Complete Genome Sequence of *Spiroplasma litorale* TN-1^T (DSM 21781), a Bacterium Isolated from a Green-Eyed Horsefly (*Tabanus nigrovittatus*)

Wen-Sui Lo,[a,b,c] Yi-Ching Lai,[a] Yun-Wei Lien,[a] Tzu-Haw Wang,[a] Chih-Horng Kuo[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 litorale TN-1^T (DSM 21781) was isolated from the gut of a green-eyed horsefly (*Tabanus nigrovittatus*), collected at Ocracoke Island in North Carolina in 1983. Here, we report the complete genome sequence of this bacterium to facilitate the investigation of its biology.

Address correspondence to Chih-Horng Kuo, chk@gate.sinica.edu.tw.

Spiroplasma litorale is a bacterium known to be associated with tabanid flies belonging to the genus *Tabanus* (1). Strains assigned to this species have been isolated from *Tabanus* species *T. americanus*, *T. atratus*, *T. gladiator*, *T. lineola*, *T. nigrovittatus*, *T. sulcifrons*, and *T. zythicolor*. The restriction of *S. litorale* to the coastal areas of the southeastern United States makes it one of the best-known examples of a geographically limited insect symbiont within the genus *Spiroplasma* (1). To facilitate the future investigation on the biology of this bacterium and to improve the taxon sampling of available *Spiroplasma* sequences for comparative genomics and evolutionary studies, we determined the complete genome sequence of *S. litorale* TN-1^T.

The strain was acquired from the German Collection of Microorganisms and Cell Cultures (catalogue no. DSM 21781). The freeze-dried sample was processed according to the manufacturer's instructions and cultured in the M1D medium (2) prior to DNA extraction using the Wizard genomic DNA purification kit (Promega, USA). PCR and Sanger sequencing were performed to verify that the 16S rRNA gene sequence matched the reference record (GenBank accession no. NR_025708.1).

The procedures for sequencing, assembly, and annotation were based on those described in our previous studies on *Spiroplasma* genomes (3–8). Briefly, the Illumina MiSeq platform was used to generate 300-bp reads from one paired-end library (~570-bp insert, 2,155,514 reads) and one mate-pair library (~4,000-bp insert, 2,024,260 reads). The initial *de novo* assembly was performed using ALLPATHS-LG release 52188 (9). Subsequently, PAGIT version 1 (10) was used to assist an iterative process for improving the assembly. For each iteration, the raw reads were mapped to the assembly using Burrows-Wheeler Aligner (BWA) version 0.7.12 (11), programmatically checked using the MPILEUP program in the SAMTools package version 1.2 (12), and visually inspected using IGV version 2.3.57 (13). 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 (14), tRNAscan-SE (15), and Prodigal (16) were used for gene prediction. The gene names and product descriptions were first annotated based on the homologous genes in other *Spiroplasma* genomes (3–8), as identified by OrthoMCL (17). Subsequent manual curation was based on BLASTp (18) searches against the NCBI nonredundant database (19) and the KEGG database (20, 21).

The circular chromosome of *S. litorale* TN-1^T is 1,225,519 bp in size and has a G+C content of 24.9%; 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), and 1,064 protein-coding genes.

Nucleotide sequence accession number. The complete genome sequence of *S. litorale* TN-1^T has been deposited at DDBJ/EMBL/GenBank under the accession no. CP012357.

REFERENCES

1. **Konai M, Whitcomb RF, French FE, Tully JG, Rose DL, Carle P, Bové JM, Hackett KJ, Henegar RB, Clark TB, Williamson DL.** 1997. *Spiroplasma litorale* sp. nov., from tabanid flies (Tabanidae: Diptera) in the southeastern United States. Int J Syst Bacteriol **47:**359–362. http://dx.doi.org/10.1099/00207713-47-2-359.

2. **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.

3. **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.

4. **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.

5. **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.

6. **Ku C, Lo W-S, Chen L-L, Kuo C-H.** 2014. Complete genome sequence of *Spiroplasma apis* B31ᵀ (ATCC 33834), a bacterium associated with May disease of honeybees (*Apis mellifera*). Genome Announc **2(1):**. http://dx.doi.org/10.1128/genomeA.01151-13.

7. **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.

8. **Lo W, Gasparich GE, Kuo C.** 2015. Found and lost: the fates of horizontally acquired genes in arthropod-symbiotic *Spiroplasma*. Genome Biol Evol, in press. http://dx.doi.org/10.1093/gbe/evv160.

9. **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 USA **108:**1513–1518. http://dx.doi.org/10.1073/pnas.1017351108.

10. **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.

11. **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.

12. **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 of Map Format and SAMtools. Bioinformatics **25:**2078–2079.

13. **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.

14. **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.

15. **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.

16. **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.

17. **Li L, Stoeckert 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.

18. **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.

19. **Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.** 2015. GenBank. Nucleic Acids Res **43:**D30–D35. http://dx.doi.org/10.1093/nar/gku1216.

20. **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.

21. **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.