_5.

4. Shamsusah NA, Md Zain BM, Dharmalingam S, Amran F, Agustar H-K. 2021. Pengesanan dan pencirian *Leptospira* spp. pada haiwan liar dan persekitaran di pusat pemuliharaan ex situ. Sains Malays 50:35–43. https://doi.org/10.17576/jsm-2021-5001-04.

5. Joint Genome Institute. BBTools package: BBDuk. https://jgi.doe.gov/data-and-tools/bbtools/.

6. Simon A. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

7. 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.

8. 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.

9. 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.

10. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955.

11. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.

12. Boonsilp S, Thaipadungpanit J, Amornchai P, Wuthiekanun V, Bailey MS, Holden MTG, Zhang C, Jiang X, Koizumi N, Taylor K, Galloway R, Hoffmaster AR, Craig S, Smythe LD, Hartskeerl RA, Day NP, Chantratita N, Feil EJ, Aanensen DM, Spratt BG, Peacock SJ. 2013. A single multilocus sequence typing (MLST) scheme for seven pathogenic Leptospira species. PLoS Negl Trop Dis 7:e1954. https://doi.org/10.1371/journal.pntd.0001954.