



Complete Genome Sequence of *Spiroplasma* sp. NBRC 100390

Mindia Haryono,^a  Wen-Sui Lo (羅文穗)^{a,b,c} Gail E. Gasparich,^{d,e}
 Chih-Horng Kuo (郭志鴻)^{a,b,f}

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; Department of Biological Sciences, Towson University, Towson, Maryland, USA^d; College of Arts and Sciences, Salem State University, Salem, Massachusetts, USA^e; Biotechnology Centre, National Chung Hsing University, Taichung, Taiwan^f

ABSTRACT *Spiroplasma* sp. NBRC 100390 was initially described as a duplicate of *S. atrichopogonis* GNAT3597^T (=ATCC BAA-520^T) but later found to be different in the 16S rDNA sequences. Here, we report the complete genome sequence of this bacterium to establish its identity and to facilitate future investigation.

The bacterial strain *Spiroplasma* sp. NBRC 100390 deposited in the National Institute of Technology and Evaluation–Biological Resource Center (NBRC) was initially described as a duplicate of *S. atrichopogonis* GNAT3597^T (1). However, subsequent investigations revealed that the 16S rDNA sequence of this strain (GenBank no. AB681165.1) shows only 96% nucleotide sequence identity to that of *S. atrichopogonis* GNAT3597^T maintained in Gail Gasparich's laboratory at Towson University (GenBank no. KR349130.1) and another duplicate deposited in the American Type Culture Collection (ATCC BAA-520^T; GenBank no. KR349131.1) (2). To establish the identity of this strain and to facilitate future comparative genomics of *Spiroplasma* species (3, 4), we determined the complete genome sequence of *Spiroplasma* sp. NBRC 100390.

The procedures for sequencing, assembly, and annotation were based on our previous studies of *Spiroplasma* genomes (2, 5–7). We utilized the Illumina MiSeq platform to obtain 301-bp sequencing reads from one paired-end library with approximately 300-fold coverage. Our preliminary analysis revealed that the sequences are almost identical to *Spiroplasma* sp. TU-14 (7). Thus, we chose a resequencing approach using the *Spiroplasma* sp. TU-14 genome (GenBank no. CP017658.1) as the reference. The raw reads were mapped to the reference using the Burrows–Wheeler alignment (BWA) tool version 0.7.12 (8), programmatically checked using the MPileUP program in SAMtools package version 1.2 (9), and visually inspected using the Integrative Genomics Viewer (IGV) version 2.3.67 (10). The programs RNAmmer (11), tRNAscan-SE (12), and Prodigal (13) were used for gene prediction. Additionally, the presence of putative clustered regularly interspaced short palindromic repeats (CRISPRs) was checked using CRISPRFinder (14).

A total of 18 polymorphic sites were found between *Spiroplasma* sp. NBRC 100390 and the reference, including 11 1-bp deletions, four 1-bp insertions, two single-nucleotide polymorphisms (SNPs), and one 12-bp deletion. All polymorphic sites were manually inspected using Artemis (15) to check if the polymorphisms affected genic regions. We found that all 15 1-bp indels were located in homopolymeric intergenic regions, while the two SNPs both resulted in a nonsynonymous substitution in the predicted functional domain of the respective genes (locus tags S100390_v1c06730 and S100390_v1c09590). However, the BLASTp (16) searches against the NCBI NR database (17) showed that these polymorphisms did not affect the protein domain

Received 3 January 2017 Accepted 5 January 2017 Published 9 March 2017

Citation Haryono M, Lo W-S, Gasparich GE, Kuo C-H. 2017. Complete genome sequence of *Spiroplasma* sp. NBRC 100390. Genome Announc 5:e00008-17. <https://doi.org/10.1128/genomeA.00008-17>.

Copyright © 2017 Haryono 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.

identification. Finally, the 12-bp deletion was located outside of the predicted functional domain (locus tag S100390_v1c07060). In conclusion, the gene content of *Spiroplasma* sp. NBRC 100390 is identical to that of *Spiroplasma* sp. TU-14 (7), and we annotated the gene names and product descriptions entirely based on the reference genome.

The circular chromosome of *Spiroplasma* sp. NBRC 100390 is 1,199,621 bp in size and has a G+C content of 28.7%; no plasmid was found. The first version of annotation includes one set of 16S-23S-5S rRNA genes, 32 tRNA genes (covering all 20 amino acids), 1,036 protein-coding genes, and four pseudogenes. No putative plectrovirus-related sequence or CRISPR element was found.

Accession number(s). The complete genome sequence of *Spiroplasma* sp. NBRC 100390 has been deposited at DDBJ/EMBL/GenBank under accession number [CP018022](https://www.ncbi.nlm.nih.gov/nuccore/CP018022).

ACKNOWLEDGMENTS

The bacterial strain was provided by the Biological Resource Center, National Institute of Technology and Evaluation (Chiba, Japan). The Sanger sequencing service was provided by the DNA Analysis Core Laboratory, Institute of Plant and Microbial Biology, Academia Sinica (Taipei, Taiwan). The Illumina sequencing library preparation service was provided by the DNA Microarray Core Laboratory, Institute of Plant and Microbial Biology, Academia Sinica (Taipei, Taiwan). The Illumina MiSeq sequencing service was provided by the DNA Sequencing Core Facility, Institute of Molecular Biology, Academia Sinica (Taipei, Taiwan).

The funding for this project was provided by the Institute of Plant and Microbial Biology at Academia Sinica and the Ministry of Science and Technology of Taiwan (NSC 101-2621-B-001-004-MY3 and MOST 104-2311-B-001-019) to C.-H.K. Also, W.-S.L. was supported by the Taiwan International Graduate Program (TIGP) in Molecular and Biological Agricultural Sciences (MBAS). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Koerber RT, Gasparich GE, Frana MF, Grogan WL. 2005. *Spiroplasma atrichopogonis* sp. nov., from a ceratopogonid biting midge. *Int J Syst Evol Microbiol* 55:289–292. <https://doi.org/10.1099/ijs.0.02465-0>.
- Lo WS, Gasparich GE, Kuo CH. 2015. Found and lost: the fates of horizontally acquired genes in arthropod-symbiotic *Spiroplasma*. *Genome Biol Evol* 7:2458–2472. <https://doi.org/10.1093/gbe/evv160>.
- Gasparich GE, Whitcomb RF, Dodge D, French FE, Glass J, Williamson DL. 2004. The genus *Spiroplasma* and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the *Mycoplasma mycoides* clade. *Int J Syst Evol Microbiol* 54:893–918. <https://doi.org/10.1099/ijs.0.02688-0>.
- Lo WS, Huang YY, Kuo CH. 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 WS, Chen LL, Chung WC, Gasparich GE, Kuo CH. 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>.
- Lo WS, Gasparich GE, Kuo CH. 2016. Complete genome sequence of *Spiroplasma turonicum* Tab4C^t, a bacterium isolated from horse flies (*Haematopota* sp.). *Genome Announc* 4(5):e01010-16. <https://doi.org/10.1128/genomeA.01010-16>.
- Lo W-S, Haryono M, Gasparich GE, Kuo C-H. 2017. Complete genome sequence of *Spiroplasma* sp. TU-14. *Genome Announc* 5(2):e01465-16. <https://doi.org/10.1128/genomeA.01465-16>.
- 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, Thorvaldsdóttir 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, 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>.
- 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>.
- Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
- 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>.
- Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualisation and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28:464–469. <https://doi.org/10.1093/bioinformatics/btr703>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://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. <https://doi.org/10.1093/nar/gkv1276>.