genomeA
Journals.ASM.org

Complete Genome Sequence of *Spiroplasma cantharicola* CC-1^T (DSM 21588), a Bacterium Isolated from Soldier Beetle (*Cantharis carolinus*)

Wen-Sui Lo (羅文穗),[a,b,c] Po-Yu Liu (劉勃佑),[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]; Genome and Systems Biology Degree Program, National Taiwan University and Academia Sinica, Taipei, Taiwan[d]; Department of Life Science, National Taiwan University, Taipei, Taiwan[e]; Biotechnology Center, National Chung Hsing University, Taichung, Taiwan[f]

***Spiroplasma cantharicola* CC-1^T (DSM 21588) was isolated from the gut of a soldier beetle (*Cantharis carolinus*) collected in Maryland, USA. Here, we report the complete genome sequence of this bacterium to facilitate the investigation of its biology.**

Received 11 September 2015 **Accepted** 15 September 2015 **Published** 22 October 2015

Citation 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. doi:10.1128/genomeA.01253-15.

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

Spiroplasma cantharicola is an insect-associated bacterium found in Maryland, USA (1). This species belongs to the XVI-1 subgroup within the genus *Spiroplasma*, which includes strains isolated from beetle (*Cantharis carolinus*; represented by strain CC-1^T) and wasp (*Monobia quadridens*; represented by strain MQ-6) (2). Other serologically related strains with less than 70% similarities in DNA-DNA hybridization tests have been assigned to subgroups XVI-2 and XVI-3 (2). Of these, XVI-2 includes strains isolated from beetle (*Cantharis bilineatus*) and mosquito (*Aedes fulvus/annulipes* and *Anopheles punctipennis*) collected in the United States. In contrast, XVI-3 strains were all isolated from the Savoy region of France. Most of the XVI-3 strains are associated with mosquito (*Aedes* spp. and *Coquillettidia richiardii*; represented by strain Ar-1357). The XVI-3 strain PI-30L was isolated from a thistle plant (*Cirsium* sp.), representing a notable exception in terms of host association. To facilitate future investigation into the biology of these bacteria, as well as to improve the taxon sampling of available *Spiroplasma* sequences for comparative genomics and evolutionary studies, we determined the complete genome sequence of *S. cantharicola* CC-1^T.

The procedures for sample processing, sequencing, assembly, and annotation were based on those described in our previous studies on *Spiroplasma* genomes (3–9). The strain was acquired from the German Collection of Microorganisms and Cell Cultures (catalogue no. DSM 21588). The freeze-dried sample was processed according to the manufacturer's instructions and cultured in the M1D medium (10) prior to DNA extraction. PCR and Sanger sequencing were performed to verify that the 16S rRNA gene sequence matched the reference record (GenBank accession no. NR_125516.1).

The Illumina MiSeq platform was used to generate 301-bp reads from one paired-end library (~580-bp insert; 1,323,240,548 reads). The initial *de novo* assembly was performed using Velvet version 1.2.10 (11). Subsequently, PAGIT version 1 (12) 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 (13), programmatically checked using the MPILEUP program in the SAMTOOLS package version 1.2 (14), and visually inspected using IGV version 2.3.57 (15). 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 (16), tRNAscan-SE (17), and Prodigal (18) were used for gene prediction. The gene names and product descriptions were first annotated based on the homologous genes in other *Spiroplasma* genomes (3–9) as identified by OrthoMCL (19). Subsequent manual curation was based on BLASTp (20) searches against the NCBI nonredundant database (21) and the KEGG database (22, 23).

The circular chromosome of *S. cantharicola* CC-1^T is 1,179,577 bp in size and has a G+C content of 25.0%; 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 1,017 protein-coding genes.

Nucleotide sequence accession number. The complete genome sequence of *S. cantharicola* CC-1^T has been deposited in DDBJ/EMBL/GenBank under the accession number CP01262.

ACKNOWLEDGMENTS

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. W.-S.L. was supported by the TIGP-MBAS program (Academia Sinica and National Chung Hsing University). P.-Y.L. was supported by the GSB program (National Taiwan University and Academia Sinica). The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

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 service was provided by the DNA Sequencing Core Facility (Institute of

Molecular Biology, Academia Sinica). We thank Hon-Tsen Yu (Department of Life Sciences, National Taiwan University) for scientific collaboration and discussion related to this project.

REFERENCES

1. **Whitcomb RF, Chastel C, Abalain-Colloc M, Stevens C, Tully JG, Rose DL, Carle P, Bové JM, Henegar RB, Hackett KJ, Clark TB, Konai M, Williamson DL**. 1993. *Spiroplasma cantharicola* sp. nov., from cantharid beetles (Coleoptera: Cantharidae). Int J Syst Bacteriol **43:**421–424. http://dx.doi.org/10.1099/00207713-43-3-421.
2. **Abalain-Colloc ML, Williamson DL, Carle P, Abalain JH, Bonnet F, Tully JG, Konai M, Whitcomb RF, Bové JM, Chastel C**. 1993. Division of group XVI spiroplasmas into subgroups. Int J Syst Bacteriol **43:** 342–346. http://dx.doi.org/10.1099/00207713-43-2-342.
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* 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.
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-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.
9. **Lo W-S, Lai Y-C, Lien Y-W, Wang T-H, Kuo C-H**. Complete genome sequence of *Spiroplasma litorale* TN-1^T (DSM 21781), a bacterium isolated from green-eyed horsefly (*Tabanus nigrovittatus*). Genome Announc **3**(5)**:**e01116-15. http://dx.doi.org/10.1128/genomeA.01116-15.
10. **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.
11. **Zerbino DR, Birney E**. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
12. **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.
13. **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.
14. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup**. The Sequence Alignment of Map Format and SAMtools. Bioinformatics **25:**2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
15. **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.
16. **Lagesen K, Hallin P, Rodland EA, Staerfeldt 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.
17. **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.
18. **Hyatt D, Chen G, 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.
19. **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.
20. **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.
21. **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.
22. **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.
23. **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.