

Complete Genome Sequence of *Spiroplasma helicoides* TABS-2^T (DSM 22551), a Bacterium Isolated from a Horsefly (*Tabanus abactor*)

Wei-Yi Shen (沈威邑)^a  Wen-Sui Lo (羅文穗)^{a,b,c} Yi-Ching Lai (賴怡靜)^a  Chih-Horng Kuo (郭志鴻)^{a,b,d}

Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan^a; Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, National Chung Hsing University and Academia Sinica, Taipei, Taiwan^b; Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan^c; Biotechnology Center, National Chung Hsing University, Taichung, Taiwan^d

***Spiroplasma helicoides* TABS-2^T (DSM 22551) was isolated from the gut of a horsefly (*Tabanus abactor*) collected near Ardmore, Oklahoma, USA, in 1987. Here, we report the complete genome sequence of this bacterium to facilitate the investigation of its biology and the comparative genomics among *Spiroplasma* species.**

Received 1 September 2016 Accepted 2 September 2016 Published 20 October 2016

Citation 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(5):e01201-16. doi:10.1128/genomeA.01201-16.

Copyright © 2016 Shen 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.

Spiroplasma helicoides is a horsefly-associated bacterium found in North America (1). The type strain TABS-2^T was isolated from the gut of a horsefly (*Tabanus abactor*) collected near Ardmore, Oklahoma, USA, in 1987 and was assigned to group XXXII within the genus. Biochemical characterization of this strain demonstrated that it is capable of catabolizing glucose but does not hydrolyze arginine or urea (1). To facilitate future investigation on the biology of this bacterium, as well as to improve the taxon sampling of available *Spiroplasma* sequences for comparative and evolutionary studies (2), we determined the complete genome sequence of *S. helicoides* TABS-2^T.

The strain was acquired from the German Collection of Microorganisms and Cell Cultures (catalogue number DSM 22551). The freeze-dried sample was processed according to the manufacturer's instructions and cultured in MID medium (3) prior to DNA extraction using the Wizard Genomic DNA purification kit (Promega, USA). PCR and Sanger sequencing was performed to verify that the 16S rRNA gene sequence matched the reference record (GenBank accession no. NR_025704.1).

The procedures for sequencing, assembly, and annotation were based on those described in our previous studies on *Spiroplasma* genomes (4–13). Briefly, the Illumina MiSeq platform was used to generate 300-bp reads from one paired-end library (~550-bp insert, 2,327,640 reads) and one mate-pair library (~4,200-bp insert, 2,828,554 reads). The initial *de novo* assembly was performed using ALLPATHS-LG release 52188 (14). Subsequently, PAGIT version 1 (15) was used to assist an iterative process for improving the assembly. For each iteration, the raw reads were mapped to the assembly using BWA version 0.7.12 (16), programmatically checked using the MPILEUP program in SAMtools package version 1.2 (17), and visually inspected using IGV version 2.3.57 (18). Polymorphic sites and gaps were corrected based on the mapped reads. The process was repeated until the complete genome sequence was obtained. The programs RNAmmer (19), tRNAscan-SE (20), and Prodigal (21) were used for gene prediction. The gene names and product descrip-

tions were first annotated based on the homologous genes in other *Spiroplasma* genomes (4–13) as identified by OrthoMCL (22). Subsequent manual curation was based on BLASTp (23) searches against the NCBI nonredundant database (24) and the KEGG database (25, 26). Putative clustered regularly interspaced short palindromic repeats (CRISPRs) were identified using CRISPRFinder (27).

The circular chromosome of *S. helicoides* TABS-2^T is 1,326,546 bp in size and has a G+C content of 26.8%; no plasmid was found. The first version of annotation includes one set of 16S-23S-5S rRNA genes, 29 tRNA genes (covering all 20 amino acids), 1,151 protein-coding genes, four pseudogenes, and one CRISPR locus (chromosomal positions 1,068,189 to 1,071,129; containing 44 spacers).

Accession number(s). The complete genome sequence of *S. helicoides* TABS-2^T has been deposited at DDBJ/EMBL/GenBank under the accession number [CP017015](https://www.ncbi.nlm.nih.gov/nuccore/CP017015).

ACKNOWLEDGMENTS

The bacterial strain was imported under permit number 103-B-001 issued by the Council of Agriculture of Taiwan. The Sanger sequencing service was provided by the DNA Analysis Core Laboratory (Institute of Plant and Microbial Biology, Academia Sinica). The Illumina sequencing library preparation service was provided by the DNA Microarray Core Laboratory (Institute of Plant and Microbial Biology, Academia Sinica). The Illumina MiSeq sequencing service was provided by the DNA Sequencing Core Facility (Institute of Molecular Biology, Academia Sinica).

FUNDING INFORMATION

This work, including the efforts of Chih-Horng Kuo, was funded by the Institute of Plant and Microbial Biology, Academia Sinica, Taiwan. This work, including the efforts of Chih-Horng Kuo, was funded by the Ministry of Science and Technology, Taiwan (NSC 101-2621-B-001-004-MY3 and MOST 104-2311-B-001-019). This work, including the efforts of Wei-Yi Shen and Yi-Ching Lai, was funded by the Summer Undergraduate Internship Program, Institute of Plant and Microbial Biology, Academia Sinica, Taiwan. This work, including the efforts of Wen-Sui Lo, was

funded by the Taiwan International Graduate Program in Molecular and Biological Agricultural Sciences, Academia Sinica and National Chung Hsing University, Taiwan. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Whitcomb RF, French FE, Tully JG, Gasparich GE, Rose DL, Carle P, Bove JM, Henegar RB, Konai M, Hackett KJ, Adams JR, Clark TB, Williamson DL. 1997. *Spiroplasma chrysopicola* sp. nov., *Spiroplasma gladiatoris* sp. nov., *Spiroplasma helicoides* sp. nov., and *Spiroplasma tabanidicola* sp. nov., from tabanid (Diptera: Tabanidae) flies. *Int J Syst Bacteriol* 47:713–719. <http://dx.doi.org/10.1099/00207713-47-3-713>.
- Lo W-S, Huang Y-Y, Kuo C-H. 2016. Winding paths to simplicity: genome evolution in facultative insect symbionts. *FEMS Microbiol Rev* [Epub ahead of print.] <http://dx.doi.org/10.1093/femsre/fuw028>.
- Whitcomb RF, Tully JG, McCawley P, Rose DL. 1982. Application of the growth inhibition test to *Spiroplasma* taxonomy. *Int J Syst Bacteriol* 32:387–394. <http://dx.doi.org/10.1099/00207713-32-4-387>.
- 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. <http://dx.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. <http://dx.doi.org/10.1093/gbe/evt084>.
- Lo W-S, Ku C, Chen L-L, Chang T-H, Kuo C-H. 2013. Comparison of metabolic capacities and inference of gene content evolution in mosquito-associated *Spiroplasma diminutum* and *S. taiwanense*. *Genome Biol Evol* 5:1512–1523. <http://dx.doi.org/10.1093/gbe/evt108>.
- 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(1):e01151-13. <http://dx.doi.org/10.1128/genomeA.01151-13>.
- Chang T-H, Lo W-S, Ku C, Chen L-L, Kuo C-H. 2014. Molecular evolution of the substrate utilization strategies and putative virulence factors in mosquito-associated *Spiroplasma* species. *Genome Biol Evol* 6:500–509. <http://dx.doi.org/10.1093/gbe/evu033>.
- Paredes JC, Herren JK, Schüpfer F, Marin R, Claverol S, Kuo C-H, Lemaitre B, Béven L. 2015. Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. *mBio* 6:e02437-14. <http://dx.doi.org/10.1128/mBio.02437-14>.
- Lo W-S, Gasparich GE, Kuo C-H. 2015. Found and lost: the fates of horizontally acquired genes in arthropod-symbiotic *Spiroplasma*. *Genome Biol Evol* 7:2458–2472. <http://dx.doi.org/10.1093/gbe/evv160>.
- 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(5):e01116-15. <http://dx.doi.org/10.1128/genomeA.01116-15>.
- Lo W-S, Liu P-Y, Kuo C-H. 2015. Complete genome sequence of *Spiroplasma cantharicola* CC-1^T (DSM 21588), a bacterium isolated from soldier beetle (*Cantharis carolinus*). *Genome Announc* 3(5):e01253-15. <http://dx.doi.org/10.1128/genomeA.01253-15>.
- Lo W-S, Gasparich GE, Kuo C-H. 2016. Complete genome sequence of *Spiroplasma turonicum* Tab4c^T, a bacterium isolated from horse flies (*Haematopota* sp.). *Genome Announc* 4(5):e01010-16. <http://dx.doi.org/10.1128/genomeA.01010-16>.
- 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. <http://dx.doi.org/10.1073/pnas.1017351108>.
- Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc* 7:1260–1284. <http://dx.doi.org/10.1038/nprot.2012.068>.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. <http://dx.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. <http://dx.doi.org/10.1093/bioinformatics/btp352>.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29:24–26. <http://dx.doi.org/10.1038/nbt.1754>.
- 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. <http://dx.doi.org/10.1093/nar/gkm160>.
- 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. <http://dx.doi.org/10.1093/nar/25.5.0955>.
- Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
- Li L, Stoekert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <http://dx.doi.org/10.1101/gr.1224503>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <http://dx.doi.org/10.1186/1471-2105-10-421>.
- Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. *Nucleic Acids Res* 44:D67–D72. <http://dx.doi.org/10.1093/nar/gkv1276>.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30. <http://dx.doi.org/10.1093/nar/28.1.27>.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38:D355–D360. <http://dx.doi.org/10.1093/nar/gkp896>.
- 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. <http://dx.doi.org/10.1093/nar/gkm360>.