



Complete Genome Sequence of *Spiroplasma alleghenense* PLHS-1^T (ATCC 51752), a Bacterium Isolated from Scorpion Fly (*Panorpa helena*)

Lin Chou (周林),^a Ting-Yi Lee (李庭儀),^a Yi-Ming Tsai (蔡乙鳴),^a  Chih-Horng Kuo (郭志鴻)^a

^aInstitute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

ABSTRACT *Spiroplasma alleghenense* PLHS-1^T (ATCC 51752) was isolated from the gut of a scorpion fly (*Panorpa helena*) collected in West Virginia. The complete genome sequence of this bacterium, which consists of a single 1,336,077-bp circular chromosome, is reported in this work.

Spiroplasma alleghenense is the representative of group XXVI within this genus of arthropod-associated bacteria (1). Strains belonging to this group were all isolated from common scorpion flies (*Panorpa helena*) collected in the Allegheny Mountain region (West Virginia). This bacterium could be observed in the blood and gut fluids of adult flies. However, no apparent pathogenicity to its insect host has been reported. To facilitate future research on this bacterium, and to improve the availability of *Spiroplasma* sp. genomes for comparative analysis (2), we determined the complete genome sequence of the type strain, PLHS-1. This strain was triply cloned from the primary culture in M1D liquid medium (1).

The procedures for sample preparation and genome sequencing were based on those described in our previous studies (3–8). Default parameters were used for all software unless otherwise specified. The strain was acquired from the American Type Culture Collection (catalog number ATCC 51752) and grown in M1D liquid medium prior to DNA extraction using the Wizard Genomic DNA purification kit (Promega, USA). The 16S rRNA gene was amplified by PCR using the 8F/1492R primer set, followed by Sanger sequencing to verify that it matched the reference record (GenBank accession number [AY189125](https://doi.org/10.1093/nar/23.11.2189)). The shotgun sequencing utilized two Illumina libraries. The paired-end library was prepared using the Kapa library preparation kit (Roche, USA) and sequenced using an Illumina HiSeq 2000 instrument (211 + 126 bp; ~1.7 million pairs; ~400-fold coverage). The mate pair library was prepared using the Nextera mate pair sample preparation kit (Illumina, USA) and sequenced using an Illumina MiSeq instrument (2 × 300 bp; ~1.5 million pairs; ~600-fold coverage). The raw reads were trimmed from the 5' end at the first position with a quality score of <20. *De novo* assembly using ALLPATHS-LG v52188 (9) produced 10 contigs, which were arranged into one scaffold based on mate pair reads. We utilized an iterative process to improve the assembly until the complete genome sequence was obtained. In each iteration, all reads were mapped using BWA v0.7.17 (10). The results were programmatically checked using SAMtools v1.9 (11) and visually inspected using Integrative Genomics Viewer (IGV) v2.1.5 (12). The gaps were filled based on overhanging reads, and polymorphic sites were resolved based on the majority rule. Gene prediction was based on RNAmmer v1.2 (13), tRNAscan-SE v1.3.1 (14), and Prodigal v2.6.3 (15). Homologous genes in other *Spiroplasma* sp. genomes available in GenBank (16) were identified using OrthoMCL v1 (17) and used as the reference for annotation. Manual curation was based on the information obtained from BlastKOALA v2.1 (18) and BLASTP v2.7.1 (19) searches against the NCBI nonredundant database (16). Putative clustered regularly interspaced

Citation Chou L, Lee T-Y, Tsai Y-M, Kuo C-H. 2019. Complete genome sequence of *Spiroplasma alleghenense* PLHS-1^T (ATCC 51752), a bacterium isolated from scorpion fly (*Panorpa helena*). *Microbiol Resour Announc* 8:e00317-19. <https://doi.org/10.1128/MRA.00317-19>.

Editor David Rasko, University of Maryland School of Medicine

Copyright © 2019 Chou 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 Chih-Horng Kuo, chk@gate.sinica.edu.tw.

Received 18 March 2019

Accepted 3 April 2019

Published 25 April 2019

short palindromic repeats (CRISPRs) were identified using CRISPRFinder v2017-05-09 (20).

The complete genome sequence of *Spiroplasma alleghenense* PLHS-1^T consists of one circular chromosome that is 1,336,077 bp with 29.5% G+C content; no plasmid was found. The annotation includes two sets of 16S-23S-5S rRNA genes, 30 tRNA genes (covering all 20 amino acids), 1,146 protein-coding genes, 1 pseudogene, and 1 CRISPR locus (chromosomal positions 1065812 to 1069081) containing 49 spacers.

Data availability. The Illumina raw reads have been deposited at the NCBI Sequence Read Archive under the accession number [PRJNA205100](https://www.ncbi.nlm.nih.gov/sra/PRJNA205100). The genome sequence reported in this work has been deposited at GenBank/ENA/DDBJ under the accession number [CP031376](https://www.ncbi.nlm.nih.gov/genbank/CP031376).

ACKNOWLEDGMENTS

We thank Tean-Hsu Chang, Ling-Ling Chen, and Wen-Sui Lo for technical assistance. The service for Sanger sequencing and Illumina sequencing library preparation was provided by the Genomic Technology Core Facility of our institute. The Illumina paired-end sequencing service was provided by Yougene Bioscience (New Taipei City, Taiwan). The Illumina mate pair sequencing service was provided by the Genomics Core Facility (Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan).

Funding for this project was provided by the Institute of Plant and Microbial Biology at the Academia Sinica and the Ministry of Science and Technology of Taiwan (grants NSC 101-2621-B-001-004-MY3 and MOST 104-2311-B-001-019) to C.-H.K. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Adams JR, Whitcomb RF, Tully JG, Clark EA, Rose DL, Carle P, Konai M, Bove JM, Henegar RB, Williamson DL. 1997. *Spiroplasma alleghenense* sp. nov., a new species from the scorpion fly *Panorpa helena* (Mecoptera: Panorpididae). *Int J Evol Microbiol* 47:759–762. <https://doi.org/10.1099/00207713-47-3-759>.
- Lo W-S, Huang Y-Y, Kuo C-H. 2016. Winding paths to simplicity: genome evolution in facultative insect symbionts. *FEMS Microbiol Rev* 40: 855–874. <https://doi.org/10.1093/femsre/fuw028>.
- Lo W-S, Chen L-L, Chung W-C, Gasparich GE, Kuo C-H. 2013. Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genomics* 14:22. <https://doi.org/10.1186/1471-2164-14-22>.
- Ku C, Lo W-S, Chen L-L, Kuo C-H. 2013. Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in *Spiroplasma*. *Genome Biol Evol* 5:1151–1164. <https://doi.org/10.1093/gbe/evt084>.
- Ku C, Lo W-S, Chen L-L, Kuo C-H. 2014. Complete genome sequence of *Spiroplasma apis* B31^T (ATCC 33834), a bacterium associated with May disease of honeybees (*Apis mellifera*). *Genome Announc* 2:e01151-13. <https://doi.org/10.1128/genomeA.01151-13>.
- Lo W-S, Lai Y-C, Lien Y-W, Wang T-H, Kuo C-H. 2015. Complete genome sequence of *Spiroplasma litorale* TN-1^T (DSM 21781), a bacterium isolated from a green-eyed horsefly (*Tabanus nigrovittatus*). *Genome Announc* 3:e01116-15. <https://doi.org/10.1128/genomeA.01116-15>.
- Shen W-Y, Lo W-S, Lai Y-C, Kuo C-H. 2016. Complete genome sequence of *Spiroplasma helicoides* TABS-2^T (DSM 22551), a bacterium isolated from a horsefly (*Tabanus abactor*). *Genome Announc* 4:e01201-16. <https://doi.org/10.1128/genomeA.01201-16>.
- Tsai Y-M, Wu P-S, Lo W-S, Kuo C-H. 2018. Complete genome sequence of *Spiroplasma floricola* 23-6^T (ATCC 29989), a bacterium isolated from a tulip tree (*Liriodendron tulipifera* L.). *Genome Announc* 6:e00302-18. <https://doi.org/10.1128/genomeA.00302-18>.
- Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* 108:1513–1518. <https://doi.org/10.1073/pnas.1017351108>.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative Genomics Viewer. *Nat Biotechnol* 29: 24–26. <https://doi.org/10.1038/nbt.1754>.
- Lagesen K, Hallin P, Rødland EA, Stærfeldt 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>.
- Lowe T, Eddy S. 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>.
- Hyatt D, Chen G-L, LoCascio P, Land M, Larimer F, Hauser L. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW. 2018. GenBank. *Nucleic Acids Res* 46:D41–D47. <https://doi.org/10.1093/nar/gkx1094>.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <https://doi.org/10.1101/gr.1224503>.
- Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden T. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
- Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57. <https://doi.org/10.1093/nar/gkm360>.