Data Article

Genomic data of Leptospira interrogans HP358 isolated from rodent captured from the human leptospirosis suspected areas of Selangor state, Malaysia

Noraini Philip, Jaeyres Jani, Nurul Natasya Azhari, Zamberi Sekawi, Vasantha Kumari Neela

© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

https://doi.org/10.1016/j.dib.2021.107261
Specifications Table

| Subject                        | Medicine and Public Health |
|--------------------------------|---------------------------|
| Specific subject area          | Microbiology, Leptospira  |
| Type of data                   | Analyzed data             |
| How data were acquired         | The genomic data was acquired by whole genome sequencing conducted on Illumina MiSeq instruments platform |
| Data format                    | Raw data and whole genome sequence data analysis. |
| Parameters of data collection  | Genomic DNA was extracted from pure Leptospira culture at room temperature. |
| Description of data collection | Kidney from rodents (Sundamys muelleri) was cultured in EMJH medium and incubated at 28°C for growth of leptospires. It was filtered until a pure Leptospira culture was obtained. |
| Data source location           | The rodent caught from forest in Hulu Perdik, Selangor, Malaysia |
| Data accessibility             | Repository name: Data is publicly available at NCBI GenBank from the following links: https://www.ncbi.nlm.nih.gov/nuccore/JAFCYY0 https://www.ncbi.nlm.nih.gov/biosample/SAMN16925409 |

Value of the Data

- WGS of Leptospira interrogans HP358 reported in Malaysia, the whole genome sequence of this strain could provide fundamental knowledge and insight towards understanding its virulence and microbial properties.
- The data, an important source towards understanding the relationship among Leptospira interrogans strains from Malaysia and other regions will benefit researchers, public health, health sector, veterinary and reference laboratories.
- The data may provide valuable information in understanding the epidemiology and diagnostics of leptospirosis, defining the virulence properties, thereby could assist in developing informed policy in the design and implementation of leptospirosis control programme.

1. Data Description

The data reported in this article is the genomic data of whole genome sequencing (WGS) of Leptospira interrogans HP358 (L. interrogans HP358) with ST number ST238, a new ST number recorded in Leptospira MLST data base [1]. L. interrogans HP358 was isolated from the kidney of rodents, Sundamys muelleri (S. muelleri), captured from the human leptospirosis suspected area, in forest environment, Hulu Perdik, Selangor, Malaysia. S. muelleri is a species of rodent commonly found in Malaysia, Indonesia, Myanmar, Thailand and Philippines. The isolate was cultured in liquid Ellinghausen-McCullough-Johnson-Harris (EMJH) medium supplemented with Leptospira enrichment medium, rabbit and calf serum and 5-fluorouracil. The genome size of the L. interrogans HP358 contained a length of 4,808,724 nucleotides with G+C content of 35.01%, 229 subsystems, 5236 coding sequences and 39 RNAs.

2. Experimental Design, Materials and Methods

2.1. Bacterial culture and DNA extraction

The genomic DNA of L. interrogans HP358 was extracted using DNeasy Blood & Tissue kits (Qiagen, German) according to the manufacturer’s instruction. The DNA library for WGS was prepared using the IGP-NGSP Illumina Library and subsequently sequenced using the MiSeq instruments platform.
2.2. Whole genome sequencing and bioinformatic analysis

Approximately 99% of the genome was sequenced using 177 × sequencing coverage and generated a total of 2,902,368 paired reads of a 151-bp paired-end (Illumina, San Diego, CA). The sequencing data was deposited in the Sequence Read Archive (SRA) (Bio-sample accession number of SAMN16925409) under the bio-project accession number PRJNA680866. The quality of the raw sequences read was assessed using FastQC and then pre-processed using BBMap version 38.43 tool [2]. The adapters were trimmed and sequences read with less than 50bp were removed based on the phred quality score (below Q30) using BBduk.sh [2]. The quality-filtered sequences were de novo assembled using SPAdes version 3.11.1 [3]. The de novo assembly of the genome generated 118 contigs with N50 of 133,176bp. The whole genome size of L. interrogans HP358 was determined with a length of 4,808,724 bp and G + C content of 35.01%. The draft genome of L. interrogans HP358 (Genome accession number of JAFYY000000000) was annotated by the Rapid Annotation of microbial genomes using Subsystem Technology (RAST) [4] and showed 229 subsystems, 5 236 coding sequences and 39 RNAs.

The availability of this genome sequence would enable the comparative genomics study and understanding the phylogenetic relationship of this isolate with other Leptospira strains. The sequence data may play important role in understanding the key difference in the virulence level of different Leptospira isolates and subsequently increase the knowledge that determines the different clinical manifestation of leptospirosis. Therefore, the availability of this genome of local isolates provides genetic and epidemiological information about prevailing Leptospira species in Malaysia.

Ethics Statement

This research work was supported by the Ministry of Education, Malaysia through the Fundamental Research Grant Scheme (FRGS/1/2019/WAB13/UPM/02/2).

CRediT Author Statement

Noraini Philip: Conceptualization, Methodology, Software, Data curation, Investigation, Writing – original draft; Jaeyres Jani: Conceptualization, Methodology, Software, Data curation, Writing – review & editing; Nurul Natasya Azhari: Data curation, Investigation; Zamberi Sekawi: Conceptualization, Funding, Supervision; Vasantha Kumari Neela: Conceptualization, Funding, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Acknowledgments

This research was financially supported by the grant from Ministry of Higher Education and Universiti Putra Malaysia through Long Term Research Grant Scheme (LRGS Phase 2/2014, UPM/700-2/7/LRGS/5526400). We take this opportunity to thank all members of the project, for their insights and critical appraisal of this manuscript during its development.
References

[1] N.N. Azhari, S.N.A. Ramli, N. Joseph, N. Philip, N.F. Mustapha, S.N. Ishak, F.S. Mohd-Taib, S. Md Nor, M.A. Yusof, S.A. Mohd Sah, M.N.B. Mohd Desa, B. Garba, C.G. Zeppelini, F. Costa, Z. Sekawi, V. Neela, Molecular characterization of pathogenic *Leptospira* sp. in small mammals captured from the human leptospirosis suspected areas of Selangor state, Malaysia, *Acta Trop.* 188 (2018) 68–77 PMID: 30145261, doi:10.1016/j.actatropica.2018.08.020.

[2] B. Bushnell, J. Rood, E. Singer, BBMerge-accurate paired shotgun read merging via overlap, *PLoS One* 12 (2017) 1e15, doi:10.1371/journal.pone.0185056.

[3] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455–477, doi:10.1089/cmb.2012.0021.

[4] R. Overbeek, R. Olson, G.D. Pusch, G.J. Olsen, J.J. Davis, T. Disz, R.A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A.R. Wattam, F. Fia, R. Stevens, The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST), *Nucleic Acids Res.* 42 (Database issue) (2014) D206–D214.